

# Computergestützte Audio- und Musikanalyse

Christof Weiß

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

**Abstract** Mit der fortschreitenden Digitalisierung sind nicht nur Textdokumente, sondern auch andere Arten von Medien in großen Mengen verfügbar. Dazu gehören Tonaufnahmen, die drei Haupttypen von Inhalten umfassen: Sprache, Umgebungsgeräusche (z. B. natürliche oder städtische Geräuschkulissen) und Musik. Zwar können alle für die theologische Forschung relevant sein, jedoch konzentriert sich dieses Kapitel auf die Verwendung von Audioaufnahmen zur Untersuchung von Kirchenmusik (computergestützte Musikwissenschaft). Nach einer Einführung in die Grundlagen von Audiodaten skizzieren wir zunächst eine Technik für die Visualisierung tonaler Verläufe (lokale Tonarten und Modulationen) in einer Musikaufnahme und wenden diese Technik auf Bachs Johannespassion BWV 245 an. Außerdem demonstrieren wir das Potenzial von Audioaufnahmen für die Korpusanalyse. Wir stellen einen Ansatz zur Untersuchung der tonalen Komplexität und ihrer Entwicklung über die Jahrhunderte vor. Mit dieser Technik untersuchen wir die tonale Entwicklung der geistlichen Musik anhand eines annotierten Audio-Korpus (5 773 Tracks), der von einem führenden Musikverlag für Chormusik, dem Carus-Verlag Stuttgart, stammt.\*

**Keywords** Audiosignalverarbeitung, Harmonieanalyse, computergestützte Musikwissenschaft, Korpusanalyse

## 1. Audiodaten und Anwendungen

Die laufenden Digitalisierungsbemühungen führen zu einer wachsenden Zahl von Archiven und Korpora kultureller Artefakte. Textdaten waren der Ausgangspunkt für die Computational Humanities (CH), bei denen statistische Methoden auf umfassende literarische Texte angewendet wurden. Heutzutage sind weitere Modalitäten in gleicher Weise verfügbar, einschließlich Audio-(Ton-)Aufnahmen. Im Gegensatz zu Text sind bei rohen Audioaufnahmen eine Reihe von Herausforderungen zu beachten: Erstens erfordert die Speicherung und Übertragung von Audiodaten aufgrund des erheblich größeren Umfangs (eine Sekunde unkomprimierten Stereotons ent-

\* Dieses Kapitel wurde inkl. fremdsprachiger Zitate von der Redaktion aus dem Englischen übersetzt.

spricht 88 200 16-Bit-Werten) mehr Ressourcen – ein Problem, das durch effiziente Audiocodierungstechnologien angegangen wurde, angefangen bei der MP3-Audio-kompression und ähnlichen Codecs. Zweitens erfordert die computergestützte Analyse von Audiodaten aufwändigere Verarbeitungstechniken.

Im Gegensatz zu Text sind explizite Symbole wie Zeichen oder Wörter (in Sprache) oder Notenergebnisse (in Musik) nicht direkt aus Audiodaten zugänglich. Um diese Informationen zu extrahieren, werden seit Jahrzehnten algorithmische Lösungen entwickelt, die Techniken aus dem Ingenieurwesen (Signalverarbeitung) und der Informatik (Mustererkennung, maschinelles Lernen und heutzutage *Deep Learning*/KI-Technologie) umfassen. Zentrale Foren für diese Forschung sind die *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*<sup>1</sup> oder die *IEEE/ACM Transactions on Audio, Speech, and Language Processing*.<sup>2</sup>

Im Allgemeinen enthalten Audiodaten eine Mischung aus verschiedenen Quellen. Nehmen wir bspw. eine Filmttonspur oder die eines Hörbuchs, die Sprache (von verschiedenen Sprecher\*innen), Hintergrundmusik sowie diegetische Geräusche (Soundeffekte, Geräusche, die von Personen oder Objekten in der Handlung erzeugt werden, oder Musik, die innerhalb der Handlung gespielt oder gesungen wird) enthalten kann. Die Trennung dieser Quellen stellt eine große rechnerische Herausforderung dar (Smaragdis, 2004). In vielen Fällen haben wir es jedoch mit reinen Audiodaten zu tun, die eine der drei Arten von Inhalten (Sprache, Umgebungsgeräusche, Musik) abdecken, die wir im Folgenden kurz zusammenfassen.

**Sprachdaten.** Da die grundlegendste Art von Audio die gesprochene Sprache ist, ist ein großer Teil der Audiotechnologie durch Anwendungen für die zwischenmenschliche Kommunikation motiviert. Folglich hat die Sprachverarbeitung die Entwicklung der digitalen Audiotechnologie vorangetrieben mit Forschungen zu grundlegenden Zeit-Frequenz-Transformationen, spezifischen Audiomeerkmalen wie den *Mel-Frequency-Cepstral Coefficients* (MFCCs), dynamischen Programmieretechniken wie *Hidden-Markov-Modellen* und in jüngerer Zeit Algorithmen des maschinellen Lernens auf der Grundlage neuronaler Netze (Bäckström et al., 2022). Ein zentraler Treffpunkt für diesen Forschungsbereich ist die jährliche INTERSPEECH-Konferenz der *International Speech Communication Association (ICSA)*.<sup>3</sup> Motiviert durch verschiedene Anwendungen umfasst die Sprachverarbeitung eine Vielzahl von Aufgaben wie Sprachkodierung und -übertragung, Sprecher\*innenidentifizierung, Speech-to-text-Transkription, Analyse von Emotionen, Prosodie oder Dialekt sowie Audioforensik. In jüngster Zeit hat sich die Erzeugung kohärenter Sprachsignale aus Text oder direkt aus einer

1 S. <https://ieeexplore.ieee.org/xpl/conhome/1000002/all-proceedings>, zuletzt aufgerufen am 21.06.2024.

2 S. <https://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=6570655>, zuletzt aufgerufen am 21.06.2024.

3 S. <https://www.isca-speech.org/jiscaweb/index.php/conferences>, zuletzt aufgerufen am 21.06.2024.

Benutzer\*innenanfrage dank enormer Fortschritte bei generativen *Deep-Learning*-Techniken weiterentwickelt. Interaktive Sprachassistenten sind eine der bekanntesten Anwendungen dieser Technologie. Für Computational Humanities sind effiziente Speech-to-text-Systeme (oder automatische Spracherkennung, ASR) (Schneider et al., 2019) von großem Interesse, da sie als Vorverarbeitungsschritt für die anschließende Anwendung textbasierter CH-Strategien verwendet werden können.

**Umgebungsgeräusche.** Neben der Sprache gibt es einen zweiten Bereich, der sich mit der Verarbeitung von Geräuschen im allgemeineren Sinne befasst. Es gibt eine spezielle Forschungsgemeinschaft zur Erkennung und Klassifizierung von akustischen Szenen und Ereignissen (orig. „Detection and Classification of Acoustic Scenes and Events“, DCASE), die sich im Rahmen des jährlichen DCASE-Wettbewerbs mit einer Vielzahl von Aufgaben zur Erkennung von akustischen Ereignissen und zur Klassifizierung akustischer Szenen befasst.<sup>4</sup> Eine besondere Anwendung dieser Technologie ist die Überwachung von Wildtieren und der Biodiversität, bei der z. B. Naturschutzgebiete mit Mikrofonen ausgestattet werden, um Tiergeräusche zu erfassen und die Anwesenheit von Arten, z. B. von Vögeln, zu analysieren (Bardeli et al., 2010). Eine weitere Anwendung ist die Überwachung städtischer Geräuschkategorien, um Probleme wie Lärmbelastung oder die Erkennung von Verbrechen oder potenziellen Gefahren gezielt anzugehen. Dies wurde z. B. in einem großen Projekt in New York City durchgeführt (Bello et al., 2019). Spezifisch für solche Anwendungen ist der Bedarf an ressourcenschonender Technologie, da viele Signale parallel über große Gebiete und sehr lange Zeiträume aufgezeichnet werden. Daher muss die (Vor-)Verarbeitung lokal auf einzelnen Sensoreinheiten (*Edge Devices*) erfolgen.

