Transitions of Tonality: A Model-Based Corpus Study

Thèse N° 9808

Présentée le 16 décembre 2019

au Collège des humanités
Digital and Cognitive Musicology Lab
Programme doctoral en humanités digitales

pour l'obtention du grade de Docteur ès Sciences

par

Fabian Claude MOSS

Acceptée sur proposition du jury

Prof. D. Gatica-Perez, président du jury
Prof. M. A. Rohrmeier, Dr M. F. J. Neuwirth, directeurs de thèse
Prof. G. Lauer, rapporteur
Prof. D. Shanahan, rapporteur
Prof. R. West, rapporteur

2019

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

The better Musick is known and understood, the more it will be valued and esteemed
— Christoper Simpson, *A Compendium of Practical Musick*, London, 1665.

Et hätt noch immer jod jejange.
— *Kölsches Grundgesetz*, Article 3.

To everyone who walked a mile or two with me.

Acknowledgements

Looking back on the past five years I cannot but feel overwhelmed by the enormous support, inspiration, and encouragement I have received from so many people.

First and foremost I want to express my deep gratitude towards my supervisor Martin Rohrmeier who, over the years, has time after time proven to be a beacon of academic rigor and creativity. Maybe the most important lesson he has taught me is that the scientific study of music must be rooted in and has to be nourished by a deep love for the subject. I cannot imagine a place more ideally suited for my personal academic interests than his Digital and Cognitive Musicology Lab (DCML) with its inherently interdisciplinary focus. In the beginnings of my doctoral studies I did anticipate that I would learn a lot about music but I would have never guessed to also learn so much about psychology, linguistics, computer and data science, probability theory, cognitive science, and epistemology, all of which are in one way or the other represented in our daily work at the DCML. I am equally grateful for the discussions with, and the feedback and encouragement I received from my co-supervisor Markus Neuwirth. His endurance and meticulous work have been and continue to be exemplary to me.

I appreciate the time and effort all members of my doctoral committee invested into reading and evaluating my thesis: Daniel Gatica-Perez (EPFL), Robert West (EPFL), Gerhard Lauer (Universität Basel), and Daniel Shanahan (The Ohio State University). Thank you for your feedback and helpful comments.

During my doctoral studies at Technische Universität Dresden (2015–2017) and École Polytechnique Fédérale de Lausanne (2017–2019), I was constantly surounded by a marvelous group of people that I am happy to call not only colleages but friends, in particular (and in alphabetical order) Christine Ahrends, Fernando Bravo, Christoph Finkensiep, Daniel Harasim, Johannes Hentschel, Robert Lieck, Wolfgang Mende, Jessica Pidoux, Tudor Popescu, Christoph Wald, Leona Wall, and Jonathan Walter. The often passionate and intense discussions—ranging from music, art, and society to academic and methodological issues—helped me to shake up and view in a new light some old beliefs that I had become too fond of to question them. It allowed my to acknowledge different perspectives and opinions on matters where I wrongly assumed to already know it all. Doing a PhD—I have come to realize—is not about having all the answers but about learning how to ask the right questions.

I am also glad to have been influenced by so many dear colleagues and mentors inside and outside of my own field that it is impossible for me to name and thank them all. Nonetheless, I want to express my special indebtedness to Johannes Schild who ignited my curiosity for 19th-century harmony; to Thomas Noll who fostered my interest in mathematical music theory during my exchange stay at ESMUC, Barcelona; to David Pesetsky for his diligent supervision during my research stay at MIT and the inspiring discussions on the syntax of music; to Timothy J. O'Donnell from whom I learned most of what I know about generative modeling and Bayesian statistics; and to Uwe Seifert who brought me in 'first contact' with music cognition and encouraged me to pursue a PhD from the start.

I am happy to have met and befriended Colin Aitken, Joshua Albrecht, Rie Asano, Vishnu Bachani, Scott Beveridge, Oren Boneh, Minna Karppanen, Harriet Lau, Mora Maldonado, Uri Rom, Oliver Schwab-Felisch, Willian Fernandes Souza, and Florina Speth.

I am also extremely grateful to my colleagues in the administration who always had an open ear and were tremendously helpful in the myriad issues that arise along the way, in particular Carolin Thiele (DMCL, TUD), Kathleen Martons Collins (Digital Humanities Institute, EPFL), Tatiana Dubi (Digital Humanities Doctoral School, EPFL), and Aurélie Nicoulaz and Estelle Quinton (DCML, EPFL).

Parts of this project could not have been realized without the help of others. I want to thank the annotators who produced the analyses studied in the "Mesoanalysis" part, namely Wendelin Bitzan, Alexander Faschon, Daniel Grote, Johannes Hentschel, Adrian Nagel, and Markus Neuwirth, as well as the student assistants at DCML who transcribed some of the scores for the analyses in the "Macroanalysis" part, namely Laurène Tribolet, Dennis Steffen, Albane Descombes, Dylan Ifergan, as well as Timothy Loayza for their great of work.

I want to thank the Konrad Adenauer Foundation for funding my doctoral studies from 2016–2017, the Graduate Academy of Technische Universität Dresden for funding my research visit to the Massachussets Institute of Technology in 2016, the *Zukunftskonzept* (ZUK 64), funded by the *Exzellenzinitiative* of the Deutsche Forschungsgemeinschaft for the financial support from 2015–2017, as well as Claude Latour for supporting the Digital and Cognitive Musicology Lab through the Latour Chair in Digital Musicology at EPFL.

I am grateful to my family for always giving me the freedom I needed. Most of all, I feel incapable of expressing how grateful I am towards all my friends at home and all over the world, who I felt were always there for me—despite the long distances to Barcelona, Dresden, Boston, or Lausanne. You incessantly provided me with the right balance of motivation and sometimes badly needed distraction. I would not want to miss your affection for a single day. It pains me to address you as a group rather than individually, but I am afraid that I have neither room nor strength left. Each one of you, in her and his own very unique ways, is dear to me and I am happy to have you in my life. You know it.

Lausanne, November 29, 2019 Fabian C. Moss

Abstract

Tonality has been the cornerstone of Western music-theoretical discourse for centuries. This study addresses the subject, using traditional music analysis, data-driven corpus methods, and computational models, concentrating on historical changes of tonality with a particular focus on the 19th century. The thesis engages three analytical levels of increasing scope—micro, meso, and macro—and is thus located between the poles of the particular and the general.

The micro-level presents a detailed analysis of Franz Liszt's *Sonetto 47 del Petrarca*, S. 161, no. 4 (1858), in order to illustrate compositional innovations testifying to the radical changes in tonality within the 19th century. The analysis exemplifies how these novelties permeate musical compositions in that period, and also expose the benefits and limitations of manual music analysis.

The meso-level examines a corpus of harmonic annotations of pieces by Beethoven, Schubert, Chopin, Liszt, Dvořák, Grieg, Tchaikovsky, Debussy, and Medtner, containing over 75,000 chord symbols. It presents a comprehensive model for the analysis of chord symbols in large corpora in order to study chords and the progressions between them. Whilst the individual composers' chord vocabularies vary considerably—paying tribute to idiomatic usages of harmony—it is shown that the overarching similarities of the chord distributions point to similarities in their harmonic language that surpass individual traits and that can be modeled by Zipf's and Heaps's laws. An entropy-based method is presented to systematically study the effect of certain features on chord prediction, revealing that suspensions are the strongest predictors. The study shows that chord progressions are largely asymmetrical and proceed mostly by fifths; however, third-based progressions become increasingly prevalent within the studied period.

The macro-level explores a corpus of nearly 3 million notes in more than 2000 pieces created by 75 composers, comprising a historical range of approximately 600 years. The encoding of the data distinguishes enharmonically equivalent notes, hence providing a larger note vocabulary than most previous approaches in empirical music research. A Principal Component Analysis (PCA) shows that the line of fifths can be derived from the co-occurrence as well as the co-evolution of tonal pitch-classes. Moreover, the hierarchical topic model known as Latent Dirichlet Allocation (LDA) is used to discover latent tonal profiles. These largely correspond to distributions on contiguous line-of-fifths segments and moreover demonstrate the elevated roles of fifths as well as major and minor thirds as intervals between the most frequent notes. This motivates to model pieces as distributions on the Tonnetz. To that end, a new model, the

Tonal Diffusion Model (TDM), is introduced. The results are obtained by fitting the model to the corpus and exhibit two trends. Over the entire historical period under consideration, notes are primarily distributed along the fifths axis of the Tonnetz. Furthermore, 19th-century composers also explore the major and minor thirds axes of the Tonnetz, extending their compositions in ever farther regions.
The diverse methodology in this study provides quantitatively grounded insights from a range of perspectives, bridging the fields of music theory, computational musicology, mathematical modeling, and the digital humanities.

Keywords: Tonality, Corpus Studies, Computational Musicology, Music Theory, Digital Humanities

Résumé

Pendant des siècles, la tonalité a été le fondement du discours théorique pout la musique occidental. Ce travail de thèse traite de ce sujet en s'appuyant sur l'analyse musicale traditionnelle, sur des méthodes guidées par un corpus de données et sur des modèles computationnels. Il se concentre sur les changements historiques de la tonalité, avec un accent particulier sur le XIXe siècle. La thèse comporte trois niveaux analytiques d'échelle croissante (micro, méso et macro) se situant ainsi entre le particulier et le général.

Le niveau micro présente une analyse détaillée du *Sonetto 47 del Petrarca*, S. 161, no. 4 (1858), de Franz Liszt, afin d'illustrer les innovations compositionnelles qui témoignent des changements radicaux de tonalité au cours du XIXe siècle. L'analyse met en lumière comment ces nouveautés imprègnent les compositions musicales de cette période, et exposent également les avantages et les limites de l'analyse musicale manuelle.

Le niveau méso étudie un corpus de plus de 75,000 annotations harmoniques d'oeuvres de Beethoven, Schubert, Chopin, Liszt, Dvořák, Grieg, Tchaïkovski, Debussy, et Medtner. Il présente un modèle général pour l'analyse des symboles d'accords dans des grands corpora avec le but d'étudier les accords et les progressions entre ceux-ci.

Bien que les vocabulaires d'accords individuels des compositeurs varient considérablement (s'adaptant aux usages idiomatique de l'harmonie), les similitudes globales des distributions d'accords indiquent des ressemblances dans leur langage harmonique. Celles-ci dépassent les traits individuels et peuvent être modelées par les lois de Zipf et de Heaps. Une méthode basée sur l'entropie est ici présentée pour l'étude systématique de l'effet de certaines caractéristiques dans la prédiction des accords. Elle révèle que les suspensions sont les plus forts prédicteurs. Les résultats de l'analyse montrent que les progressions d'accords sont largement asymétriques et se développent principalement par quintes. La fréquence des progressions basées sur des tierces augmentent néanmoins progressivement durant la période étudiée.

Le niveau macro étudie un corpus de près de 3 millions de notes dans plus de 2,000 oeuvres composées par 75 compositeurs, sur une période historique d'environ 600 ans. L'encodage des données distingue les notes équivalentes de type enharmonique. Il fournit ainsi un vocabulaire de notes plus large que la plupart des approches de la recherche empirique musicale. Une analyse en composantes principales (ACP) montre que la ligne de quintes peut être déduite de la co-occurrence ainsi que de la co-évolution des classes de hauteur tonale. De plus, le Topic Model connu sous le nom *Latent Dirichlet Allocation* (LDA) est utilisé pour découvrir des profils tonaux latents. Ceux-ci correspondent en grande partie à des répartitions sur des segments contigus de la ligne des quintes. Ils montrent en outre les rôles élevés

des quintes et des tierces majeures et mineures comme intervalles entre les notes les plus fréquentes. Ce constat mène à une modélisation des oeuvres comme des distributions sur le Tonnetz. À cette fin, un nouveau modèle, le *Tonal Diffusion Model* (TDM), est introduit. Les résultats sont obtenus en ajustant le modèle au corpus et présentent deux tendances. D'une part, sur l'ensemble de la période historique considérée, les notes sont principalement réparties le long de l'axe des quintes du Tonnetz. D'autre part, les compositeurs du XIXe siècle explorent également les axes des tierces majeures et mineures du Tonnetz, étendant ainsi leurs compositions à des régions toujours plus éloignées.
La méthodologie mixte de cette étude apporte des connaissances empiriques quantitatives à partir d'un large éventail de perspectives, réunissant les domaines de la théorie musicale, de la musicologie computationnelle, de la modélisation mathématique et des humanités digitales.

Zusammenfassung

Tonalität hat über die Jahrhunderte eine zentrale Stellung im musiktheoretischen Diskurs eingenommen. Die vorliegende Arbeit behandelt dieses Thema unter Berücksichtigung traditioneller Musikanalyse, datenbasierter Korpusmethoden, so wie von Computermodellen im Hinblick auf historische Veränderungen von Tonalität unter besonderer Berücksichtigung des 19. Jahrhunderts. Diese Studie ist in drei analytische Levels aufsteigenden Umfangs gegliedert – Mikro, Meso und Makro – und steht somit im Spannungsfeld zwischen dem Partikulären und dem Allgemeinen.

Das Mikrolevel präsentiert eine detaillierte Analyse von Franz Liszts *Sonetto 47 del Petrarca*, S. 161, no. 4 (1858), zur Veranschaulichung der kompositorischen Innovationen, die von den radikalen Veränderung der Tonalität im 19. jahrhundert Zeugnis ablegen. Die Analyse belegt exemplarisch wie diese Neuheit die Kompositionen dieser Zeit durchdringen und zeigt zudem auch die Vorzüge und Grenzen manueller Musikanalyse auf.

Das Mesolevel analysiert einen Korpus von mehr als 75.000 harmonischen Annotationen von Stücken von Beethoven, Schubert, Chopin, Liszt, Dvořák, Grieg, Tschaikowsky, Debussy und Medtner. Ein umfassendes Modell zur Analyse von Akkordsymbolen in großen Korpora wird vorgestellt und auf Akkorde und Progressionen von Akkorden angewandt. Während die verschiedenen Komponisten große Differenzen im Hinblick auf die verwendeten Akkorde vorweisen und dadurch ihren idiomatischen Gebrauch von Harmonik demonstrieren, so wird auch gezeigt, dass Ähnlichkeiten in den Verteilungen der Akkorde auf Gemeinsamkeiten in ihrer harmonischen Sprache verweisen, welche mit Hilfe von Zipfs und Heaps' Gesetzen modelliert werden können. Eine auf Entropie basierende Methode wird vorgestellt, um den Einfluss bestimmter Akkordeigenschaften auf die Vorhersage von Akkorden systematisch zu untersuchen. Es wird gezeigt, dass Vorhalte die stärksten Prädiktoren sind. Diese Arbeit zeigt überdies, dass Akkordprogressionen überwiegend asymmetrisch sind und vornehmlich in Quinten fortschreiten, aber auch, dass Schritte in Terzen im Untersuchungszeitraum mehr und mehr zunehmen.

Im Makro-Teil dieser Arbeit wird ein Korpus von etwa drei Millionen Noten in über 2000 Stücken von 75 Komponisten untersucht, welcher insgesamt einen Zeitrahmen von fast sechshundert Jahren abdeckt. Die Datenkodierung unterscheidet zwischen Tönen, welche enharmonisch äquivalent sind und stellt demnach ein größeres Tonvokabular zur Verfügung als in den meisten empirischen Musikstudien. Mittels *Principal Component Analysis* (PCA) wird gezeigt, dass die Quintenlinie aus den Verteilungen von Tönen und ihrer Ko-Evolution hergeleitet werden kann. Zudem wird ein Topic Model namens *Latent Dirichlet Allocati-*

on (LDA) verwandt, um latente Tonprofile zu entdecken. Es stellt sich heraus, dass diese zusammenhängenden Segmenten auf der Quintenlinie entsprechen und zudem die hervorgehobene Bedeutung von Quinten sowie großer und kleiner Terzen zwischen den häufigsten Tönen betonen. Dies motiviert die Modellierung von Tonverteilungen auf dem Tonnetz. Dazu wird ein neues Modell vorgestellt, das *Tonal Diffusion Model* (TDM). Die Ergebnisse weisen zwei Tendenzen auf. Im gesamten historischen Untersuchungszeitraum sind Töne primär entlang der Quintenachse des Tonnetzes verteilt. Darüber hinaus explorieren Komponisten des 19. Jahrhunderts auch die Groß- und Kleinterzachsen des Tonnetzes und breiten ihre Kompositionen so in immer weitere Regionen aus.

Die Methodenvielfalt dieser Arbeit liefert quantitative Einsichten aus einer Reihe von Blickwinkeln und baut Brücken zwischen Musiktheorie, Computational Musicology, mathematischer Modellierung, und den Digital Humanities.

Contents

List of Figures

List of Tables

List of Examples

List of Abbreviations

This list specifies the acronyms and abbreviations used throughout the thesis and in the overview of the data sources in the tables in the appendix.

ABC Annotated Beethoven Corpus

BWV Bach-Werkverzeichnis

CCARH Center for Computer Assisted Research in the Humanities

CPDL Choral Public Domain Library

D. Deutsch catalogue (Schubert's works)

DB Daniel Bernhardsson

DCML Digital and Cognitive Musicology Lab

DFT Discrete Fourier Transform

ELVIS Electronic Locator of Vertical Interval Successions

HWV Händel-Werkverzeichnis

ICA Independent Component Analysis

IMSLP International Music Score Library Project

K Kirkpatrick catalogue (Scarlatti's works)

KPM Kantorei Pro Musica

KV Köchel-Verzeichnis (Mozart's works)

L. Lesure catalogue (Debussy's works)

LDA Latent Dirichlet Allocation

LOWESS Locally Weighted Scatterplot Smoothing

M. Marnat catalogue (Ravel's works)

m. measure

MCMC Markov Chain Monte Carlo

MDS Multidimensional Scaling

MIDI Musical Instrument Digital Interface

mov. movement

MS MuseScore

NLP Natural Language Processing

no. numero

Op. Opus

OSLC Open Score Lieder Corpus

PCA Principal Component Analysis

RCT Rameau Catalogue Thématique

RISM Répertoire International des Sources Musicales

RNA Roman Numeral Analysis

S. Searle catalogue (Liszt's works)

SE Spectral Embedding

SLSQP Sequential Least Squares Programming

Sz. Szőllősy catalogue (Bártok's works)

t-SNE t-distributed Stochastic Neighbor Embedding

TDM Tonal Diffusion Model

TPS Tonal Pitch Space

TTR type-token ratio

WoO Werk ohne Opusnummer (work without Opus number)

YCAC Yale–Classical Archives Corpus

Introduction

Music has engaged the bodies and riddled the minds of humans for millenia. Its origins can be traced back at least 40,000 years (Conard et al., 2009), and theoretical reflection on the intricate relations between elements of music is documented since antiquity (Christensen, 2002). While ‘music’ seems impossible to define, one of its most general characterizations is ‘sound organized in time’. Throughout the ages, one particular aspect of this organization has undeniably occupied center stage: the coordination of pitches, or *harmony* (Damschroder, 2008). In the 19th century, music theorists become aware that this organization is not only guided by mathematical and hence universal principles but is also imbued with cultural, in particular historical factors (Choron, 1810; Fétis, 1844; Polth, 2016). They understand harmony not anymore only as a given system that composers discover and elaborate but also acknowledge that creativity, innovation, and invention can lead to different systems of harmonic organization. Systems for the organization of musical tonal material are nowadays commonly subsumed unter the term ‘tonality’ (Hyer, 2001), which is the principal subject of this thesis. It is examined under a primarily historical perspective by studying several corpora of Western classical music with a certain emphasis on the 19th century.

The title of this study—*Transitions of Tonality*—is inspired by Samson’s *Music in transition* (1977). It alludes to the fact that tonality is not only a description of the relations between tones, but also that these relations are in constant change. While a purely mathematical description might fall short of acknowledging these dynamics, an exclusively historical account is in danger of neglecting constants which result from physical, psychoacoustical, and cognitive invariances. This study embraces both perspectives, placing itself under the interdisciplinary umbrellas of the Digital Humanities in general and computational musicology in particular. The subtitle *A Model-Based Corpus Study* points to the epistemological stance taken here, namely that the creation of musical corpora enables to address the old issue of tonality on a larger scale with modern methods, and that the study of the historical development of tonality and its music theoretical implications can substantially benefit from computational modeling.

Structure of the thesis. First, the notion of tonality is introduced in detail, discussing the concept and its origins as well as providing a definition and an account of the two main narratives regarding its historical development (Chapter 1). The discussion then proceeds to the scope and aims of computational musicology, reviews a number of corpus studies relevant to the present work, and discusses difficulties and challenges that musical corpus research faces (Chapter 2). Chapter 3 introduces the basic representation for musical pieces that is adopted here and presents the corpora constituting the empirical basis for the subsequent analyses.

The analytical part of this thesis is divided in three levels of increasing scope—Microanalysis, Mesoanalysis, and Macroanalysis—insinuating a tight connection to a number of recent publications in various domains within the Digital Humanities (e.g., Jockers, 2013; Moretti, 2013; Cook, 2014; Weitin, 2017; Underwood, 2019). The increasing scale of these three levels pertains to several aspects. First, the corpora studied in these parts cover increasing *historical ranges*. The Microanalysis studies a single piece, the Mesoanalysis covers a corpus of 19th-century composers, and the Macroanalysis studies musical pieces from the 14th to the 20th century. Second, the different levels refer to the *representation* of the data. The piece analyzed on the micro level is given as a printed score. Although optical music recognition (the automated inference of musical structure from sheet music) has made great advances in recent years (Rebelo et al., 2012; Calvo-Zaragoza et al., 2019), the subtle inferences a proficient musician can draw from them are still out of reach. The harmonic annotations on which the Mesoanalysis part is based could be seen as a bridge between manual and computational analyses. The mathematically well-defined syntax to which all of them conform was developed in order to be parseable by a computer but at the same time be readily understood by humans who are trained in music theory. The representation of the pieces in the Macroanalysis part as bags-of-notes (see Section 3.2) is the most abstract one used in this thesis and also the most remote representation from both printed music and chord symbols. For this reason, the algorithmic analysis of this dataset is inevitable. Closely related to this point is the issue of the *methodology* used. The manual analysis on the micro level is based on traditional and recent music theoretical concepts that involves reduction and interpretation and hence potentially subjective decisions by the author. The meso level of analysis studies a large corpus of harmonic annotations. While these annotations have been manually entered by music theory experts, the entire dataset is already too large for manual analysis and requires computational methods. The more than 2000 pieces of the corpus studied in the Macroanalysis part extend the analytical focus even more, making automated analyses not only suitable but necessary. A natural consequence of the different historical ranges, data representations, and methodological approaches (manual vs. computational) are the research questions that arise, as well as the specific methods that are used to address them.

Part II (Microanalysis) presents a thorough analysis of Franz Liszt's *Sonetto 47 del Petrarca*, S. 161, no. 4, that concentrates on the issue of tonality in this piece and exemplifies a traditional approach to music analysis (Chapters 4 and 4.5). It moreover provides initial observations about the tonality in the 19th century that are subsequently connected to the results in the following parts.

In Part III (Mesoanalysis), a model for corpora of harmonic labels is presented (Chapter 5) and applied to a set of annotated corpora of pieces by nine 19th-century composers, namely

Ludwig van Beethoven, Franz Schubert, Frédéric Chopin, Franz Liszt, Antonín Dvořák, Edvard Grieg, Piotr Tchaikovsky, Claude Debussy, and Nikolai Medtner. The analyses focus in particular on the distributions of chords (Chapter 6) and on transitions between them (Chapter 7). Chapter 8 summarizes the findings and contextualizes them against the backdrop of music theory by discussing the benefits of corpus-based methods.

Finally, Part IV (Macroanalysis) extends the scope of the analyses even further by studying a corpus covering more than 2,000 pieces spanning a historical range of almost 600 years. Chapter 9 presents the corpus, introduces the fundamental bag-of-notes model along with the representation of musical notes as tonal pitch-classes. On this basis, Chapter 10 demonstrates that imporant aspects of tonality can be inferred both directly from the occurrences of pitch classes in the corpus, as well from the tonal pitch-class co-evolution, a newly introduced concept. In Chapter 11, the corpus is studied under the Latent Dirichlet Allocation (LDA) model—one of the most commonly used topic models—and under the interpretation of topics as tonal profiles. Finally, Chapter 12 introduces a novel model, the Tonal Diffusion Model (TDM), to study distributions of tonal pitch-classes on the Tonnetz. This model is based on historical and cognitive considerations about tonal space and the results in this chapter allow to draw conclusions about the historical usage of tonal pitch-classes.

Part V concludes this study by drawing implications of the results for our understanding of tonality and outlines how future research may build upon the present work. This study contributes to computational musicology and musical corpus studies in that it presents a multi-level approach to music analysis, addressing the subject of tonality from a variety of perspectives and employing a wide range of methods.

Contributions. The main contribution of this thesis is threefold: First, it demonstrates how music theoretical questions can be reformulated and formalized in order to study tonality with quantitative methods. Second, it broadens the state-of-the art in computational musicology by applying these methods to two new large corpora. The corpus of chord labels of compositions by 19th-century composers (Part III, "Mesoanalysis") focuses on a musical period that has so far not been studied comprehensively in its own right, and the corpus of pieces in XML format (Part IV, "Macroanalysis") transcends previous research in its historical extend as well as in the representation of notes as tonal pitch-classes, entailing a much larger note vocabulary than all studies that are based on MIDI data. Third, this thesis clearly demonstrates that historical and empirical musicology are not necessarily mutually exclusive. On the contrary, it is the author's deep conviction that both approaches substantially benefit from each other. Empirical research can gain a lot from historical considerations, e.g. in modeling decisions, in the formulation of hypotheses and research questions, as well as in the interpretation and contextualization of results. Historical questions, on the other hand, can not only be investigated quantitatively, but the necessity to develop formal concepts and models for empirical research is vital to resolve ambiguities and to study larger amounts of historical sources than one could do with traditional methods.

Apart from these more general contributions, this thesis introduces several new concepts and measures that might prove useful for future research as well. Among those are a comprehensive model of chord morphology (based on Neuwirth et al., 2018, Section 5.2) a list of chords shared by all 19th-century composers in the sample (Section 6.1), and measures for the richness, typicality, and conformity of a corpus with respect to a reference corpus (Section 6.2). Further, it is shown how the normalized conditional entropies of chord progressions can be used to determine the effect of certain chord features on chord prediction (based on Moss et al., 2019b, Section 7.1). With respect to notes, it is shown that their representation of tonal pitch-classes allows to infer important aspects of the underlying tonal space, namely the line-of-fifths (Section 10.1). The concept of tonal pitch-class coevolution is introduced and used in order to reveal large-scale historical trends (Section 10.3), Finally, note distributions are visualized on the Tonnetz, revealing characteristic patterns which, in turn, motivate a novel model, the Tonal Diffusion Model (based on Lieck et al., in preparation, Section 12.3).

Musical notation. Throughout the thesis, capital letters are used to describe pitch classes, e.g. C, F, and D, potentially with an appropriate number of flats (♭) or sharps (♯), e.g. A♯, D♭♭, and G♯♯.[1] Keys are specified with the name of the tonic note followed by the mode, e.g. G major, E minor, or C♭ major. The notation of chords is explained in great detail in Section 5.2 because they are in the center of the analyses in Part III.

Software. This thesis relies heavily on open-source software, namely the Python programming language, in particular the numerical and scientific computing libraries *NumPy* (van der Walt et al., 2011) and *SciPy* (Jones et al., 2001), as well as on the packages *pandas* (McKinney, 2010) for dealing with tabular data and *statsmodels* (Seabold and Perktold, 2010) for the calculation of certain regression curves. The plots have mainly been generated using *Matplotlib* (Hunter, 2007), and occasionally with *seaborn* (Waskom et al., 2017) and *pitchplots* (Moss et al., 2019a).

[1] The LaTeX-typesetting system does not allow for the proper sign for double-sharps so that they are expressed as two consecutive single-sharps.

Background Part I

1 Histories of tonality

> Few terms in music theory are more profound and more enigmatic than 'tonality.'
>
> Matthew Brown, *Explaining Tonality: Schenkerian Theory and Beyond.*

1.1 What is tonality?

'Tonality' is a concept that is difficult to define. It is often associated with other terms such as 'key', 'mode', 'scale', or 'harmony' but these are not less problematic to delineate. For instance, what is understood by 'mode' depends on the musical context, e.g. the historical period. Modes in Renaissance motets are different from modes in Mozarts Sonatas which, in return, differ from the usage of the term by Messiaen (1944). The notion of 'tonality' itself is relatively young. It originated in the 19th century (Christensen, 2019), supposedly first introduced by Alexandre-Étienne Choron (1810) and then popularized by François-Joseph Fétis who defined tonality as the "necessary relations between the notes of the scale" (Fétis, 1844, p. iii). Hugo Riemann understands tonality as "the particular meaning which chords receive through their relationship to one principal clang, that of the *tonic*" (Riemann, 1896, p. 796). In contrast to Fétis, Riemann's conception is built around chords, not notes. Summarizing his elaborate review of Fétis's and Riemann's positions on tonality, Dahlhaus (1968, pp. 9–18) recognizes that the question whether tonality is exclusively given by the relation of chords that are organised around a central tonic (according to Riemann) or whether it can also by constituted by other relation of tones (according to Fétis) is an open problem (Dahlhaus, 1968, p. 14).

There are numerous concurrent usages of the term 'tonality'. For instance, Zbikowski describes it as "a way of understanding musical organization" (Zbikowski, 2002, p. 76), putting emphasis on the cognitive-perceptual aspect of music. One common usage in musicology and music theory is to designate with 'tonal music' the music composed in the so-called Common-Practice period (Piston, 1948, ca. 1650–1900). A comprehensive account of the core components of tonality in this understanding is given by Samson (1977, p. 2) and is worth being restated in some length:

> In its most fully realized form classical tonality may be distinguished by its use of two modes only, each of them transposable to any pitch level, and by its total clarification of the relationships existing between pitches grouped around a single tonic [...]. The central harmonic unit of classical tonality is the major triad, its fundamental tonal-harmonic progression I—V—I. However widely ranging the harmonic movement within this fundamental progression, it takes place against a background of hierarchical relationships between diatonic triads grouped around a tonic triad and between secondary tonal regions grouped around a central tonality.

Many elements of this definition are crucial for this thesis, in particular for the definition and the study of chords and chord progressions in Part III. In this understanding, 'tonality' is synonymous with 'classical tonality' or 'harmonic tonality', the tonal language of the Common-Practice period, and it is used to distinguish it from earlier 'modal' and later 'atonal' music. This conception is problematic in several respects. First of all, neither of the three periods thus defined are as homogenous as this distinction suggests. Moreover, it is problematic to anachronistically define both modal and atonal music as a negation of tonality, as 'not yet tonal' and 'not tonal anymore', pretending that the period from Bach to Brahms were somehow privileged. To avoid this partition, Tymoczko (2011) uses the term "Extended Common Practice" and includes the time periods both before and after the Common Practice period.

The musical language of the 19th century, after the Common Practice period, is often called 'extended tonality' (Schoenberg, 1969, pp. 76–113) or the 'Second Practice' (Kinderman and Krebs, 1996). One of its characteristics is supposedly the loss of centrality, a crucial component of the earlier harmonic tonality, which allowed that "[r]emote transformations and successions of harmonies were understood as remaining within the tonality" (Schoenberg, 1969, p. 76). While Schoenberg attributes these changes to extra-musical influences, above all to the composers' desire for expressiveness as well as repercussions from drama and poetry, Samson views the emergence of extended tonality as being caused by inner-musical factors, in particular the increasing usage of chromaticism: "It is a law of the evolution of harmony that notes which had originally functioned as dissonant passing-notes or suspensions should eventually discover new contexts as self-sufficient harmonic resources" (Samson, 1977, p. 4). Of course, both views are compatible. They differ only with respect to what is seen as the cause and what as the effect. Did the aesthetic and philosophic demands of 19th-century Romantic composers urge them to search for new ways to express their musical ideas, or did the creative explorations and elaborations of composers gradually and continuously produce new sounds and relations between these sounds that were not yet associated with conventional meanings inherited from classical tonality and could thus serve new compositional purposes? Naturally, it is impossible to answer this question conclusively but it seems most likely that a complex interaction of both causal directions lead to the appearance of extended tonality. Moreover, the compositional innovations of the Romantic era are not confined to the 19th-century but are still prevalent in 20th-century and modern music, for instance in the highly chromatic harmony of Jazz (Tymoczko, 2011; Rohrmeier, accepted), some non-classical progressions in Rock (Temperley, 2018), 'very late' Romantic composers such as Richard Strauss, or the symphonic brainchilds of Wagner such as Erich Wolfgang Korngold, John Williams, and others in Hollywood film music (Lehman, 2014, 2018). Extended tonality thus does not only describe a historical period but also stylistic properties of writing music.

1.2 Two narratives

As diverse as the different meanings of 'tonality' are, so are the stances with respect to its historical development. Despite this variety, the many accounts for the history of tonality can largely be subsumed under two concurrent positions. On one hand a teleological one that posits that the development of tonality reached its peak in the first half of the 19th century—often associated with Beethoven—and that further developments merely contributed to its dissolution and decay, and on the other hand a view that sees a paradigmatic change in the harmonic innovations of the 19th century, giving rise to an entirely new tonal language.

The teleological position mainly relies on the understanding of 'tonality' as classical tonality. The transition to atonality in the early 20th century is then, in this view, a logical consequence of eroding tendencies of extended tonality, such as chromaticism and enharmonicism (Gauldin, 2004, p. 754), that ultimately lead to the dissolution of the concept of a closed musical work (Niemöller, 1991). The vocabulary used to describe this process is often drastic, ranging from "disintegration" to tendencies to "self-destruct" and "dissolve" the established tonality (Dahlhaus, 1989, p. 379). Some even perceive the changes in musical composition as the "end of tonal harmony" because certain compositions, such as the highly chromatic music by Liszt, "defy any attempt to produce systematic analysis" (de la Motte, 1991, p. 313). However, this speechlessness towards late 19th-century music can also be traced back to the inadequacy of the analytical arsenal that is unable to generalize to the new practices (Kopp, 2002, p. 1). A less radical view is advocated by Meyer (1989, p. 300) who speaks of the "weakening of tonal syntax" (his notion of Common Practice tonality). What happened in the 19th century were—according to Meyer—not fundamental changes of the musical language *per se* but a redistribution in terms of importances and roles, such as the centrality of the tonic, within the language.

A different view is articulated by Fétis (1844) who describes the history of tonality as a succession of different stages, each of which introduces new elements to the art of composition. He calls the tonal stage of his contemporaries the *ordre omnitonique* (Fétis, 1844, p. 184) in which composers have the liberty to modulate to any key. This newly acquired freedom is thus not understood as the end of tonality, but as something that has replaced the old tonality with something new (Schild, 2010, p. 314), a development that Polth (2016) considers to be a paradigm change.[1] Since not all composers follow these new trends, the possibilities for musical expression have become extended, allowing to write music in both the old or the new ways. Cohn (2012, 195ff.) calls this the “double syntax”. This perspective has generated a number of recent new music theoretical approaches. Two prominent theoretical frameworks for the analysis of 19th century music are Neo-Riemannian theory (Cohn, 1998) and Tonfeld theory (Haas, 2004). Part II (Microanalysis) draws mainly on these two methods and applies them for the analysis of Franz Liszt’s *Sonetto 47 del Petrarca*, S. 161, no. 4.

1.3 A compromise

Some of the tensions between the two narratives of the history of tonality can be resolved if one adopts a more general notion of tonality, for instance the one described by Hyer (2001, p. 583) who states that

> the term is sometimes used as an equivalent for what Rousseau called a *sistême musicale*, a rational and self-contained arrangement of musical phenomena: accordingly, Sainsbury, who had Choron translated into English in 1825, rendered the first occurrence of *tonalité* as a ‘system of modes’ before matching it with the neologism ‘tonality’. While tonality *qua* system constitutes a theoretical (and thus imaginative) abstraction from actual music, it is often hypostatized in musicological discourse, converted from a theoretical structure into a musical reality. In this sense, it is understood as a Platonic form or prediscursive musical essence that suffuses music with intelligible sense, which exists before its concrete embodiment in music, and can thus be theorized and discussed apart from actual musical contexts.