**Musikaufnahmen.** Die dritte große Kategorie umfasst Musikaufnahmen. Im Allgemeinen liegen Musikdaten in einer Vielzahl digitaler Datentypen vor, darunter (neben Audio) grafische Notenblätter oder symbolische (d. h. maschinenlesbare) Partituren, die explizit musikalische Symbole kodieren und in der Regel die detailliertesten Analysen ermöglichen (Temperley 1997; Bellmann 2012; White 2013; Nakamura & Kaneko 2019). Für eine Vielzahl von Musiktraditionen und -stilen, einschließlich improvisierter (z. B. Orgel Improvisation in der Kirche), elektronisch erzeugter oder mündlich überlieferter Musik, sind jedoch keine Partituren verfügbar. Darüber hinaus ermöglichen audiobasierte Ansätze die Untersuchung von Aufführungsaspekten wie dem Verhalten des Gemeindegesangs in der Kirche. Schließlich sind symbolische Partituren schwer zu beschaffen, da die manuelle Erstellung zeitaufwändig ist und die automatische Umwandlung von Notenbildern (*Optical Music Recognition*, OMR, siehe Calvo-Zaragoza et al. 2020) oder Audioaufnahmen (*Automatic Music Transcription*, AMT, siehe Benetos et al. 2019) in symbolische Partituren oft unbefriedigend ist und erhebliche manuelle Nachbearbeitung erfordert. Aus diesem Grund sind Audio-

4 S. <https://dcase.community>, zuletzt aufgerufen am 21.06.2024.

aufnahmen eine vielversprechende Alternative, da sie eine effiziente Skalierung der computergestützten Musikanalyse auf große Korpora ermöglichen (Scherbaum et al. 2017; Mauch et al. 2015; Weiß et al. 2018; 2019). Dies erfordert fortgeschrittene computergestützte Techniken, die die Daten in semantisch sinnvolle Darstellungen umwandeln, die von Musikexpert\*innen direkt interpretiert werden können. Diese Technologie wird innerhalb einer interdisziplinären Forschungsgemeinschaft entwickelt, die sich um die *International Society for Music Information Retrieval* (ISMIR) gruppiert,<sup>5</sup> die eine jährliche Konferenz und eine Zeitschrift anbietet.<sup>6</sup> *Music Information Retrieval* (MIR) umfasst eine Vielzahl von Aufgaben und Anwendungen, darunter Musiksynchronisation, Harmonieanalyse (Erkennung von Akkorden und Tonarten), Takt- und Tempoverfolgung, Genre- und Stilklassifizierung, Audiozerlegung und Musiktranskription (Müller 2021). Neben diesen Aufgaben der Audioanalyse spielen auch Aufgaben der Musikgenerierung und andere musikalische Datentypen eine wichtige Rolle im MIR. Im Folgenden konzentrieren wir uns auf das Potenzial der MIR-Technologie für die musikwissenschaftliche Forschung, die spezifische Datensätze und Analysetechniken erfordert. Wir stellen zwei analytische Studien vor. Die erste (Abschnitt 2) befasst sich mit der Visualisierung von Harmonie (lokale Tonarten und Skalen) zur Analyse der tonalen Organisation großer musikalischer Werke am Beispiel der Johannespassion BWV 245 von Johann Sebastian Bach. Die zweite (Abschnitt 3) demonstriert eine audiobasierte Korpusanalyse des musikalischen Stils in der westlichen Sakralmusik auf der Grundlage eines Datensatzes eines führenden Verlags für Chor- und Sakralmusik, dem Carus-Verlag Stuttgart.

## 2. Visualisierung tonaler Strukturen – Eine Fallstudie zu Bachs *Johannespassion*

In diesem Abschnitt wird ein algorithmischer Ansatz zur Visualisierung tonaler Informationen im Verlauf einer Audioaufnahme vorgestellt. In Anlehnung an Weiß & Müller (2021) führen wir grundlegende Begriffe des Tons, grundlegende Verarbeitungstechniken und unsere Visualisierungsstrategie am Beispiel des Chorals Nr. 22 „Durch Dein Gefängnis“ aus J. S. Bachs *Johannespassion* BWV 245 ein. Abschließend wenden wir diese Technik auf die gesamte *Johannespassion* an und zeigen ihr Potenzial für die Untersuchung der tonalen Organisation groß angelegter Werke.

5 S. <https://www.ismir.net>, zuletzt aufgerufen am 21.06.2024.

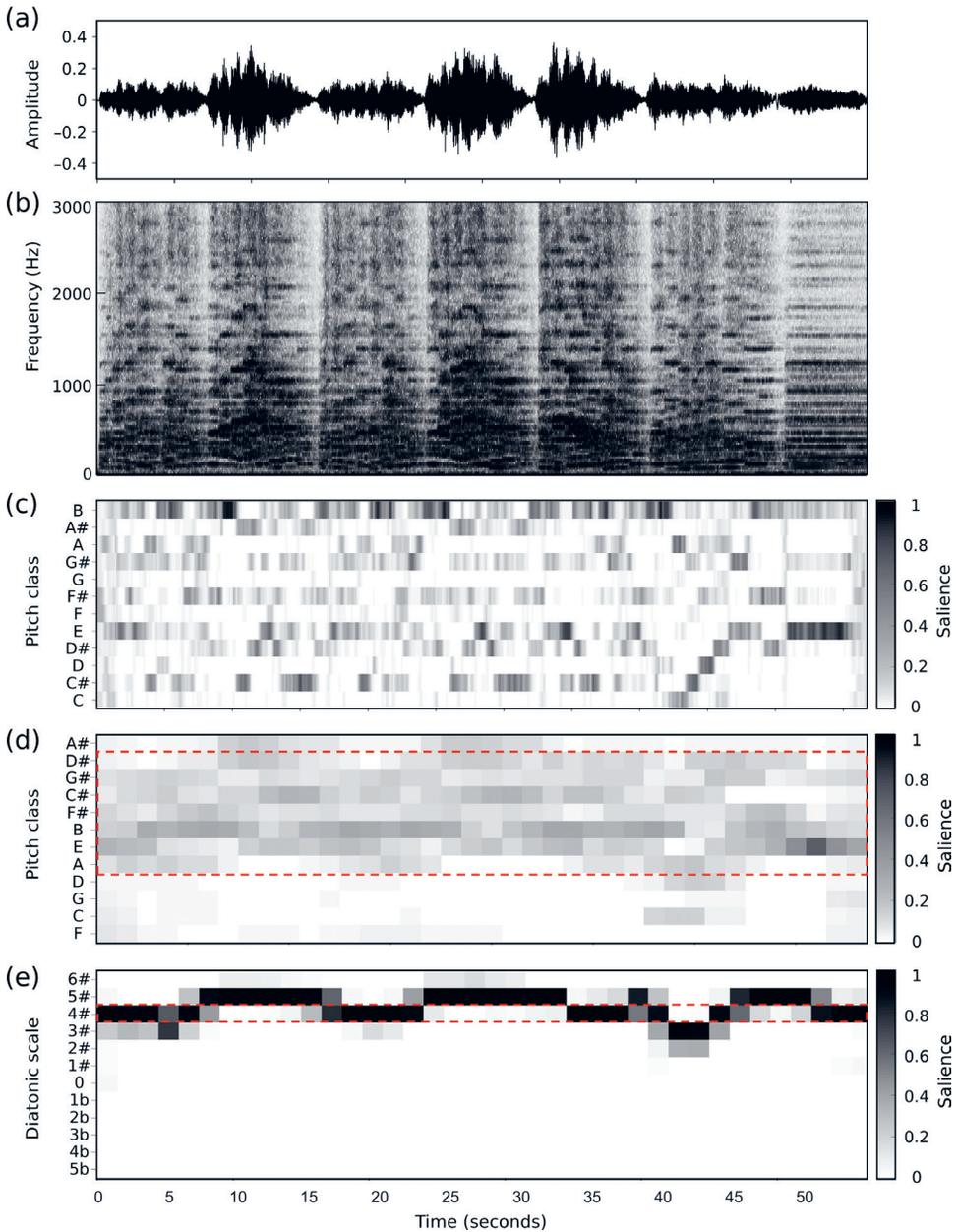
6 S. <https://transactions.ismir.net>, zuletzt aufgerufen am 21.06.2024.