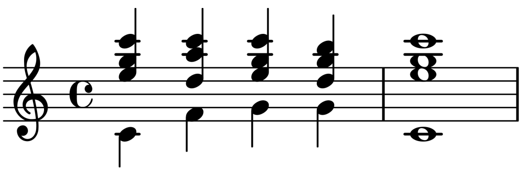
The central aspect in this verbose definition is that tonality is understood as a system of musical relations, or as Piston concisely puts it: “Tonality is the organized relationship of tones in music” (Piston, 1948, p. 29).[2] This broad conception of tonality is criticized by Dahlhaus (1968, pp. 17–18) because it is alledgedly redundant with ‘tone system’ and it would lead to “linguistic circuitousness” but his critique addresses only terminology and not the actual subject-matter. There is no reason not to generalize the notion of tonality to tone system, or “sistême musical” (Rousseau, 1768), and understand the tonality of the Common Practice period—‘harmonic tonality’ in Dahlhaus’s terms—as a concrete instantiation of it in a particular historical period. Moreover, the distinction in Hyer’s definition between the “prediscursive musical essence” existing prior to “its concrete embodiment in music” is most relevant for the present study since it can be translated into the distinction between formal

[1] “Systemwechsel” in the German original.

[2] His concept of tonality is, in fact, less general as this quote suggests. He also requires that tonality implies a central tone “with all other tones supporting it or tending toward it, in one way or another” and equates ‘tonality’ with ‘key’ (Piston, 1948, p. 29).

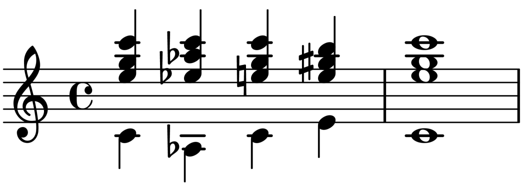
models of tonality, as described by mathematical structures, and the data extracted from pieces in musical corpora. This differentiation thus implicitly provides a link to empirical corpus research. By studying distributions of musical objects one can draw inferences about their mutual relations, the “rational and self-contained arrangement of musical phenomena”, i.e. about tonality.

From now on, we adopt this general notion of tonality as the relations between musical elements, e.g. notes or chords in a corpus. To illustrate this , consider Examples 1.1 and 1.2. The first example shows the chord sequence C Dm^7 G^6_4 G C.[3] Being a cadence, this chord sequence establishes the key of C majorand clearly identifies it as an instance of classical tonality since all chords are related to the final tonic chord, the C-major triad. All notes in this example are contained within a diatonic scale, so that the tonality established by the relations between these notes is a diatonic tonality.



Example 1.1 – Diatonic tonality established by a cadence in the key of C major.

The second example also presents a chord sequence, namely C A♭ C E C. The roots of these chords form a major third cycle, something that is unusual for classical tonality. The notes of these chords taken together form a hexatonic scale, thus establishing a hexatonic tonality,[4] but not establishing a key (Riemann, 1900, p. 1143). In other words, this chord sequence is an instance not of classical but of extended tonality.



Example 1.2 – Hexatonic tonality established by a non-diatonic triadic sequence; reproduced from Riemann (1900, p. 1143).

The general definition adopted here also allows to study the diachronic development of tonality since it is not dependent on any particular historical situation. The subsequent analyses in Part II (Microanalysis), Part III (Mesoanalysis), and Part IV (Macroanalysis) extend the above considerations and demonstrate how this definition of tonality can be used to study its historical evolution empirically. The specific aspects of tonality that are investigated are described and defined in detail on the following pages. Providing formal definitions, operationalizations,

[3] The notation ‘G^6_4’ indicates that the third chord in the sequence is a G-major chord with six-four suspension. A more comprehensive notation for chord symbols is introduced in Part III (Mesoanalysis).

[4] This conception is also related to Tymoczko’s definition of *macroharmony* that he defines as “the total collection of notes heard over moderate spans of musical time” (Tymoczko, 2011, p. 4).

and models for a variety of aspects of tonality is one of the main contributions of this work. However, before entering the analytical parts of this study, the next chapter reviews previous studies within computational musicology and discusses a number of challenges for musical corpus research (Chapter 2). Subsequently, we introduce the main representation and the used corpora for this study (Chapter 3).

2 Computational musicology

> Coordonner les sons dans des rapports qui développent des sensations et des idées plus ou moins vives, plus ou moins élévées, plus ou moins agréables, plus ou moins capables de réaliser les vues de l'artiste est l'objet de l'art; découvrir les lois de ces rapports, est celui de la science. L'harmonie est donc à la fois un art et une science.[1]
>
> François-Joseph Fétis, *Traité complet de la théorie et de la pratique de l'harmonie*

This study is situated within *computational musicology*, a rather novel approach to the study of music. Being relatively young, the boundaries of the field are not yet clearly defined, nor is its denomination. Other names include 'mathematical music theory', 'computer music', 'systematic musicology', 'music information retrieval', 'digital musicology', 'sound and music computing', and 'music informatics' (see Meredith, 2016, preface), which are tightly interconnected but not identical. According to Schaffer (2016), computational musicology is

> the use of computational methods and statistics to analyze musical structures (notes, chords, rhythms, etc., and patterns thereof). This combination of computation, statistics, and a domain of knowledge makes computational musicology a form of *data science*. However, due to the music theoretical aspect, computational musicology sits firmly within the *digital humanities* and focuses on the same kinds of questions as traditional humanities research.

He further lists four core activities of computational musicology, namely corpus research,

[1] "To arrange tones according to those relations that evoke sensations and ideas which are more or less vivid, more or less elevated, more or less pleasant, more or less able to express the intentions of the artist, that is the objective of the [musical] art; to discover the laws of these relations is the objective of science. Harmony is thus at the same time an art and a science." (Fétis, 1844, p. 1); translation by the author.

modeling, music encoding, and music information retrieval. While music information retrieval is not of great relevance for the present study, the other three areas are of core importance. All three analytical parts of this thesis are based on the study of corpora, and a number of models are employed in order to attain the results. Moreover, the study touches upon the issue of music encoding since the data used for the analyses comes in specific representations that have an impact on both the results and their interpretation.

2.1 Previous work

Early approaches in the computational study of music were based on small corpora, gathered and evaluated manually. For example, Budge (1943) conducted a study on chord frequencies in 18th and 19th century music. Later decades have witnessed discussions on the usefulness of probability and information theory to study musical styles (Meyer, 1957; Youngblood, 1958; Knopoff and Hutchinson, 1983; Snyder, 1990), but rather focused on theoretical and methodological aspects, and less on the study of particular corpora. Whereas the rampant development of computing capabilities and algorithmic methods during the 20th century gave rise to new research areas such as computational linguistics, and sparked an enormous body of research, the computational study of music has not seen a similar development (Neuwirth and Rohrmeier, 2016). One reason for this belated development has been the lack of suitable datasets in machine-readable formats. Another reason is the fact that musicology and music theory are traditionally situated within the humanities where quantitative approaches were and are not always received with great enthusiasm—to say the least (Morehen and Bent, 1979). As Underwood (2019, p. 66) puts it:

> Numbers are widely associated with a quest for objectivity in the physical sciences. [...] Humanists, for their part, have grown proud of unsettling that claim to objectivity. So any suggestion that numbers might illuminate human history immediately calls forth a well-rehearsed script, where math is expected to define objective patterns and humanists sigh that things are more complicated and depend on the observer's assumptions.

This phenomenon is not unique to the study of music, e.g. Moretti (2005), Jockers (2013), and Herrmann and Lauer (2018) make a similar case for literary studies. Only in recent decades have researchers begun to express larger interest in musical corpus studies, which has also accelerated the creation and publication of new datasets suitable for this particular purpose. This trend is reflected in a number of recent special issues in central journals for the empirical study of music, namely *Music Perception: An Interdisciplinary Journal*, Vol. 31, Nos. 1 and 3 (Temperley and VanHandel, 2013; VanHandel and Temperley, 2014), and *Empirical Musicology Review*, Vol. 11, Nos. 1 and 2 (Shanahan, 2016, 2017), as well as *The Oxford Handbook of Music and Corpus Studies* (Shanahan et al., forthcoming). Nonetheless, the amount of digitally encoded musical data is still small as compared to the more than one million resources captured by the database of the Répertoire International des Sources

Musicales[2] (RISM; Pugin, 2015). The following review of a number of recent studies provides an overview of the state-of-the-art in musical corpus research on harmony or tonality, without claiming that this survey is comprehensive or exhaustive, since the number of musical corpus studies has, luckily, grown fast and continues to do so.

To the date, the largest body of corpus studies by far is concerned with Western classical music. Within this scope, one can recognize several more specific research avenues. Studies focused on pattern recognition for the analysis of musical styles are provided, for instance, by Backer and Van Kranenburg (2005) and Bellmann (2011). Brinkman et al. (2014) use stylometry in order to distinguish the composition styles of Josquin and Bach in an author attribution study, and Yust (2019) employs the Discrete Fourier Transform (DFT) to extract information about the musical information in pitch-class distributions. The study of harmony in Western classical music more generally is another active area of study, aiming to put traditional music theory to the test and provide empirical support or criticism (e.g. Temperley, 2009). One of the central questions in this regard is the problem of key-finding (e.g. Temperley, 1999; Shmulevich and Yli-Harja, 2000; Temperley, 2002; Madsen and Widmer, 2007; Temperley and Marvin, 2008; Albrecht and Shanahan, 2013; Farbood et al., 2013, to name only a few). Some of these studies focus on a specific repertoire, e.g. Bach's chorales (Rohrmeier and Cross, 2008), or Haydn's (Cortens, 2014) and Beethoven's (Moss et al., 2019b) string quartets. Most recently, a number of studies have taken on the challenge to trace large-scale historical developments in Western classical music. Zivic et al. (2013), White (2014), and Weiß et al. (2018) have studied stylistic transitions from the 17th to the 20th century, Albrecht and Huron (2014) and Harasim et al. (submitted) have studied the historical evolution of the major and the minor mode in a wide historical timespan. Other approaches concentrate on very specific music theoretical questions, for instance the occurrence of parallel fifths in Bach's chorales (Fitsioris and Conklin, 2008), the discovery of certain harmonic cycles predicted by Neo-Riemannian theory (Bragg et al., 2011), the particular role of the sixth scale-degree (Brinkman and Huron, 2017), the counterpoint patterns in pieces by di Lasso (Schubert and Cumming, 2015), or the evaluation of a number of root-motion theories (Hedges and Rohrmeier, 2011). Others have addressed methodological issues such as the alleged transpositional invariance of musical keys (Quinn and White, 2017), or the regularities in frequency distributions of musical elements (Manaris et al., 2005; Zanette, 2006; Rohrmeier and Cross, 2008). The large amount of corpus studies on Western classical music—and within its scope the dominance of canonic repertoires—reveal a bias that computational musicology has inherited from traditional musicology and not yet overcome.

This is not to say that musical corpus research is solely focussed on Western classical music. A growing number of researchers have employed corpus methods to study popular music genres as well and addressed questions regarding tonality in these styles. These corpus studies on Western popular music include Burgoyne (2011), Burgoyne et al. (2011), Smith et al. (2011), and Burgoyne et al. (2013) who provide a large dataset and analyses of harmonic labels of

[2] http://www.rism.info

songs from the charts in the *Billboard* magazine between 1958 and 1991. Serrà et al. (2012) investigate the central features of pitch, timbre, and volume in more than 450,000 audio recordings across a range of approximately 50 years and such diverse genres as Rock, Pop, Hip-hop, Metal, and electronic dance music from the second half of the 20th century. Focusing on particular Popular music genres, de Clercq and Temperley (2011), Temperley and de Clercq (2013), Temperley (2018), and Tan et al. (2019) study the harmonic language of Rock, as well as harmony or melody in Jazz (Shanahan and Broze, 2012; Broze and Shanahan, 2013; Pfleiderer et al., 2017) and Blues (Katz, 2017). A study on specific harmonic characteristics in Popular music from the 1960's is given by Gauvin (2015), based on the *Billboard* dataset. Harte et al. (2005), Mauch et al. (2007), and Harte (2010) study the idiomatic usage of certain chords in Beatles songs and the Jazz *Real Books*, while Mauch et al. (2015) investigate harmonic and timbral features in more than 17,000 audio recordings of songs that were in the US-charts between 1960 and 2010. Two recent studies address traditional and contemporary Brazilian popular music, in particular the questions of genre classification (Wundervald and Zeviani, 2019), and harmony and form (Moss et al., submitted). Panteli et al. (2018) gather a wide range of statistical approaches in ethnomusicology, both manual and computational, and provide a large, comprehensive review. The rising number of Popular music corpus studies gives reason to hope that this trend will continue. Since Popular music is almost never notated but recorded or digitally produced, strong and reliable audio-based methods are indispensable. A tighter exchange of methods and knowledge between computational musicology and music information retrieval will thus hopefully boost quantitative research on Popular music in the future.

2.2 Methodological considerations

The computational study of musical corpora comes not without its challenges. For instance, while the number of available corpora is increasing, they are often created for specific projects without considering other use cases (but see Serra, 2014). Another challenge is the inherently interdisciplinary nature of the field, entailing different encodings, formalisms, forms of notation, and terminology that may hinder more rapid progress (Volk and Honingh, 2012).

Another, rarely discussed issue in the context of musical corpus studies is that of the *work concept*. Commonly, the notion of a 'work' is used more or less synonymously with 'composition' or 'piece', presupposing that these concepts are well-defined. This notion is tightly related with the self-perception of composers as creative artists (Goehr, 1994; Erauw, 1998) and thus tied to sociological changes in the 19th century (Meyer, 1989). The work concept assumes that there exists a definitive, authorized, and notated version of a composition (Dahlhaus, 1978; Butt, 2015) that constitutes some kind of aesthetic unity (Hanslick, 1922). While this is generally a valid assumption, one can easily think of borderline cases that call it into question. For instance, several versions of the same piece can exist in different keys or arranged for different instruments. Should all be included in a corpus? Which version of a piece should be chosen if there are several, potentially contradicting sources? Are Liszt's piano transcriptions of Beethoven's symphonies works on their own right and should thus be counted as well? The present study opts for a pragmatic answer to these questions but acknowledges that this issue deserves broader attention. We assume here that each datapoint in the used corpora (see Section 3.1) is a work and we follow the convention of using the aforementioned notions interchangeably. Different versions of pieces are not taken into account and neither are different instrumentations.

Yet the greatest challenge for computational musicology, as acknowledged by many authors, is the issue of representativity of the corpora and the quesion of the right sampling methods. For instance, Huron (2013, p. 5) stresses that "[l]arge data sets necessitate more careful consideration of issues of representative sampling" and discusses the implications for statistical inference. London (2013) addresses this issue by proposing a model case for a representative corpus (see below). Neuwirth and Rohrmeier (2016, 175ff.) discuss a range of related challenges in musical corpus research, namely the dangers of studying "skewed and too small samples", and falling into the traps of "availability bias" (only taking data into account that was easily accessible; see also Huron, 2013) and "confirmation error", e.g. only considering examples that confirm one's hypothese which can, for instance, occur when the hypotheses are formed and evaluated on the same data.

These challenges also affect more traditional approaches to the study of music, where observations are grounded in "intuitive statistics" (Neuwirth and Rohrmeier, 2016). Take as an example four volumes on 19th-century music of approximately the same lenght, namely *Style and music: Theory, history, and ideology* (Meyer, 1989), *Harmonic function in chromatic music: A renewed dualist theory and an account of its precedents* (Harrison, 1994), *Chromatic transformations in nineteenth-century music* (Kopp, 2002), and *Audacious euphony: Chromatic harmony and the triad's second nature* (Cohn, 2012). All of them are exceptionally insightful studies on 19th-century music. However, each of them is unique in the specific theoretical approach to this century, reading it through the lense of style analysis (Meyer), extended functional harmony (Harrison), chromatic transformations (Kopp), and Neo-Riemannian theory (Cohn). Thus, a comparison of the theories proposed in these books seems appealing since they supposedly treat the same subject—19th century music—with different methods.[3] But is this really the case? Figure 2.1 compares the relative frequencies of musical examples used in the four texts, grouped by composer. Table B.1 in the appendix provides the actual counts.

As this simple tally reveals, the 'empirical basis' in each of the four studies is quite different from the others. Meyer takes most examples from Mozart, Harrison focuses on Reger and Strauss, Kopp and Cohn both emphasize Schubert but differ, for example, with respect to Liszt. It appears that, while the subject of all four books is 19th-century music, they are, in fact, talking about very different things. Any disagreement between their conclusions might thus not necessarily be related to their theories but could just be attributed to the specific

[3] Meyer's approach is not strictly confined to the 19th century. While he uses it as an example for ideological influences on musical styles, his theory of style analysis consideres other centuries as well.

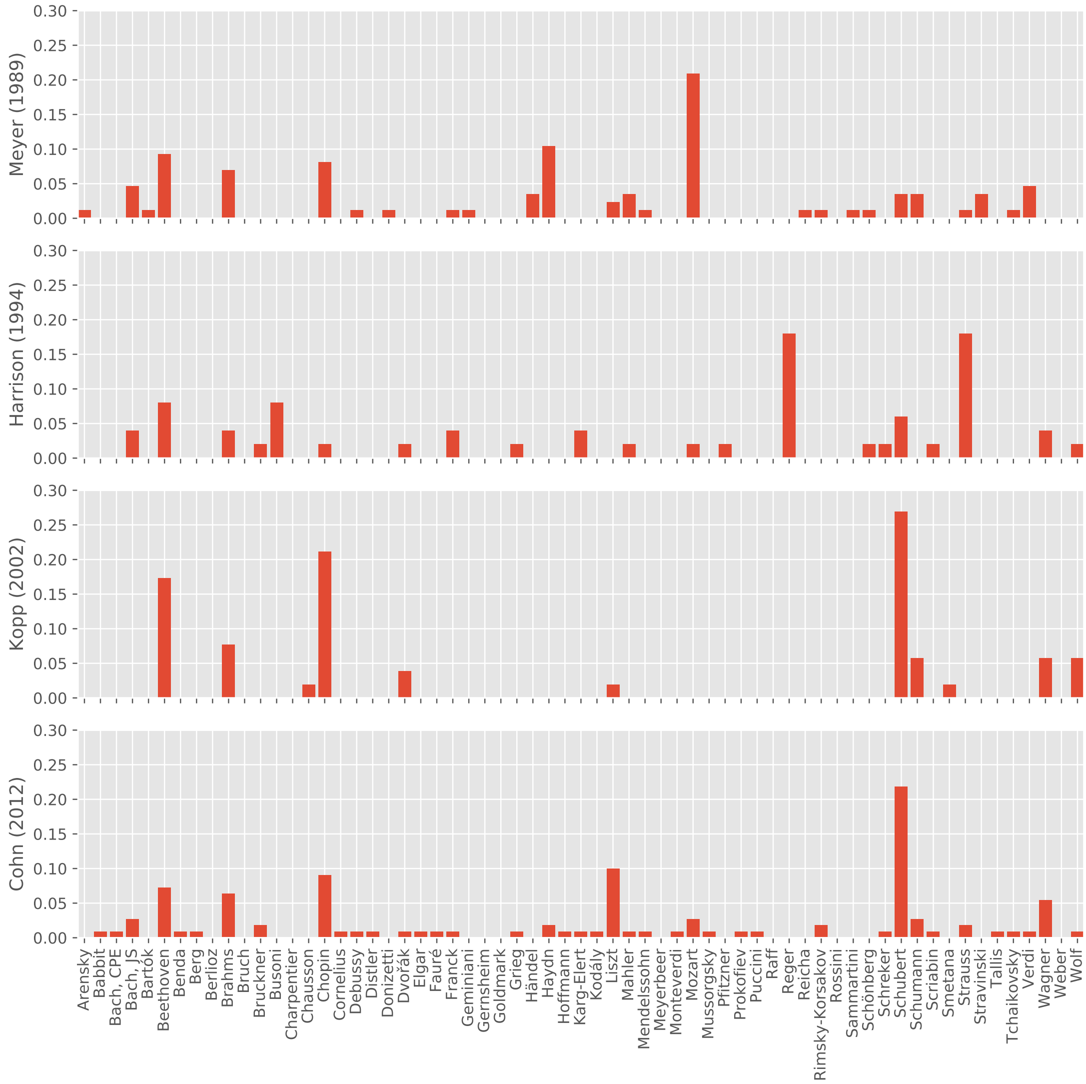


Figure 2.1 – Frequency of mentioned works per composer in Meyer (1989), Harrison (1994), Kopp (2002), and Cohn (2012). The counts are provided in Table B.1 in the appendix.

selection of examples that they use. One has to assume, of course, that the examples found in the texts are only a fraction of the examples that these scholars know and have analyzed. Nonetheless, the arguments supporting their theories are based on very different samples of musical examples which might raise the criticism that these were chosen *because* they fit the respective theories (e.g. Buchler, 2016).

Evidently, the selection of these four volumes also constitutes merely "anecdotal evidence" (Jockers, 2013), although the statistics are not "intuitive" (Neuwirth and Rohrmeier, 2016) but based on empirical data, the counts of the composer names. The argument raised here could thus be turned against this selection itself. Why have these four volumes been chosen here? This is a fair question and the answer is that they were explicitly chosen to highlight the difference between the respective approaches to the music of the 19th century. Certainly, one can make no claims of representativity of this observation. This brings us to the aforementioned problem of *sampling strategies* for corpus-based research. As White and Quinn (2016) note,

> [w]hen compiling a musical corpus, a researcher must decide which composers and historical eras to sample, and which and how many pieces from those composers and eras to include, as well as along and what the criteria for consideration are [...]. The intended goals and uses of a corpus will inform these decisions (White and Quinn, 2016, p. 52).

What are the criteria according to which these decisions can be made? One approach could be to aim for a *balanced* dataset, in wich all classes, e.g. composers, are equally represented. Balancing datasets is a common technique for observational studies on data that cannot be experimentally controlled, for example when studying the effectiveness of seat belts in preventing deaths (Rosenbaum, 2010). The composer counts in Figure 2.1 are all highly imbalanced. Not only do they differ from each other, each of them is also biased towards canonical composers, whereas others occur much less frequently and most are absent. The two basic strategies to introduce balance in datasets are *undersampling* (e.g. Liu et al., 2009) and *oversampling* (e.g. Chawla et al., 2002). Undersampling reduces the number of items in each category, e.g. the number of examples for each composer in a corpus, so that all composers are approximately equally represented. The drawbacks are obvious. If there are categories that contain only a few data points, one throws away a lot of the data from the other categories, drastically reducing the overall size of the corpus and discarding valuable information. Oversampling, on the other hand, creates new artificial datapoints based on the available data. It could, in principle, be used to synthetically increase the number of pieces of some composers to match the oeuvre of others, but one has to ask whether balancing is desirable for observed historical data in the first place. Not all composers across time did write the same number of works, not even approximately. Furthermore, the observed influence of a composer or of certain pieces on the history of tonality will be skewed if one balances a corpus. In conclusion, while both undersampling and oversampling are useful for a wide range of applications, neither of them is appropriate here because a corpus with a balanced number of pieces per composer would not reflect the historical reality. Based on similar observations, Koplenig (2017) even suggests to abandon statistical hypothesis-testing in corpus studies because, in most cases, the necessary conditions are not met. While this is accurate for some corpus studies, it is not true for all of them since one can create corpora that conform to the assumptions of hypothesis testing. An exemplary case (London, 2013) is summarized below.

The so-called *bootstrap* sampling method (Efron, 1979) has a different goal.[4] Instead of aiming at class balance in a dataset, its objective is “assigning measures of accuracy to statistical estimates” (Efron and Tibshirani, 1993, p. 10) for a given sample, e.g. a corpus. The corpora studied here are not experimentally controlled samples but rather constitute “found data” (Arppe et al., 2010) so that certain assumptions for many statistical tests are not met. The bootstrap method is well-suited for this situation. The major advantage of this method is that it does not rely on any assumptions about the underlying distribution that generated the sample and provides a means of estimating uncertainty in the datasets (Piotrowski, 2019). The method is explained in more detail in Section 7.3 where it is used for the first time in this study to investigate how certain chord features affect chord prediction. Later, it is used to estimate the variance in root progressions as well as in the historical trends underlying the growing tonal material (Section 10.2) and the relative importance of certain intervals (Section 12.6) in musical pieces.

It has become clear that the criteria for compiling a corpus largely depend on what it should represent (Temperley and VanHandel, 2013). The following paragraphs present a selection of recently published corpora, highlighting what they aim to represent as well as the sampling procedures with which they were assembled.

The objective of London (2013) is “to build a corpus that is broadly representative of the classical composers, styles, and genres that are most familiar to the 21st century listener” (London, 2013, p. 68). This corpus thus aims at representing what contemporary listeners understand by “classical music” and which pieces are most representative for this concept. Accordingly, he does not only consider musicological accounts of Western classical music for the construction of the corpus, but also resources that help to estimate the consumer behavior of present-day listeners of classical music, namely the Naxos music library,[5] Amazon,[6] and the *Orchestral Repertoire Reports*[7] from the League of American Orchestras from 2000 to 2007, a concert program database.

The Yale–Classical Archives Corpus (YCAC; White and Quinn, 2016) consists of approximately 14,000 files in MIDI format, gathered from the community website *ClassicalArchives*.[8] The authors acknowledge that the YCAC is “rather a survey of the musical priorities of a group of individuals committed to converting their favorite pieces into a digital format” (White and

[4] I am indebted to Timothy J. O’Donnell who brought this method to my attention.
[5] http://www.naxos.com
[6] http://www.amazon.com
[7] https://americanorchestras.org/knowledge-research-innovation/orr-survey.html
[8] http://classivalarchives.com

Quinn, 2016, p. 53). It has a high bias towards piano compositions since this is the most common medium to create MIDI files. Consequently, composers who wrote many pieces for piano are represented more frequently than others (de Clercq, 2016).

In order to study statistical information about musical style in pitch-class distributions, Yust (2019) uses a subset of the YCAC. His sampling criteria are based both on statistical as well as on musical considerations. Only composers having at least five pieces in both the major and the minor mode that begin and end in the same key were included. A number of composers were subsequently exluded because the majority of their pieces consists of etudes which the author considers to serve primarily pedagogical rather than artistic purposes. Furthermore, he specifically excludes Bach's chorales, based on the findings of Albrecht and Shanahan (2013) who argue that the chorales deserve a special treatment because of their short duration and because of their fast harmonic rhythm. He explicitly mentions biases inherited from the YCAC, in particular the bias towards piano music and German composers, e.g. Bach, Mozart, Haydn, and Beethoven, all of which contribute a large proportion of piano pieces to the corpus.

The Annotated Beethoven Corpus (ABC; Neuwirth et al., 2018; Moss et al., 2019b) consists of Beethoven's string quartets and is thus homogenous with respect to the composer and the instrumentation. Apart from understanding it as a sample from, say, all classical music, it can also be conceived of as the entire population of the composer's works in this genre.[9] In this case, statistical analyses remain descriptive since the whole population is accessible. Research questions exceeding the scope of the string quartets, e.g. in order to draw conclusions about tonality or stylistic traits in general, must remain somewhat speculative and can only be answered in a comparative fashion, as will be done subsequently in Part III.

[9] Bearing in mind that one quartet, the *Große Fuge*, op. 133, is not included in this corpus.

3 Data

3.1 Corpora used in this study

The analyses in the respective parts of this study are based on three corpora of different sizes, formats, and historical ranges, and correspond to the three levels Microanalysis, Mesoanalysis, and Macroanalysis. An overview is given in Table 3.1.

Table 3.1 – Overview of the three levels of analysis as well as the size and historical range of the datasets used in this thesis.

Level	Description	No. of pieces	Historical range
Microanalysis	Score	1	1858
Mesoanalysis	Labeled corpora	289	1799–1932
Macroanalysis	XML Corpus	2012	1361–1942

Microanalysis. Part II presents a detailed analysis of Franz Liszt's *Sonetto 47 del Petrarca*, S. 161, no. 4 (1858). This piece was selected to emphasize certain phenomena that are supposedly pertinent to the tonality of the mid- and late 19th century. The score was obtained from the International Music Score Library Project (IMSLP)[1] and is shown as Example C.1 in the appendix. Methodologically, the analyses rely on traditional music analysis as well as on concepts from the Neo-Riemannian and Tonfeld theories. Part II motivates the other parts of this thesis that operate on larger scales of analysis and that address some of the observations made in this part in the light of larger datasets.

Mesoanalysis. Part III proceeds to the mid-range level of analysis and studies a corpus of harmonic analyses of selected works of nine composers of Western classical music from the 19th century. The works in this corpus are the 70 movements of Ludwig van Beethoven's string

[1] https://imslp.org/wiki/Ann%C3%A9es_de_p%C3%A8lerinage_II%2C_S.161_(Liszt%2C_Franz)

quartets (the Annotated Beethoven Corpus (ABC)), the 19 pieces in Schubert's *Winterreise*, 57 of Chopin's *Mazurkas*, 9 pieces from the second volume ("Italie") of Franz Liszt's *Années de pèlerinage* (including the supplement *Venezia e Napoli*), Antonín Dvořák's *Silhouettes* (12 pieces), the 66 *Lyrical Pieces* by Edvard Grieg, the 12 numbers in Tchaikovsky's *Seasons* (one for each month of the year), the four movements of Claude Debussy's *Suite bergamasque*, and 38 *Fairy tales* by Nikolai Medtner. A full overview of all pieces in this corpus, their composition dates, and the sources from which they were obtained is shown in Table B.2 in the appendix.

Macroanalysis. Finally, Part IV broadens the scope to the macroscopic scale by both covering a much wider range of compositions from the 14th to the 20th century and by including a larger dataset than in the parts before. The corpus contains 2,012 pieces (2,962,952 notes) by 75 composers. This corpus was assembled from various resources. Many files have been taken from published scientific datasets such as Renaissance scores from the Electronic Locator of Vertical Interval Successions (ELVIS) project,[2] and the Humdrum **kern scores of the Center for Computer Assisted Research in the Humanities (CCARH)[3]. Other scores have been added from public repositories such as the Choral Public Domain Library (CPDL),[4] or the community webpage of the music notation software MuseScore,[5] while others have been transcribed by student assistants at the Digital and Cognitive Musicology Lab (DCML).[6] A particular set of scores, Koželuh's Piano Sonatas, have been transcribed by Daniel Bernhardsson. An overview of all pieces in this corpus is given in Table B.3 in the appendix.

3.2 The bag-of-words model

The *bag-of-words model* is a widely used representation for documents in Natural Language Processing (NLP; Manning and Schütze, 2003; Jurafsky and Martin, 2009). It describes a document by simply counting all the words that it contains. In the context of the present study, this model is adopted in the Meso- and Macroanalysis parts where the documents correspond to the musical pieces in the analyzed corpora. Accordingly, words correspond to the smallest units in these pieces. In Part III these are chord symbols in the corpus of harmonic labels, and in Part IV these are pitch classes[7] in the pieces of the corpus.

Let us consider the case of the *bag-of-notes model* to illustrate its implications. It represents pieces by simply counting the frequencies of occurrence of pitch classes without taking into account the order in which they appear in the piece. Figure 3.1 shows two bags of notes, i.e. the pitch-class counts for two pieces. Let us presume for a moment not to know from which

[2] https://elvisproject.ca
[3] http://kern.ccarh.org
[4] http://www.cpdl.org
[5] http://www.musescore.com
[6] http://dcml.epfl.ch
[7] Pitch classes consider all notes as equivalent that are octave-related (see Section 9.2 for a more detailed account).

pieces these counts were obtained. What can one discern about the tonality in these pieces by only looking at the frequency of occurrence of the notes?

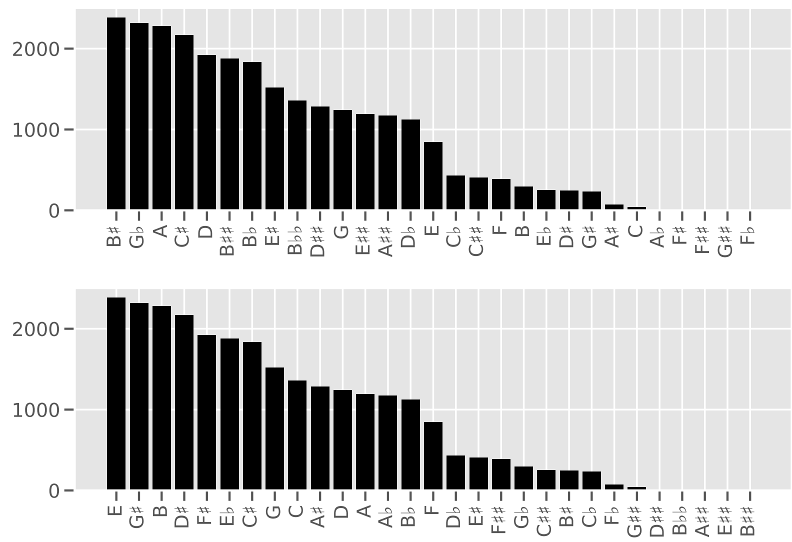
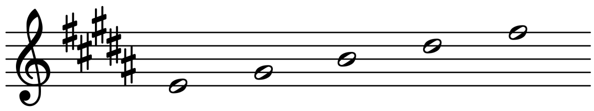


Figure 3.1 – Tonal pitch-class counts for two compositions. Bottom: Alkan's *Concerto for Solo Piano*, op. 39, No. 8, mov. 1; top: a composition generated by randomly reassigning the weights from this piece to its tonal pitch-classes.

A few observations can be made by just looking at the two plots. First, both pieces contain exactly the same pitch-classes. This can be seen by comparing the labels of the horizontal axes. In total, there are 29 distinct pitch classes, a considerably large number taking into account that there are only twelve distinct keys on the piano. Second, the frequencies of these pitch classes in the two pieces are very different. While the most frequent pitch class in the piece shown in the top panel is B♯, the most frequent pitch class in the other piece is E. No pitch class has the same frequency in both pieces. However, the counts with respect to the rank of the pitch classes in the two pieces are identical. Let us now have a closer look at the pitch classes in these two pieces and focus on the most frequent ones. In the first piece, the five most frequent pitch classes are B♯, G♭, A, C♯, and D, in the second piece the first five pitch classes are E, G♯, B, D♯, and F♯. While there is no evident relation between these pitch classes in the first piece, the top five pitch classes in the second piece form a chain of diatonic thirds in the keys of either E major, C♯ minor, B major or G♯ minor as shown in Example 3.1.



Example 3.1 – A chain of diatonic thirds.

The relations between the pitch classes in the second piece are thus much more interpretable with respect to tonality than in the first one. The reason for this is that the second piece, shown in the bottom panel of Figure 3.1, is an actual composition, namely the first movement of

Alkan's *Concerto for Solo Piano*, op. 39, no. 8, in G♯ minor. The piece shown in the top panel is an artificially created bag of notes, generated by randomly reassigning the pitch-class counts in Alkan's composition. It is thus no surprise that the pitch-class counts in this randomly generated 'piece' are not interpretable. It is a fundamental assumption of the present work that the relation between the notes in a piece and their frequency of occurrence is largely determined by tonality. Consequently, the random reassignment of frequencies to pitch classes breaks the underlying tonal relations. There is an important conclusion to be drawn from this which is essential for this study. While both pieces are represented as a bag of notes, only the real composition allowed to infer aspects of its tonality from its pitch-class counts under this representation. This may indicate that tonality is a latent factor influencing the counts of the elements in musical compositions, a hypothesis that we are going to investigate more rigorously in the subsequent chapters, both for chord symbols and for pitch classes.

Microanalysis Part II

4 Franz Liszt, *Sonetto 47 del Petrarca*

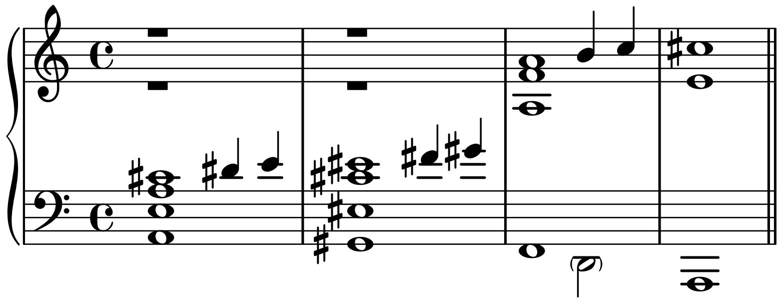
Franz Liszt (1811–1886) wrote *Sonetto 47 del Petrarca* between 1837 and 1839 during his stay in Italy. It was published in 1858 as the fourth piece in the second volume ("Italie") of the piano cycle *Années de pèlerinage*, S. 161 (Liszt, 1978, p. III). The score is reproduced as Example C.1 in the appendix. It sets to music a sonnet by Petrach, one of the most reknowned Italian Renaissance poets. This piece was chosen for the analysis in the Microanalysis part because it exhibits a range of artistic devices that are characteristic for 19th-century tonality, such as the usage of extended triadic sequences, harmonies based on equal-divisions of the octave, and expanded key relations. In particular, the analysis will provide evidence for the usage of non-traditional forms of tonality in this piece, which motivates the analyses on a larger scale in the subsequent parts Mesoanalysis (Part III) and Macroanalysis (Part IV).

The overall structure of the 95-measure piece is based on the sonnet's structure of two stanzas with four verses each, followed by two stanzas with three verses each, but is not as rigid as the poem with respect to its form. The piece consists of several sections, partly separated by double bars, key signature changes, or metrical changes. The first section (mm. 1–5) is entitled "Preludio con moto" and serves as an introduction. It is followed by a two-part theme, consisting of cadential motions in the key of D♭ major (mm. 6–11), which prepare the melodic theme of the piece ("Sempre mosso con intimo sentimento", mm. 12–21). Following the strophic structure, this double theme reoccurs several times in different keys.

4.1 The introduction

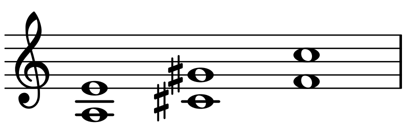
We begin the analysis with the "Preludio". It consists of five bars of largely triadic harmonies that prolong an A-major triad from mm. 1 to 5 by means of progressions by major thirds: from A major to C♯ major via F major back to A major. A harmonic reduction is shown in Example 4.1.

The triads are shown as whole notes and are related in a very systematic fashion. The transitions between them are realized by a combination of triadic transformations (Cohn, 1996;



Example 4.1 – Harmonic reduction of *Sonetto 47 del Petrarca*, mm. 1–5.

Gollin, 2005). Having roots in 19th-century music theory (Riemann, 1880), these transformations are now extensively studied within *Neo-Riemannian theory* (Klumpenhouwer, 1994; Cohn, 1998; Fiore and Noll, 2011; Gollin and Rehding, 2011; Rings, 2011). The three basic transformations are *parallel* (P), *relative* (R), and *leading-tone exchange*. In the Preludio, the transformations are combinations of P and L. The former transforms a major triad into the minor triad with the same root, e.g. it transforms A major to A minor. The latter transformation exchanges the root of a major triad with its leading tone, e.g. A major is transformed to C♯ minor by exchanging the tones A and G♯. The transformations are reversed if applied to minor triads, e.g. A minor is mapped to A major via P and to F major via L. In the chord sequence in the *Preludio*, A major is transformed into C♯ major by an implicit transformation to C♯ minor which is then mapped to its parallel triad, C♯ major. This combined transformation can be expressed as PL where, by convention, the transformations are read from left to right (first L, then P) because of the relation of these transformations to mathematical group theory (Fiore et al., 2013; Harasim et al., 2016). Taking all three triads together, their constituent notes form a *hexatonic* collection: A, C, C♯, E, E♯/F, and G♯—assuming the enharmonic equivalence of E♯ and F—, which can also be partitioned into two augmented triads consisting of the root notes A, C♯, F, and of E, G♯, C, the fifths of the triads. This collection of notes (and its transpositions) is called *Konstrukt* within the Tonfeld theory (Haas, 2004; Schild, 2010; Polth, 2018) and is shown in Example 4.2.



Example 4.2 – Hexatonic collection of notes, characterized by the intervals of the perfect fifth (vertically) and the major third (horizontally).

It is characterized by the intervals of the perfect fifth (vertically) and the major third (horizontally). Since we have defined tonality as the relations between notes (see Section 1.1), we can say that the prelude establishes a hexatonic tonality. The regularity of this hexatonic sequence is, however, somewhat distorted. The notes forming part of the melodic line (shown as quarter notes) on beat 3 (D♯, F♯, and B) in the first four measures are outside the hexatonic collection, and so is the D in the bass in m. 3 (shown in parentheses in Example 4.1) that bisects the interval from F to A. It supports a half-diminished chord (B, D, F, and A) that contains with B and D two of the notes outside the hexatonic collection. The intervals forming the melodic lines in each measure are a whole step and a half step, whose combination cannot lie in a hexatonic collection. It seems that the melody in the *Preludio* implies other aspects of tonality that we have not yet taken into account. We will come back to this issue further below.

It is noteworthy that this prelude does not have a key signature. This is not coincidental but reflects the composer's conception of tonality in these five measures. Liszt had an extensive exchange with Fétis (Móricz, 1993) who coined the term 'tonality' and regarded his period to be in the *ordre omnitonique* (see Section 1.2). Liszt's album-leaf *Prélude omnitonique*, S. 166e, from the year 1844—only a few years after the composition of the *Sonetto*—as well as his later *Bagatelle sans tonalité*,[1] S. 216a (1885), epitomize the composer's engagement with Fétis's theories. The *Bagatelle* is shown in Example 4.3.

Example 4.3 – Franz Liszt, Album-leaf *Prélude omnitonique*, S. 166e (1844).

This aphoristic album-leaf consists entirely of eight chromatically ascending fully diminished seventh chords, a motion that could in principle be continued perpetually. Neither the *Prélude omnitonique* nor the *Bagatelle sans tonalité* nor the prelude of the *Sonetto 47 del Petrarca* have a key signature, but the tonalities established in these cases is very different. While the former two are largely constructed around fully diminished seventh chords, the first five measures of the *Sonetto* embody hexatonic tonality.

4.2 The musical task

After this five-bar introduction follows another short section (mm. 6–11, "Ritenuto") as indicated by the enclosing double bars. The excerpt is shown in Example 4.4. The new key signature of five flats implies that the main key of the piece is D♭ major and that the hexatonic tonality of the prelude is suspended or abandoned. It seems as if the piece does really begin here. The segment consists of a twofold predominant–dominant motion where the dominants in mm. 7 and 9 are given as $^6_4$-suspensions. Only the second one in m. 9 resolves to a proper dominant seventh chord in the subsequent measure. The resolution to the tonic triad D♭ major that is implied by these dominants, is denied, or rather delayed—literally in

[1] "Sans tonalité" should be translated as "without key" rather than "without tonality"; see also Riemann (1900).

Example 4.4 – Franz Liszt, *Sonetto 47 del Petrarca*, S. 161 (1885), mm. 6–11.

the last minute—through the reversal of the descending melodic motion in m. 11. Instead of continuing from E♭, the second scale degree in D♭ major, to the tonic D♭, the sense of melodic direction is altered by the chromatic passing note E—emphasized by the fermata—and diverted to the third scale degree F (m. 12, not shown in Example 4.4). Hence the harmonic tension built up so far is not fully resolved but further delayed.

This delay can be understood as determining the 'task' of this piece. A musical task is defined as a "technique of delaying for artistic reasons the resolution of certain tendencies possessed by material in the expository section of a work" (Marco, 1958), a strategy that many composers employ (see, for instance, Neuwirth, 2015). Throughout the piece, several attempts are made to achieve the resolution and fulfill the task in variants of these initial cadential figures. For instance, m. 60 brings a similar cadential motion but in the key of C♯ minor, the enharmonically equivalent of the key parallel to D♭ major. Abbreviated versions of this motion can be found in mm. 75–77, and an exaggerated version ("***f*** *con somma passione*") appears shortly before the conclusion of the piece in mm. 85–89. However, even there, the resolution to the tonic does not immediately follow the dominant but occurs only after a shortened form of the theme of the piece, which is discussed in the next section.

The two predominants in mm. 6 and 8 consist of quickly arpeggiated eight notes in the left hand that sound very similar but look very different. The first contains the notes B♭♭, F♭, and G, the second consists of the notes G, E, and B♭. While the latter can be identified as a diminished chord with root E, the former is a so-called German augmented sixth chord in the key of D♭ major. In both cases, the melody in the right hand features the notes D♭, E♭, and E, forming the intervals of a whole step and a half step. These melody fragments are enharmonically equivalent to the one in m. 1 (C♯, D♯, and E). Retrospectively, the melodic lines in the introduction can now be understood as forecasting the melody in the second segment from mm. 6 to 11.

To summarize, the musical task set up in mm. 6–11 is to achieve cadential closure in the tonic key. This is not an exceptional task. In fact, it is not too far-fetched to state that this is the objective in virtually all tonal compositions (Schenker, 1935; Markus Neuwirth and Bergé, 2015). In principle, one could accomplish this task by jumping directly from m. 7 to m. 89, deleting everything in between to the effect that the $^6_4$-suspension is properly resolved and the piece concludes shortly thereafter. The appeal of assuming such a task is that it allows to analyze the ways in which this goal is denied, delayed, and ultimately fulfilled in a particular musical work.

4.3 The theme

After the statement of the task and a two-measure long unfolding of the tonic harmony, Liszt introduces in mm. 14–21 the thematic material in the melody that shapes large parts of the piece. The theme has a periodic structure, consisting of an antecedent phrase (mm. 14–17) and a modulating consequent phrase (mm. 18–21).[2] Example 4.5 shows a rhythmically simplified rendition of this period.[3]

Example 4.5 – *Sonetto 47 del Petrarca*, theme in mm. 14–21 with simplified rhythm.

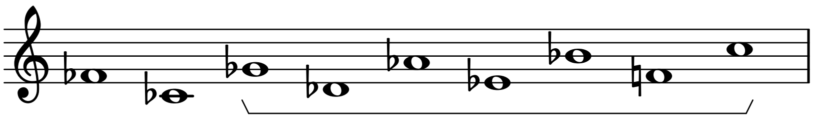
The basic idea (first line in Example 4.5) unfolds the D♭-major tonic triad melodically and uses almost all notes of the key of D♭ major, except C. The contrasting idea (second line) exhibits the same contour since it is largely a transposed version of the basic idea, and introduces the note C, extending the tonal material of the melody in the entire antecedent phrase to the diatonic scale in the tonic key. The first part of the consequent phrase (m. 18 in the example) repeats the basic idea of the antecedent phrase but introduces E natural, the leading tone to F, thus anticipating the subsequent modulation to F minor and closing only weakly on the triadic third in m. 12. Finally, the last two measures execute the modulation, introducing B natural as a new tone that forms an augmented sixth with the bass D♭ and creates a double

[2]One could also argue that it is a *hybrid* (Caplin, 1998) between a period and a sentence because the antecedent phrase does not conclude with a weak cadence, and the consequent phrase modulates and brings no cadential closure.

[3]In this chapter, “period” denotes a specific formal arrangement of music and not a temporal span.

leading-tone, forcing both B and D♭ to resolve into C, the dominant in the new key.

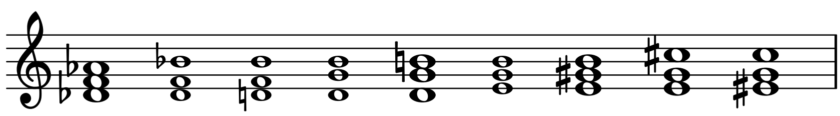
The expansion of tonal material during the unfolding of the theme can be summarized in a systematic arrangement of the notes, namely as a series of eight consecutive fifths that is called the *series* or *line of fifths* (Weber, 1851; Temperley, 2000; Haas, 2004). Example 4.6 shows that all notes used in the theme can be ordered linearly. Note that, in order to do so, the notes E natural and B natural were replaced by their enharmonic equivalents F♭ and C♭.



Example 4.6 – Series of eight consecutive fifths. The notes belonging to the key of D♭ major are indicated by the bracket.

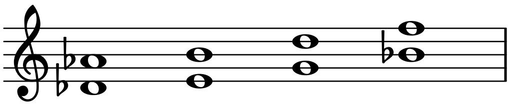
In total, this series spans nine notes from F♭ to C. Subsets of this collection are important structures for tonal music, e.g. diatonic and pentatonic scales spanning six and four consecutive fifths, respectively, as well as the musical *tetractys* (de Jong and Noll, 2011) consisting of three consecutive notes on this line.

As mentioned at the beginning of this chapter, the theme occurs four times over the course of the entire piece, reflecting the underlying strophic structure of the eponymous sonnet. These variants of the theme feature different keys, namely G major (mm. 36–43), D major (mm. 48–52), E major (mm. 62–68), and again D♭ major towards the end of the piece (mm. 90–95). In each case, the key signatures change accordingly, except for the reappearance of the theme in D major. In this instance, the key signature still indicates the previous G major, which expresses the fact that the region in D major is subordinate to the G major section as its dominant key. The relation between the remaining key regions is of astonishing uniformity. Consider Example 4.7 which shows the tonic triads of these keys, interspersed with other triads that reveal that they are connected by a regular chain of transformations.



Example 4.7 – Octatonic sequence of triads. Keys that do not occur in the *Sonetto* are shown with smaller notes.

The triads in this chain are related by relative and parallel transformations. Recall that the parallel transformation maps a major triad to its minor version and *vice versa.* The relative transformation, on the other hand, raises the fifth of a major triad to the sixth, e.g. mapping C major to A minor. The triads in the chain shown in Example 4.7 are related by alternatingly applying the relative and parallel transformations. The relative harmony of the initial D♭ major is B♭ minor, and the parallel of this chord is B♭ major, etc. Note that this chain arrives at the end at C♯ major, the enharmonically equivalent chord to D♭ major, revealing that the chain is a cycle. The collection of notes in this cycle is called *octatonic* and is shown in Example 4.8.

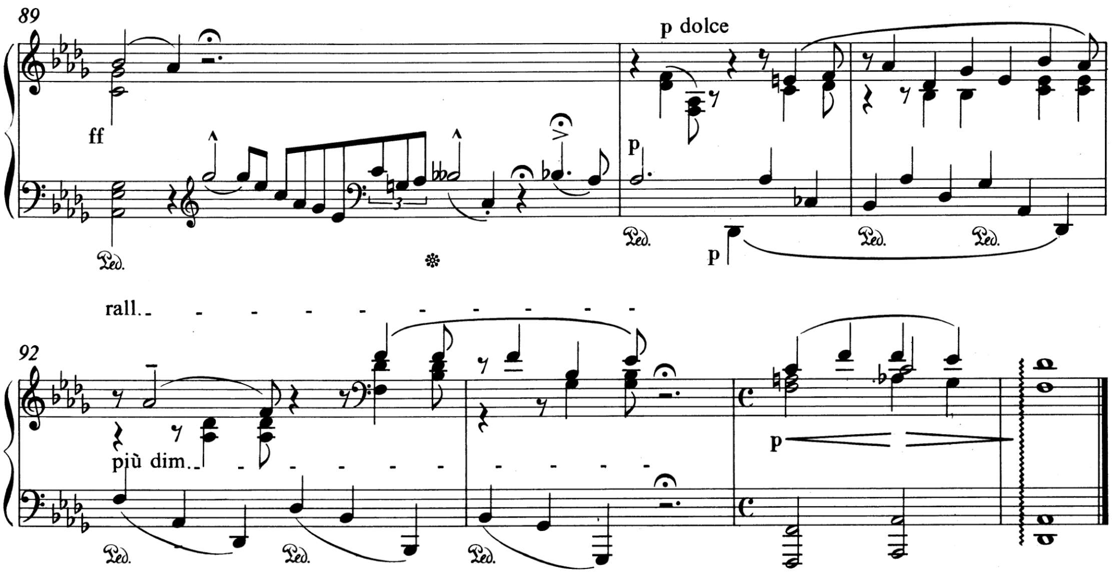


Example 4.8 – Octatonic collection of notes, characterized by the intervals of the perfect fifth (vertically) and the minor third (horizontally).

Similarly to the hexatonic collection explained above, the octatonic collection is characterized by the interval of the perfect fifth (vertically) and, contrary to the hexatonic collection, by the interval of the minor third (horizontally). In the terminology of Tonfeld theory, the octatonic collection is called *Funktion* and regarded as a generalization of Riemannian harmonic functions (Haas, 2004; Schild, 2010; Polth, 2018). Whereas the hexatonic collection established the tonality for the five measures of the introduction, the octatonic collection establishes the tonality for a much wider range in the piece, from the first occurrence of the theme in mm. 14–21 to its final appearance in mm. 91–93.

4.4 The resolution

After the many delays of the cadential closure and digression into a range of different keys, the piece reaches its final resolution in the last seven measures of the piece (mm. 89–95), displayed in Example 4.9.



Example 4.9 – Franz Liszt, *Sonetto 47 del Petrarca*, mm. 89–95.

The A♭ dominant seventh chord at the beginning of m. 89 reiterates the interrupted attempts at cadential closure in m. 7 and elaborates this dominant melodically throughout this measure, embellishing its root with its neighbors G, B♭♭ (with an accent mark), and B♭ (with a fermata). This clearly demarcates it as the final dominant, constituting the climax of the accrued harmonic tension. However, this tension is still not resolved immediately. Liszt continues, at first glance, the same way as at the beginning of the piece (m. 11ff.) and restates the theme, albeit in an altered form. It is only half as long as in the original version and also exhibits some harmonic changes. The first of these changes is the brief tonicization of the subdominant G♭ major in mm. 90–91, a classical signal that the end of a piece is reached, at least since Baroque times. The second alteration is that of the underlying harmonies in the theme. While the contrasting idea of the original version (mm. 16–17) describes a ii V^7 I progression in D♭ major over a tonic pedal, the abbreviated version here inserts a triad on the sixth scale degree (second half of m. 92) and thus elongates the sequence of falling fifths, another indicator of cadential closure. This cadence to the final tonic—B♭ minor, E♭ minor, A♭^{7}, D♭ major—is again interrupted by a long fermata (second half of m. 93) and the insertion of an F-major triad before the dominant in the penultimate measure. This F-major triad gleams in the surrounding context of D♭ major, in particular because of its major third, the out-of-scale A. This harmony constitutes a reminiscence of the F major from the hexatonic context of the *Preludio* that launched the piece (mm. 1–5), where it was juxtaposed with C♯ major. At the same time, one can observe that F major and A♭ major are related by the combination of a parallel (P) and a relative (R) transformation. The entire content of m. 94 is part of one and the same octatonic scale and can thus be seen as an extended dominant. In this sense, one can analyze the final two measures of the *Sonetto* as its summary, containing the three central tonalities in this piece: the hexatonic tonality from the opening measures symbolized by the F-major triad, the octatonic tonality manifest in the relations between the keys of the appearances of the theme, and the diatonic tonality as embodied in the dominant–tonic progression, forming part of the line of fifths. This ending features the harmonic resolution of the dominant seventh chord to the tonic and the resolution of the melody that descends to the tonic. Hence, it has both of the desired properties that the solution to the musical task requires, and the piece ends here.

4.5 Discussion

The analysis of Franz Liszt's *Sonetto 47 del Petrarca*, S, 161, no. 4, has shown how the composer delays the musical task of cadential closure by a variety of strategies. Moreover, it was revealed that different relations between notes—different tonalities, see Section 1.1—govern different parts of this composition, both on a local and on a more global level. It has been shown in particular that tonality is established by notes that are related by consecutive perfect fifths (e.g. in the case of the diatonic scale), by perfect fifths and major thirds (the hexatonic collection), or by perfect fifths and minor thirds (the octatonic collection). The questions is whether these findings are particular to this piece, particular to Liszt's composition style, to the 19th century in general, or whether their relevance even exceeds this temporal range. The analyses in the subsequent parts will provide insights into these issues.

Comprehensive analyses of musical pieces at this level of detail are to the date beyond reach for computational approaches. How would one, for instance, formally define the musical task of a composition? The difficulty for computational methods to address such intricate music theoretical questions can, however, only partially be attributed to a methodological lack of sophistication. It is likewise caused by incomplete, inconsistent, and even incomprehensible descriptions on the musicological side, where theories are often formulated in a verbose manner that defies formalization and thus, at least to a certain degree, rationalization. This study attempts to offer a bridge between traditional historical and music theoretical descriptions on the one hand, and statistical and computational modeling on the other hand. The two subsequent parts build on some of the insights gained here and address the case of tonality on larger scales. The Mesoanalysis part directly bridges traditional music analysis and corpus studies by employing a dataset of harmonic labels produced by music theory experts and by using statistical methods for the analysis of these annotations to draw inferences about tonality in the 19th century. The Macroanalysis part addresses some of the questions raised here that *can* be formalized, such as the relations between notes in musical compositions. Applying the measures and models used in this part to a large corpus also allows to draw conclusions about the historical development of tonality. Apart from using computational methods to study traditional music-theoretical questions, it will also be shown that the corpus approach creates entirely novel questions that cannot be addressed in a traditional manner. The subsequent corpus studies are thus not only be shown to test music analytical case studies against a larger database, but also to enable also to open up new avenues of empirical music research that, in turn, would be impossible to approach with traditional methods. A central concern of this thesis is to emphasize that manual and computational music analysis can complement each other and together advance the empirical study of tonality.

Mesoanalysis Part III

5 Modeling harmonic annotations

So what's the problem this time, knucklehead?
There is no problem! It's just that... He closes his eyes and calls out, downbeat by downbeat: *Tonic. Subdominant. Dominant.*
Those guys need to learn some new chords.

Richard Powers, *Orfeo*

5.1 Harmonic analysis with chord symbols

Analyzing musical corpora based on chord symbols is one of the most common approaches to the computational study of harmony (see Section 2.1). It has the advantage that labeled corpora already provide a segmentation of the stream of musical events into harmonic units (Hanninen, 2012; Abdallah et al., 2016), one of the central ways of thinking about harmony for centuries (Lester, 2002; Damschroder, 2008). Moreover, the the labels in such corpora are given in machine-readable formats that allow for large-scale pattern analysis. Naturally, harmonic analyses given by human annotators are prone to ambiguities, subjective biases, disagreement between annotators, as well as plain errors (Temperley, 2001; Temperley and VanHandel, 2013) and they thus need to be taken with a grain of salt. There are strategies and initiatives to evaluate and enhance inter-annotator consistency (McFee et al., 2017; Koops et al., 2019) but these usually work on a much coarser level than the harmonic analyses studied here. Although there is a wide range of approaches to automatic chord recognition (e.g. Sheh and Ellis, 2003; Mauch, 2010; Humphrey and Bello, 2012; Masada and Bunescu, 2019), the performance of these approaches has not yet reached the level of detail and accuracy a human music theory expert can achieve. For this reason, this part of this study relies on the harmonic annotations of human music theory experts, henceforth called 'annotators'. Six annotators were carefully selected, all of whom are in the possession of a university-level degree in music theory or musicology, and—in most cases—teach the subject in academic

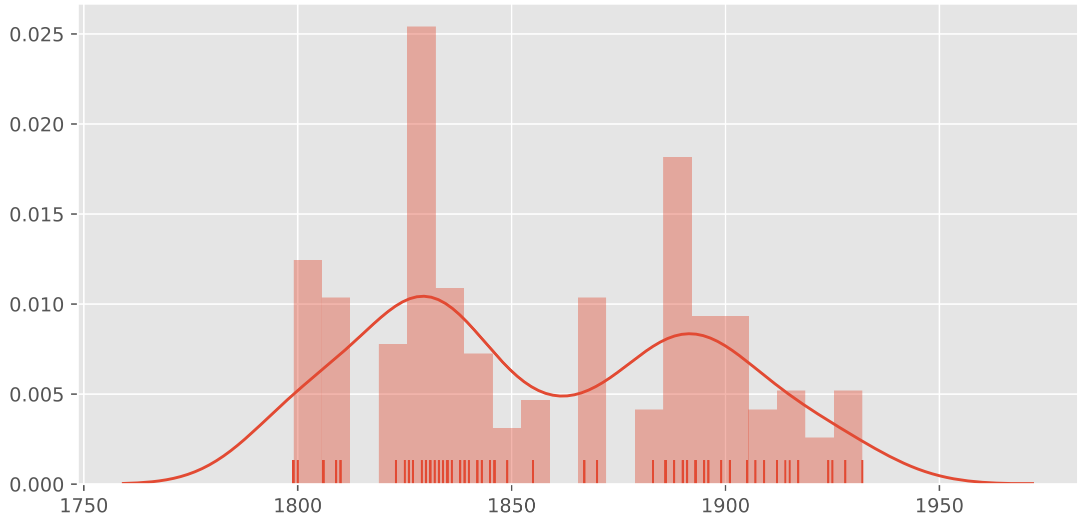


Figure 5.1 – Temporal distribution of pieces in the labelled corpora shown by ticks on the x-axis that represent the years of the pieces, by a histogram and a density estimate showing that most pieces lie in the first half of the 19th century and in the transition from the 19th to the 20th century. Note that, for Beethoven's string quartets, movements are counted as pieces.

contexts. Nonetheless, their analyses are potentially biased and varied. Annotators may differ with respect to the analysis of a particular situation (inter-annotator inconsistency) but both might be syntactically correct so that this discrepancy might not have been detected in the automatic parsing of the annotations. Moreover, one annotator might analyze similar passages differently, rendering his or her analyses inconsistent (intra-annotator inconsistency). These and other biases introduce variance in the data that one has to be aware of when interpreting the results.

Corpus selection. The corpus studied in this part consists of harmonic analyses of nine sub-corpora by different 19th-century composers, namely Ludwig van Beethoven's string quartets, Franz Schubert's *Winterreise*, Frédéric Chopin's *Mazurkas*, the first volume "Italie" from Franz Liszt's *Années de pèlerinage* (except the *Tarantella*), Piotr Tchaikovsky's *Seasons*, Edvard Grieg's *Lyrical pieces*, Antonin Dvořák's *Silhouettes*, Claude Debussy's *Suite bergamasque*, and Nikolai Medtner's *Fairy tales*. A full overview of all the pieces is given in Table B.2 in the appendix. Together, these datasets cover a range of more than a hundred years (1799–1932) and thus cover the entire 19th century and the early as well as the late Romantic period. The distribution of all pieces in this corpus is shown in Figure 5.1. Note that we count different movements of pieces separately. Otherwise, they would be considerably longer than most pieces in the other sub-corpora. Moreover, most of the movements are tonally closed.

Annotation procedure. All labelled corpora used for this study were created within the scope of several research projects of the *Digital and Cognitive Musicology Lab* (DCML) at EPFL.

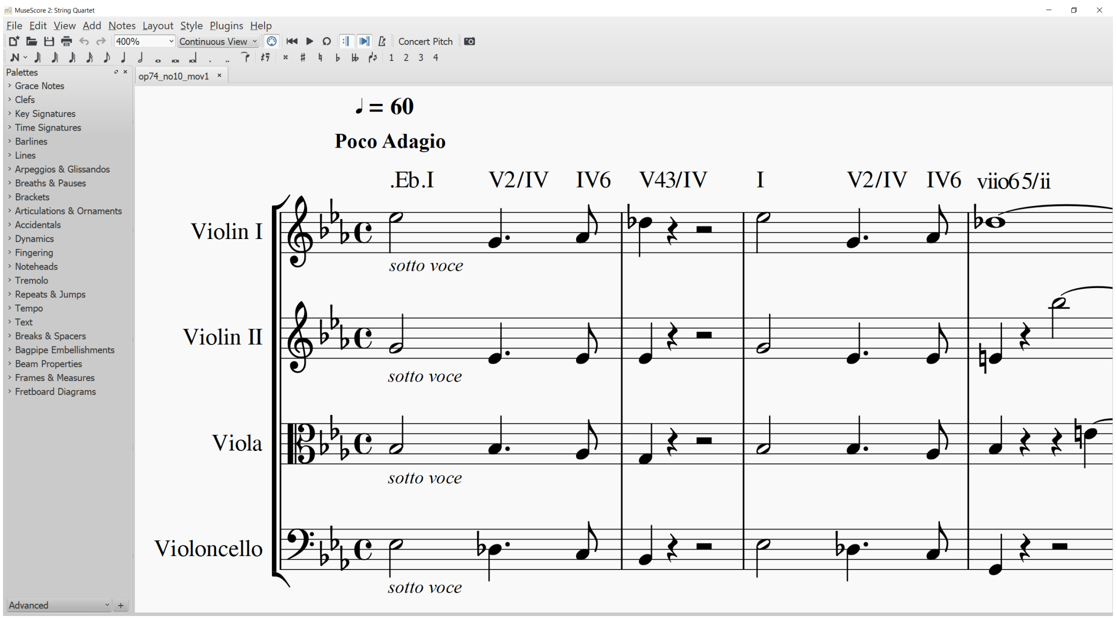


Figure 5.2 – Screenshot of MuseScore (v2.0.2) interface as used by annotators. The example shows the first four measures of the first movement of Beethoven's op. 74.

The creation process as well as the annotation standard is documented in the data report accompanying the Annotated Beethoven Corpus (ABC; Neuwirth et al., 2018) and also applies to the other sub-corpora. The harmonic analyses were provided by six expert annotators, as explained above. The annotation standard itself that was used for the harmonic analyses is detailed further below in Section 5.2. In order to facilitate the annotation procedure for those corpora for which digital encodings of the scores were available, the graphical interface of the open-source notation software MuseScore[1] (v.2.x; Bonte, 2009) was used. Annotators could import the provided XML files into MuseScore and enter the labels directly into the score. The XML files for Beethoven's string quartets were obtained from Project Gutenberg[2] and the ones for the *Winterreise*, *Seasons*, *Suite bergamasque*, and some of the pieces of the *Années de pèlerinage* from the MuseScore community website.[3] The remaining pieces of the *Années de pèlerinage* as well as the *Silhouettes* were transcribed by the DCML. An extract of a labelled score in MuseScore is shown in Figure 5.2. The underlying XML data for the first measure is shown in Figure 5.3.

The harmonic analyses for the other corpora, namely the *Mazurkas*, the *Lyrical Pieces*, and the *Fairy Tales*, were provided in separate text files, specifying the harmonic labels as well as their position in the score by measure and beat numbers. An example is given in Figure 5.4 that shows the analysis of the first five measures of Nikolai Medtner's *Ophelia's Song*, op. 14, no. 1, in this format.

[1] https://musescore.org
[2] http://www.gutenberg.org
[3] http://musescore.com

```
<?xml version="1.0" encoding="UTF-8"?>
    <museScore version="2.06">
        <Score>
        ...
            <Measure number="1">
              ...
              <Harmony>
                <name>.Eb.I</name>
              </Harmony>
              <Chord>
                <durationType>half</durationType>
              </Chord>
              <Harmony>
                <name>V2/IV</name>
              </Harmony>
              <Chord>
                <dots>1</dots>
                <durationType>quarter</durationType>
              </Chord>
              <Harmony>
                <name>IV6</name>
              </Harmony>
              <Chord>
                <durationType>eighth</durationType>
              </Chord>
            </Measure>
            ...
        </Score>
    </museScore>
```

Figure 5.3 – The chord annotations for m. 1 in the first movement of Beethoven's op. 74 in the MuseScore XML format (see also Fig. 5.2). Many lines of code have been omitted.

```
@piece: op.14_no.1
# Nikolai Medtner: Skazka in F minor, op. 14 No. 1 "Ophelia's Song" (1906-1907)
# Annotator: Wendelin Bitzan

@meter: 2/4
@tempo: Andantino con moto
@length: 85
@key: f

@form: A
m000  2.1 .f.IV65  2.2 i64 @alt: V(6)
m001  1.1 i(9)  1.2 i6  2.1 IV65  2.2 i64 @alt: V(6)
m002  1.1 i(9)  1.2 i6  2.1 IV65  2.2 i64(4) @alt: V7(6)
m003  1.1 i6(9)  1.2 i6  2.1 V43  2.2 i
m004  1.1 IIIM7  1.2 #vi2 2.1 v 2.2 i64 @alt:  V(6)m005 1.1 i7(9) 1.2 IV64 2.1 i 2.2 V43/VII
```

Figure 5.4 – Harmonic analysis in .txt format. The example shows the first five measures of Nikolai Medtner's *Ophelia's Song*, op. 14, no. 1.

Table 5.1 – Data frame representation of the first four measures of Beethoven's string quartet op. 74, mov. 1. The table shows a selection of all columns.

op.	no.	m.	b.	timesig	global_key	local_key	root	form	inversion	suspension	applied
74	10	1	1	4/4	Eb	I	I	–	–	–	–
74	10	1	3	4/4	Eb	I	V	–	2	–	IV
74	10	1	4	4/4	Eb	I	IV	–	6	–	–
74	10	2	1	4/4	Eb	I	V	–	43	–	IV
74	10	3	1	4/4	Eb	I	I	–	–	–	–
74	10	3	3	4/4	Eb	I	V	–	2	–	IV
74	10	3	4	4/4	Eb	I	IV	–	6	–	–
74	10	4	1	4/4	Eb	I	vii	o	65	–	ii

Parsing and preprocessing. The annotators submitted their analyses in MuseScore's .mscz format which encodes a score in the MuseScoreXML dialect of XML or in text files, depending on the corpus as mentioned above. The submitted annotations have subsequently undergone several stages of checking. All analyses were automatically checked for syntactic validity and randomly checked for semantic correctness and consistency by three members of the DCML, including the author. At the time of writing of this thesis, all corpora have passed the basic stages of quality control, but only the ABC has been published yet.[4] The other corpora will follow soon. The scores were subsequently parsed into the tabular representation of a data frame, where each chord symbol in a corpus is given in a row, and the columns are defined by the features (such as metrical position, inversion etc.). Table 5.1 shows the the first four measures of Beethoven's string quartet op. 74, mov. 1 in this representation (Figure 5.2 shows the first four measures in MuseScore). Only a selection of the columns is shown. This data frame representation of the annotations have also been made public for Beethoven's string quartets as the ABC (Neuwirth et al., 2018) and will be complemented by the other corpora in due time.

5.2 Chord morphology

Roman numeral analysis. A widespread system for the analysis of tonal music is Roman Numeral Analysis (RNA).[5] It is used in many contemporary music theory textbooks (e.g., Caplin, 1998; Gárdonyi and Nordhoff, 2002; Gauldin, 2004; Aldwell et al., 2010; Laitz, 2012) and is also adopted by music theorists across musical styles (Lerdahl and Jackendoff, 1983; Lerdahl, 2001; Temperley, 2001; Tymoczko, 2011; Doll, 2017; Temperley, 2018). RNA could arguably be conceived as the *de facto* standard in academic music theory, although there are individual differences in typography and level of detail. The fundament of RNA is a musical scale. Each note of the scale in diatonic order is called a scale-degree and labeled with a Roman numeral, the tonic being the first. Notation systems differ, for example, whether to only use uppercase Roman numerals (e.g. Schoenberg, 1969) or also lowercase ones (e.g.

[4]https://github.com/DCMLab/ABC

[5]This chapter is an elaborate version of the definitions given in Neuwirth et al. (2018) and Moss et al. (2019b).

Gárdonyi and Nordhoff, 2002) to indicate whether the chords are built on major (uppercase) or minor (lowercase) triads. Further alterations of the Roman numerals are common. A variety of them is introduced below.