## 2.1 Extraktion von Spektralinformationen aus Tonaufnahmen

Ausgangspunkt der Audioanalyse ist eine akustische Wellenform (auch als Signal bezeichnet), wie in Abb. 1a für eine Aufnahme des Bach-Chorals dargestellt. In einem ersten Schritt führen wir eine Spektralanalyse durch (Müller 2021, Kap. 2). Dazu unterteilen wir das Signal zunächst in lokale Zeitfenster (*Frames*). Die Breite des Zeitfensters (angegeben in Sekunden) ist ein kritischer Parameter, der an die jeweiligen Anwendungsanforderungen angepasst werden muss, da es einen Zielkonflikt zwischen Frequenz- und Zeitauflösung gibt. Innerhalb eines Zeitfensters wird die Ausprägung verschiedener Frequenzen berechnet, was z. B. durch die Fourier-Transformation realisiert werden kann.<sup>7</sup> Dieses Zeitfenster wird nun über das Signal geschoben, sodass man für jeden Frame eine lokale Frequenzverteilung erhält. Daraus ergibt sich eine Zeit-Frequenz-Darstellung, ein sogenanntes *Spektrogramm*, das in Abb. 1b für das Bach-Choralbeispiel dargestellt ist.

Für die tonale Analyse fassen wir diese spektrale Information weiter nach Tonhöhen zusammen. Zu diesem Zweck gehen wir vereinfachend davon aus, dass der Tonhöheninhalt durch die gleichschwebende Zwölftton-Skala gut genug beschrieben werden kann. Außerdem gehen wir davon aus, dass die Information über die Tonhöhenklasse (ohne Berücksichtigung der Oktavlage einer Tonhöhe) für unsere tonalen Analyseaufgaben ausreicht. So erhalten wir die zwölf chromatischen Tonhöhenklassen  $c, c\#, d, d\#, \dots, h$ . Eine enharmonische Unterscheidung von Tonhöhen wie  $c\#$  und  $d\flat$  ist hier nicht möglich. Für jedes Zeitfenster des Spektrogramms werden die Frequenzkomponenten nach diesen zwölf Tonhöhenklassen aggregiert. Daraus ergibt sich eine Zeit-Chroma-Darstellung oder ein Chromagramm. Abb. 1c zeigt ein solches Chromagramm (Zeitauflösung 10 Hz) für das Beispiel des Bach-Chorals. Diese Darstellung erfasst die Energieverteilung des Audiosignals über die zwölf chromatischen Tonhöhenklassen im Zeitverlauf. Die Umwandlung von Musikaufnahmen in Chromagramme als Zwischenschritt in der Verarbeitungskette ist ein grundlegender Ansatz für verschiedene MIR-Aufgaben wie Tonartschätzung, Tonleiteranalyse und Akkorderkennung. Für Details der zugrundeliegenden Signalverarbeitung verweisen wir auf Müller (2021).

7 Zusammen mit dem zuvor beschriebenen Fensterverfahren wird diese spezielle Variante als (diskrete) Kurzzeit-Fourier-Transformation (orig. „Short-Time Fourier Transform“, STFT) bezeichnet. Andere Transformationen wurden für spezielle Anwendungen entwickelt, wie die Konstant-Q-Transformation (orig. „Constant-Q Transform“, CQT) für die Tonhöhenanalyse oder die modifizierte diskrete Kosinustransformation (orig. „Modified Discrete Cosine Transform“, MDCT) für die Audiocodierung.



**Abb. 1** Choral Nr. 22 „Durch Dein Gefängnis“ aus J. S. Bachs Johannespassion BWV 245, Aufnahme von *The Scholars Baroque Ensemble* (Naxos 1994). (a) Wellenform des Audiosignals. (b) Spektrogramm. (c) Chromagramm. (d) Geglättetes Chromagramm, neu geordnet nach reinen Quinten. Usualisierung der Wahrscheinlichkeiten für diatonische Skalen.

## 2.2 Visualisierung diatonischer Tonleitern

Die Beobachtung und Messung von Energieverteilungen in Tonhöhenklassen liefern nur begrenzte Informationen für tonal komplexe mehrstimmige Musik. Nützlichere Kategorien für diesen Zweck sind Intervalle, Akkorde oder Skalen, die durch weitere Verarbeitungsschritte beschrieben werden müssen. Im Folgenden betrachten wir die Messung der Tonhöheninhalte nach den zwölf diatonischen Skalen. Dazu glätten wir zunächst das Chromagramm (durch lokale Mittelung), um der größeren zeitlichen Auflösung musikalischer Skalen Rechnung zu tragen (die Tonhöhen einer Tonleiter treten in der Regel nicht innerhalb einer so kurzen Zeitspanne wie unsere 100 ms-Chromagrammframes auf). Dies wird in Abb. 1d deutlich, wo wir eine Filterlänge von 45 Frames (d. h. 4,5 Sekunden) gewählt haben. Zweitens wird ein Chromagramm für jedes dieser geglätteten Frames mit binären Templates verglichen, die den Skalen entsprechen. Die Diatonik  $0$  (Tonhöheninhalt C-Dur oder A-Moll, keine Vorzeichen) wird bspw. durch ein Template modelliert, in der die Werte für die sieben Tonhöhenklassen  $c, d, e, f, g, a, h$  auf  $1$  und für die restlichen fünf Tonhöhenklassen  $c\#, d\#, f\#, g\#, a\#$  auf  $0$  gesetzt werden. In Abb. 1d ist der Tonhöheninhalt der  $4\#$ -Tonleiter (entsprechend E-Dur, der Grundtonart des Chorals) rot hervorgehoben. Aufgrund der Ähnlichkeit der diatonischen Skalen, die um eine reine Quinte verwandt sind (die sechs von sieben Tonklassen gemeinsam haben), ordnen wir die Skalen nach dem Quintenzirkel an.

Durch den lokalen Abgleich der Chromagrammframes mit den zwölf verschiedenen Templates können Wahrscheinlichkeiten für das Auftreten dieser Skalen im Zeitverlauf berechnet werden. Dies führt zu einer verallgemeinerten Zeit-Diatonik-Darstellung, bei der die Wahrscheinlichkeiten über ein Graustufenschema visualisiert werden können (hier entspricht Schwarz der Wahrscheinlichkeit  $1$  und Weiß der Wahrscheinlichkeit  $0$ ). Abb. 1e zeigt eine solche Visualisierung für eine Aufnahme des Bach-Chorals. Um wichtige Strukturen besser hervorzuheben, können zusätzliche Verarbeitungsschritte erforderlich sein. Zum Beispiel kann eine weitere zeitliche Glättung oder Anhebung der höheren Energiewerte und die Unterdrückung niedriger, rauschartiger Werte zu klareren visuellen Strukturen führen. Letzteres wird durch eine exponentielle Umskalierung der Energiewerte (ähnlich der *Softmax*-Funktion) realisiert. Die aus einer Musikaufnahme abgeleitete Zeit-Diatonik-Darstellung ist zunächst nach physikalischen Zeitschritten (Sekunden) organisiert. Für einige Anwendungen, wie z. B. die Bewertung der musikalischen Form von großen Werken, kann dies nützlich sein. Für den Vergleich mit Partiturdarstellungen oder anderen Interpretationen eines Werkes (versionsübergreifende Analyse) ist es jedoch sinnvoll, musikalische Zeitinformationen (z. B. Strukturgrenzen oder Takte/Schläge) zu berücksichtigen. Wenn solche Informationen verfügbar sind, können die zeitlichen Komponenten der Zeit-Diatonik-Darstellung musikalisch geglättet werden, um z. B. eine Darstellung mit Viertelnotenauflösung zu erhalten.