Formalized harmonic annotation systems. A few recent approaches for formal representations of chord symbols are reviewed in this section. Mauch et al. (2008) use a chord representation that includes the chord root as an absolute note along with a variety of chord types, encoded as interval patterns above the root. Temperley (2009) uses a more abstract representation of chords in Western classical music and encodes only the root of the harmonies as Roman numerals. A more elaborate chord representation for Popular music was introduced by Harte et al. (2005) and Harte (2010) who define a chord ontology that includes the chord root as an absolute note, a list of intervals on top of that root, and a bass note relative to the root from which the chord inversion can be inferred. In the context of studying symbolic corpora of Rock music, de Clercq and Temperley (2011) and Temperley and de Clercq (2013) devise a system that is able to express the key, root, type, inversion of chords, and applied chords, but they acknowledge that they "did not attempt to fully standardise our use of symbols following the Roman numerals" (de Clercq and Temperley, 2011, p. 55). Studying Jazz, Broze and Shanahan (2013) create an encoding of chord symbols that contains root, bass, and note suspensions. Cambouropoulos et al. (2014) have a broader stylistical scope and encode chords as interval vectors which renders their encoding more universal but comes with a loss of readability. Based on this previous work, we developed a formal system for harmonic annotations that is capable of expressing a rich set of harmonic features while at the same time being close to traditional music theoretic notation.

A comprehensive model of chord morphology. Given the many historical and contemporary attempts to deal with vertical co-ocurring harmonic events, the question arises whether a comprehensive system is possible or even desirable. We have already discussed that the choice of representation depends in part on the goals of the researcher or analyst, and on the underlying model of what a chord is. While the most general approach takes 'everything that sounds together' as a definition, more theoretically motivated approaches require that chords are constituted by stacked thirds (e.g. Rameau, 1722). Moreover, a formal representation of chords needs an exact specification that allows it to be read by computers. Consequently, the trade-off between machine- and human-readability is a serious challenge. The representation system used here and published in Neuwirth et al. (2018) proposes one solution. It establishes a chord annotation system that makes relatively few theoretical assumptions in order to be as general as possible but also prevents arbitrary individual choices by formally defining the morphology for chord symbols. Its assumptions and their formalizations are stated below in detail. In general, the harmonic annotation system is based on RNA. This choice was made because of its ubiquity in historic and contemporary music theory. The present system incorporates the assumption that harmonies are based on a root note, expressed as a Roman numeral, on which thirds are stacked. In the following paragraphs it is explained which aspects of harmony

level a:	C											
level b:	C							G				
level c:	C				E			G				
level d:	C		D		E	F		G		A		B
level e:	C	D♭	D	E♭	E	F	F♯	G	A♭	A	B♭	B

Figure 5.5 – Schematic representation of a key as a hierarchy of notes after Lerdahl (2001).

can be expressed by the current annotation system and how.

Global keys. One of the most central music theoretical concepts for tonal harmony is that of a *key*. A key is defined by a collection of notes, a *scale*, and an associated hierarchy of prevalence in this scale, with one note, the *tonic*, being the most important one. For example, the key of F major is defined by the scale {F, G, A, B♭, C, D, E} and a hierarchy that states that the most prominent note is F and the second prominent one is the note on the fifth scale degree, C, (sometimes called the *dominant*). This model of a key corresponds to the *Tonal Pitch Space* model (TPS; Lerdahl, 1988, 2001) which is schematically depicted in Figure 5.5.[6]

The hierarchy defines also the *mode*, i.e. whether it is a major or a minor key. The scale mentioned above can either constitute the basis for the key of F major or D minor, depending on the tonal hierarchy. Major keys are denotated by uppercase letters and minor keys by lowercase letters. It is thus sufficient to specify the tonic note and whether it is capitalized or not in order to know the key. For instance, G designates G major and g means G minor. We can thus specify the *natural key symbols* as

$$\mathcal{A}_{\text{nat}} = \{\mathsf{A}, \mathsf{B}, \mathsf{C}, \mathsf{D}, \mathsf{E}, \mathsf{F}, \mathsf{G}, \mathsf{a}, \mathsf{b}, \mathsf{c}, \mathsf{d}, \mathsf{e}, \mathsf{f}, \mathsf{g}\}. \tag{5.1}$$

Accordingly, the *global key alphabet* is

$$\mathcal{A}_{\text{glob}} = \left(\mathcal{A}_{\text{nat}} \times \{\mathsf{b}\}^*\right) \cup \left(\mathcal{A}_{\text{nat}} \times \{\#\}^*\right), \tag{5.2}$$

where the asterisk (*) denotes that arbitrary many accidentals can follow the letter—e.g. for the keys F♯ major or b♭ minor—but not in combination. A key symbol such as C♯♭ is not contained in $\mathcal{A}_{\text{glob}}$. Accidentals are represented by the symbols b and #, respectively. One assumption of the present model is that a tonal piece always has exactly one global key which has to be stated in the first chord symbol of a piece and cannot be changed throughout the piece. All

[6]Note that this schematic should not be understood as a bar plot. While the vertical dimension represents the different hierarchical levels, it is the horizontal dimension on each level that describes which notes belong to which level.

other chord symbols are defined in relation to it. Because of the theoretically arbitrary number of accidentals, the size of $\mathcal{A}_{\text{glob}}$ is infinite. However, in practice keys with more than one accidental are extremely rare.

Local keys. Over the course of a musical piece, the key can change from the global to one or several subordinate *local keys*. This can happen by either an instantaneous change or a gradual process called *modulation*. It is not always clear where a new local key begins exactly, for instance because one or several *pivot chords* are involved. A pivot chord is a chord that has two different functions, one in the old and another in the new key. The annotators were encouraged to indicate key changes as early as it made sense to them. Local keys vary with respect to their distance to the global key. In general, one can specify the relation of a local key to the global key as the Roman numeral of that scale degree in the global key which is equal to the scale degree of the tonic in the local key. For example, if the global key is F♯ minor and the local key is A major, the tonic of the local key is *A* which is the third scale degree in F♯ minor. The local key would consequently be denotated as III. The fact that the local key is in the major mode is expressed in the uppercase symbol. The alphabet of *Roman numerals* is thus defined as

$$\mathcal{A}_{\text{RN}} = \{\mathsf{I}, \mathsf{II}, \mathsf{III}, \mathsf{IV}, \mathsf{V}, \mathsf{VI}, \mathsf{VII}, \mathsf{i}, \mathsf{ii}, \mathsf{iii}, \mathsf{iv}, \mathsf{v}, \mathsf{vi}, \mathsf{vii}\} \tag{5.3}$$

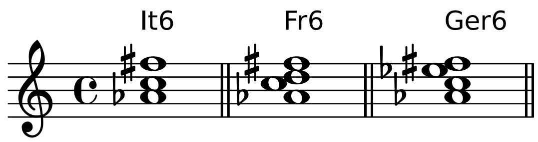
Sometimes the tonic of the local key is not among the scale-degrees of the global key. In this case, alterations have to be applied. For instance, if the global key is again F♯ minor, and the local key is C major, the local key symbol would be bV because the fifth scale degree in F♯ minor is C♯ and C is its flattened version. Accordingly, the alphabet for *local keys* is

$$\mathcal{A}_{\text{loc}} = \left(\{\mathsf{b}\}^* \times \mathcal{A}_{\text{RN}}\right) \cup \left(\{\#\}^* \times \mathcal{A}_{\text{RN}}\right). \tag{5.4}$$

Again, uppercase letters stand for major keys and lowercase letters stand for minor keys. Note that, by convention and in contrast to the global key alphabet, the accidentals precede the Roman numeral. If a global key is given in a chord symbol, the symbol from $\mathcal{A}_{\text{RN}}$ has to be enclosed between two periods (.) to prevent MuseScore from parsing and rewriting it to its internal chord representation. If a local key is specified, it also has to be separated from the chord information by a period but only needs to be preceded by a period if it starts with a flat (b). If a chord symbol does not contain the indication of a global or local key, it is implicitly assumed that it refers to the key that was last specified.

Chord roots. As mentioned before, the chord model employed here presupposes that chords always have a root. Accordingly, the root is an obligatory part of any chord symbol, in fact the only one. The root of a chord is specified in relation to the tonic of the local key in which it occurs. Because both local keys and chord roots are specified in relation to some tonic, the same alphabet $\mathcal{A}_{\text{loc}}$ can be used for them. Local keys are described in relation to the absolute, global key of a piece or movement, and chords are described in relation to the local tonic. As with local keys, chord roots can be chromatically altered by a number of sharps (♯) or flats (♭), e.g. bII or #vii but these accidentals precede the Roman numeral by convention. Moreover, it is assumed that chords are stacks of thirds. The two most basic chord types are major and minor triads, which are expressed as uppercase and lowercase Roman numerals, respectively.

Although it is a fundamendal assumption of the chord representation outlined here that every chord has a root, it is clear that not every harmonic event can be considered as a chord in this strict sense. Harmonic events can also result from simultaneous contrapuntal phenomena and thus constitute an accidental instance rather than an event in itself. It may also be impossible to rearrange co-occurring notes to a stack of thirds which renders the assignment a root note difficult. Musical passages also might be motivated rather by voice leading or exhibit chromatic lines. For these and other cases where the annotator was not able to assign a chord symbol the special symbol @none could be used. In order to not only formalize the chord symbols but also to accommodate the system to the needs of music theorists, a small number of further symbols is included, namely the *augmented sixth chords* (Aldwell et al., 2010). The most common variants are called the *Italian*, *French*, and *German augmented sixths chords*, encoded as It6, Fr6, and Ger6, respectively. The three variants are shown in Example 5.1. They have been added to the annotation system because they are widely used in music analysis. The 6 in the respective symbols does not mean the same thing as with other roots. The three augmented sixth chord symbols should be understood as immutable. These chord symbols also have to be preceded by a dot to prevent reinterpretation by MuseScore.

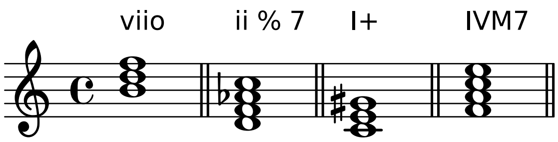


Example 5.1 – The three variants of augmented sixth chords. The Italian (It6), the French (Fr6), and the German (Ger6) augmented sixth chord.

Chord forms. Chords are defined as stacks of thirds on top of the chord root (Rameau, 1722). Since diatonic (in-scale) thirds are either major or minor, there are four possibilities to stack two thirds. A *major triad* has a minor third on top of a major third, and the reverse is the case for a *minor triad*. A *diminished triad* consists of two stacked minor thirds, and an *augmented triad* consists of two major thirds. While major and minor triads are implicitly encoded in the uppercase and lowercase spelling of the root, respectively, diminished and augmented triads are indicated by o and +, respectively. Two other common chord forms are the half-diminished seventh chord and the major-seventh chord. For half diminished chords, the root is followed by %. For major-seventh chords, it is followed by M. For the latter, it is customary to denote it with M7 which is technically speaking redundant but was nevertheless employed to make it easier to read for humans. For the former, % was chosen as symbol for half-diminished chords because it loosely resembles a struck-through circle, which is the common symbol in music theory textbooks for half-diminished chords. The alphabet of *chord forms* is thus

$$\mathcal{A}_{\text{form}} = \{\text{o}, +, \%, \text{M}\}. \tag{5.5}$$

Instances for each chord type are shown in Example 5.2.

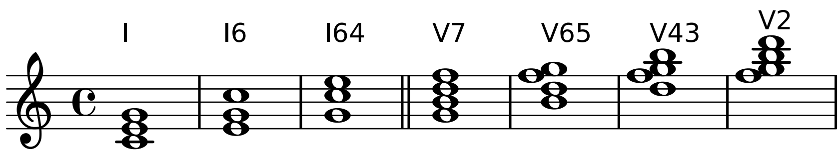


Example 5.2 – Different chord forms in C major. The diminished triad on the seventh scale degree (viio), the half diminished seventh chord on the second scale degree (ii%7, the augmented triad on the first scale degree (I+), and the major seventh chord on the fourth scale degree (IVM7).

Chord inversions. For all chords, the root defines its relation to the local key. Although 'root' suggests that it is also the lowest note and thus coincides with the bass, this is not always the case. Permutating the vertical order of notes is called *chord inversion*. Chord inversions are indicated by the interval pattern the inverted chords form, counted from the bass note upwards. The alphabet of chord inversions for triads and seventh chords is

$$\mathcal{A}_{\text{inv}} = \{6, 64\} + \{7, 65, 43, 2\}, \tag{5.6}$$

denoting the first and second inversion of triads (6 and 64) as well as the root position, first, second, and third inversion of seventh chords (7, 65, 43, and 2). The + in Equation 5.6 denotes the union of disjunct sets. Example 5.3 shows all inversions of a C major triad and of a G7 dominant seventh chord.



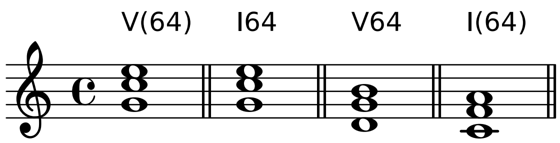
Example 5.3 – The C major triad in root position (I), first (I6) and second (I64) inversion, as well as the G dominant seventh chord in root position (V7), first (V65), second (V43), and third (V2) inversion.

Pedal tones. Sustained notes during longer sequences of chords are called *pedal tones*. Formally, pedal notes are indicated with square brackets around the chord sequence they support. For instance, in the chord sequence V[I I6 ii6 V2 I6 V43] there are six chords on top of the pedal note on the fifth scale degree of the local key. Since each scale degree can in principle function as a pedal note, the pedal note alphabet is the same as for local keys and chord roots.

Suspensions and added notes. Chord notes can be replaced by adjacent notes, called *suspensions*. Also, *added notes* can be included in a chord. Their alphabet is

$$\mathcal{A}_{\text{susp/add}} = \{\flat, \sharp\}^* \times \mathbb{N}_+ \tag{5.7}$$

and they are understood as distances from the root of the chord. All suspensions and added notes have to be included in parentheses, and added notes have to be preceded by a $+$ in order to distinguish them from chord suspensions. While, technically, any positive integer can be used for suspensions or added notes, chord symbols commonly only use integers up to 13. The notation with parentheses allows for an important analytical distinction between chord inversion and suspensions that is sometimes not disambiguated in the previously mentioned textbooks. Compare the four chords in Example 5.4. The first one (V(64)) denotes the so-called 'cadential six-four' which occurs in cadential contexts. The analysis expresses the interpretation that the root of the chord is V but the fifth and third of the chord have been temporarily displaced ('suspended') by the sixth and the fourth. The second chord (I64) on the other hand is identical in terms of the notes in the score but differs with respect to the analysis. In this case, it is interpreted as the second inversion of the tonic triad which can, for instance occur in a chord arpeggiation or in a passing-chord context. The third and fourth chords illustrate the reverse pattern in the symbols, leading to the second inversion of a chord on the fifth scale degree (V64) and a six-four suspension of the tonic triad in root position (I(64)).



Example 5.4 – Syntactic and semantic differences between chord inversions and suspensions. V(64) is a dominant triad with 64-suspension, I64 is a tonic triad in second inversion, V64 is a dominant triad in second inversion, and I(64) is a tonic triad with 64-suspension.

Applied chords. As stated before, the roots of chords are defined in relation to the tonic of the local key. One could make this implicit relation explicit by appending /I or /i to the chord symbol. For example, a V chord is five diatonic steps above the tonic I of the current local key which could be stated as V/I. But sometimes one wishes to express that a chord refers to a harmony other than the tonic on a very local level, so that one does not want to assume a key change. The most prominent examples are *applied dominants*. Applied chords create a local reference to another key, e.g. V2/iii, the minor key of the third scale degree, without defining that this is the new local key for the subsequent events. To be precise, although applied chords are commonly conceived as being applied to a chord other than the tonic, they are modeled here as applied to the tonic of the key in which the respective chord would be the tonic. For this reason, only uppercase or lowercase Roman numerals, possibly with alterations, can be used after the slash /.

Special characters. A small set of special symbols has already been introduced in the explanations above. If a key is stated in a chord symbol, it has to be separated from the chord information by a period (.); a plus sign (+) distinguishes added notes from suspensions, and the forward slash (/) indicates an applied chord.[7] Hence, we have

$$\mathcal{A}_{\text{sp}} = \{., +, /\}. \tag{5.8}$$

Assembling the parts. In order to check each chord symbol for syntactic validity, the restrictions elaborated above have been captured by CHORD,[8] a Perl-compatible *regular expression* (Wintner, 2002; Manning and Schütze, 2003; Jurafsky and Martin, 2009). CHORD is shown in Figure 5.6. The semantic validity of the annotations is ensured by the expertise of the annotators and correctors. The annotation process itself is detailed in section 5.1.

```
r"""^
(\.)?
((?P<key>[a-gA-G](b*|\#*)|(b*|\#*)
    (VII|VI|V|IV|III|II|I|vii|vi|v|iv|iii|ii|i))\.)?
((?P<pedal>(b*|\#*)(VII|VI|V|IV|III|II|I|vii|vi|v|iv|iii|ii|i))\[)?
(?P<root>(b*|\#*)
    (VII|VI|V|IV|III|II|I|vii|vi|v|iv|iii|ii|i|Ger|It|Fr|@none))
(?P<form>[\o+M%])?
(?P<inversion>(64|6|7|65|43|2))?
(\((?P<susp_add>(\+?(b*|\#*)\d)+)\))?
(/\.?(?P<applied>(b*|\#*)(VII|VI|V|IV|III|II|I|vii|vi|v|iv|iii|ii|i)))?
(?P<pedalend>\])?$
"""
```

Figure 5.6 – Regular Expression CHORD for chord symbols. The different named groups enclosed in angle brackets specify the regular expressions for the respective parts of the chord symbols.

5.3 Corpora of harmonic labels

A hierarchical model of corpora of harmonic labels. The regular expression CHORD introduced in section 5.2 captures the syntax of harmonic analyses for harmonic segments. For music analysis, it is not only of interest which harmonic units exist but also how they relate to each other. The following sections specify a model that expresses a hierarchical structure of corpora, pieces, and segments in which chords are the smallest units. The model specification proceeds from the definition of the set of all syntactically correct chords.

Definition 5.1 (Chord Universe) *The* chord universe $\mathcal{U}$ *is the set of all possible harmonic*

[7]The annotated corpora contain also phrase ending markers that occur, for instance, at cadences or ends of movements and sections. They have been excluded here because their usage varies greatly between annotators.

[8]CHORD is a slightly altered version of the regular expression that has been developed by Daniel Harasim and was employed in Neuwirth et al. (2018). For instance, phrase boundaries are not taken into account here.

annotations that match the regular expression CHORD*:*

$$\mathcal{U} = \{c \mid \text{match}(\textsc{Chord}, c) = \textsc{True}\}. \tag{5.9}$$

In particular, this definition implies that the chord universe is infinitely large. Although some parts of CHORD are supported by a finite alphabet (e.g. chord forms, chord inversions, and special characters), all parts based on Roman numerals or the global key symbols, as well as the alphabet for suspensions and added notes allow for an arbitrary number of preceding or following accidentals. Theoretically, the specification of a root $\#\#...\#\#$VI is possible although not very likely to occur in any actual analysis of a tonal piece. Note that the chord universe is not defined as a concrete set of chord labels but instead consists of all potential harmonic annotations that conform to CHORD.

In general, chords do not occur in isolation but within some musical context $\mathcal{K}$. A context can, for instance, be a piece, a movement, or a chord sequence. In such a context $\mathcal{K}$, a chord $c_i \in \mathcal{U}$ is preceded or followed by other chords,

$$...c_{i-2}\,c_{i-1}\,c_i\,c_{i+1}\,c_{i+2}... \tag{5.10}$$

There are numerous ways to create musical contexts (Hanninen, 2012). In the following, we define three fundamental musical contexts for this study: segments, pieces, and corpora. The most basic chord context used in this part of the thesis is the local key region in which a chord occurs. These regions are also called local key segments, or just *segments*.

Definition 5.2 (Segment) *A* segment *$\mathcal{S}$ is an ordered sequence of S chord symbols $c_1, ..., c_S$ from the chord universe $\mathcal{U}$ in which the first and the last chord of the sequence define the boundaries of a local key region:*

$$\mathcal{S} = (c_1, c_2, ..., c_S). \tag{5.11}$$

The size S of a local key segment $\mathcal{S}$ is given by the number of chords it contains, $|\mathcal{S}| = S$. Note also that, according to this definition, segments can contain multiple occurrences of the same chord symbol. Moreover, segments can be classified into either major segments $\mathcal{S}^{maj}$ or minor segments $\mathcal{S}^{min}$. This means that each chord has a unique function relative to the local key that is explicitly given by its relation to the local key of the segment in which it occurs. This key, in turn, is specified in relation to the global key, so that a chord's function with respect to the global key is given by the hierarchcal representation. All other functions in relation to other keys are implicit and have to be inferred. It follows that segments are non-overlapping and contiguous, which allows for the following definition of a musical piece.

Definition 5.3 (Piece) *A* piece $\mathcal{P}$ *is a partition of* $N_{\mathcal{P}}$ *segments:*

$$\mathcal{P} = \{\mathcal{S}_1, \mathcal{S}_2, ..., \mathcal{S}_{N_{\mathcal{P}}}\} = \mathcal{S}_{\mathcal{P}}^{maj} + \mathcal{S}_{\mathcal{P}}^{min}, \tag{5.12}$$

where "+" denotes the disjunct union of sets and $\mathcal{S}_{\mathcal{P}}^{m}$ denotes the set of segments with local mode m in piece $\mathcal{P}$. Note that, in this definition of a piece, the order of segments is not taken into account. This simplification is justified because the order of segments has no impact on either the chord frequencies contained in a piece and the intra-segment chord transition frequencies which are the main objects of the present study. This definition implies also that a piece is not merely a set of segments because the same segment can occur multiple times. A piece $\mathcal{P}$ is thus modeled as a multiset or bag of $N_{\mathcal{P}}$ segments.

Here in this part of the thesis, on the Mesoanalysis level, the focus lies not only on particular musical pieces but on collections of pieces, also called musical *corpora*. Most corpora are not mere collections of pieces but have an internal structure which may or may not correspond to the temporal order of composition, first performance, or publication. An obvious example are the different movements of a sonata of symphony. In almost all cases the order of movements is prescribed by the composer, although there are instances where a composer changed his or her mind later and reordered movements, e.g. in the case of Gustav Mahler's sixth symphony ("Tragische") where the order of the second (*Scherzo*) and third movement (*Andante*) from the first published version was exchanged in subsequent editions. Ralph Vaughan Williams' *Songs of Travel* contain a final ninth song ("I Have Trod the Upward and the Downward Slope") that is only to be performed to conclude the cycle as a whole.

With older pieces in particular, dates of composition or first publication can sometimes not be reconstructed. This applies for instance to the Köchelverzeichnis that contains all known compositions by Wolfgang Amadeus Mozart, or to posthumous publications of works by Frédéric Chopin. This highly complex issue is simplified here and a corpus is modeled as an unordered collection of pieces disregarding the internal structure, sometimes not even containing all pieces (as in the case of the Mazurkas) because digital scores were unavailable. For the corpus of Beethoven's string quartets it was decided to count movements as pieces because multimovement pieces would be much longer than the pieces in the other corpora. This model assumes also that all pieces are different from each other, otherwise modeling a corpus as a set would not contain all pieces.

Definition 5.4 (Corpus) *A* corpus $\mathcal{C}$ *is a set of* $N_{\mathcal{C}}$ *pieces:*

$$\mathcal{C} = \{\mathcal{P}_1, \mathcal{P}_2, ..., \mathcal{P}_{N_{\mathcal{C}}}\}. \tag{5.13}$$

The definitions above specify a hierarchical model of musical contexts, namely collection of corpora, individual corpus, piece, local key segment, and chord. Corpora are modeled as sets of pieces, each piece is a bag of segments, and segments are chord sequences within local key boundaries.

Chord types and chord tokens. All contexts on all hierarchical levels in the corpus model can contain multiple occurrences of the same chord symbol. The number of occurrences of a chord symbol c in a context $\mathcal{K}$, i.e. a corpus $\mathcal{C}$, a piece $\mathcal{P}$, or a segment $\mathcal{S}$, is given by the *count function*

$$\#_{\mathcal{K}} : \mathcal{U} \longrightarrow \mathbb{N}_0, \tag{5.14}$$

that assigns each chord symbol in the chord universe $\mathcal{U}$ its number of occurrence in $\mathcal{K}$. This function effectively maps most chords in $\mathcal{U}$ to 0 because they never occur in any actual corpus, such as the chord $\flat\flat\flat\flat\flat\flat ii$. The set $\mathbb{N}_0$ denotes the set of the positive integers including 0. The number $\#_{\mathcal{K}}(c)$ of occurrences of the chord symbol c in context $\mathcal{K}$ is also called the number of *chord tokens* for chord c in context $\mathcal{K}$. The number of chord tokens of a chord c in two contexts $\mathcal{K}_1$ and $\mathcal{K}_2$ is given by

$$\#_{\mathcal{K}_1,\mathcal{K}_2}(c) = \#_{\mathcal{K}_1}(c) + \#_{\mathcal{K}_2}(c) \tag{5.15}$$

and, accordingly, the number of chord tokens of a chord c in K contexts $\mathcal{K}_1, \ldots, \mathcal{K}_K$ is given by

$$\#_{\mathcal{K}_1,\ldots,\mathcal{K}_K}(c) = \sum_{i=1}^{K} \#_{\mathcal{K}_i}(c). \tag{5.16}$$

The support of $\#_{\mathcal{K}}$ consists of all the unique chords in context $\mathcal{K}$, also called the set of *chord types $t_{\mathcal{K}}$*. It is given by

$$t_{\mathcal{K}} = \text{support}(\#_{\mathcal{K}}) = \{c \in \mathcal{U} \mid \#_{\mathcal{K}}(c) \neq 0\} \tag{5.17}$$

and its size corresponds to the number of chord types in a context $\mathcal{K}$, also called the *vocabulary size*

$$V_{\mathcal{K}} = |\text{support}(\#_{\mathcal{K}})| = |\{c \in \mathcal{U} \mid \#_{\mathcal{K}}(c) \neq 0\}|. \tag{5.18}$$

The *vocabulary size* of context $\mathcal{K}$, the total number $N_{\mathcal{K}}$ of chord tokens in a context $\mathcal{K}$, is given by summation over all chords:

$$N_{\mathcal{K}} = \sum_{c \in \mathcal{U}} \#_{\mathcal{K}}(c). \tag{5.19}$$

The chord vocabulary and core. The previous sections defined a hierarchical model of chords in corpora and introduced the notion of chord types and chord tokens. This section introduces two central sets of chords, the chord vocabulary and the chord core.

Definition 5.5 (Chord vocabulary) *The* chord vocabulary $\mathbb{V}$ *is defined as the union over the sets of chord types in all corpora.*

$$\mathbb{V} = \bigcup_{\mathcal{C}} \{c \in \mathcal{U} \mid \#_{\mathcal{C}}(c) \neq 0\} \tag{5.20}$$

The cardinality V of the chord vocabulary is given by the number of chords in $\mathbb{V}$.

Definition 5.6 (Chord core) *For all corpora of harmonic annotations the (empirical)* chord core $\mathbb{C}$ *is given by the intersection of the sets of chord types over all corpora.*

$$\mathbb{C} = \bigcap_{\mathcal{C}} \{c \in \mathcal{U} \mid \#_{\mathcal{C}}(c) \neq 0\} \tag{5.21}$$

This definition presupposes that all corpora share at least one chord. The intersection would otherwise be empty because there would be no common core of chord types at all. Obviously, the size and content of the chord vocabulary and core depend only on the empirical corpora that are taken into account.

Chord frequencies and chord ranks. The count function $\#_{\mathcal{K}}$ maps a chord type to its frequency count in the context $\mathcal{K}$. In order to compare the frequencies of chords in multiple corpora of different size, it makes more sense to base this comparison not on the absolute but on the *relative frequencies* $f_{\mathcal{K}}(c)$ of chords c in context $\mathcal{K}$. This normalization is achieved by dividing the absolute frequency count $\#_{\mathcal{K}}(c)$ by the total number $N_{\mathcal{K}}$ of chord tokens in $\mathcal{K}$:

$$f_{\mathcal{K}}(c) = \frac{\#_{\mathcal{K}}(c)}{N_{\mathcal{K}}} = \frac{\#_{\mathcal{K}}(c)}{\sum_{c' \in \mathcal{U}} \#_{\mathcal{K}}(c')}. \tag{5.22}$$

These relative frequencies can be interpreted as estimators for the probabilities of occurrence of chord c, given a context $\mathcal{K}$, namely

$$p(c \mid \mathcal{K}) \approx f_{\mathcal{K}}(c). \tag{5.23}$$

The numbers of chord tokens or, equivalently, the relative chord frequencies induce a *ranking function* $r_{\mathcal{K}}$ that allows for an ordering of the chords in context $\mathcal{K}$ according to their (relative) frequencies:

$$r_{\mathcal{K}}(c) = \left|\left\{c' \in \mathcal{U} \;\middle|\; f_{\mathcal{K}}(c') > f_{\mathcal{K}}(c), \forall c \in \mathcal{U}\right\}\right| + 1. \tag{5.24}$$

Note that, following this definition, chords with the same frequencies will also have the same rank, meaning that a rank can be occupied by more than one chord. Note also that this study focuses on the corpus level, i.e. $\mathcal{K} = \mathcal{C}$. If the context to which the counts, relative frequencies, and ranks are referring to is clear, the subscripts of the functions defined above are omitted.

To summarize, a corpus is a set of pieces which can each be partitioned into segments here defined as to local key areas. Each segment thus defines a chord sequence in a local key. In contrast to the chord universe, which is of infinite size, corpora, pieces, and segments contain a finite number of chord labels. For the hierarchical levels of corpus, piece, and segment, one can also count the number of occurrences of each chord token that it contains, and also consider the set of unique chord types and their relative frequencies and ranks.

Chord progressions. The corpus model thus far only addresses occurrences of chords within specified contexts without taking into account the temporal order in which the chords appear in a segment. This section introduces the notion of a *chord progressions* which is central for music theory (e.g. Gárdonyi and Nordhoff, 2002; Aldwell et al., 2010). The most basic class of formal models that incorporate sequential patterns are so-called n-gram models (Manning and Schütze, 2003; Jurafsky and Martin, 2009). Applied to chords, the underlying assumption of an n-gram model is that the probability of a chord c_i from a sequence $c_1...c_{i-1}c_i$ does not depend on its full history $c_1 \dots c_{i-1}$, but only on the $n-1$ chords immediately preceding it,

$$p(c_i \mid c_1 \dots c_{i-1}) \approx p(c_i \mid c_{i-n+1} \dots c_{i-1}). \tag{5.25}$$

This approximation is also known as the Markov assumption. Because chord progressions describe the transition from one chord to another, the corresponding model is the *bigram model* ($n = 2$):

$$p(c_i | c_1 \dots c_{i-1}) \approx p(c_i | c_{i-1}). \tag{5.26}$$

To simplify the notation for the transition probability from a chord A to a chord B in a given mode $m \in \{\text{major}, \text{minor}\}$, we will use the notation $p_m(\mathsf{A} \to \mathsf{B})$. For example, if there is a probability of 40% of a V7 chord being followed by a I chord in a major context, we write $p_{\text{major}}(\mathsf{V} \to \mathsf{I}) = .4$. Naturally, chord progressions only express relations between pairs of chords. Larger n-grams such as trigrams, quadrigrams, etc. describe longer sequences and may be able to capture recurrent patterns on a larger scale. But there is a tradeoff as longer patterns are also less likely to occur many times. Moreover, for larger patterns in music, such as cadences, the constituent elements do not have to be strictly locally adjacent. In order to find such patternings, more sophisticated models such as *skip-grams* can be employed (Sears, 2016; Finkensiep et al., 2018). Here, we focus on the distributions of chords and chord progressions for which the simpler n-gram models are well-suited.

6 Empirical chord frequencies

An important aspect for the characterization of tonal harmony is the used chord vocabulary.[1] All subsequent analyses of the corpora of chord labels depend on the (empirical) *chord vocabulary*, or *chord lexicon*, $\mathbb{V}$, the set of all chords that they contain. The terms ‘lexicon’ and ‘vocabulary’ are used interchangeably. Here, we employ a unigram model to investigate structural regularities in the chord lexicon as well as a bigram model for the analysis of chord transitions.

6.1 Chord frequencies in the vocabulary and the core

First, we compare the counts of chord types and tokens as well as the numbers of segments in the different sub-corpora which are listed in Table 6.1. The number of segments ranges from only 11 in Debussy’s *Suite bergamasque* (four movements) to around one thousand in Beethoven’s string quartets (seventy movements), Chopin’s *Mazurkas*, and Medtner’s *Fairy tales*. These 19th-century composers seem to favor minor over major keys as witnessed by the number of minor segments. The cardinality of the set of minor segments $\mathcal{S}_{\text{min}}$ is generally larger than the number of major segments—leading to a major-to-minor segment ratio smaller than 1— or at least approximately the same, e.g. in the case of Grieg’s *Lyrical pieces*, were the ratio between major and minor segments is 57 to 55. Due to the large variance in the number of segments, the numbers of tokens and types vary as well, ranging from only 197 chord types (1,534 chord tokens) in Dvořák’s *Silhouettes* to 1,594 chord types (15,351 chord tokens) in Medtner’s *Fairy tales*. The ratio between the number of chord types and chord tokens (*type-token ratio;* TTR), which we take as an approximate measure of lexical diversity in the subcorpora (Milička, 2009, 2012), varies from Beethoven’s string quartets, the least diverse sub-corpus (TTR = .0402), to Debussy’s *Suite bergamasque*, the most diverse sub-corpus (TTR = .2717). The smaller the TTR value, the fewer types account for the tokens in the data. The TTR is maximal with TTR = 1 in the hypothetical case when each token appears only once

[1] Parts of this and the next chapter are based on the analyses in Moss et al. (2019b) that examines the *Annotated Beethoven Corpus* (ABC), the harmonic annotations of Beethoven’s string quartets. This study builds upon and extends that work by analysing and comparing corpora by other composers (see Table 6.1).

Table 6.1 – Overview of the number of segments ($\mathcal{S}$) in both modes ($\mathcal{S}_{\text{maj}}$ and $\mathcal{S}_{\text{min}}$), the number of types (V) and tokens (N), and the type-token ratio (TTR) in all sub-corpora as well as the respective values for the chord union and chord core.

Name	$\lvert\mathcal{S}\rvert$	$\lvert\mathcal{S}_{\text{maj}}\rvert$	$\lvert\mathcal{S}_{\text{min}}\rvert$	$\lvert\mathcal{S}_{\text{maj}}\rvert/\lvert\mathcal{S}_{\text{min}}\rvert$	$V_{\mathcal{K}}$	$N_{\mathcal{K}}$	TTR
Beethoven	929	358	571	0.63	1130	28095	0.0402
Schubert	95	25	70	0.36	312	3097	0.1007
Chopin	1040	394	646	0.61	654	11418	0.0573
Liszt	154	54	100	0.54	447	3377	0.1324
Dvořák	71	32	39	0.82	197	1534	0.1284
Grieg	112	57	55	1.04	928	8464	0.1096
Tchaikovsky	221	60	161	0.37	261	3028	0.0862
Debussy	11	4	7	0.57	276	1016	0.2717
Medtner	1105	161	944	0.17	1594	15351	0.1038
Union	–	–	–	–	3185	75380	0.0423
Core	–	–	–	–	43	–	–

and the number of types is equal to the number of tokens creating maximal diversity.

The union of all sub-corpora contains $V = 3,185$ chord types. We call it the empirical chord vocabulary $\mathbb{V}$. The size of the empirical chord core $\mathbb{C}$ based on the used sub-corpora is $|\mathbb{C}| = 43$. That is, only 43 out of 3,185 chord types are shared by all sub-corpora. This proportion of core chord types that is used in all sub-corpora is vanishingly small:

$$\frac{|\mathbb{C}|}{|\mathbb{V}|} = \frac{43}{3,185} \approx 0.0135. \tag{6.1}$$

Only slightly more than one percent of all chord types in the empirical chord vocabulary is contained in the core. These chord types are

$$\begin{aligned}\mathbb{C} = \{&\mathsf{I, I6, I64, i, i6, i64,}\\ &\mathsf{ii, ii6, ii7, ii65, iio, ii\%7, ii\%43, ii\%2,}\\ &\mathsf{iii, iii6, III,}\\ &\mathsf{IV, IV6, iv, iv64,}\\ &\mathsf{V, V6, V64, V7, V7(4), V65, V43, V2,}\\ &\mathsf{vi, vi6, VI, VI6,}\\ &\mathsf{viio6, \#viio6, \#viio43,}\\ &\mathsf{V65/III, V7/IV, V2/IV, V65/V, V43/V, V2/V, \#viio7/vi}\}.\end{aligned} \tag{6.2}$$

The chord core contains chords built on prevailingly diatonic scale degrees with several variants, the most varied being scale degrees ii and V. The core also contains a number of applied dominants, shown in the last line of Equation 6.2 Interestingly, there are no chords with flattened scale degrees in the core but a few instances of the sharpened seventh scale degree vii which can occur, for instance, in the minor mode, or in the case of an applied diminished chord such as #viio7/vi. It is surprising that the chord core contains neither the diminished triad on the seventh scale degree viio (in-scale in major segments) nor the first inversion of the major chord on the third scale degree (in-scale in minor segments) but inversions of these. Acknowledging that set intersection is a very strict operation—if only one of the sub-corpora does not contain a chord symbol, it will not appear in the core—it is astonishing that the empirical chord core resembles very much the small vocabulary of chords that is commonly treated in introductory music theory textbooks. One goal of this study is to show that, while the miniscule proporation of core chords do account for immense proportions of tonal music, the harmonic language is much richer. The computational methods employed here can help us to understand this richness better.

6.2 Similarity of corpora based on chord frequencies

How do the chord vocabularies of the sub-corpora relate to each other and to the union and the core? In this section, we compare the respective chord collections under several measures. These are the *Jaccard similarity*, the *weighted Jaccard similarity*, and the *cosine similarity* (Levandowsky and Winter, 1971; Manning and Schütze, 2003). The Jaccard similarity of two sets A and B is defined as the ratio between their intersection $A \cap B$ and their union $A \cup B$,

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}, \tag{6.3}$$

where $|\cdot|$ denotes the cardinality of sets. The union contains all chord types that occur in either corpus, the intersection only those that are contained by both. We define $J(\emptyset,\emptyset) = 1$. This measure equals 0 if the two sets are disjunct and 1 if they are identical. The Jaccard similarities for all corpora are shown in Figure 6.1.[2]

Since J is defined on sets of chord types, we can also compare the sub-corpora to the chord type union and core. The similarities are relatively low which is not surprising since this measure is based on the intersection of chord types. The similarities with respect to the union and the core will be discussed further below. The Jaccard similarity is based on the corpus vocabularies, i.e. the chord types in the corpora but it does not account for the token frequencies of the chord types. To compare these, one can use the *weighted Jaccard similarity* between two sets A and B. It is defined as

$$J_w(A,B) = \frac{\sum_i \min(a_i, b_i)}{\sum_i \max(a_i, b_i)} \tag{6.4}$$

and is based on the relative frequencies of chord types in a corpus. These relative frequencies

[2]Note that the similarity matrices shown in Figures 6.1–6.3 are symmetrical because the similarity metrics are symmetric.

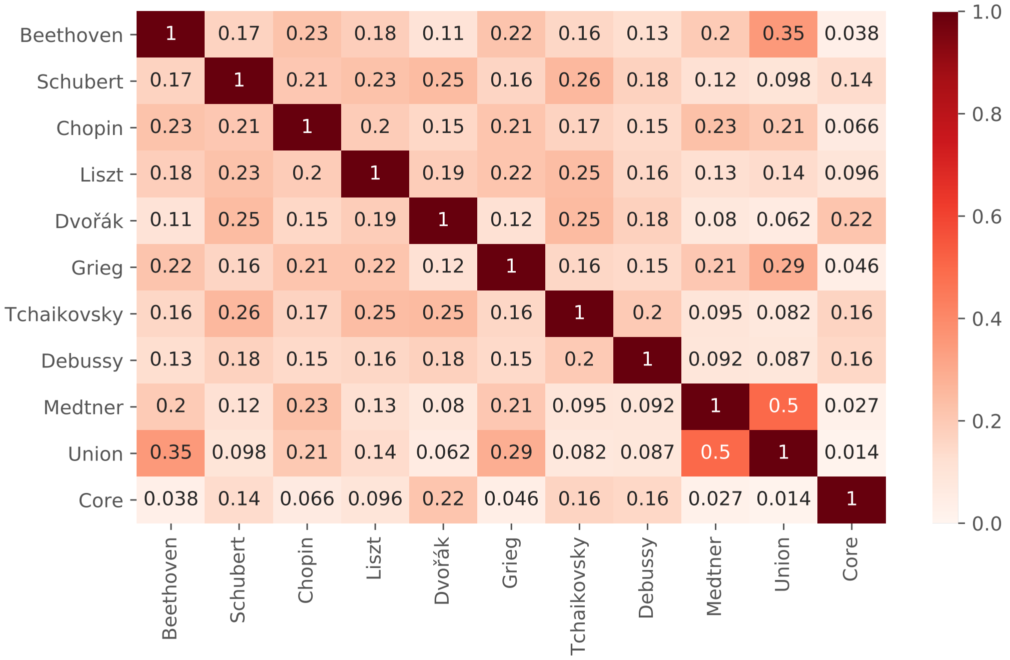


Figure 6.1 – Jaccard similarity J between corpora.

of a corpus can be represented as a vector $A \in \mathbb{R}^V$, where V is the vocabulary size, the size of the union of all corpora. The weighted similarities between all corpora are shown in Figure 6.2. Note that the union and the core are not shown since they are defined as sets of chord types from which we cannot derive relative frequencies.

The overall similarities are relatively low, $J_w(A, B) \leq .5$ for all sub-corpora A and B. Contrary to the simple Jaccard similarity J, the weighted measure J_w shows that the sub-corpora become generally more dissimilar over time, with the exception of Dvořák's *Silhouettes* and Tchaikovsky's *Seasons* which are most similar to Beethoven's string quartet under this measure. Broadly speaking, the corpora become less similar to each other with respect to chord usage over the course of the 19th century.

To add further support to this observation and to see whether this is an artefact of the specific choice of measurement, we employ yet another similarity measure and calculate the *cosine similarity* between two vectors given by the relative frequencies of chords. The *cosine similarity* between two vectors $A, B \in \mathbb{R}^V$ is defined as the cosine of the angle θ between them,

$$\begin{aligned} \mathrm{cossim}(A, B) &= \cos(\theta) \\ &= \frac{A \cdot B}{\|A\| \cdot \|B\|} \\ &= \frac{\sum_{i=1}^{V} a_i \cdot b_i}{\sqrt{\sum_{i=1}^{V} a_i^2} \cdot \sqrt{\sum_{i=1}^{V} b_i^2}}. \end{aligned} \tag{6.5}$$

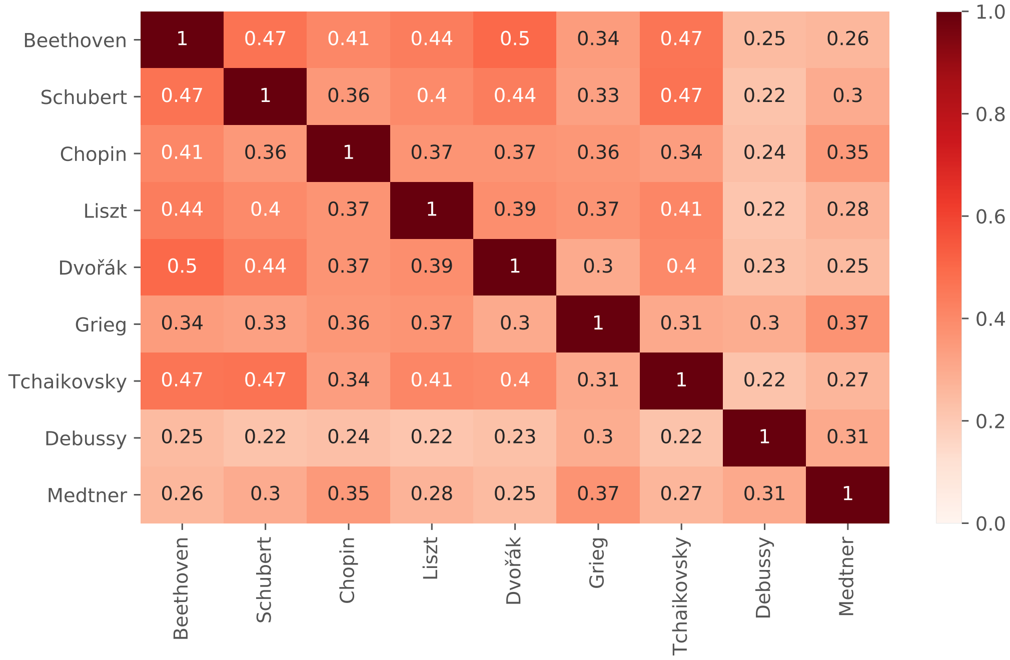


Figure 6.2 – Weighted Jaccard similarity J_w between corpora.

The results are shown in Figure 6.3. No cosine similarity value is less than .63 but the relative differences between the sub-corpora are similar to the ones under the weighted Jaccard similarity in Figure 6.2 which we take as a corroboration of the historical trend towards more diverse chord vocabularies.

It is important to note that the magnitude of the absolute values given by the different similarities J, J_w, and cossim should not be compared directly. Rather, one should compare the relative differences between them in order to attest the robustness of the findings under different metrics.

What is the relation between the tokens and types of a sub-corpus with the chord vocabulary and the core? More precisely, we can ask how much the chord types $t_{\mathcal{K}}$ from one specific sub-corpus contribute to the overall empirical chord vocabulary. A corpus vocabulary is richest if it contains all chords from the chord vocabulary $\mathbb{V}$ and the above ratio is equal to one. In other words, one can define the vocabulary *richness* for a corpus $\mathcal{K}$ as

$$\frac{|t_{\mathcal{K}} \cap \mathbb{V}|}{|\mathbb{V}|}. \tag{6.6}$$

These proportions are shown in the top panel of Figure 6.4 and correspond to the numerical values in the row second to the bottom in Figure 6.1 ("Union"). It is evident that the proportion of chord types from one corpus in the overall empirical chord vocabulary depends to some degree on the absolute number of types in that corpus. A smaller corpus contributes less

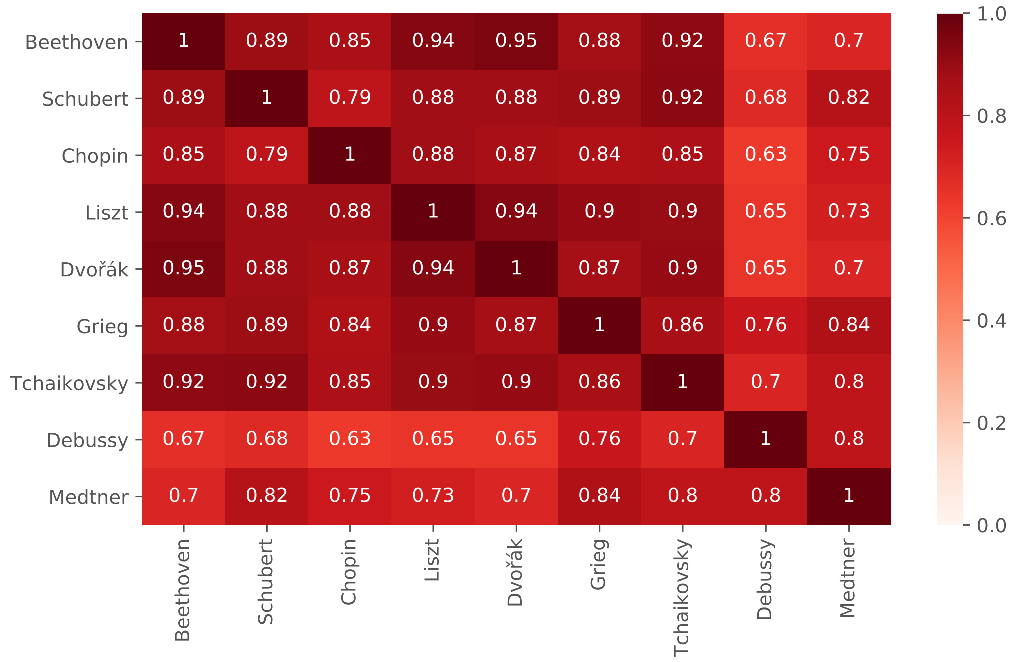


Figure 6.3 – Cosine similarity cossim between corpora.

than a large one to the chord vocabulary, and the chord types of the four largest corpora also have the four largest proportions of chord types in the overall chord vocabulary. The chord vocabulary was, after all, defined as the union of chord types over all sub-corpora. Note that the proportions displayed in the top panel of Figure 6.4 do not imply that the chord types contained in these proportions are shared between the different sub-corpora. For example, since Beethoven's string quartets contain 35% of all chord types in the union and Medtner's *Fairy tales* contain 50% of all chord types in the union, it would in principle be possible that these two sets of chord types do not share a single chord. That this is actually not true has been shown above and can be seen, for instance, in Figure 6.1 that shows that both sub-corpora share 20% of the union of their chord types.

Conversely, we can ask which proportion of the chord types $t_{\mathcal{K}}$ of a sub-corpus is taken from the chord types contained in the core, i.e. by the 43 chord types shown in Equation 6.2, and can define the vocabulary *typicality* as the ratio of core chords in the corpus vocabulary,

$$\frac{|t_{\mathcal{K}} \cap \mathbb{C}|}{V_{\mathcal{K}}}. \tag{6.7}$$

The vocabulary of a sub-corpus is most typical if all of its chords are shared by the other corpora, i.e. if all of its chords are contained in the core $\mathbb{C}$. This is shown in the center panel of Figure 6.4. The numerical values correspond to those in the bottom row of Figure 6.1. As can be seen, the proportions of the core in the respective sub-corpora are marginal. In the

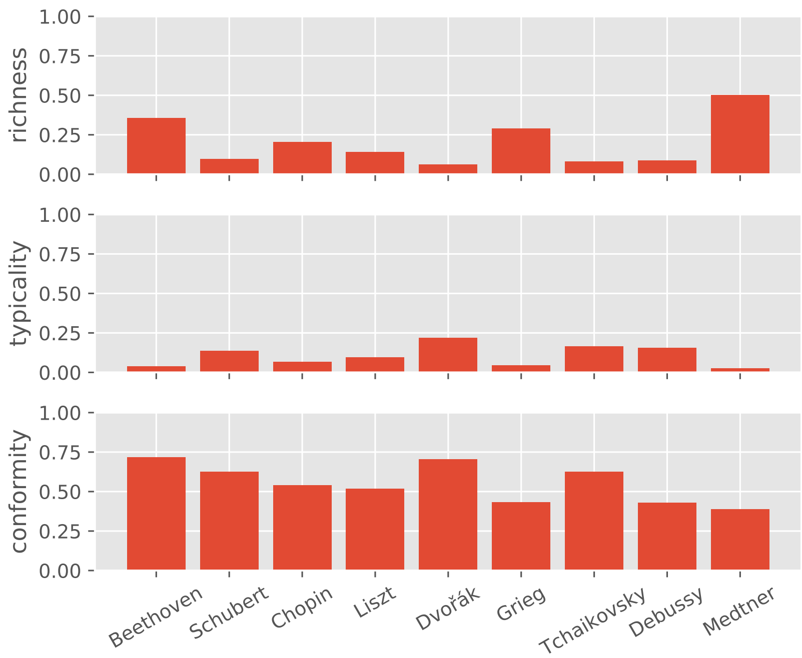


Figure 6.4 – Percentage of chord types of the respective corpora in the chord type union (top), of the core chord types in the respective corpora (center), and of core types among the corpus tokens (bottom).

case of Dvořák's *Silhouettes*, less than a quarter of its chord types are from the core. These proportions are particularly small in Beethoven, Chopin, Grieg, and Medtner. This means that the shared vocabulary between all corpora only constitutes a small fraction of the whole set of chord types in any sub-corpus. Note that, in contrast to the top panel of Figure 6.4 where the ratios were taken relative to the union of chord types of all corpora, the ratios in the center panel are calculated with respect to the set of types in the individual corpora separately. This is obviously also dependent on corpus size. The most striking difference between the top and the center panel of Figure 6.4 is shown for the Medtner corpus. Its chord types constitute a larger proportion of the chord vocabulary than all other corpora (50%) but the proportion of core chord types in this corpus is only 2.7%, smaller than in any other corpus. To conclude, the chord core only constitutes a fraction of the set of chord types of any sub-corpus.

This picture changes drastically if one takes not only the chord types but also their frequency of occurrence into account. One can define the *conformity* of a corpus as the amount of its tokens that are shared with all other corpora,

$$\frac{\sum_{c\in\mathbb{C}} \#_{\mathcal{K}}(c)}{N_{\mathcal{K}}}. \tag{6.8}$$

Accordingly, a corpus is most conform if all of its tokens are taken from the chord core $\mathbb{C}$. In this case, it is also most typical according to Equation 6.7. Compare the striking difference between the center and the bottom panel of Figure 6.4. The bottom panel shows which proportion of the chord tokens in each sub-corpus consists of chord tokens from chord core. For instance, almost 75% of the chord tokens of all of Beethoven's string quartets consists of chords from the chord core which itself only consists of approximately 1% of chords from the empirical chord vocabulary. In other words, 1% of the chord vocabulary from all sub-corpra accounts for almost three quarters of the chords in Beehoven's string quartets. In other words, an extremely small set of chords accounts for a vast majority of chords in these pieces.

6.3 Chord types and chord tokens: Heaps' law

The type-token ratio (TTR) is a common measure for lexical diversity (Milička, 2009, 2012) but its definition is not unproblematic. Recall that it is defined as the ratio of the number of chord types $V_{\mathcal{K}}$ and the number of chord tokens $N_{\mathcal{K}}$ in a given context $\mathcal{K}$,

$$\mathrm{TTR}(\mathcal{K}) = \frac{V_{\mathcal{K}}}{N_{\mathcal{K}}}. \tag{6.9}$$

The TTR values for all sub-corpora were reported in the rightmost column of Table 6.1. This value is always positive, and maximally equal to 1 if there are as many chord types as there are chord tokens, i.e. if each chord is used exactly once. It is smaller than 1 for the corpora studied here, reflecting that many chord symbols are used multiple times (see Tables 6.1, 6.2, and 6.3). The TTR suggests that the number of types in a context $\mathcal{K}$ is proportional to the number of tokens in that corpus. Rearranging the previous equation estimates the number of tokens as

$$V_{\mathcal{K}} = \mathrm{TTR}(\mathcal{K}) \cdot N_{\mathcal{K}}, \tag{6.10}$$

hence modeling the relationship between types and tokens as a linear one. Because of its dependence on the corpus size (the number of tokens $N_{\mathcal{K}}$), comparing the TTR of several corpora makes only sense if they have a similar size. Bearing this in mind, we can better interpret the TTR values in Table 6.1 and state that the lexical diversity in the Debussy's *Suite bergamasque* is higher than in Liszt's *Années de pélerinage*, that of Tchaikovsky's *Seasons* is higher than that of Schubert's *Winterreise*, that of Grieg's *Lyrical pieces* is higher than Chopin's *Mazurkas*, and that of Medtner's *Fairy tales* is higher than Beethoven's string quartets. It is hence supported that the composers towards the end of the 19th century employ a richer chord vocabulary than their early 19th-century predecessors.

A more general approximation of the relation between the number of types $V_{\mathcal{K}}$ and number of tokens $N_{\mathcal{K}}$ is given by *Heaps' law* (Heaps, 1978; van Leijenhorst and van der Weide, 2005). It models the relation between types and tokens as a power law,

$$V_{\mathcal{K}} \approx a \cdot (N_{\mathcal{K}})^b, \tag{6.11}$$

with a proportional parameter $a \in \mathbb{R}$ and an exponential parameter $b \in \mathbb{R}$. For $b = 1$, Heaps'

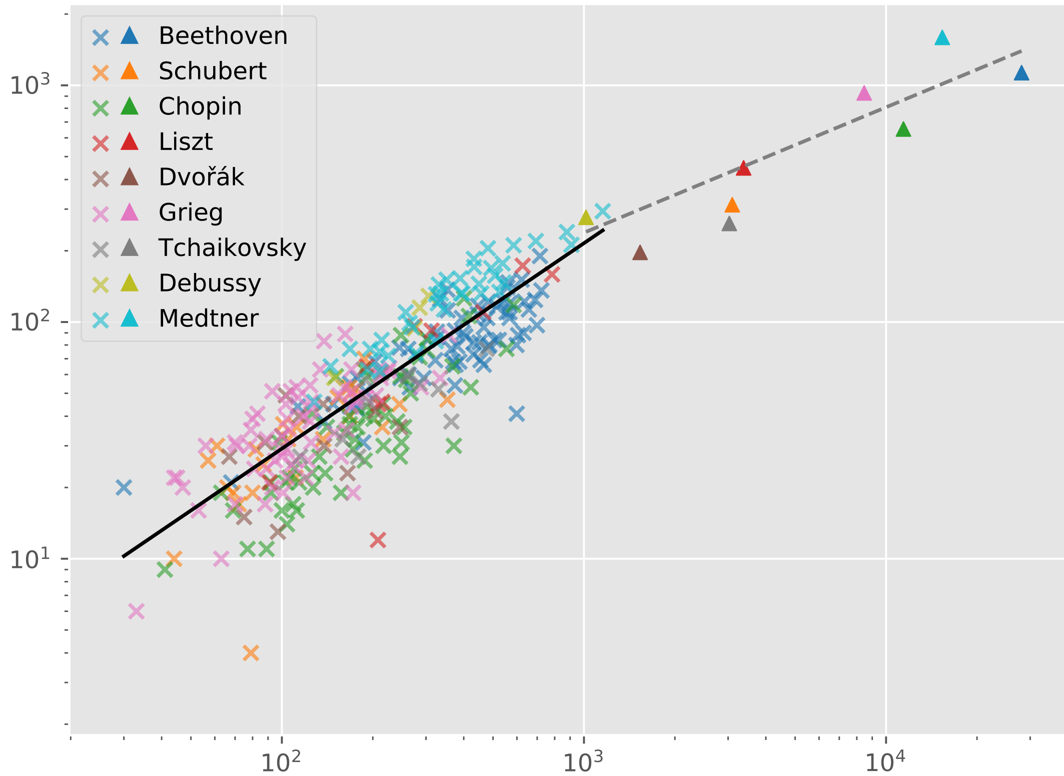


Figure 6.5 – Number of types vs. number of tokens in all pieces (crosses) and corpora (triangles) on a log-log scale. The lines show Heaps fits for the corpora (dashed gray, $a = 6.13, b = .53$) and all pieces in them (solid black, $a = .563, b = .868$).

law simplifies to the relation described by the TTR.

The empirical frequencies of chord tokens $N_{\mathcal{K}}$ versus the number of types $V_{\mathcal{K}}$ for all sub-corpora (triangles) as well as for all pieces in these sub-corpora (crosses) is plotted in Figure 6.5. The dashed gray and solid black lines show the Heaps fits for all corpora ($a = 6.13, b = .53$) and all pieces in them ($a = .563, b = .868$), respectively. The exponential parameter b is smaller than 1 for both the chords in the entire sub-corpora (triangles) and in the individual pieces (crosses). One can conclude that the relation between chord tokens and chord types in these datasets is better modeled by Heaps' power law (Equation 6.11) than by a linear model, such as the TTR (Equation 6.9).

6.4 Chord tokens in major and minor segments

In the previous section we looked at the size of the chord vocabularies employed by the different composers. Here, we take a closer look at the actual distributions of chords in the sub-corpora. We also study the differences between chord distributions in major and minor segments separately in order to get a more fine-grained picture of the composers' idiomatic usage of chords in the two modes and hence of the composers' treatment of tonality. We will see that few of the most frequent chords account for most of the chord tokens within the sub-corpora and that these chords can accordingly be identified as being central for the respective sub-corpora. To assess the importance of these central chords, we will compare their relative frequencies rather than their absolute occurrences since the sub-corpora are of very different sizes (see Table 6.1).

Chord tokens in major segments. The relative frequencies $f(c)$ and ranks $r(c)$ of the 25 most frequent chord types c in the major segments for all nine sub-corpora are displayed in Table 6.2. The frequencies are calculated relative to the respective sub-corpus size N_C, e.g. 15.9% of all chord tokens in Beethoven's string quartets are of type I but only 9.8% in Grieg's *Lyrical Pieces* have this chord type.

In Beethoven's string quartets, it is striking that the first ten chord types are tonics, dominants, and the subdominant, i.e. chords with roots I, V, or IV. While tonic chords occur as variants I and I6 (root position and first inversion), the subdominants occur only in root position. The dominants, in contrast, exhibit a far bigger variety, consisting of dominant seventh chords in root position (V7), and all three inversions (V43, V65, V2), as well as dominant triads in root position (V) and first inversion (V6). These first ten chords together account for 55%, that is more than half of all chords in the major segments. The following chords include variants of chords on the second (ii and ii6), sixth scale degree (vi and vi6), and seventh scale degree (viio6 and viio). Moreover, the minor tonic i occurs on rank 16, but only with a relative frequency of 1.4%. Many of the next ranking chords belong to the class of applied chords (V2/IV, V65/V, V7/V, V7/IV, V65/IV).

The chord distribution in *Winterreise* is largely similar but also shows some differences. Again, one can see that the top ranking chords have either root I or V, but the triad on scale degree vi, while being less frequent, ranks higher in Beethoven. Moreover, its applied dominant V/vi has rank 14 with a relative frequency of 1.1%. Together this might indicate that local harmonic progressions in the relative key are more common in *Winterreise* than in Beethoven's string quartets but are not so prominent as to lead the annotator to assign a modulation to a new key. Another observation is that suspension chords are more common than in Beethoven's string quartets. In both the string quartets and the *Winterreise*, the cadential V(64) suspension chord ranks among the first 25 chords in major, but in Schubert we also see suspensions of sixths, ninths, fourths, and even minor ninths (V7(6), V7(9), V(4), V7(b9)), strikingly all on the fifth scale degree.

In the case of Chopin's *Mazurkas*, the highest ranks are also occupied by chords with roots I, V, IV, or ii, but the triad on scale degree iii appears on rank 12 in root position. It is as frequent as viio7 chords and much more frequent than in the previous corpora. Its applied dominant seventh chord V7/iii, although ranked only 21st, is more frequent than in the other corpora as

Table 6.2 – 25 most frequent chord types in major segments.

	Beethoven		Schubert		Chopin		Liszt		Dvořák		Grieg		Tchaikovsky		Debussy		Medtner	
$r(c)$	c	$f(c)$	c	$f(c)$	c	$f(c)$	c	$f(c)$	c	$f(c)$	c	$f(c)$	c	$f(c)$	c	$f(c)$	c	$f(c)$
1	I	0.159	I	0.176	I	0.188	I	0.165	I	0.235	I	0.098	I	0.172	I	0.070	I	0.123
2	V7	0.081	V7	0.095	V7	0.114	V7	0.085	I6	0.086	V	0.060	V7	0.071	iii	0.052	V	0.054
3	I6	0.068	V	0.088	V	0.046	V	0.061	V7	0.082	V7	0.047	V	0.047	IV6	0.048	V7	0.043
4	V	0.063	I6	0.072	@none	0.041	vi	0.034	V	0.071	IV	0.032	V(64)	0.040	V7(+2)	0.048	IV	0.023
5	IV	0.037	V(64)	0.054	IV	0.037	IV	0.033	IV	0.037	vi	0.028	I6	0.034	V7	0.043	vi7	0.022
6	V43	0.033	IV	0.030	I64	0.030	I6	0.026	viio6	0.031	I6	0.023	V6	0.033	IV	0.043	I6	0.020
7	V65	0.032	vi	0.021	V7/V	0.028	I(9)	0.023	V2	0.025	ii7	0.019	V43	0.030	ii	0.039	ii7	0.019
8	V6	0.027	I64	0.020	V9	0.024	ii	0.020	vi	0.016	V7(9)	0.018	I64	0.025	ii7	0.030	I64	0.018
9	V2	0.026	V2	0.020	ii	0.020	IV6	0.016	vi43	0.014	V43	0.018	ii	0.021	iii6	0.026	vi	0.016
10	V(64)	0.026	V43	0.018	vi	0.016	V7(+6)	0.015	vii%43	0.014	I64	0.018	vi7	0.018	viio6	0.026	ii	0.016
11	ii6	0.026	ii6	0.017	V7(6)	0.016	I(+9)	0.014	vii%43/V	0.013	V7/V	0.015	I(6#4)	0.016	I6	0.026	IV64	0.014
12	ii	0.025	IV6	0.013	iii	0.013	V7(4)	0.014	ii	0.010	ii	0.013	@none	0.016	IV64	0.022	IVM7	0.014
13	vi	0.023	V7/IV	0.013	viio7	0.013	V7(9)	0.014	i/iv	0.010	V64	0.012	vi	0.015	V	0.017	V7/IV	0.014
14	I64	0.020	V7/vi	0.011	V7/ii	0.012	I64	0.013	V6	0.009	IM7	0.011	viio2	0.014	vi	0.017	V9	0.012
15	IV6	0.019	bVI	0.011	V7/IV	0.011	V7(6)	0.012	V(6)	0.009	vi7	0.010	V7(9)	0.014	V43/IV	0.017	IV6	0.012
16	i	0.014	V7(6)	0.010	ii6	0.010	V65	0.011	V65/V	0.008	iii	0.010	V2	0.013	IVM7	0.017	I(4)	0.011
17	viio6	0.010	V7(9)	0.009	I6	0.010	I(+6)	0.010	V65	0.008	IVM7	0.009	V65	0.013	V2	0.017	V6	0.010
18	viio	0.008	V(4)	0.007	V65	0.009	viio	0.008	V(64)	0.008	ii65	0.009	bVI+	0.013	vi43	0.017	V7/V	0.009
19	V2/IV	0.007	V43/V	0.007	V(6)	0.009	iv	0.008	V43(4)	0.008	IV6	0.008	viio43	0.012	V6	0.017	V43	0.009
20	V65/V	0.007	V7(b9)	0.007	viio7(4)	0.008	viio7	0.008	ii%43	0.008	V2	0.008	vi7(4)	0.010	vii%65	0.017	vi6	0.008
21	V7/V	0.007	ii	0.007	V7/iii	0.008	ii7	0.007	ii7	0.007	V7(6)	0.007	IV6	0.010	ii65	0.013	iii	0.008
22	V64	0.006	ii%43	0.007	IV64	0.007	IM7(+6)	0.007	viio	0.007	IM2	0.007	ii%2	0.010	V65	0.013	V7(4)	0.008
23	V7/IV	0.005	viio7	0.006	I(42)	0.006	V7/IV	0.007	vii%7	0.007	ii6	0.007	V(4)	0.010	IV(+6)	0.013	vi65	0.008
24	vi6	0.005	i	0.006	ii7	0.006	vi(9)	0.007	viio43	0.007	V6	0.006	V64	0.009	III6	0.013	@none	0.008
25	V65/IV	0.005	iio6	0.006	V/V	0.006	ii%43	0.007	V43	0.007	V7/IV	0.006	#viio2/iii	0.009	vi7	0.013	iii6	0.007

well. It appears that the relative chord of the dominant (iii) is locally more prominent in the *Mazurkas*. Another observation is that the applied dominant seventh chord of the triad on the second V7/ii and the fourth (V7/IV) scale degrees have ranks 14 and 15, respectively. Overall, subordinated local tonicizations of iii and ii are more frequent than in the other corpora. Note also that the symbol @none is extremely frequent in this sub-corpus. Being ranked fourth, it constitutes 4.1% of all chords in the *Mazurkas*. This can be attributed to the composer's frequent use of chromatic lines and sometimes underspecified textures.[3]

Liszt's *Années de pèlerinage* likewise ranks tonic and dominant chords highest but also features the vi chord relatively prominently on rank 4. Moreover, Liszt employs certain chord suspensions and extensions much more frequently than the composers of the other sub-corpora, in particular ninths (e.g. I(9) on rank 7 and I(+9) on rank 11) and sixths (e.g. V7(6) on rank 15 and V7(+6) on rank 10). As the corpus similarities in Figures 6.1–6.3 have shown, the chord distributions in Dvořák's *Silhouettes* are most similar to the ones in Beethoven's string quartets. And indeed, there is a large overlap also among the 25 most common chords. A noteable difference is that Dvořák much more frequently employs half-diminished chords in several inversions on the second and seventh scale degrees ii% and vii%, respectively, that do not appear among Beethoven's most frequent chord choices.

Tchaikovsky's *Seasons* diverge from the observations we have made so far in that they prominently feature an augmented chord on the diminished sixth scale degree bVI+. This chord type occurs for the first time amongst the most common chords, both with respect to its root bVI and its quality (augmented, +). Another new chord type comes to the surface in this dataset: The half-diminished seventh chord on the second scale degree in third inversion ii%2. Both chords do not belong to the major scale and might indicate that *modal mixture* (Schenker, 1906), resulting in *chromaticism*, is common in the *Seasons*. A particular frequent use of chromaticism can also be seen in the relatively high frequency (rank 11) of I(6#4) chords that feature a double suspension, a diatonic (6) and a chromatic one (#4). We see also again a relatively high proportion (1.6%) of @none chords.

Grieg's *Lyrical Pieces* seem to be less chromatic when compared to *Winterreise* and *Seasons*. The 25 top ranking chords in major do not include any chromatic alterations that might reflect the general folk-like tone of these pieces. Nonetheless, three major seventh chord types, IM7, IVM7, and IM2, appear which are not to be found among the top ranking chords in the other corpora, hence revealing a harmonic particularity of the *Lyrical Pieces*. It is further noteable that applied chords do not occur with the exception of the double dominant V7/V. Recall that the frequencies in Table 6.2 are calculated with respect to major segments. Having this in mind, one can interpret the sparsity of applied chords with local harmonic homogenity of the major segments in this corpus.

Debussy's *Suite bergamasque* has a remarkably different distribution of the top chords. While the most frequent one is I as in all other sub-corpora, the second and third ranking chords

[3]Recall that the annotators were encouraged to use this symbol not too often.

are iii and IV with frequencies of f(iii) = .052 and f(IV6) = .048, respectively. Moreover, the frequency of the most frequent chord in major segments is lower than in any other corpus (f(I) = .07), accounting for the fact the this sub-corpus is most dissimilar to most other corpora.

Chord tokens in minor segments. For the minor segments in all sub-corpora, the relative frequencies $f(c)$ and ranks $r(c)$ of the 25 most frequent chord types c are shown in Table 6.3.

In general, they show a similar picture than the major segments: the first ranks are occupied by tonic and dominant chords, while other scale degrees occur in the lower ranks. The most frequent chord type in the minor segments of Beethoven's string quartets is i with a relative frequency of 9.2%. Note that this is substantially less than the 15.9% frequency of occurrence of I in major. Moreover, the fact that both I and I6 have quite similar frequencies as i and i6 is puzzling. It seems that the minor mode is more open to modal mixture than the major mode. This would also explain the occurrence of the major subdominant IV on rank 15 with a relative frequency of 1.5%. Noteable is further the first occurrence of a variant of the augmented sixth chord, namely the *German sixth* (see Section 5.2), Ger6 which we have not seen among the top 25 chords in the major segments.