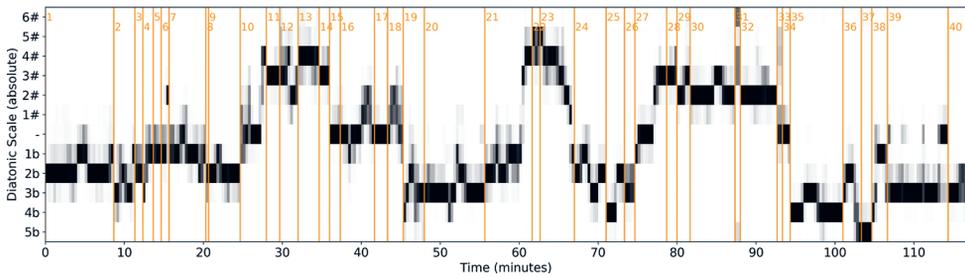
### 2.3 Bach-Choral-Beispiel

Schauen wir uns nun das Ergebnis der Visualisierung für unseren Bach-Choral genauer an (Abb. 1e). Der erste Teil beginnt in der Grundtonart des Chorals (4♯) und moduliert ab etwa 8 Sekunden in die Oberquinte oder Dominanttonart (5♯). Eine interessante Beobachtung ist die Abweichung von der 4♯-Ebene bei etwa 5 Sekunden, wo Veränderungen (hier das d im Akkord g♯-d-e-h) das Ergebnis beeinflussen. Tatsächlich wird im gesamten ersten Takt kein d♯ gefunden, sodass auch für die 3♯-Diatonik eine hohe Wahrscheinlichkeit erkennbar ist. Dies zeigt, dass unser Verfahren keine explizite *Erkennung* von Tonarten vorsieht, sondern nur den lokalen Tonhöhengehalt in Form von diatonischen Skalen beschreibt. Auf der Grundlage der Zeit-Diatonik-Darstellung lässt sich der grobe harmonische Verlauf des Bach-Chorals problemlos nachvollziehen. Die Anfangsphase (4♯ mit Modulation zu 5♯) wird mit anderem Text wiederholt. Der Choral setzt sich mit komplexeren Harmonien fort. Hier führt die chromatische Passage im Bass (Text: „Unsere Knechtschaft“) bei etwa 40 Sekunden zu mehreren Skalen, die Wahrscheinlichkeiten ungleich Null erhalten. Schließlich endet der Choral in der Grundtonart E-Dur (4♯). Neben dem einfachen Zugang zu Musikaufnahmen und deren unkomplizierter Verarbeitung ist die direkte Anwendbarkeit auf Tonaufnahmen von besonderem Vorteil, da die Analyse direkt mit dem akustischen Eindruck verknüpft werden kann, z. B. durch die Verwendung eines laufenden Cursors als Animation.

### 2.4 Großflächige tonale Visualisierung der *Johannespassion*

Einer der größten Vorteile der vorgestellten Analysestrategie ist ihre Skalierbarkeit. Umfangreiche Werke wie Opern, Oratorien oder Sinfonien können auf kompakte und konsistente Weise visualisiert werden und ermöglichen so einen guten Überblick über die tonale Konzeption – ein sehr wichtiger Aspekt, da die Tonalität ein zentrales Mittel zur formalen Strukturierung langer Werke ist. Wir demonstrieren dies nun, indem wir unsere Analyse vom einminütigen Choral auf die gesamte *Johannespassion* BWV 245 ausdehnen, was (in der Aufführung von *Scholars Baroque*) einer Gesamtspielzeit von rund zwei Stunden entspricht. Abb. 2 zeigt das Ergebnis dieser Analyse, wobei das gesamte Werk zu einer einzigen Grafik verdichtet wurde.

Um zu vermeiden, dass Mikrofluktuationen die grobe tonale Struktur verdecken, entscheiden wir uns jetzt für eine viel größere Fenstergröße bei der Mittelwertbildung der Chroma-Merkmale vor dem *Template-Matching*. Im Gegensatz zur feingranularen Analyse des Chorals (Abschnitt 2.2) mit einer Filterlänge von 4,5 Sekunden haben wir nun eine Filterlänge von 60 Sekunden gewählt. Dies führt zu einer Unterdrückung von Details, erhöht aber die Robustheit der Methode und hilft, die tonale Gesamtstruktur zu betonen.



**Abb. 2** J. S. Bachs *Johannespassion* BWV 245, Gesamtaufnahme von *Scholars Baroque*. Visualisierung der diatonischen Tonleiter für alle Sätze.

Lassen Sie uns nun die Ergebnisse diskutieren. Zur besseren Lesbarkeit enthält Abb. 2 nur die Nummern der einzelnen Sätze. Die Form (Instrumentierung) und Textanfänge sind als Referenz in Tab. 1 aufgeführt. Zunächst beobachten wir eine gewisse lokale Stabilität. Innerhalb eines Satzes und zwischen benachbarten Sätzen sind entfernte Modulationen selten, sodass der nächste Satz meist in derselben oder einer eng verwandten Tonart beginnt. Zweitens identifizieren wir eine gewisse tonale Geschlossenheit. Das Werk beginnt in der  $-2$ -Diatonik (Nr. 1, Chor „Herr, unser Herrscher“ in  $g$ -Moll) und endet mit zwei Nummern in der  $-3$ -Diatonik (Nr. 39, Chor „Ruht wohl, ihr heiligen Gebeine“ in  $c$ -Moll und Nr. 40 (Choral „Ach Herr, lass dein lieb Engelein“ in  $Es$ -Dur), nur eine Quinte voneinander entfernt. Dazwischen wird dieser Tonartbereich deutlich verlassen, mit dem Höhepunkt unseres oben besprochenen Chorals Nr. 22 „Durch Dein Gefängnis“, der mit seinen  $4\sharp$  etwa sieben Quinten entfernt liegt. Dies ist nicht nur bemerkenswert, da  $E$ -Dur im Barock mit dem Tod in Verbindung gebracht wurde – Johann Mattheson schrieb dieser Tonart in seinem Werk *Das Neu-eröffnete Orchester* von 1713 eine „tödliche Traurigkeit“ zu. Tatsächlich ist der Choral in mehrfacher Hinsicht auch das Zentrum der Passion. Was die Spielzeit angeht, so geschieht dies mitten im Werk, nach etwa einer Stunde (was nicht der Beginn des zweiten Teils ist). Darüber hinaus gibt es eine klare Symmetrie um diesen Choral, die in unserer Visualisierung gut zu erkennen ist. Die Nummern 21 und 23 (beides eine Reihe von Rezitativen und Chören, die mit unterschiedlichem Text wiederholt werden) enthalten Übergänge zu bzw. von  $4\sharp$ . Die Arien Nr. 20 und 24 bestehen jeweils aus den Skalen  $3b$  und  $2b$ . Nr. 15–17 und 27 betonen die  $0$ -Diatonik. Diese tonalen Beziehungen zeigen die ausgefeilte Konzeption von Bachs *Johannespassion* und seinen bewussten Einsatz der Tonalität zur Strukturierung des Werks und zur Betonung der theologischen Botschaften der Passionserzählung.

**Tab. 1** J. S. Bachs *Johannespassion* BWV 245, Übersicht über Nummern (Sätze), Formen und Textanfänge.

Nr.	Form	Text-Hinweis
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	Aria	Von den Stricken meiner Sünden
8	Rezitativ	Simon Petrus aber folgte Jesu nach
9	Aria	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	Aria	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	Aria	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 schrienen und sprachen
24	Aria	Eilt, ihr angefochtenen 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	Aria	Es ist vollbracht
31	Rezitativ	Und neiget das Haupt
32	Aria	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	Aria	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

### 3. Audiobasierte Korpusanalyse

Die Skalierung von Analysen auf umfassende Werke ist ein großer Vorteil computergestützter Methoden. Eine weitere interessante Möglichkeit besteht darin, ganze Korpora musikalischer Werke zu analysieren. Dabei kann sich ein Korpus auf einen geschlossenen Satz von Werken beziehen, z. B. alle Choräle von J. S. Bach oder alle Streichquartette von L. van Beethoven. Es könnte sich jedoch auch auf eine offene Teilmenge einer gesamten Zeitspanne beziehen. Um ein langfristiges Ziel zu skizzieren, stellen wir uns die Analyse der Entwicklung der westlichen Kirchenmusik vor, die mehr als tausend Jahre vom einstimmigen Gesang bis zu den heutigen Avantgarde-Kompositionen umfasst. In Anlehnung an Weiß & Müller (2023) präsentieren wir nun einen ersten Schritt zu einer solchen Korpusanalyse. Zu diesem Zweck betrachten wir einen Audiodatensatz, der von einem führenden Verlag für Chor- und Kirchenmusik bereitgestellt wurde.

#### 3.1 *Carus Audio Corpus*

Der 1972 in der Nähe von Stuttgart gegründete Carus-Verlag ist ein Familienunternehmen mit Schwerpunkt auf Vokal- und Kirchenmusik. Ihre Notenausgaben umfassen rund 45 000 Werke (die meisten davon Vokalkompositionen) und spiegeln die Entwicklung von fünf Jahrhunderten Chormusik wider, von gregorianischen Gesängen, Madrigalen und Motetten der Renaissance bis hin zu zeitgenössischer Chormusik und Werken für Jazz- und Pop-Chor.<sup>8</sup> Carus bietet historisch-kritische Musikausgaben der bedeutendsten Oratorien, Messen und Kantaten der Musikgeschichte, orientiert an der historischen Aufführungspraxis. Carus ist auch als Plattenlabel aktiv und veröffentlicht Referenzaufnahmen auf Basis eigener Editionen.