Looking at Schubert's Winterreise, we see a largely similar distribution. Here, too, the major tonics I and I6 permeate the local minor key and occur relatively often (ranks 6 and 12, respectively). The relative chord to the tonic, III, occurs on rank 24 with a relative frequency of 0.6% and somewhat mirrors the occurrence of the relative chord to the major tonic vi in the major segments. The occurrence of the in-scale minor chord on the fifth scale degree constitues another instance of modal mixture that we already witnessed. The German sixth chord occurs also on rank 25.

In the minor segments of Chopin's Mazurkas, we see the only time that the first rank is not occupied by a tonic chord but by the dominant seventh V7. Again, we see a relatively high frequency of @none chord labels, and that the major tonic chord I is relatively frequent (rank 5). Compared to Schubert, the relative to the tonic chord III is even more frequent, as is its applied dominant seventh chord V7/III, and the parallel of the dominant v with its applied dominant seventh chord V7/v. Another noteworthy occurrence is the bII chord on rank 15, also called the *Neapolitan chord*. In contrast to the previous corpora, the German sixth chord is not amongst the top 25 chords. Some of the rather unusual chords in the minor segments in Liszt's *Années de pèlerinage* are the major versions of the chord on the fourth scale degree, IV on rank 13 and IV6 on rank 19 that are more frequent than the in-scale chord on the fourth scale degree iv6 on rank 20. The *Seasons* share with the other corpora that the top ranks consist of tonics, including the major tonics I and I6, and dominants (V7, V65, V2, V43, V(64), V6, V7(6)). They also exhibit a number of noteable differences. First, a number of half-diminished seventh chords on the second scale degree are relatively frequent (ii%43, ii%65, ii%7). Further, many chords with chromatically altered roots are among the top chords (#viio2, #vi%2, #viio6, and #viio43). In contrast to the other corpora, the relative to the tonic chords III does

Table 6.3 – 25 most frequent chord types in minor segments.

	Beethoven		Schubert		Chopin		Liszt		Dvořák		Grieg		Tchaikovsky		Debussy		Medtner	
$r(c)$	c	$f(c)$	c	$f(c)$	c	$f(c)$	c	$f(c)$	c	$f(c)$	c	$f(c)$	c	$f(c)$	c	$f(c)$	c	$f(c)$
1	i	0.092	i	0.158	V7	0.105	i	0.074	i	0.142	i	0.095	i	0.157	i	0.072	i	0.074
2	I	0.081	V	0.099	i	0.088	V	0.056	V7	0.097	V	0.057	V7	0.076	III	0.030	V	0.034
3	V	0.074	V7	0.062	@none	0.062	I	0.049	V	0.082	V7	0.029	I	0.061	IV	0.025	i6	0.031
4	V7	0.073	V(64)	0.057	V	0.061	V7	0.047	I	0.067	i6	0.024	i6	0.052	V7/III	0.025	@none	0.027
5	i6	0.038	i6	0.053	I	0.060	i6	0.044	i6	0.049	III	0.023	V	0.049	VI	0.023	V7	0.025
6	I6	0.036	I	0.031	V7/V	0.021	V(64)	0.024	I6	0.034	iv	0.021	V65	0.044	iv	0.021	i64	0.023
7	V65	0.031	iv	0.025	V9	0.019	bII6	0.020	V(64)	0.030	VI	0.018	V2	0.027	V7	0.021	iv	0.022
8	V43	0.026	V43	0.021	III	0.018	I6	0.020	iv	0.027	v	0.015	V43	0.026	I	0.021	VI	0.017
9	V(64)	0.025	iv6	0.020	v	0.016	V6	0.019	V2	0.021	I	0.013	V(64)	0.021	i6	0.020	III	0.016
10	V2	0.024	V(4)	0.016	i64	0.016	#viio7	0.017	I/III	0.018	#viio2	0.012	VI	0.020	v	0.015	v	0.014
11	V6	0.022	VI	0.015	V7(6)	0.015	Ger6	0.016	IV	0.018	iv7	0.012	I6	0.017	V	0.015	ii%7	0.014
12	iv	0.020	I6	0.014	V7/III	0.012	I64	0.015	V64	0.016	ii%43	0.011	V6	0.016	V7/VII	0.013	I	0.013
13	VI	0.018	V6	0.012	iv	0.012	IV	0.013	VI	0.015	i64	0.011	iv	0.015	V43	0.012	iv6	0.013
14	i64	0.017	V2	0.011	i6	0.011	V43	0.012	VI6	0.015	VI6	0.010	ii%65	0.013	i43	0.011	III6	0.011
15	IV	0.015	V65	0.011	IV	0.010	bII	0.011	V6	0.015	V64	0.009	Ger6	0.013	iii	0.011	iv7	0.010
16	vi	0.012	i64	0.010	bII	0.010	V7(9)	0.011	V7/III	0.010	i(+6)	0.009	ii%43	0.013	iv7	0.011	ii%65	0.010
17	iv6	0.012	V64	0.010	V64	0.010	III6	0.011	iv6	0.010	v6	0.009	V7(6)	0.013	V/III	0.011	V7/III	0.009
18	ii6	0.012	#viio43	0.008	V43	0.009	#viio43	0.010	iio6	0.010	bII	0.009	ii%7	0.011	i(2)	0.011	V6	0.009
19	I64	0.012	ii%65	0.007	iv64	0.009	IV6	0.010	V43(4)	0.009	VII	0.008	#viio2	0.010	i64	0.011	iv64	0.009
20	ii	0.011	IV6	0.007	VI	0.009	iv6	0.009	V/V	0.009	iv6	0.008	#vi%2	0.010	ii	0.009	i7	0.009
21	#viio7/V	0.008	ii%43	0.006	V7/v	0.009	VI6	0.009	ii%7	0.009	V43	0.008	i64	0.008	V2	0.009	IV	0.007
22	IV6	0.008	Ger6	0.006	I64	0.008	V2	0.008	I6/III	0.009	iv64	0.008	iv(+6)	0.008	i%7	0.009	VII	0.007
23	V64	0.008	#viio2	0.006	iv6	0.008	bII64	0.008	ii%65	0.009	III6	0.008	i(9)	0.008	III6	0.009	v6	0.007
24	#viio43	0.006	III	0.006	V7/iv	0.008	i64	0.008	V7/V	0.009	ii%7	0.007	#viio6	0.008	ii65	0.009	V9	0.007
25	Ger6	0.006	V7/bII	0.005	V6	0.008	ii	0.008	ii	0.007	V(6)	0.007	#viio43	0.007	vii7	0.008	VI6	0.007

not appear.

The chord distribution in Grieg's *Lyrical Pieces* is again less chromatic. With the exception of the major tonic I (rank 9), and the Neapolitan chord bII (rank 18), all top 25 chords are, in fact, in-scale chords.[4] An unusual chord among the top 25 chords in the minor segments of Grieg's *Lyrical pieces* is the triad v on rank 8 and its first inversion v6 on rank 17. It is the minor dominant on the fifth scale-degree that does not contain the leading-tone to the tonic and hence alludes a rather modal than tonal harmony. These observations fit well together with the folkloristic setting of the *Lyrical Pieces*.

The frequencies of the top three chords in the minor segments of Debussy's *Suite bergamasque* are similar to the ones in the major segments in that the tonic is less frequent than in the other corpora (f(i) = .072) and that the second and third most frequent chords are III and IV. The first dominant chord (to the local tonic) is ranked seventh, much lower than in any other corpus, with a frequency of only f(V7) = .021.

A final observation can be made about third scale degree in almost all sub-corpora. Aldwell et al. (2010) attest that there is a tendency in the minor mode to modulate towards the third scale degree but that the usage of chords on the third scale degree in major contexts is rare. Tymoczko (2003, p. 6) even suggests that the "iii chord is not a part of basic diatonic harmonic syntax" and could be neglected without much loss, in partiular because transitions from iii to I are supposedly "extremely rare"—although he acknowledges that they are slightly less so in the 19th century, the focal point of this study. Our data-driven approach does not contradict the theoretical assessment that chords on scale degree iii occure rarely but puts them in a different light. Although the relative frequency of iii chords is low, their ranks are relatively high, among the top 25 chords for most sub-corpora. This is especially meaningful considering that the distributions have very large vocabularies (see Table 6.1). Large-scale quantitative analyses thus can not only provide more accurate descriptions of the actual frequencies of harmonies but can also contextualize them by inspecting the shapes of the distributions within which they occur. This is the issue of the next section.

6.5 Chord frequencies and ranks: Zipf's law

The previous observations about chord frequencies and ranks were rather informal and pointed out salient observations regarding the 25 most frequent chords in all corpora, relating the findings to prior knowledge about the behavior of tonal chords. In Tables 6.2 and 6.3 one could already see that the relative frequency of chords decreases rapidly with increasing rank. In other words, there are few chords that occur many times while there are many chords that occur rarely, oftentimes only once. For instance, inspecting the chord symbol frequencies and ranks in major segments (Table 6.2), one can see, despite respective different relative frequencies, that the most frequent (rank 1) chord symbol is I for all sub-corpora. On the other

[4]This observation assumes the melodic minor scale that includes the dominant chords on root V or chords with root #vii.

hand, the the second ranked chord symbol is either V or V7, except for Dvořák where it is I6 and Debussy where it is iii. The ranks of the chord symbol vi, the relative minor chord to the tonic I, differ more: in Beethoven's string quartets, the rank of vi is 13; in the *Winterreise*, it is on rank 7; in Chopin's Mazurkas and Tchaikovsky's *Seasons* it has rank 10 and 13, respectively; and in Grieg's Lyrical pieces it appears on rank 5. Despite these differences, it is notable that this chord occurs always amongst the first 15 chord types. Recall that all corpora have much larger vocabulary sizes than 25.

Naturally, one wishes to see whether there is a systematic relationship between chord frequencies and chord ranks, given a corpus of chord labels. More precisely, one can inspect the frequency f of chords depending on their rank r. The 'long-tailed' behavior of the resulting rank-frequency distribution $f(r)$ of chords types is often modeled by a power law (Newman, 2005; Clauset et al., 2009; Broido and Clauset, 2018). This is a well-known behavior of datasets in computational musicology as well as linguistics and other domains (Manaris et al., 2005; Zanette, 2006; Mauch et al., 2008; Rohrmeier and Cross, 2008; Yang, 2013; Mehr et al., 2019), and it is customary to model this relationship with the *Zipf-Mandelbrot law* (Zipf, 1949; Mandelbrot, 1953). More explicitly, given the frequency rank r of chords, their frequency f can be approximated by a Zipf-Mandelbrot curve $\hat{f}$, given by

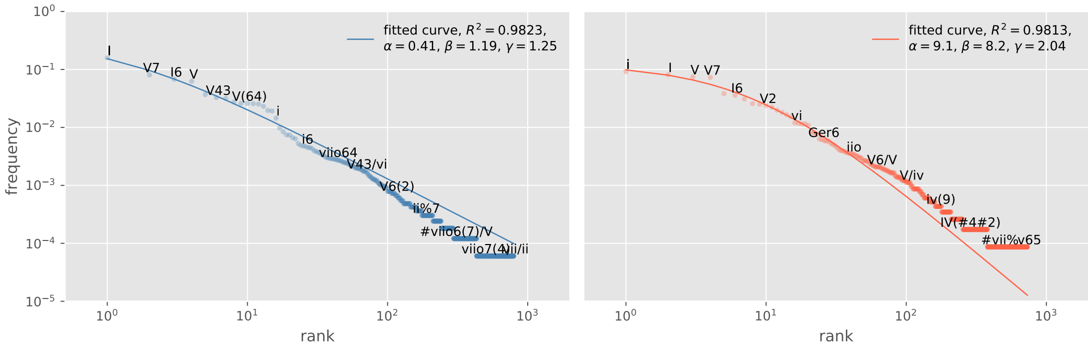
$$\hat{f}(r) = \frac{\alpha}{(\beta + r)^{\gamma}}. \tag{6.12}$$

The parameter α is a normalizing constant, ensuring that the estimated frequencies sum to one; the Mandelbrot offset parameter β determines the curvature of the fitted curve in log-log space; and the exponent γ determines the rate of the decay (Baayen, 2001). These parameters can be estimated empirically. Figure 6.6 plots the empirical frequency $f(r)$ and rank r of all chords in the five corpora (dots) in major segments (left, blue) and minor segments (right, red) on a log-log scale. Chords with identical frequencies occupy a range of ranks that corresponds to the horizontal arrangement of dots.

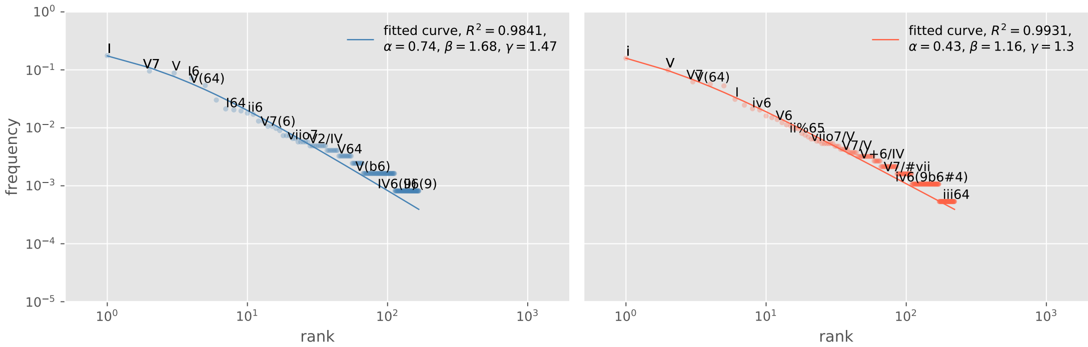
The solid line corresponds to the Zipf-Mandelbrot curve with the best fit. The accuracy of the fit between the empirical data and the estimated Zipf-Mandelbrot law is measured by the *coefficient of determination* R^2 (Izenman, 2008) with

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}, \tag{6.13}$$

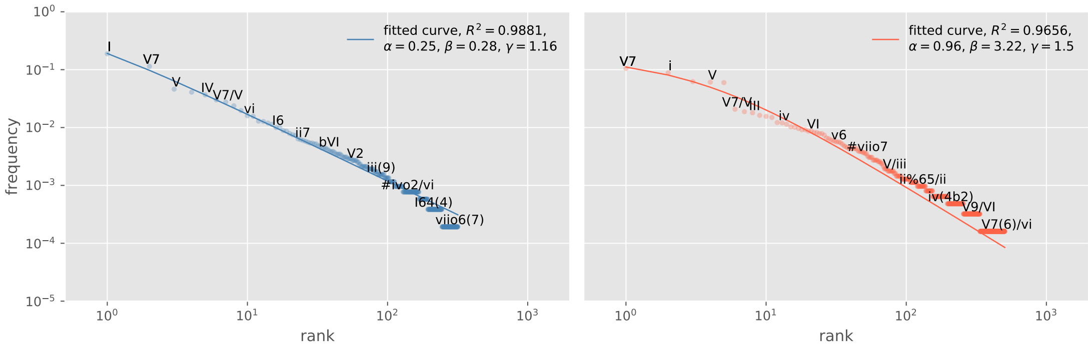
where $SS_{res} = \sum_r (f(r) - \hat{f}(r))^2$ is the *squared sum of residuals* measuring the sum of the squared distances of empirical to estimated frequency values $\hat{f}$. The *total sum of squares* is given by $SS_{tot} = \sum_r (f(r) - \bar{f})^2$ and corresponds to the sum of the squared differences of the empirical values to the mean empirical frequencies $\bar{f}$. The coefficient of determination is a suitable and widely used measure for the appropriateness of the curve fit because of its relation to the ratio of unexplained variance, despite the existence of more robust methods (Renaud and Victoria-Feser, 2010). Figure 6.6 shows that the curves are close to the actual values.



(a) Beethoven: String quartets.

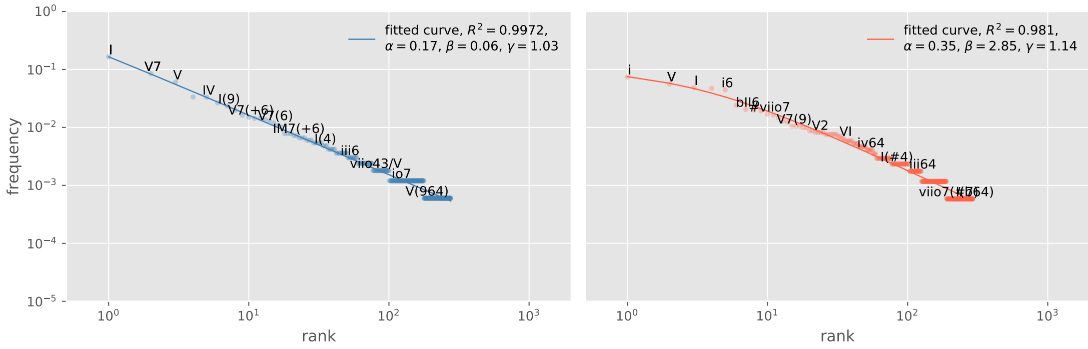


(b) Schubert: *Winterreise*.

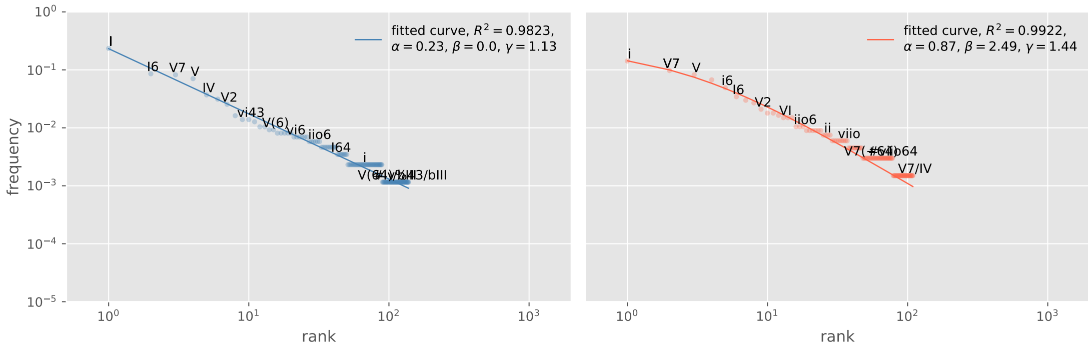


(c) Chopin: *Mazurkas*.

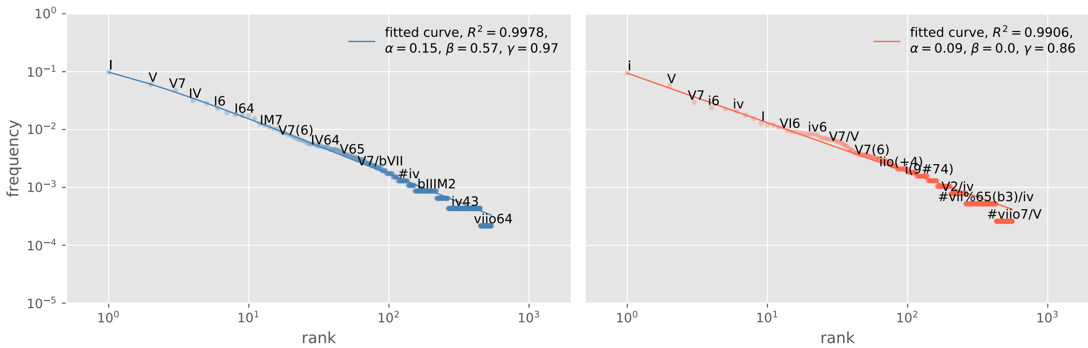
Figure 6.6 – Log-log plot of rank versus frequency of chords in major (left, blue) and minor (right, red). The solid line shows the best fit of a Zipf-Mandelbrot distribution as determined by the coefficient of determination R^2.



(d) Liszt: *Années de pèlerinage*.

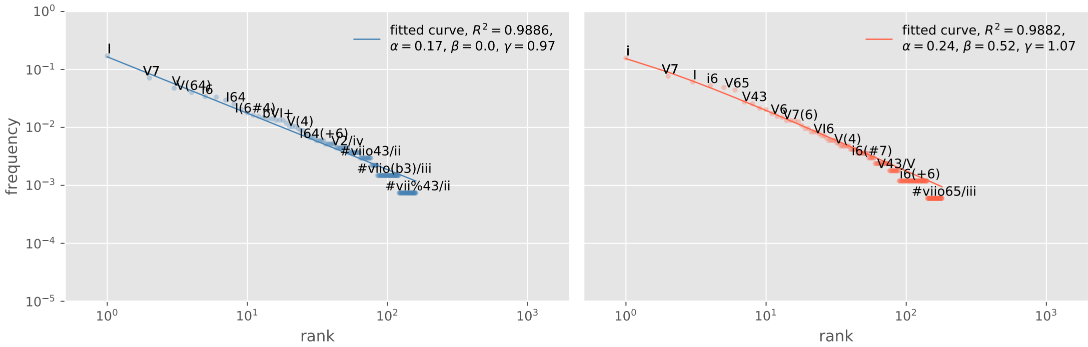


(e) Dvořák: *Silhouettes*.

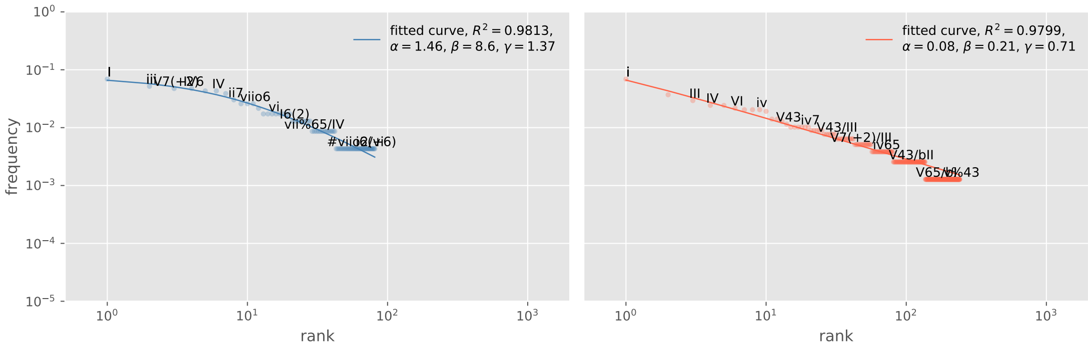


(f) Grieg: *Lyrical pieces*.

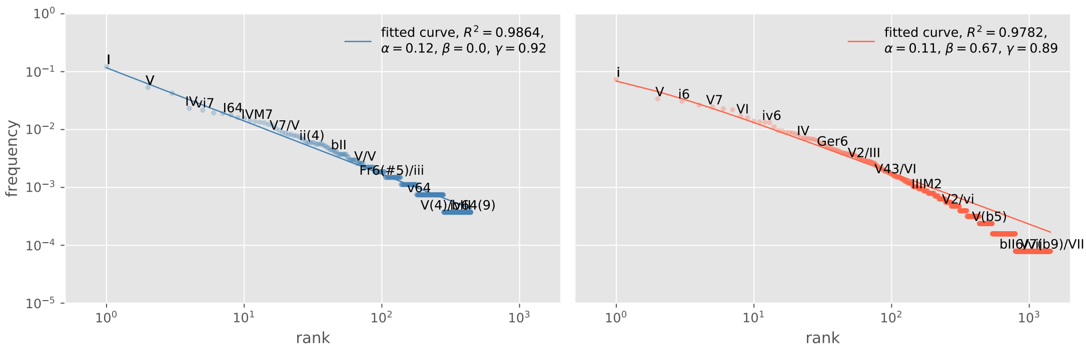
Figure 6.6 – Log-log plot of rank versus frequency of chords in major (left, blue) and minor (right, red). The solid line shows the best fit of a Zipf-Mandelbrot distribution as determined by the coefficient of determination R^2 (cont.).



(g) Tchaikovsky: *Seasons*.



(h) Debussy: *Suite bergamasque*.



(i) Medtner: *Fairy tales*.

Figure 6.6 – Log-log plot of rank versus frequency of chords in major (left, blue) and minor (right, red). The solid line shows the best fit of a Zipf-Mandelbrot distribution as determined by the coefficient of determination R^2 (cont.).

Table 6.4 – Comparison of the parameters α, β, and γ of the Zipf-Mandelbrot fit as well as the coefficients of determination R^2 for all labelled corpora.

Corpus	mode	α	β	γ	R^2
Beethoven	major	0.41	1.19	1.25	0.982
	minor	9.1	8.2	2.04	0.981
Schubert	major	0.74	1.68	1.47	0.984
	minor	0.43	1.16	1.30	0.993
Chopin	major	0.25	0.28	1.16	0.988
	minor	0.96	3.22	1.5	0.966
Liszt	major	0.17	0.06	1.03	0.997
	minor	0.35	2.85	1.14	0.981
Dvořák	major	0.23	0.00	1.13	0.982
	minor	0.87	2.49	1.44	0.992
Grieg	major	0.15	0.57	0.97	0.998
	minor	0.09	0.00	0.86	0.991
Tchaikovsky	major	0.17	0.00	0.97	0.989
	minor	0.24	0.52	1.07	0.988
Debussy	major	1.46	8.6	1.37	0.981
	minor	0.08	0.21	0.71	0.980
Medtner	major	0.12	0.0	0.92	0.986
	minor	0.11	0.67	0.89	0.978

Nonetheless, one can observe differences, in particular in the lower ranks, corresponding to many chords that occur only once or a couple of times.

Table 6.4 summarizes the parameters of the fitted curves and the coefficients of determination. The accuracies given by R^2 lie between 97.8% and 99.8% for each corpus. The parameters α, β, and γ vary considerably. For instance, α ranges from 0.08 in the minor segments of Debussy's *Suite bergamasque* to 9.1 in the minor segments of Beethoven's string quartets. The parameter β varies between 0 and 8.6, and the exponent γ ranges from .71 to 2.04. While all parameters are positive, this diversity points toward individual differences in the chord distributions in the respective sub-corpora. Apart from these divergences in the model's parameters and the disparity regarding the lower-ranking chords, the overall shape of the frequency distributions is well-approximated by a Zipf-Mandelbrot curve and confirms that, despite their respective differences, all of them follow an approximately regular shape. While one should be cautious not to over-interpret the shape of these distributions (Stumpf and Porter, 2012; Piantadosi, 2014), the unigram distribution does reveal the elevated roles of certain central chords in the top ranks of the chord frequency distributions.

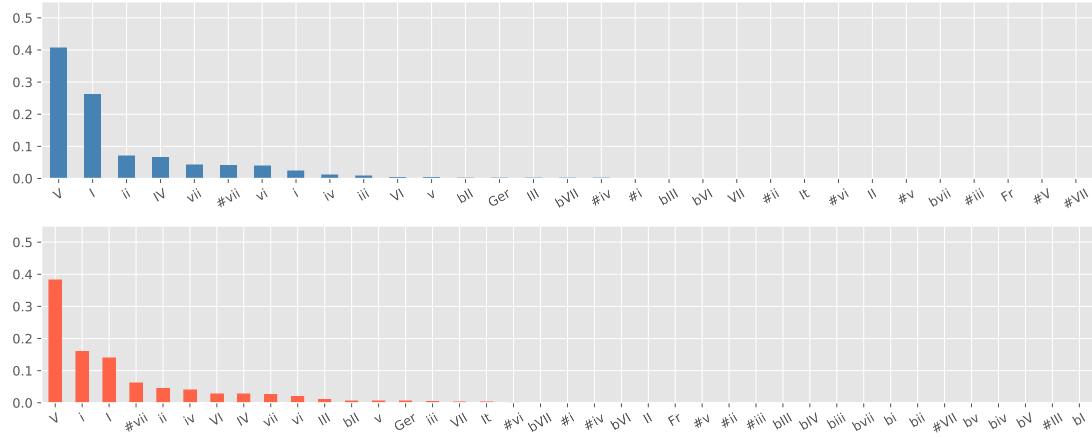
A glance at the chord distributions in Tables 6.2 (chords in major segments) and 6.3 (chords in minor segments) as as well as the rank-frequency distributions in Figure 6.6 shows that the tonics, I in major and i in minor, are by far the most common chords. The V chord and its variants, such as V7, V43, V65, and V(64), govern most of the top ranks in both major and minor segments. The central importance of these chords is clearly a common property of these corpora. Chords with other roots such as IV and ii are much less frequent, irrespective of their forms of appearance (e.g. inversion). The unigram distribution sets tonal harmony apart from Rock/Pop tonality that generally favors IV chords over V chords (de Clercq and Temperley, 2011; Temperley, 2011; Temperley and de Clercq, 2013; Doll, 2017; Temperley, 2018), although there are also exceptions. For instance, Osborn (2017) demonstrates *Radiohead*'s reliance on rather traditional tonal structures.

6.6 Distributions of chord roots

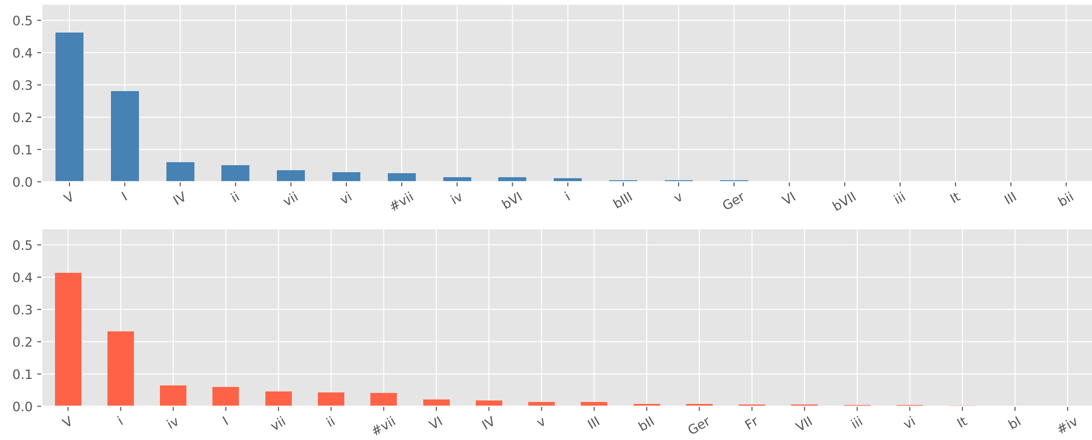
The distributions of chords in major and minor segments have so far treated each chord type separately and have not taken into account that some chords can be conceived as variants of a larger category. For example, the chords V7, V43, and V7(+b9) that can be considered to be variants of a more basic V chord. Grouping all chord tokens by their roots changes the picture. While for all sub-corpora the chord type I is the most frequent one in major segments and i is the most frequent one in minor segments (except for Chopin's *Mazurkas* where V7 is ranked first), the majority of chords in all sub-corpora have root V regardless of the mode. The distribution of chord roots in both the major (blue, top panels) and the minor (red, bottom panels) mode are shown in Figure 6.7. It can also be seen that the reduced vocabularies of chord roots vary between the sub-corpora and range from 11 chord roots in the major mode in Debussy's *Suite bergamasque* (Figure 6.7i) to 38 chord roots in Medtner's *Fairy tales* in the minor mode segments (bottom panel of Figure 6.7c).

The chord roots in all sub-corpora also seem to follow Zipf's law. It follows that only a small fraction of chords constitutes the main proportion of the data. In particular, I, i, and V chords account on average for more than 60% of all chords in major, and for more than 50% of all chords in minor, clearly showing the primacy of harmonies on these two scale degrees in tonal music. The proportions of chords with roots I and V in the major segments and chords with roots i and V for the minor segments are shown in Table 6.5 for all sub-corpora. Together, the proportions of chords in major segments with roots I or V range from 43% of all chords in Debussy's *Suite bergamasque* to 74% of all chords in Schubert's *Winterreise*. In the minor segments, the proportions of chords with roots i and V range from ca. 44% in Medtner's *Fairy tales* to 65% in Schubert's *Winterreise*. One can also make a number of interesting observations when looking at the most frequent chord roots in the respective corpora. For instance, the (natural) seventh scale degree vii is more frequent in the minor segments of Schubert's *Winterreise* than #vii which lies in the harmonic and melodic minor scales. The difference between the natural and the altered seventh scale degree is even stronger in the minor segments of Dvořák's *Silhouettes* and Debussy's *Suite bergamasque*.

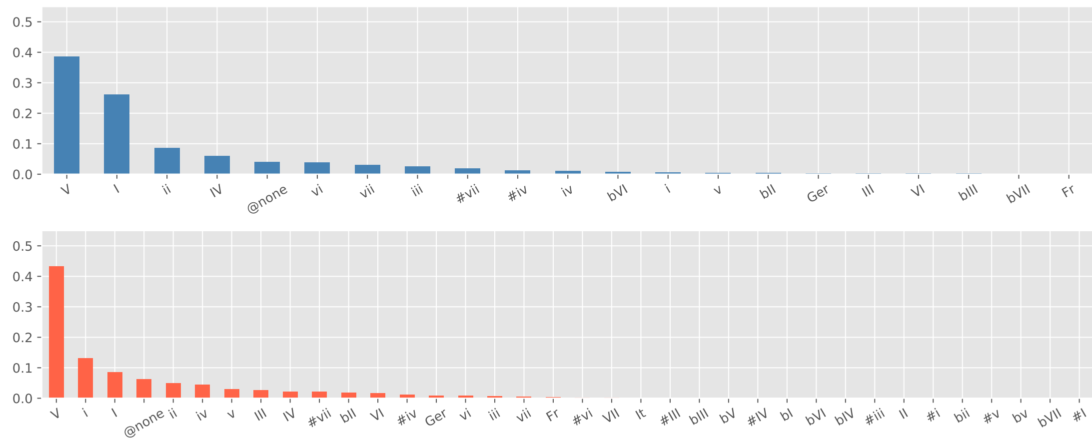
The distribution of chord roots strongly emphasizes the central roles of tonics and dominants for these corpora. Harmonies built on the first and fifth scale degree in both modes are overwhelmingly present in tonal music throughout the 19th century. This finding supports



(a) Beethoven: String quartets.

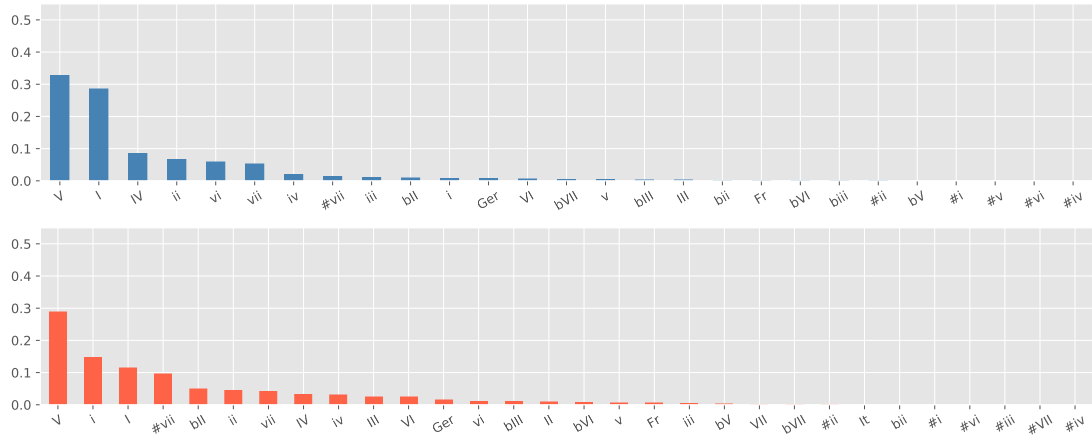


(b) Schubert: *Winterreise*.

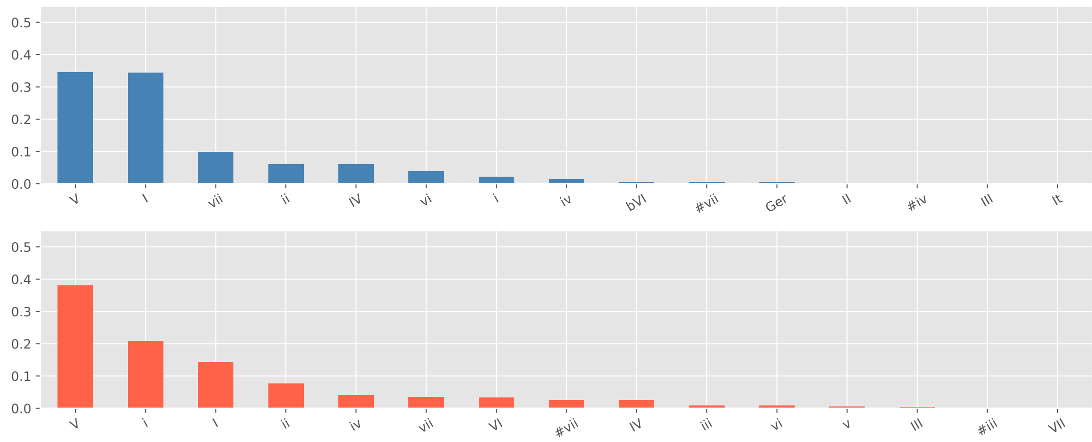


(c) Chopin: *Mazurkas*.

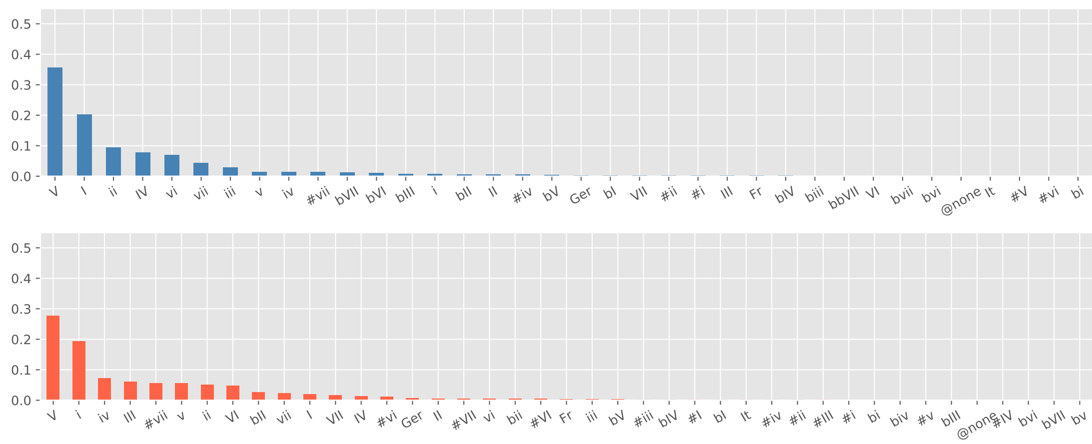
Figure 6.7 – Frequencies of chord roots in the nine sub-corpora.



(d) Liszt: *Années de pèlerinage.*

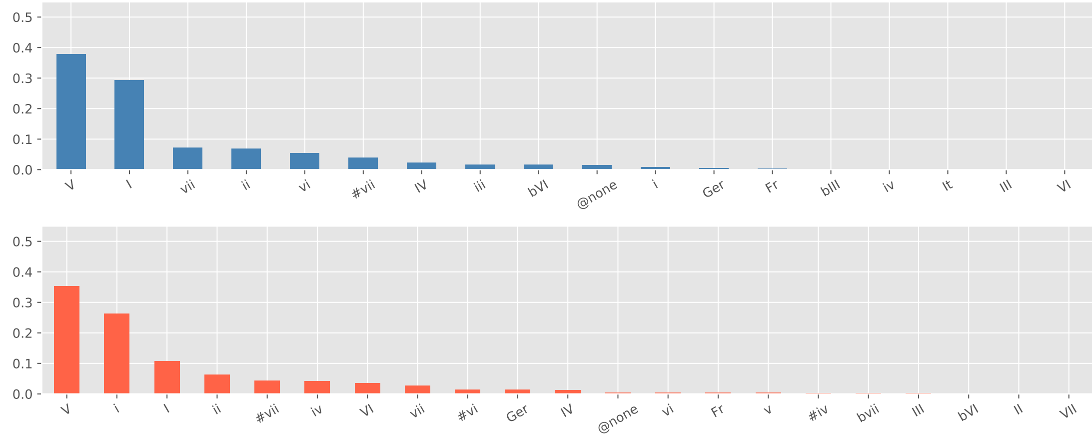


(e) Dvořák: *Silhouettes.*

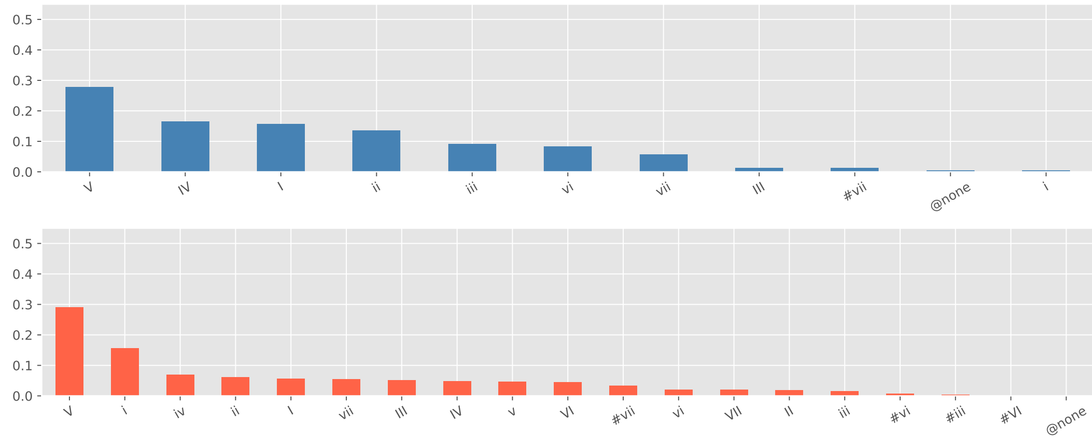


(f) Grieg: *Lyrical pieces.*

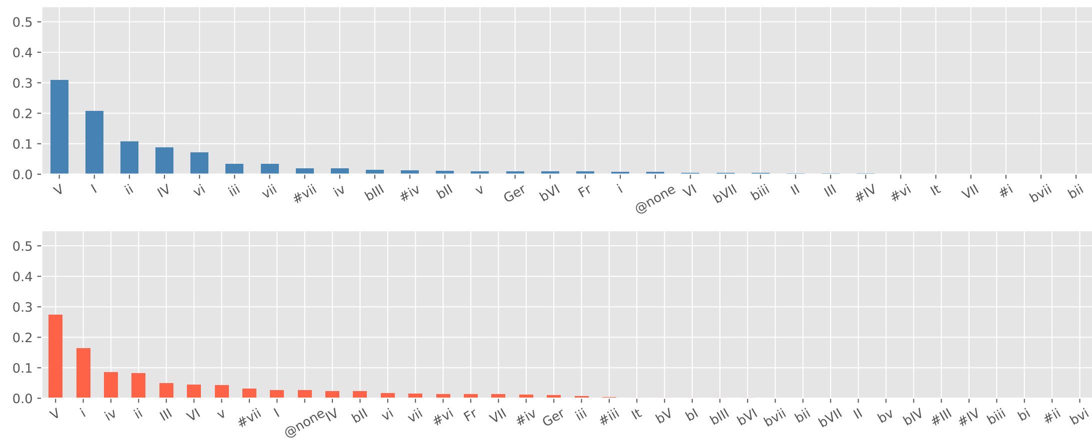
Figure 6.7 – Frequencies of chord roots in the nine sub-corpora (cont.).



(g) Tchaikovsky: *Seasons*.



(h) Debussy: *Suite bergamasque*.



(i) Medtner: *Fairy tales*.

Figure 6.7 – Frequencies of chord roots in the nine sub-corpora (cont.).

Table 6.5 – Proportions of I, i, and V chords in all sub-corpora.

	major			minor		
	I	V	sum	i	V	sum
Beethoven	0.2615	0.4053	0.6668	0.1604	0.3829	0.5433
Schubert	0.2795	0.4613	0.7408	0.2316	0.4139	0.6455
Chopin	0.2610	0.3852	0.6462	0.1312	0.4325	0.5637
Liszt	0.2857	0.3289	0.6146	0.1485	0.2887	0.4372
Dvořák	0.3434	0.3457	0.6891	0.2078	0.3812	0.5890
Grieg	0.2018	0.3568	0.5586	0.1930	0.2771	0.4701
Tchaikovsky	0.2931	0.3782	0.6713	0.2624	0.3536	0.6160
Debussy	0.1552	0.2759	0.4311	0.1505	0.2993	0.4498
Medtner	0.2045	0.3040	0.5085	0.1637	0.2719	0.4356
mean	0.2540	0.3601	0.6141	0.1832	0.3446	0.5278

the view that, despite radical changes in the way composers write music over the course of Western music history, there is a common core that ties together many of these strands to what Tymoczko (2011) calls the "Extended Common Practice".[5]

[5] The notions "Extended Common Practice" and "Extended Tonality" (Section 1.1) differ. The former describes a retrospective view on tonality as extensions of the Common Practice period (ca. 1650–1900) in both historical directions, while the latter describes the expansion of tonal harmony in the 19th century along the historical time.

7 Chord progressions

> Die entscheidenden Fragen für die stilkritische Analyse heißen daher nicht nur: welche Akkorde kommen vor, sondern auch: welche Akkordverbindungen sind charakteristisch?[1]
>
> Zolt Gárdonyi and Hubert Nordhoff, *Harmonik*

All music unfolds in time. As a result, regularities in the transitions between chords, also called *chord progressions*, are an important factor for the statistical characterization of tonality in a given musical style. In their harmony textbook, Gárdonyi and Nordhoff (2002) assume that, from the beginning of the 18th until far into the 19th century, there was no substantial change in the chord vocabulary but rather considerable developments with respect to the transitions between chords. They even regard the study of chord progressions to be the essential core of music analysis (Gárdonyi and Nordhoff, 2002, p. 20). We have seen in the previous chapter that the chord vocabularies of 19th-century composers do indeed vary which renders their first assumption at least questionable. The usage of the term 'chord progressions' here differs from Schoenberg's (1969, pp. 6–8) who does not consider all chord transitions to be progressions. For him, only diatonic (in-scale) root motions are progressions and he moreover associates them with different strengths, depending on their magnitude and direction. He classifies progressions into "strong", "descending", and "superstrong" progressions, carefully avoiding the notion of 'weak' progressions to emphasize his belief that "*[w]eak* qualities have no place in an artistic structure" (Schoenberg, 1969, p. 6, fn. 1). The directedness of harmonic progressions is without doubt an important factor for harmony. It will be analyzed further below. In this chapter, we study progressions between pairs of chords, or chord bigrams, as described in Section 5.3. As mentioned there, we consider only chord transitions within segments defined by a local key so that all chords are expressed in reference to the same local tonic. Chord transitions between local key segments are not taken into account because the

[1] "The decisive questions for critical style analysis are thus not only: which chords do occur, but also: which chord progressions are characteristic?" (Gárdonyi and Nordhoff, 2002, p. 20); translation by the author.

same symbol would have different meanings. A iii chord in C major is a different chord than a iii chord in A♭ minor. The statistical analysis of chord progressions requires a formal model, analogous to language models in Natural Language Processing (NLP; Manning and Schütze, 2003; Jurafsky and Martin, 2009). In the previous chapter, we have employed a unigram model ($n = 1$) to exhibit structural regularities in the chord vocabulary (see Chapter 6). To model chord progressions, we use a bigram model ($n = 2$) in this chapter and study in particular the most frequent chord transitions, following established approaches (e.g. Rohrmeier and Cross, 2008). The transition frequencies from a concrete chord symbol a to another concrete chord symbol b are used as estimates of the transition probability from chord a to chord b. As mentioned in Chapter 5.3, we notate a progression from chord a to chord b by a → b and its probability estimate by

$$p(\mathsf{a} \to \mathsf{b}) = p(c_i = \mathsf{b} \mid c_{i-1} = \mathsf{a}). \tag{7.1}$$

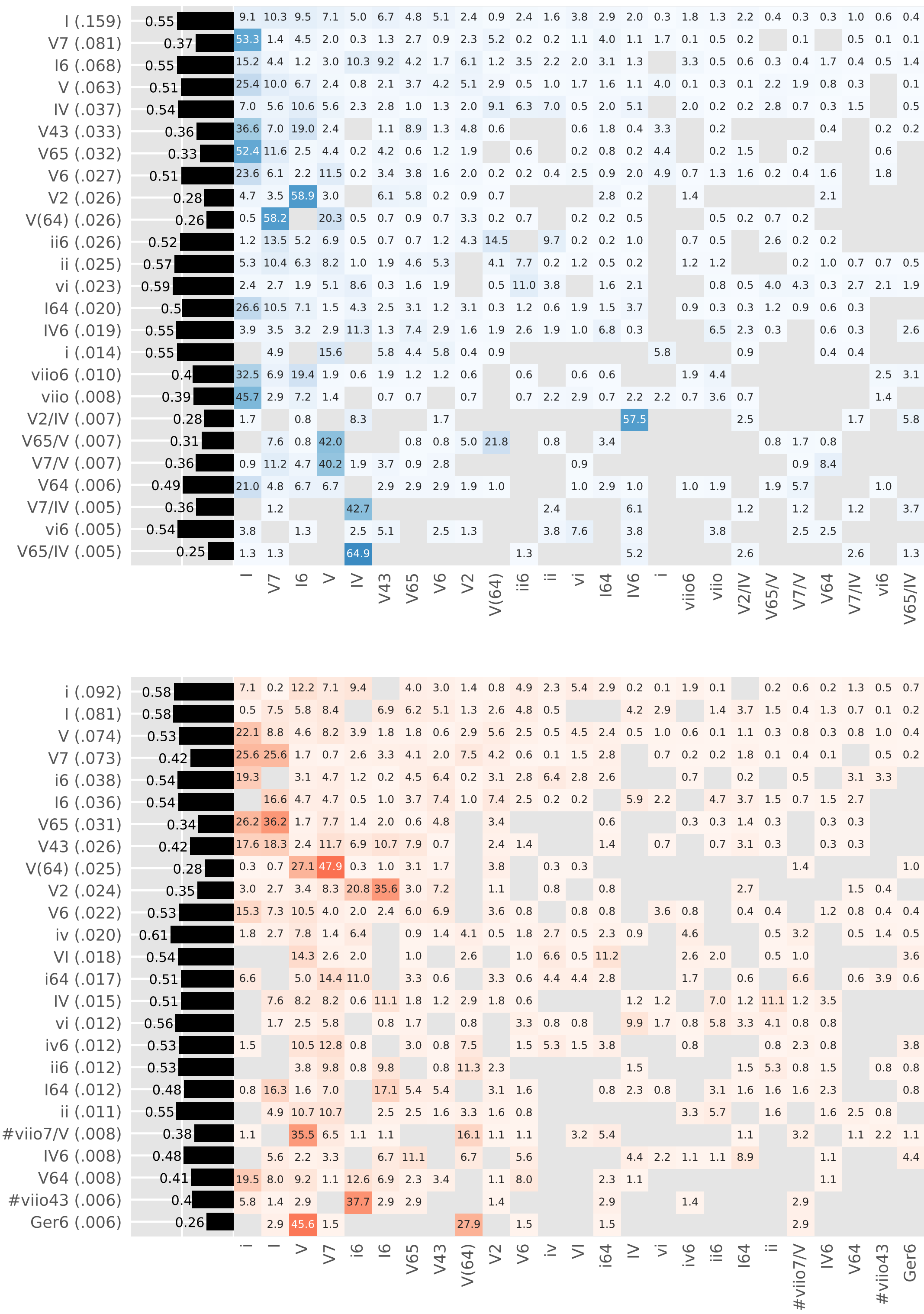
While the statistical description of chord progressions does not account for the entire makeup of musical compositions, it does provide interesting insights into the local relations between pairs of chords. Moreover, looking at specific features of the involved chords also entails information about root progressions and voice-leading, as will be the case in the subsequent analyses.

7.1 Distributions of chord progressions

The first step towards a characterization of chord progressions is to get an overview of their distributions in the respective sub-corpora. The following sections will subsequently address more specific questions, namely the symmetry of chord transitions (Section 7.2), inhowfar certain chord features impact on the predictability of chords under the bigram model (Section 7.3), and what the proportions of chord root progressions are (Section 7.4). Figure 7.1 displays several statistics of chords and chord progressions in the major segments (left, blue) and in the minor segments (right, red) of all sub-corpora as heatmaps. The chord symbols on both heatmap axes correspond to the 25 most frequent chords in the respective modes and are identical to the ones already presented in Tables 6.2 and 6.3. The complete bigram tables are very large because in a corpus with N chord types, there are N^2 possible bigram types. Consequently, the numbers of potential bigrams in the sub-corpora are very large, namely $1,130^2 = 1,276,900$ for Beethoven's string quartets, $312^2 = 97,344$ for Schubert's *Winterreise*, $654^2 = 427,716$ for Chopin's *Mazurkas*, $447^2 = 199,809$ for Liszt's *Années de pèlerinage*, $197^2 = 38,809$ for Dvořák's *Silhouettes*, $928 = 861,184$ for Grieg's *Lyrical Pieces*, $261^2 = 68,121$ for Tchaikovsky's *Seasons*, $276^2 = 76,176$ for Debussy's *Suite bergamasque*, and $1,594^2 = 2,540,836$ for Medtner's *Fairy tales*. Since only the chord progressions for the top 25 chords are shown, there are $25^2 = 625$ cells in each heatmap. The coloring and the numerical values in the heatmaps correspond to the transition frequencies between the 25 most frequent chords as percentages. For example, in the major segments in Beethoven's string quartets, 64.9% of the time V65/IV is followed by IV as shown by the dark blue cell in the last row of the left heatmap in Figure 7.1a. The black bars on the left side of the heatmaps represent the average normalized conditional entropies of the distributions in the rows of the transition tables. They are defined and studied more closely in Section 7.3. It is very likely that a corpus does not contain all possible progressions, in particular due to the fact that the chords follow a Zipf distribution and many chords occur only once or twice which also limits the number of bigrams. If a chord occurs only once it can have at most one preceding and at most one consequent chord. This means that most of all possible bigrams mentioned above do, in fact, never occur and the probability of such a chord transition is zero. Zero-probability chord transitions are represented as empty gray cells in the heatmaps in Figure 7.1. Because language models are often used for prediction tasks, zero-probabilities are highly undesirable for corpus studies. A common strategy to address this issue is to use *smoothing* (Manning and Schütze, 2003) and to distribute a small fraction of the overall probability mass over the zero-probability items. This was, for instance, used in Landnes et al. (2019) where the goal was to predict chords in the Annotated Beethoven Corpus (ABC). Since we are interested in the analysis of the particularities of tonality in the nine sub-corpora, no smoothing is performed. Another reason not to smoothen the bigram probabilities has to do with the fact that it slightly distorts the overall distribution of transition probabilities which is is problematic for some of the analyses below, for instance the analysis of symmetry (see Section 7.2).

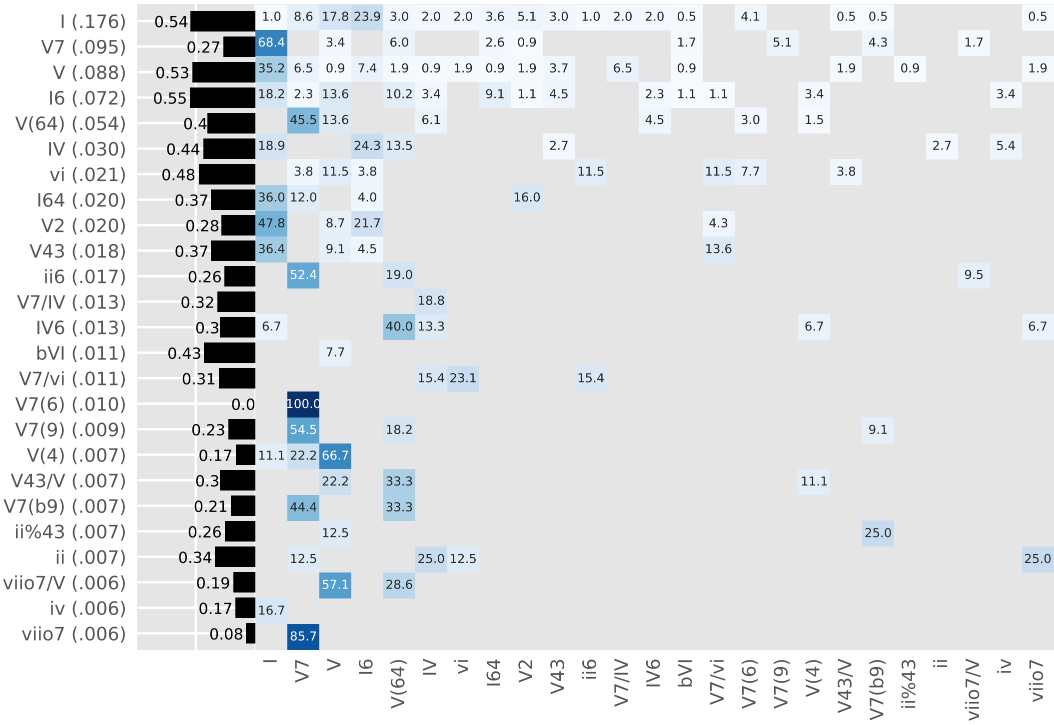
One can observe that the larger the vocabulary size of a sub-corpus, the more spread out the probability mass over the bigrams in the corpus. Naturally, the more chords a corpus contains the more transition to and from these chords this entails. For instance, the larger corpora of Beethoven's string quartets, Chopins's *Mazurkas*, Grieg's *Lyrical pieces*, and Medtner's *Fairy tales* all have relatively few zero-probability bigrams among the 25 most frequent chords, while smaller corpora like Schubert's *Winterreise*, Liszt's *Années de pèlerinage*, Dvořák's *Silhouettes*, Tchaikovsky's *Seasons*, and Debussy's *Suite bergamasque* are much more sparse. Moreover, the distribution of bigrams is most dense in the upper left corner of the heatmaps, indicating that the most frequent bigrams indeed occur between the most frequent chords (unigrams) as well. A somewhat noteable exception is Debussy's *Suite bergamasque* where there are surprisingly few chord transitions from the most frequent chord I in the major segments.

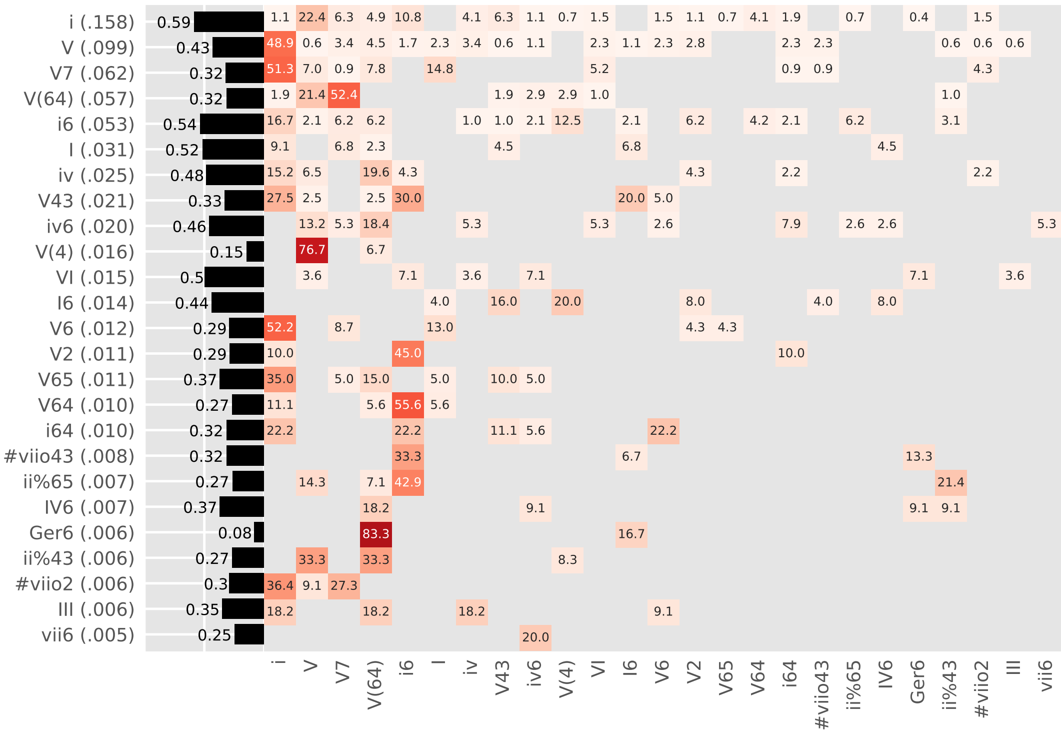
Two further observations can be made: First, some of the transition tables appear to be roughly symmetrical, at least with respect to the non-zero probabilities but not necessarily with respect to the actual frequencies of the transitions. This impression will be investigated in more detail in Section 7.2. Second, the average normalized conditional entropies $\overline{\mathrm{H}}_{avg}$ shown as black bars next to the rows of the heatmaps are particularly low or even zero for chords with suspensions such as $\mathsf{V7(4)}$. Consider for example the progression $\mathsf{I(6\#4)} \rightarrow \mathsf{I}$ in Tchaikovsky's *Seasons*. Both chords do occur frequently. The I chord has is ranked first and the $\mathsf{I(6\#4)}$ chord has rank 11. But importantly, the former is *always* followed by the latter, resulting in a transition probability of 1 and, consequently, in an average normalized conditional entropy of 0 and no black bar in that row. Other chord features, such as inversion, seem not to exhibit a general trend. Whether or not certain features do affect the prediction of subsequent chords is investigated further in Section 7.3.



(a) Beethoven: String quartets.

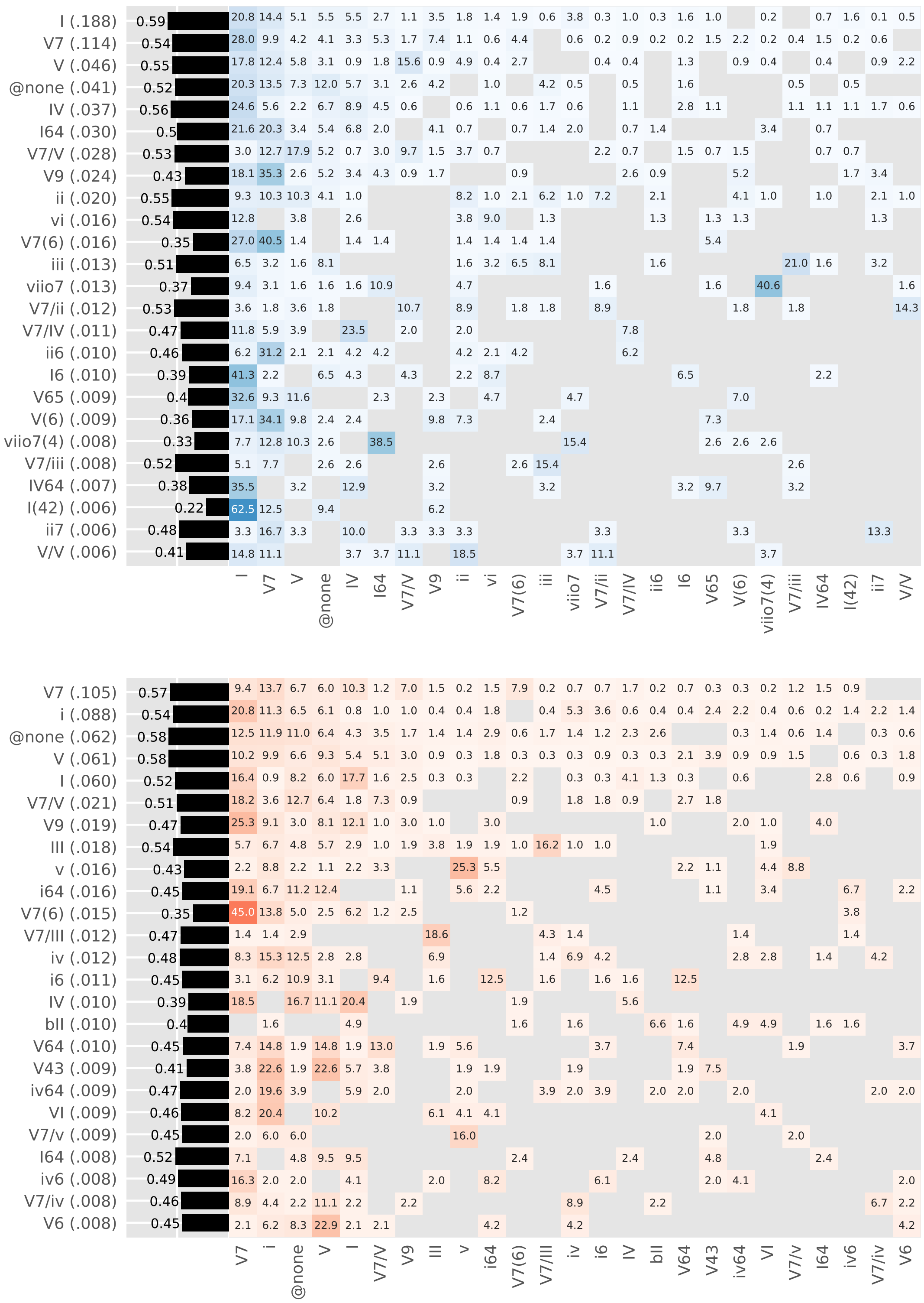
Figure 7.1 – Transition probabilities between the 25 top ranking chords in major (top, blue) and minor segments (bottom, red). Heatmap values are percentages. The black bars show the normalized conditional entropies.





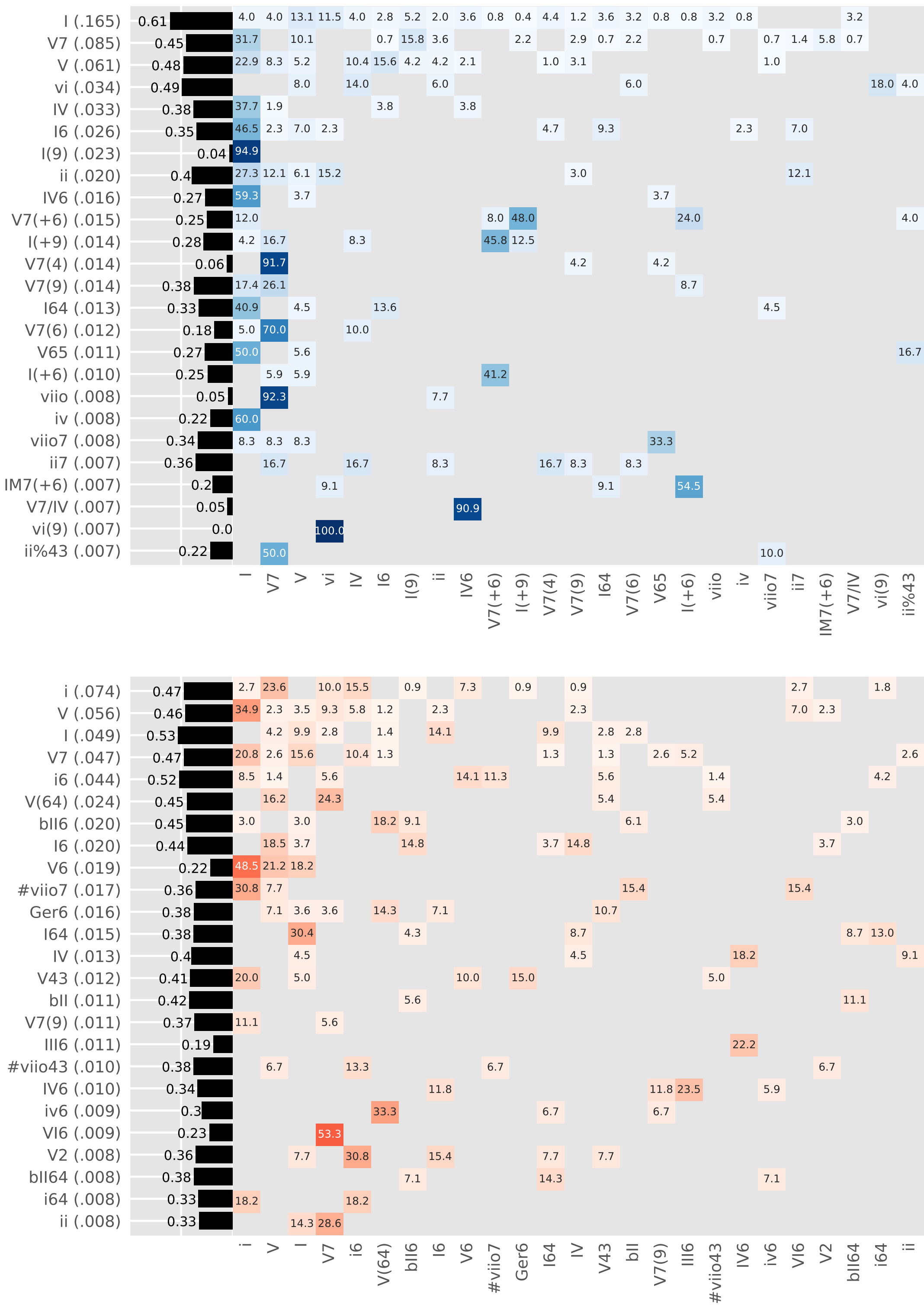
(b) Schubert: *Winterreise*.

Figure 7.1 – Transition probabilities between the 25 top ranking chords in major (top, blue) and minor segments (bottom, red). Heatmap values are percentages. The black bars show the normalized conditional entropies (cont.).



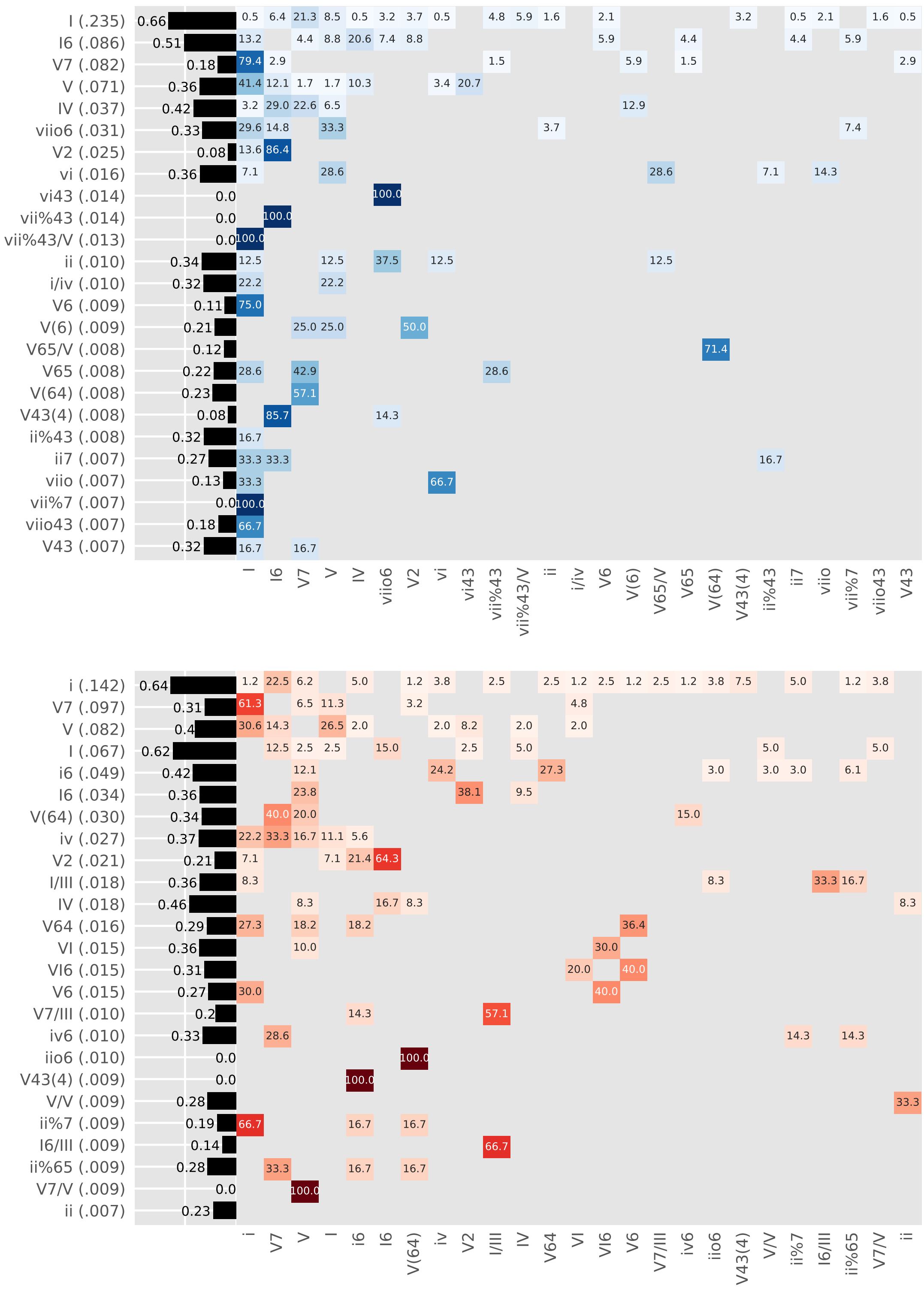
(c) Chopin: *Mazurkas*.