Das Carus Audio Corpus (CAC) umfasst den Großteil der Carus-CD-Veröffentlichungen (Stand 2019) mit insgesamt 7 115 Titeln, die einzelnen Werken oder Sätzen (bei mehrsätzigen Werken) entsprechen. Da wir uns auf originale Kunstmusikkompositionen konzentrieren möchten, führen wir einen ersten Bereinigungsschritt durch, bei dem wir Werke ohne Komponist\*innen, Werke ohne Lebensdaten der Komponist\*innen, Arrangements, Popmusik, Kinderlieder und Weihnachtslieder entfernen. Danach verbleiben 5 773 Titel (Sätze) zu 2 409 verschiedenen Werken mit einer Gesamtdauer von 389:52:20 (hh:mm:ss). Im Durchschnitt hat ein Werk 2,4 Sätze und eine Dauer von 9:43 (mm:ss). Wir stellen jedoch fest, dass die Anzahl der Sätze pro Werk sehr unausgewogen ist, mit vielen einsätzigen Werken einerseits und vielen großformatigen Werken (wie der *Johannespassion*) andererseits.

Tab. 2 enthält Statistiken über die Annotationen des CAC auf Werkebene, wobei sich Informationen wie Tonart oder Instrumentierung immer auf die übergreifende

8 S. <https://www.carus-verlag.com/en/ueber-carus>, zuletzt aufgerufen am 21.06.2024.

**Tab. 2** Statistik des CAC und seiner Annotationen. Alle Zahlen beziehen sich auf Gesamtwerke (nicht auf einzelne Sätze).

Annotationstyp	Anzahl der Werke
– Alle –	2 409
Werkdatum	1 151
Instrumentierung	1 964
instrumental	200
vokal	1 764
Chor	1 400
Solo	364
Tongeschlecht	1 166
Dur	673
Moll	348
Sonstiges	145

Komposition beziehen. Bei etwa der Hälfte der Werke sind Angaben zum Entstehungsjahr (Werkdatum) enthalten. Der Großteil ist mit Annotationen zur Instrumentierung versehen. Wie gesagt liegt ein starker Fokus auf Vokalmusik im Allgemeinen sowie auf Chormusik.<sup>9</sup> Für die tonale Analyse ist die Verfügbarkeit zentraler Annotationen zu rund der Hälfte der Werke (1 166 von 2 409) von besonderer Relevanz. Es gibt eine Tendenz zu Dur-Tonarten sowie eine beträchtliche Anzahl anderer Tonarten (Kirchentonarten).

Wie oben erwähnt, erstreckt sich CAC über etwa 450 Jahre und deckt den Zeitraum von etwa 1570 bis 2020 ab. Insgesamt stammen die Werke von 234 verschiedenen Komponist\*innen. Abb. 3 zeigt eine historische Ansicht der Daten von Komponist\*innen mit mindestens fünf Werken.

F. Mendelssohn-Bartholdy, J. S. Bach oder W. A. Mozart tragen einen wesentlichen Teil dazu bei. CAC umfasst jedoch auch weniger bekannte Komponist\*innen wie H. Schütz oder M. Reger. Carus unternimmt große Anstrengungen, fast vergessene Werke von G. A. Homilius oder J. G. Rheinberger wieder in den Fokus der Chorszene zu rücken. Besonders interessant ist die gute Abdeckung des späten 15. und 16. Jahrhunderts. Im 20. Jahrhunderts finden wir dagegen eine geringere Zahl an Werken, fast schon eine Lücke um 1950.

<sup>9</sup> Es gilt zu beachten, dass aufgrund der werkbezogenen Anmerkungen einzelne Sologesangssätze (z. B. eine Arie) innerhalb eines Chorwerks (z. B. eines Oratoriums) zu den Chorwerken gezählt werden.

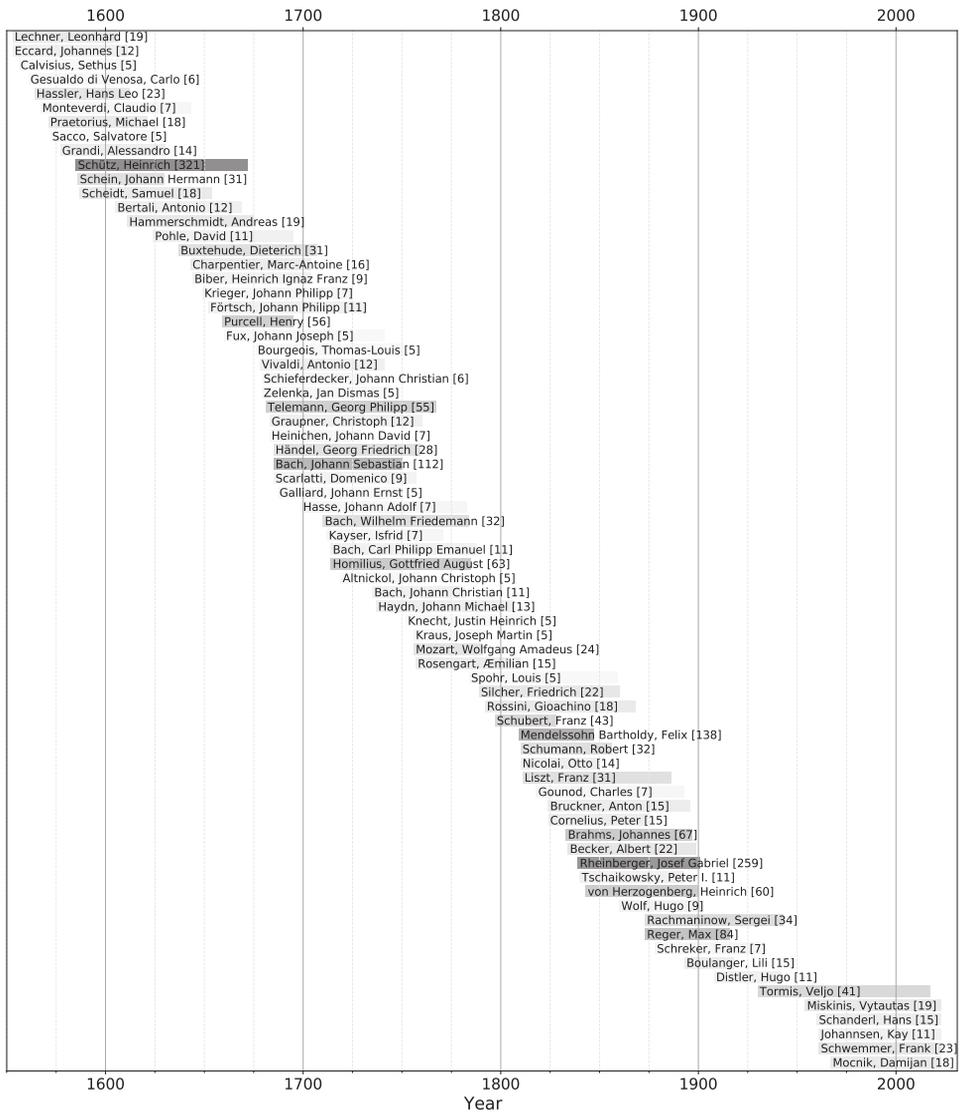


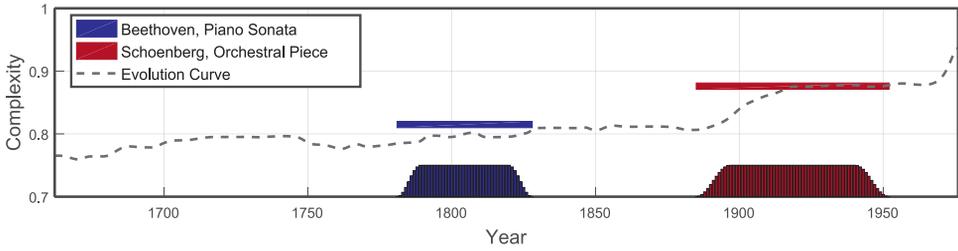
Abb. 3 Historische Darstellung des CAC unter Beruckichtigung aller Komponist\*innen mit mindestens funf Werken. Die Anzahl der Werke der jeweiligen Komponist\*innen ist in eckigen Klammern angegeben und durch den Grauwert der Balken kodiert.

### 3.2 Werkzahlkurven und Approximationsstrategie

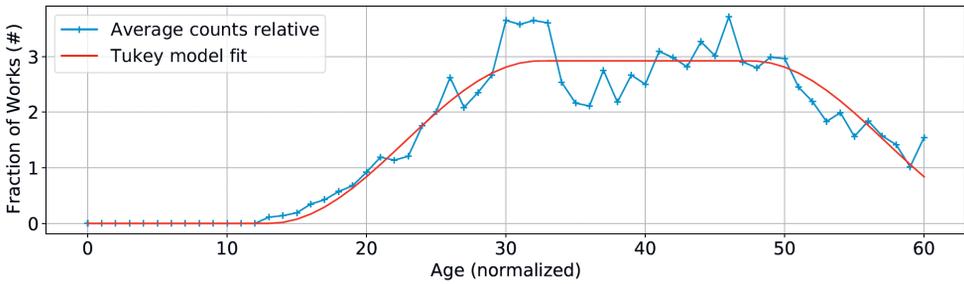
Um Musikstile in ihrem historischen Kontext zu analysieren, verfügt man idealerweise über Informationen über die *tatsächlichen Entstehungsdaten*, d. h., das Jahr, in dem eine Komposition fertiggestellt wurde. Musikstile können sich schnell weiterentwickeln und das Komponieren unterliegt Trends, dem Einfluss anderer Komponist\*innen, dem Geschmack des Publikums oder außermusikalischen Einflüssen. Man könnte an Komponist\*innen mit mehreren *Schaffensperioden* denken wie etwa L. van Beethoven oder A. Schönberg. Das Sammeln verlässlicher Anmerkungen zu Werkdaten für größere Datensätze erfordert jedoch einen erheblichen manuellen Rechercheaufwand und diese Informationen sind für eine ganze Reihe von Werken unbekannt oder zweifelhaft. Selbst wenn man alle Erstellungsdaten kennt, wird es schwierig, einen Datensatz mit einer ausgewogenen Abdeckung aller Jahre zu erstellen.

Aufgrund dieser Probleme haben wir in unserer vorherigen Arbeit (Weiß et al., 2019) einen pragmatischen Ansatz gewählt, indem wir Werke auf der historischen Zeitachse basierend auf *Komponist\*innendaten* projizierten, d. h. die Angaben zum Geburts- und Sterbejahr, die wesentlich schneller zu erfassen sind. Zur Annäherung an die Anzahl der Werke im Laufe des Lebens der jeweiligen Komponist\*innen haben wir angenommen, dass sie erst ab einem bestimmten (festgelegten) Alter mit dem Komponieren beginnen. Für die verbleibenden Jahre (Alter) haben wir mithilfe eines sogenannten *Tukey-Fensters* (dargestellt in Abb. 4) eine ungefähr flache Verteilung mit glatten Kanten berechnet. In Weiß et al. (2019) wurden die Parameter (Startalter und Tukey-Parameter  $\alpha$ ) heuristisch gewählt. Da CAC für etwa die Hälfte der Arbeiten solche Anmerkungen enthält (vergleiche Tab. 2), können wir die Approximationsstrategie validieren und nach optimalen Werten der Parameter suchen. Wir tun dies schrittweise, indem wir zunächst das optimale Startalter bestimmen und einen Wert von 13 erhalten. Anschließend bestimmen wir den optimalen Tukey-Parameter zu  $\alpha = 0,72$ . Für Details zum Tukey-Fenster und der Anpassungsstrategie verweisen wir auf Weiß & Müller (2023).

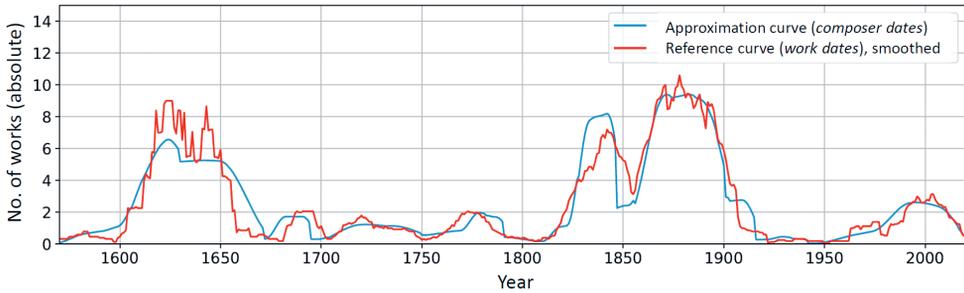
Die resultierende Kurve ist in Abb. 5 für einen hypothetischen Komponisten dargestellt, der im Alter von 60 Jahren gestorben ist. Mit diesen optimierten Fensterparametern validieren wir nun die Approximationsstrategie für die Werkzahlkurve. Dazu berechnen wir zunächst die Referenzkurve anhand der Werkdaten von 1151 Werken, die diese Anmerkungen aufweisen. Wir bearbeiten die Kurve mit einem Durchschnittsfilter der Länge von 15 Jahren nach (rote Kurve in Abb. 6). Anschließend vergleichen wir diese Referenzkurve mit unserer Näherungskurve basierend auf *Komponist\*innendaten* und unserem optimierten Tukey-Fenster (blaue Kurve in Abb. 6). Insgesamt scheint die Näherung geeignet zu sein. Nur in einigen Zeiträumen (z. B. um 1680) liegt die Näherungskurve vor, in anderen (z. B. um 1770) hinkt sie der Referenzkurve hinterher. Wir kommen zu dem Schluss, dass die auf Tukey-Fenstern basierende Näherung eine geeignete Strategie ist, um fehlende Annotationen zum Kompositionsdatum zu kompensieren.



**Abb. 4** Entwicklung der Klangkomplexität anhand von Komponist\*innendaten, schematisches Beispiel.



**Abb. 5** Kurvenanpassungsverfahren zur Bestimmung der optimalen Fensterparameter.



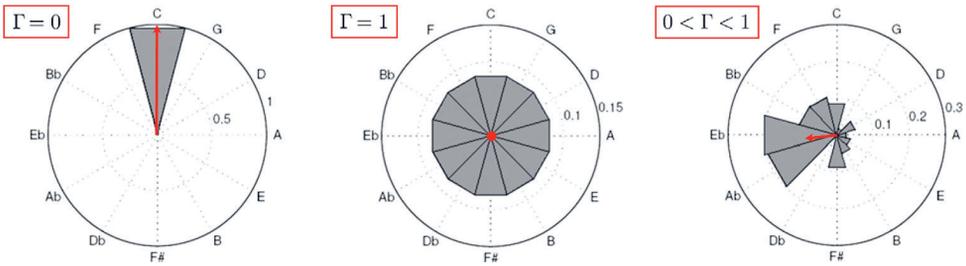
**Abb. 6** Werkzahlkurven basierend auf Lebensdaten (Approximationskurve, blau) bzw. basierend auf Kompositionsdaten (Referenzkurve, rot).

### 3.3 Untersuchung der Klangentwicklung geistlicher Musik

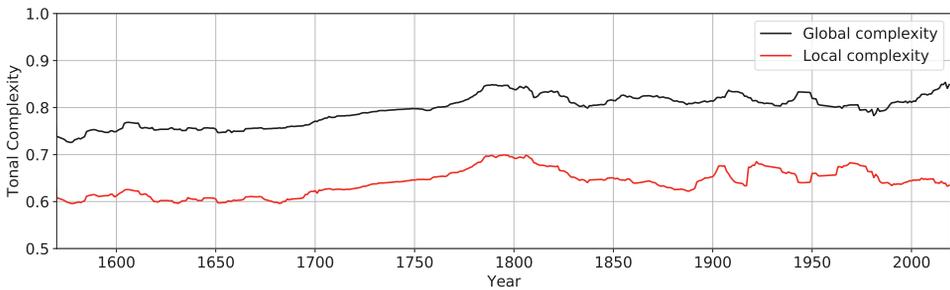
Mit der vorgestellten Strategie untersuchen wir nun die klangliche Entwicklung der Chormusik im CAC. Zu diesem Zweck betrachten wir einen computergestützten Ansatz zur Messung der tonalen Komplexität aus Audioaufnahmen (Weiß et al. 2019). Musikalische Komplexität ist ein hochrelevanter Begriff für die Analyse, der mehrere Aspekte wie akustische, klangliche oder rhythmische Eigenschaften umfasst (Streich 2007). In unserer früheren Arbeit (Weiß & Müller 2014) haben wir tonale Komplexitätsmaße eingeführt, die lokale Energieverteilungen über die zwölf im westlichen Tonsystem verwendeten chromatischen Tonhöhenklassen beschreiben. Hier haben wir die Vielfalt der Tonhöhenklassen berücksichtigt, sodass flache Verteilungen (z. B. chromatische Cluster) zu hohen Komplexitätswerten führen, während scharfe Verteilungen (z. B. einzelne Noten) zu niedrigen Komplexitätswerten führen (siehe Abb. 7). Hierzu greifen wir wiederum auf eine Chroma-Darstellung der Audioaufnahme zurück. Zur Berechnung der Komplexitätswerte bilden wir jeden Chromagrammframe (Chromavektor)  $c \in \mathbb{R}^{12}$  auf den Quintenzirkel ab. Zu diesem Zweck sortieren wir zunächst die Chromawerte nach perfekten Quintintervallen um. Basierend auf dem umsortierten Vektor berechnen wir zirkuläre Statistiken unter Verwendung des resultierenden Vektors. Dann bezieht sich unser Komplexitätsmaß  $\Gamma(c) \in [0, 1]$  auf die inverse Länge dieses resultierenden Vektors. Dieses Maß beschreibt die Ausbreitung der Tonhöhen um den Quintenzirkel und berücksichtigt somit auch die tonale Beziehung aktiver Tonklassen. Abb. 7 veranschaulicht die Definition des Komplexitätsmaßes und des resultierenden Vektors (in Rot) und zeigt Beispiele für drei Chroma-Vektoren. Für einen dünn besetzten Vektor (links) ist die Komplexität minimal. Für einen flachen Vektor (Mitte) erhalten wir maximale Komplexität. Andere Chroma-Vektoren ergeben mittlere Komplexitätswerte.

Wir stellen fest, dass es unterschiedliche Aggregationsstrategien gibt, um Werte nachzuverfolgen. Ein lokales Maß erhält man, indem die Komplexität jedes Chromagrammrahmens berechnet und dann über diese Werte gemittelt wird. Eine globale Chroma-Statistik kann berechnet werden, indem zuerst die Chromagramm-Frames gemittelt werden und dann ein einzelner Komplexitätswert für jede Spur (Bewegung) berechnet wird. Die Aggregation zu Werken erfolgt dann durch Mittelung der Komplexitätswerte für alle Sätze. In Abb. 6 haben wir die Gesamtzahl der Werke im CAC im Laufe der Jahre (Werkzahlkurven) anhand der Werkdaten oder unserer auf Komponist\*innendaten basierenden Näherungsstrategie untersucht. Wir wenden diese Strategien nun auf unsere Messungen der tonalen Komplexität an. Während die Fenster für jedes Werk mit dem Wert eins gewichtet wurden, um die Gesamtzahl der Werke zu berücksichtigen, verwenden wir nun den Komplexitätswert des jeweiligen Werks zur Gewichtung.

Wir summieren alle gewichteten Fenster und dividieren zur Normalisierung durch die jeweilige Werkzahlkurve. Wir erhalten eine sogenannte *Evolutionskurve* (orig. „evolution curve“, EC), die die durchschnittliche Komplexität der Werke ent-



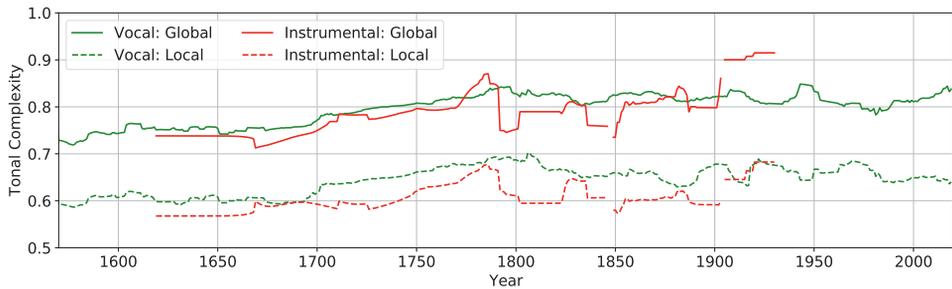
**Abb. 7** Komplexitätsmaß  $\Gamma$  basierend auf dem Quintenzirkel. Es werden Werte für einen dünn besetzten Chroma-Vektor (links), einen flachen Chroma-Vektor (Mitte) und einen realistischeren Chroma-Vektor (rechts) angezeigt. Die roten Pfeile bezeichnen die resultierenden Summenvektoren.



**Abb. 8** Vergleich von Entwicklungskurven hinsichtlich globaler und lokaler Komplexität.

lang der historischen Zeitachse angibt. Auf diese Weise trägt jedes Werk zu dem Teil der Zeitachse bei, der seinem Kompositionsdatum (für die Referenzkurve) oder den Lebensdaten seiner Komponist\*innen (für die Approximationskurve) entspricht.

Wir wenden diese gemischte Strategie nun an, um die Entwicklung der tonalen Komplexität im CAC zu untersuchen. Abb. 8 zeigt die resultierenden ECs, eine basierend auf der lokalen und eine auf der globalen Komplexität. Betrachtet man die globale EC (schwarz), so beobachten wir im Laufe des 17. und 18. Jahrhunderts eine Zunahme der Komplexität. Interessanterweise beobachten wir um 1750 keinen Rückgang, im Gegensatz zu Weiß et al. (2019), wo die Forderung nach mehr „Einfachheit“ nach dem Barock deutlich sichtbar war. Dagegen ist der in Weiß et al. (2019) beobachtete Anstieg im 19. Jahrhundert für CAC nicht sichtbar. Noch bemerkenswerter ist, dass CAC im 20. Jahrhundert keinen wesentlichen Anstieg der Komplexität verzeichnete. Der Modernismus in der Tonalität, der von expressionistischen und zwölfstimmigen Komponist\*innen wie A. Schönberg oder I. Strawinsky vorangetrieben wurde, scheint sich in der Chormusik nicht in gleichem Maße widerzuspiegeln. Dies könnte auf unterschiedlichen Stilrichtungen in der Chormusik beruhen, aber auch eine Eigenschaft des CAC sein, wo komplexe atonale Werke möglicherweise nicht im Fokus stehen, da sie von Laienchören nur schwer aufzuführen sind.



**Abb. 9** Vergleich von ECs hinsichtlich globaler und lokaler Komplexität, getrennt nach Vokal- und Instrumentalmusik.

Abschließend zeigen wir, wie die Korpusanalyse für hypothesengesteuerte Forschung genutzt werden kann. Dazu untersuchen wir die Hypothese, dass Instrumentalmusik komplexer ist als Vokalmusik. Wir könnten ein solches Verhalten erwarten, da Gesangskompositionen die höhere Schwierigkeit bei der Erzeugung von Tonhöhen beim Singen berücksichtigen müssen, insbesondere bei großen und komplexen Intervallen. Darüber hinaus behaupten Musikwissenschaftler\*innen meist, dass kompositorische „Revolutionen“ häufig in kompakten Instrumentalgattungen wie dem Streichquartett stattfanden. Um unsere Hypothese zu testen, verwenden wir die Instrumentierungsanmerkungen und berechnen eine vokale sowie eine instrumentale Evolutionskurve (Abb. 9). Als Nachteil des CAC stellen wir eine unausgewogene Situation fest (vgl. Tab. 2), die dazu führt, dass für die Instrumental-EC nur eine geringe Anzahl an Werken zur Verfügung steht. Dennoch beobachten wir eine klare Tendenz, die unserer Hypothese widerspricht: Vokalmusik scheint in den meisten Zeiträumen komplexer zu sein als Instrumentalmusik. Insbesondere ist der Abstand für die lokale Komplexität groß (gestrichelte Linien). Dies kann auf eine interessante Beobachtung hinweisen, könnte aber auch einen technischen Grund haben. Unsere Chromagramme basieren auf einem Signalverarbeitungsansatz, der aus Audioaufnahmen extrahierte Frequenzkomponenten den zwölf Tonhöhenklassen zuordnet. Bei aufgenommener Vokalmusik führt dieser Prozess häufig zu erheblichen Artefakten, da die Tonhöhenstabilität viel geringer ist als bei Instrumenten und Effekte wie Vibrato oder Portamento die Chromagramme erheblich verwischen. Dies kann zu quasichromatischen Artefakten führen, die die Komplexitätsmessungen sogar lokal verschieben.

Um solche Probleme zu überwinden, sind neuere Chroma-Extraktionsstrategien auf Basis tiefer neuronaler Netze sehr vielversprechend, da sie sich bei der Ableitung von Toninformationen aus Gesangsaufnahmen durch die Reduzierung typischer Artefakte als erfolgreich erwiesen haben (Weiß und Peeters 2021).

## 4. Fazit

Diese Erkenntnisse korpusbasierter Strategien zeigen das hohe Potenzial von Audioaufnahmen für die Forschung in der Computational Theology. In der Zusammenfassung dieses Kapitels betonen wir die Herausforderungen, die mit der Analyse roher Tonaufnahmen verbunden sind. Eine genaue Erkennung von Symbolen (Transkription) ist schwierig und oft nicht realisierbar. Dennoch gibt es computergestützte Ansätze zur Ableitung weicher, probabilistischer Beschreibungen, die stark mit dem menschlichen Verständnis und der Intuition korrelieren. In zwei Fallstudien haben wir gezeigt, wie solche Ansätze zur Untersuchung groß angelegter Musikwerke (wie der *Johannespassion*) oder ganzer Korpora aus mehreren Jahrhunderten (wie beim CAC) eingesetzt werden können. Mit der rasanten Entwicklung von Verarbeitungstechniken auf Basis tiefer neuronaler Netze ist in naher Zukunft mit einer erheblichen Verbesserung solcher Strategien zu rechnen. Um jedoch verlässliche Erkenntnisse zu theologischen und geisteswissenschaftlichen Fragestellungen zu gewinnen, bleibt ein interdisziplinärer Dialog zwischen Expert\*innen beider Fachgebiete (Theologie und Informatik) unerlässlich.

## Danksagungen

Dieses Kapitel basiert zu einem großen Teil auf früheren Arbeiten zusammen mit Meinard Müller und Kolleg\*innen der AudioLabs Erlangen. C. W. dankt dem Carus-Verlag Stuttgart (Johannes Graulich und Ester Petri) für die Ermöglichung des Studiums des Carus-Audiokorpus.

## Literaturverzeichnis

- Bäckström, T., Räsänen, O., Zewoudie, A., Zarazaga, P. P., Koivusalo, L., Das, S., Melado, E. G., Mansali, M. B., Ramos, D., Kadiri, S., & Alku, P. (2022) Introduction to Speech Processing. 2. Aufl. [Computersoftware]. *Zenodo*. <https://doi.org/10.5281/zenodo.6821775> [zuletzt aufgerufen am 21.06.2024].
- 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. <https://doi.org/10.1016/j.patrec.2009.09.014> [zuletzt aufgerufen am 21.06.2024].
- 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. <https://doi.org/10.1145/3224204> [zuletzt aufgerufen am 21.06.2024].
- Benetos, E., Dixon, S., Duan, Z., & Ewert, S. (2019). Automatic music transcription. An overview, *IEEE Signal Processing Magazine*, 36, 20–30. <https://doi.org/10.1109/MSP.2018.2869928> [zuletzt aufgerufen am 21.06.2024].
- Calvo-Zaragoza, J., Hajič Jr., J., & Pacha, A. (2020). Understanding optical music recognition, *ACM Computing Surveys*, 53(4), 1–35. <https://doi.org/10.1145/3397499> [zuletzt aufgerufen am 21.06.2024].
- 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. <https://doi.org/10.1098/rsos.150081> [zuletzt aufgerufen am 21.06.2024].
- Müller, M. (2021). *Fundamentals of Music Processing. Using Python and Jupyter Notebooks*. 2. Aufl. Cham: Springer. <https://doi.org/10.1007/978-3-030-69808-9> [zuletzt aufgerufen am 21.06.2024].
- Nakamura, E., & Kaneko, K. (2019). Statistical evolutionary laws in music styles, *Scientific Reports*, 9, 0. S. <https://doi.org/10.1038/s41598-019-52380-6> [zuletzt aufgerufen am 21.06.2024].
- Scherbaum, F., Müller, M., & Rosenzweig, S. (2017). Analysis of the Tbilisi State Conservatory recordings of Artem Erkomaishvili in 1966. In *Proceedings of the 7<sup>th</sup> International Workshop on Folk Music Analysis (FMA)* (S. 29–36). Málaga.
- Smaragdīs, P. (2004). Non-negative matrix factor deconvolution. Extraction of multiple sound sources from monophonic inputs. In C. G. Puntonet & A. Prieto (Hrsg.), *Proceedings of the International Conference on Independent Component Analysis and Blind Signal Separation ICA* (S. 494–499). Berlin/Heidelberg: Springer. [https://doi.org/10.1007/978-3-540-30110-3\\_63](https://doi.org/10.1007/978-3-540-30110-3_63) [zuletzt aufgerufen am 21.06.2024].
- 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)* (S. 3465–3469). Graz. <https://doi.org/10.21437/Interspeech.2019-1873> [zuletzt aufgerufen am 21.06.2024].
- 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 19<sup>th</sup> International Society for Music Information Retrieval Conference (ISMIR)* (S. 416–423). Paris. <https://doi.org/10.5281/zenodo.1492439> [zuletzt aufgerufen am 21.06.2024].

- Weiß, Ch., Mauch, M., Dixon, S., & Müller, M. (2019). Investigating style evolution of Western classical music. A computational approach, *Musicae Scientiae*, 23, 486–507. <https://doi.org/10.1177/1029864918757595> [zuletzt aufgerufen am 21.06.2024].
- Weiß, Ch., & Müller, M. (2014). Quantifying and visualizing tonal complexity. In *Proceedings of the Conference on Interdisciplinary Musicology (CIM 14)* (S. 184–187). Berlin. URL: [https://www.audiolabs-erlangen.de/content/05\\_fau/professor/00\\_mueller/03\\_publications/2014\\_WeissMueller\\_TonalComplexity\\_CIM.pdf](https://www.audiolabs-erlangen.de/content/05_fau/professor/00_mueller/03_publications/2014_WeissMueller_TonalComplexity_CIM.pdf) [zuletzt aufgerufen am 21.06.2024].
- Dies. (2021). Computergestützte Visualisierung von Tonalitätsverläufen in Musikaufnahmen. Möglichkeiten für die Korpusanalyse. In S. Klauk (Hrsg.), *Instrumentalmusik neben Haydn und Mozart. Analyse, Aufführungspraxis und Edition* (S. 107–130). Würzburg: Königshausen & Neumann [= *Saarbrücker Studien zur Musikwissenschaft*, 20].
- Dies. (2023). Studying tonal evolution of Western choral music. A corpus-based strategy. In A. Šela, F. Jannidis & I. Romanowska (Hrsg.), *Proceedings of the Computational Humanities Research Conference* (S. 687–702). Paris. URL: <https://ceur-ws.org/Vol-3558/paper7862.pdf> [zuletzt aufgerufen am 21.06.2024].
- 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)* (S. 754–761). Online: Zenodo. <https://doi.org/10.5281/zenodo.5624358> [zuletzt aufgerufen am 21.06.2024].
- White, Ch. W. (2013). *Some Statistical Properties of Tonality, 1650–1900* [Diss.]. New Haven, Connecticut: Yale University.

## Bildnachweise

Die Abbildungen 1–9 wurden vom Autor selbst generiert. Die Abb. 1 und 2 wurden hier erstveröffentlicht. Die übrigen wurden wie folgt in früheren Publikationen bereits publiziert:

Abb. 3: Weiß & Müller 2023, 691.

Abb. 4: Weiß et al. 2019, 8.

Abb. 5: Weiß & Müller 2023, 694.

Abb. 6: Weiß & Müller 2023, 695.

Abb. 7: Weiß et al. 2018, 417.

Abb. 8: Weiß & Müller 2023, 699.

Abb. 9: Weiß & Müller 2023, 699.