Figure 7.1 – Transition probabilities between the 25 top ranking chords in major (top, blue) and minor segments (bottom, red). Heatmap values are percentages. The black bars show the normalized conditional entropies (cont.).



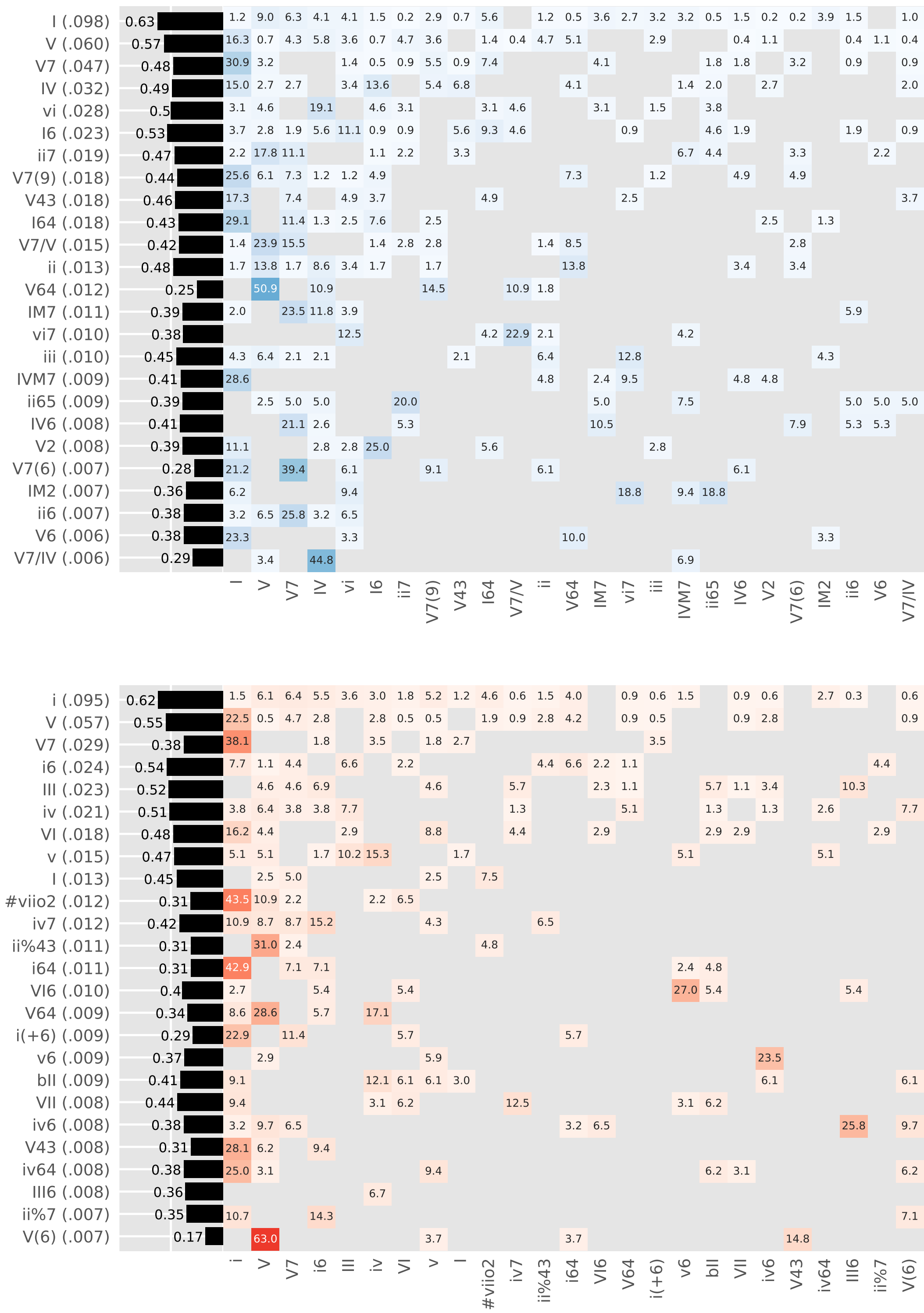
(d) Liszt: *Années de pèlerinage.*

Figure 7.1 – Transition probabilities between the 25 top ranking chords in major (top, blue) and minor segments (bottom, red). Heatmap values are percentages. The black bars show the normalized conditional entropies (cont.).



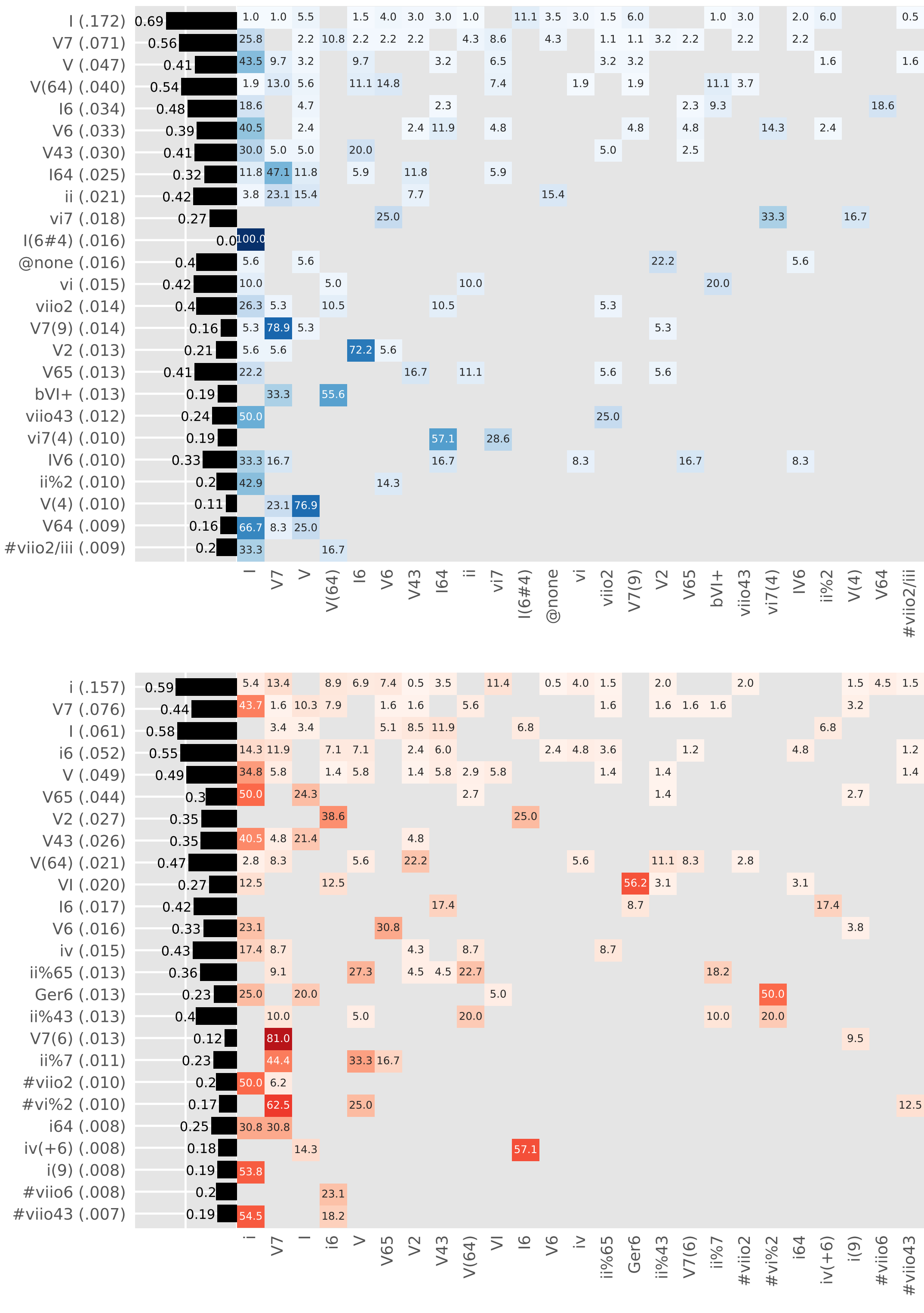
(e) Dvořák: *Silhouettes.*

Figure 7.1 – Transition probabilities between the 25 top ranking chords in major (top, blue) and minor segments (bottom, red). Heatmap values are percentages. The black bars show the normalized conditional entropies (cont.).



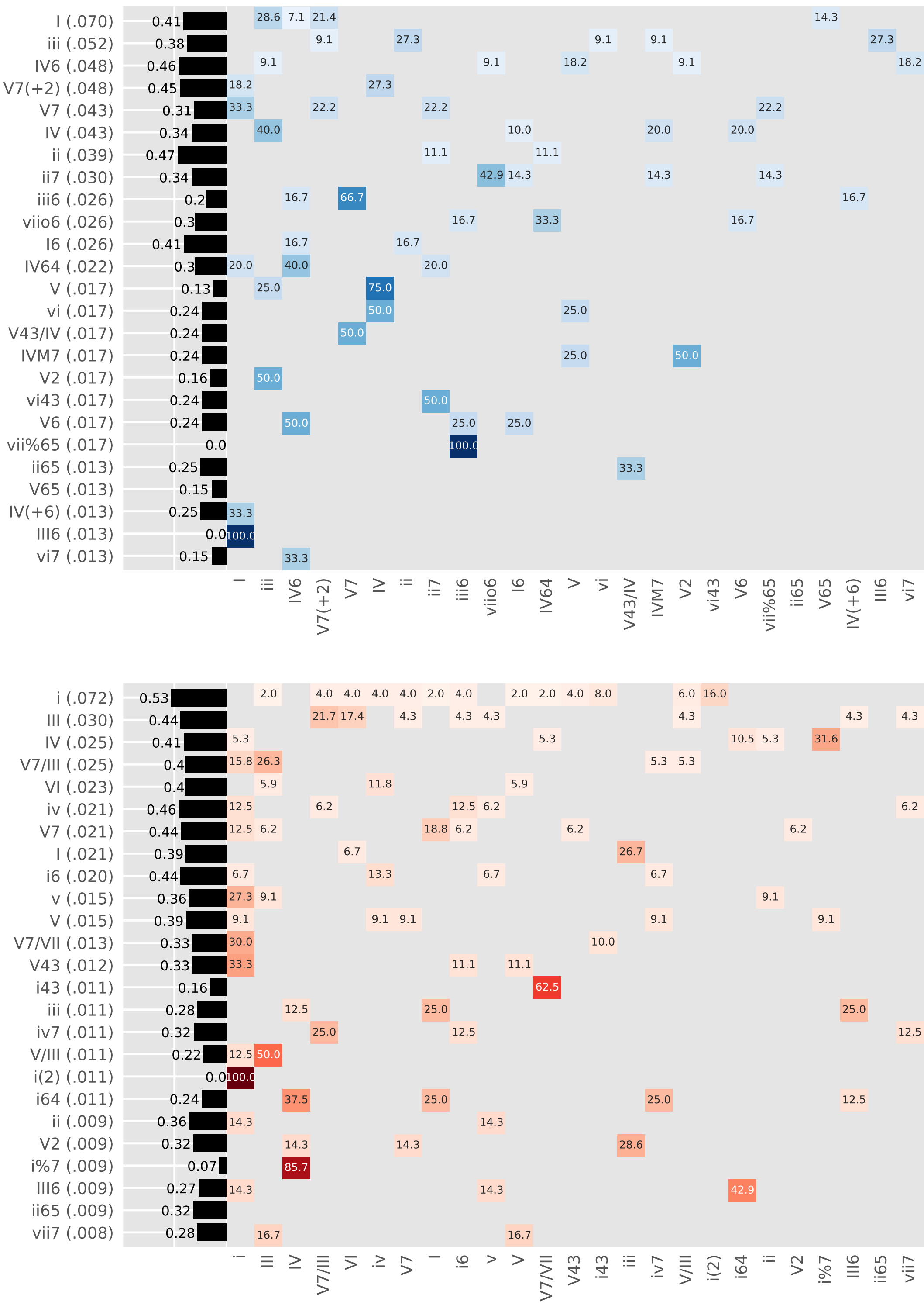
(f) Grieg: *Lyrical pieces.*

Figure 7.1 – Transition probabilities between the 25 top ranking chords in major (top, blue) and minor segments (bottom, red). Heatmap values are percentages. The black bars show the normalized conditional entropies (cont.).



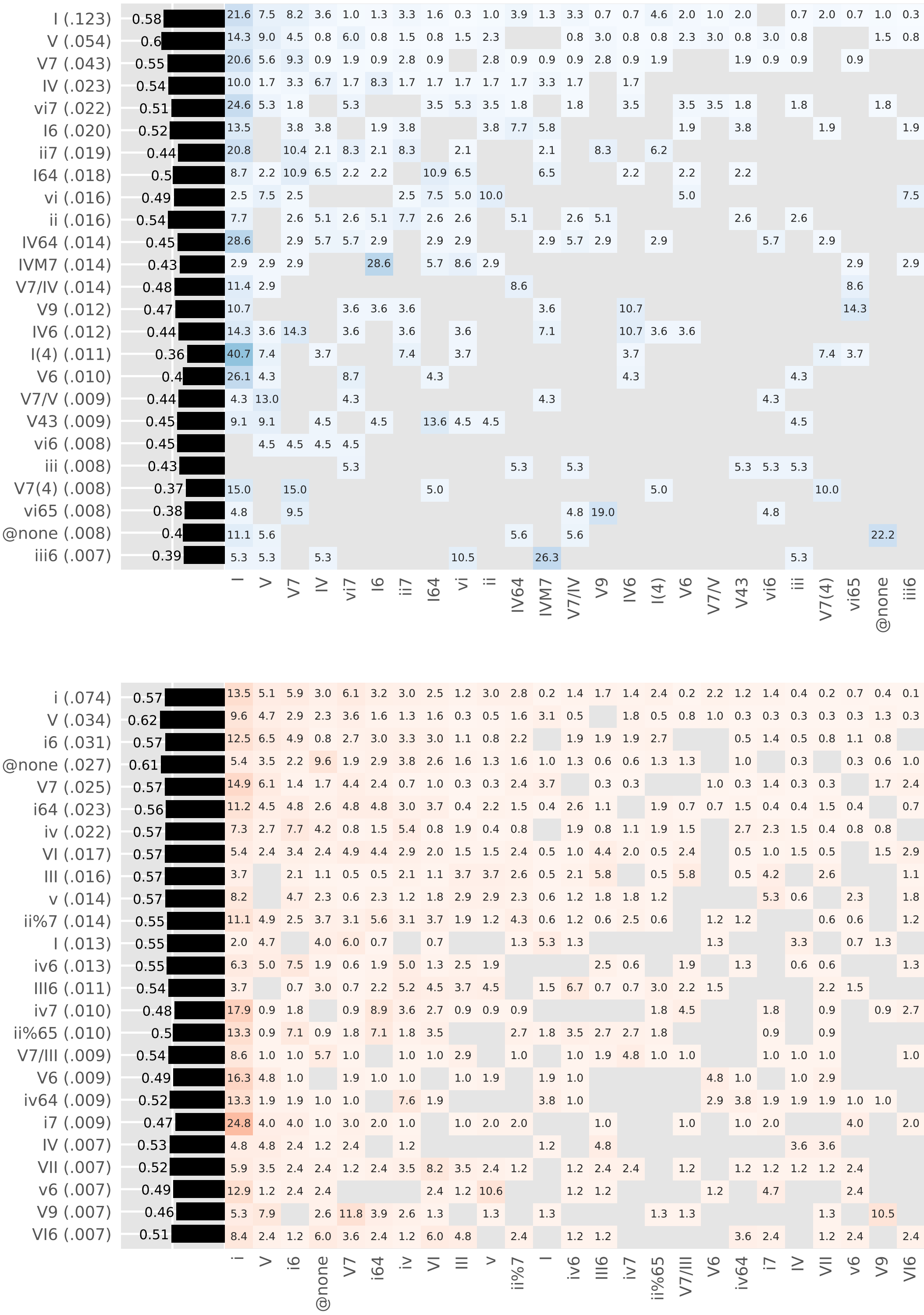
(g) Tchaikovsky: *Seasons*.

Figure 7.1 – Transition probabilities between the 25 top ranking chords in major (top, blue) and minor segments (bottom, red). Heatmap values are percentages. The black bars show the normalized conditional entropies (cont.).



(h) Debussy: *Suite bergamasque*.

Figure 7.1 – Transition probabilities between the 25 top ranking chords in major (top, blue) and minor segments (bottom, red). Heatmap values are percentages. The black bars show the normalized conditional entropies (cont.).



(i) Medtner: *Fairy tales.*

Figure 7.1 – Transition probabilities between the 25 top ranking chords in major (top, blue) and minor segments (bottom, red). Heatmap values are percentages. The black bars show the normalized conditional entropies (cont.).

7.2 Asymmetry of chord progressions

The last section has examined the distributions of chord progressions for the respective sub-corpora. Here we investigate the symmetry and therefore the directedness of these transitions and distinguish also between the two modes, major and minor. For a given chord progression $\mathsf{a} \to \mathsf{b}$ the transition probability in mode $m \in \{\text{major}, \text{minor}\}$ is then given by $p_m(\mathsf{a} \to \mathsf{b})$. It is estimated on the basis of all chord progressions $\mathsf{a} \to \mathsf{b}$ in the segments of mode m in a sub-corpus. A chord progression is asymmetric if the probability of a chord progression $p_m(\mathsf{a} \to \mathsf{b})$ is not equal to that of its reversal $p_m(\mathsf{b} \to \mathsf{a})$ within the same mode m. If these probabilities were equal the progression would be perfectly symmetric.

The degree of symmetry is measured by the *bigram symmetry* for two chords a and b in mode m that is given by

$$\text{sym}_m(\mathsf{a}, \mathsf{b}) = \min\left\{\frac{p_m(\mathsf{a} \to \mathsf{b})}{p_m(\mathsf{b} \to \mathsf{a})}, \frac{p_m(\mathsf{b} \to \mathsf{a})}{p_m(\mathsf{a} \to \mathsf{b})}\right\} = \text{sym}_m(\mathsf{b}, \mathsf{a}) \tag{7.2}$$

for non-zero values of p_m. If either $\mathsf{a} \to \mathsf{b}$ or $\mathsf{b} \to \mathsf{a}$ do not occur in mode m, the bigram symmetry sym_m is undefined for this chord pair. As mentioned before, no smoothing was applied prior to the calculation of the transition probabilities because it would also have introduced a large number of infrequent but symmetric transitions and hence greatly affect the results for mode symmetry. Since the analysis of chord progression symmetries does not take into account how frequent the chord progressions are but only considers the ratio between the frequency of one progression and that of its reversal, smoothing would have biased the results towards greater symmetry. A bigram symmetry value of 1 means perfect symmetry and implies that the chord transitions $\mathsf{a} \to \mathsf{b}$ and $\mathsf{b} \to \mathsf{a}$ occur equally often in all segments of a sub-corpus in mode m,

$$\text{sym}_m(\mathsf{a}, \mathsf{b}) = 1 \implies p_m(\mathsf{a} \to \mathsf{b}) = p_m(\mathsf{b} \to \mathsf{a}), \tag{7.3}$$

whereas lower values indicate asymmetrical behaviour for this particular pair of chords. Bigram symmetry values greater than 1 are not possible because of the min function in its definition. In order to study the symmetries of chord progressions we do not count chord repetitions because $\mathsf{a} \to \mathsf{a}$ is symmetrical by definition for each chord $\mathsf{a} \in \mathcal{U}$. This section is thus more specific than the previous one in that it does not consider all chord progressions (including chord repetitions, e.g. at phrase boundaries) but only those where the chord types are different.

Based on the bigram symmetries for all chord progressions in segments of a corpus in mode m, we can calculate the average symmetry for this mode. For a given mode m, the *mode symmetry* is defined as the average over all bigram symmetries of chord progressions in segments in this mode,

$$\overline{\text{sym}}(m) = \sum_{\mathsf{a} \in \mathcal{U}} \sum_{\mathsf{b} \in \mathcal{U}} p_m(\mathsf{a} \to \mathsf{b}) \cdot \text{sym}_m(\mathsf{a}, \mathsf{b}), \tag{7.4}$$

Table 7.1 – Mode symmetries for both modes and all corpora.

composer	major	minor	ratio
Beethoven	.500	.524	0.954
Schubert	.589	.592	0.996
Chopin	.694	.708	0.981
Liszt	.576	.709	0.813
Dvořák	.711	.634	1.120
Grieg	.616	.634	0.972
Tchaikovsky	.587	.508	1.155
Debussy	.802	.762	1.052
Medtner	.665	.665	1.000

where a and b are arbitrary chord types such that a ≠ b, and both a → b and b → a are bigrams in segments with mode *m*. Note that, since the bigram symmetries are at most 1, so is the mode symmetry. Hence, the lower the mode symmetry value, the less symmetrical are the bigrams in a given corpus and mode on average. The mode symmetries for both the major and the minor mode in all sub-corpora are shown in Table 7.1 and visualized in Figure 7.2 as a bar plot. The relatively large error bars represent the standard deviations of the bigram symmetries with respect to the two modes and show that there are no significant differences between the mode symmetries. The third colum in Table 7.1 indicates the major-to-minor ratio between the two mode symmetries for all sub-corpora.

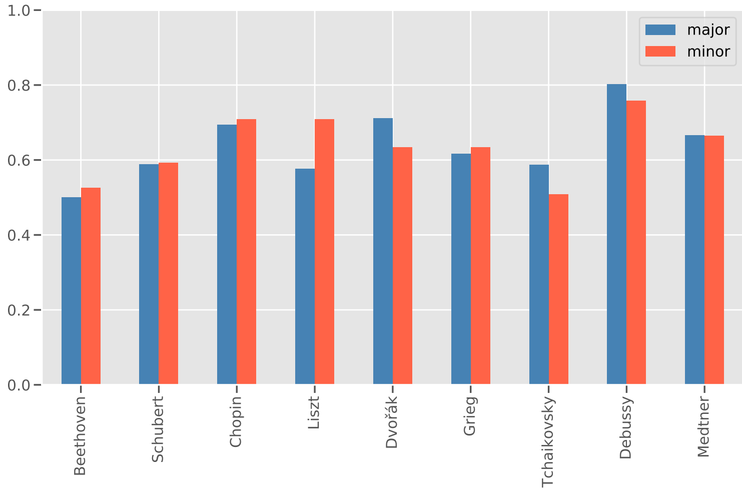


Figure 7.2 – Mode symmetries for the major (red) and minor mode segments (blue) in all corpora

For most of the sub-corpora the major and the minor mode do not substantially differ with respect to the average bigram symmetry in segments in this mode, the mode symmetry. Exceptions are Liszt's *Années de pèlerinage*, where the mode symmetry of the major mode is higher than the minor mode, and Dvořák's *Silhouettes*, Tchaikovsky's *Seasons*, and Debussy's

Suite bergamasque where the mode symmetry of the minor mode is higher. One can also note that there is a general trend in the corpora of earlier composers (Beethoven, Schubert, Chopin, Liszt) that the mode symmetries for the major mode are smaller than for the minor mode, whereas it is larger for some of the later corpora (Dvořák, Tchaikovsky, Debussy). This can be seen by the ratios shown in the third column in Table 7.1 which are in general smaller than 1 for earlier corpora and larger than 1 for the later ones.

These differences in mode symmetries are too small to draw any meaningful conclusions about the implications on tonality from them. However, the differences in mode symmetries allow to compare the respective composers' corpora in terms of the overall directedness of their chord progressions. It will be interesting to compare a wider range of corpora and composers under this measure once a broader range of annotated datasets is available to the research community. Moreover, the fact that the mode symmetries for both the major and the minor mode are considerably smaller than 1 (the largest being .802 for the major mode segments in Debussy's *Suite bergamasque*) corroborates that harmonic progressions are fundamentally asymmetrical, i.e. directed, which also retrospectively justifies Schoenberg's (1969) assertion that directedness is the determining factor for chord progressions. Simply put, tonal music would substantially change its character when played backwards but this trend is weakened in all later corpora to some degree. One can interpret this finding as a trace of the increasing usage in plagal progressions in the 19th century (e.g., Bárdos, 1979; Gárdonyi and Nordhoff, 2002, see also Table 7.3 in Section 7.4) that weakens the overall directedness of chord progressions in Baroque and Classical music (Tymoczko, 2003; Rohrmeier and Cross, 2008; Moss et al., 2019b).

7.3 Chord features affect chord progressions

Chord progressions can not only be studied in terms of their frequency or symmetry but also with respect to their predictability. The bigram frequencies in Figure 7.1 can be used as estimates to answer questions like "How likely is it that a V chord is followed by a I chord in a specific corpus?" by looking at the appropriate cell in the heatmaps. A related but more general question is "Given a certain chord a, how certain are we about which chord is to come next?". This question addresses the shape of the distribution of chords conditioned on a given chord. More formally, we measure the randomness in the distribution over all chords c_i that can follow a fixed chord a in mode m, $P_m(c_i \mid c_{i-1} = \mathsf{a})$ with the *conditional entropy* (Cover and Thomas, 2006). The conditional entropy H of a probability distribution over subsequent chords c_i given a fixed chord a is defined as

$$\mathrm{H}(c_i \mid c_{i-1} = \mathsf{a}) = -\sum_{\mathsf{b} \in c_i} p(\mathsf{a} \to \mathsf{b}) \log_2 p(\mathsf{a} \to \mathsf{b}). \tag{7.5}$$

Since some chords can be followed by hundreds of different chords but others only by a handful, the support for c_i can vary greatly and it is sensible to normalize the entropies for comparison. Given the first chord symbol a of a chord progression, the *normalized conditional* *entropy* $\overline{\mathrm{H}}$ is defined as

$$\overline{\mathrm{H}}(c_i \mid c_{i-1} = \mathsf{a}) = \mathrm{H}(c_i \mid c_{i-1} = \mathsf{a}) / \log_2(|c_i|), \tag{7.6}$$

where $|c_i|$ denotates the size of the support of c_i. This is sometimes also called *efficiency* and its complementary quantity is called *redundancy* (MacKay, 2003). The normalized conditional entropies for the 25 most frequent chord types for each sub-corpus are shown as black bars next to the heatmaps in Figure 7.1. For example, the normalized conditional entropy given the I chord in the major segments of Medtner's *Fairy tales* is demonstrated by the black bar next to the first row in the left heatmap in Figure 7.1i with a value of $\overline{\mathrm{H}}(c_i \mid c_{i-1} = \mathsf{I}) = .58$. The higher such a bar, the larger the randomness of $P_m(c_i \mid c_{i-1} = \mathsf{a})$ and the less certain one can be which chord is to follow the given chord. Consider as a striking example the difference of the normalized conditional entropies of the distributions given the tonic i with normalized conditional entropy $\overline{\mathrm{H}}(c_i \mid c_{i-1} = \mathsf{i}) = .64$ and the double dominant V7/V with normalized conditional entropy $\overline{\mathrm{H}}(c_i \mid c_{i-1} = \mathsf{V7/V}) = 0$ in the minor segments of Dvořák's *Silhouettes* (Figure 7.1e). Because the i chord can be followed by a large variety of other chords, we are less certain about which one it will be than in the case of the double dominant V7/V that is always (in the minor segments of this corpus) followed by the dominant V so that we are absolutely certain which chord will follow if we take the frequencies of occurrence as an estimate of the probabilities.[2]

The fluctuation between the normalized conditional entropy bars furthermore indicates a certain variability between chord types with respect to some of their features. Recall that the only mandatory part for a chord symbol is a root but that chord symbols can express a variety of features such as inversions, suspensions, being applied chords, or occurring over a a pedal tone (see Section 5.2). In the following, we investigate this variability more systematically, and inspect different chord features such as inversions (e.g. V65, IV6), suspensions, and added notes (e.g. V(64), ii7(+b9)). Moreover, the chord symbols can express an implied chord (e.g. V7/vi, iv/bVI) which does not necessarily need to follow directly, and they may appear over pedal notes (e.g. all chords in the brackets of V[vi7 ii V7 I] occur on the pedal V). The role of these five chord features is illuminated by analyzing how they affect the prediction of subsequent chords as measured by the normalized conditional entropies. Using this measure of predictability, we compare its values $\overline{\mathrm{H}}$ for chords having a certain feature to random samples of chords from the respective sub-corpora. For each of the five chord features, we want to know whether chords having a certain feature are statistically different from random chords with respect to their predictability of subsequent chords.

In order to investigate this question, a *one-sample bootstrap hypothesis test* (Efron and Tibshirani, 1993) is performed. The fundamental assumption of this resampling approach is that the relationship between an unknown population F and a sample $\mathbf{z}$ is analogous to the

[2]As mentioned before, computational approaches aiming at chord prediction use smoothing in order to avoid this situation (e.g. Landnes et al., 2019). Here, the focus lies on the study of the actual frequencies in the corpora.

relationship between the sample $\mathbf{z}$ and its bootstrap resamples $\mathbf{z}^*$:

$$F \to \mathbf{z} \approx \mathbf{z} \to \mathbf{z}^*. \tag{7.7}$$

The normalized conditional entropies for all chords in a corpus are such a sample $\mathbf{z} = \{z_1, \ldots, z_N\}$ from an unknown population F. Let $\mu(\mathbf{z})$ be the mean of $\mathbf{z}$. We wish to test whether the mean μ_0 of the normalized conditional entropies of only those chords having a certain feature is significantly different from the mean μ_F of F. The null hypothesis is then

$$H_0 : \mu_F = \mu_0, \tag{7.8}$$

to some fixed significance level α. Since F is unkown so is μ_F and both have to be estimated. The goal is to have a distribution $\hat{F}$ that estimates the population under the assumptions of H_0. The sample $\mathbf{z}$ can, in general, not be used as the empirical distribution because it does not conform to the null hypothesis, i.e. it has a mean $\mu(\mathbf{z})$ different from μ_0. A simple solution proposed by Efron (1979) is to transform $\mathbf{z}$ in such a way that it does conform to H_0. The sample $\mathbf{z}$ is first centered at zero by subtracting the sample mean $\mu(\mathbf{z})$ from each element $z_i \in \mathbf{z}$ and then shifted to have its mean equal to the mean assumed by the null hypothesis μ_0 by adding it to each $z_i \in \mathbf{z}$,

$$\tilde{z}_i = z_i - \mu(\mathbf{z}) + \mu_0, \text{ for } i = 1, \ldots, N. \tag{7.9}$$

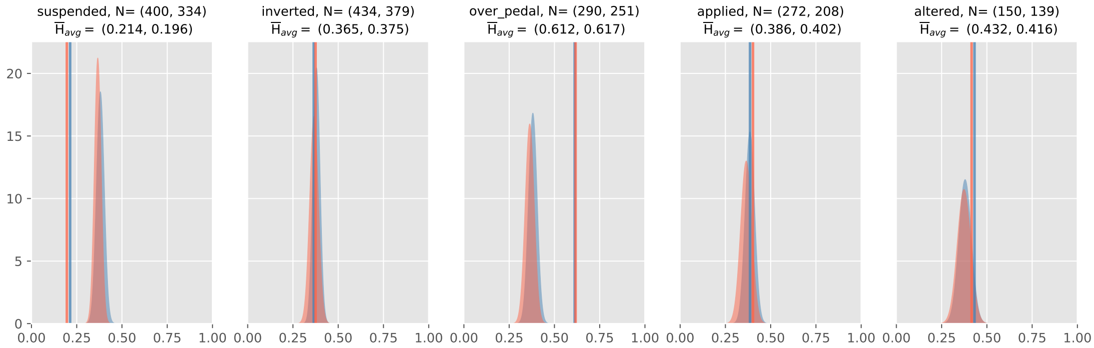
The resulting translated sample $\tilde{\mathbf{z}} = \{\tilde{z}_1, \ldots \tilde{z}_N\}$ has mean $\mu(\tilde{\mathbf{z}}) = \mu_0$ by definition and can be used as an estimate $\hat{F}$ of the unknown population F. The bootstrap procedure now generates a large number B of *bootstrap samples* $\mathbf{z}_b^*$ of size N, $b = 1, \ldots B$, from $\hat{F}$ by sampling from it B times with replacement and calculating their respective means $\mu_b^* = \mu(\mathbf{z}_b^*)$. The proportion of the bootstrap sample means that is more extreme than the actual sample statistic $\mu(\mathbf{z})$ determines whether H_0 can be rejected or not with respect to α,

$$p(\mu_0) = \frac{1}{B} \sum_{b=1}^{B} \mathbf{1}\left(\mu_b^* \leq \mu(\mathbf{z})\right), \tag{7.10}$$

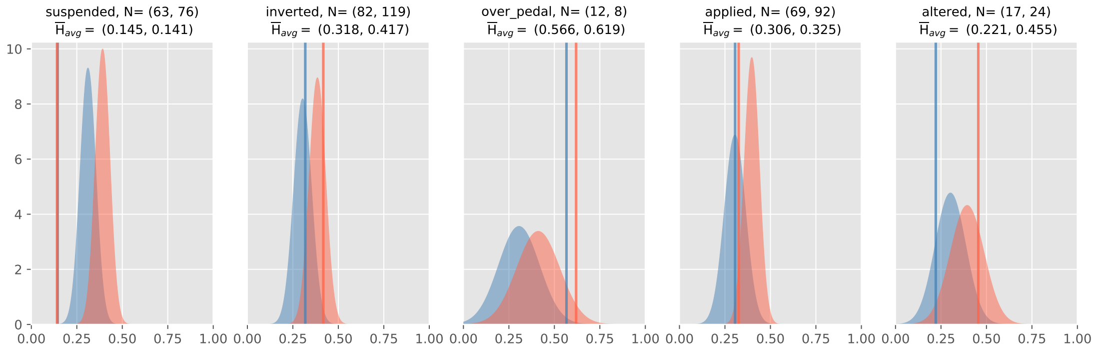
where $\mathbf{1}$ is the indicator function and the alternative hypothesis is taken to be $H_1 : \mu(\mathbf{z}) > \mu_0$. In the opposite case, the $\leq$ sign in Equation 7.10 becomes $\geq$. A major advantage of this method is that it does not require any specific assumptions about the population distribution and the test statistic. In particular, one does not have to assume that the population is normally distributed. For all subsequent analyses, the number of bootstrap resamples is $B = 100{,}000$ and the significance level is set to $\alpha = .01$.

The results for all labelled corpora are shown in Figure 7.3 for both modes (blue for the major mode, red for the minor mode). The shown densities are normal distributions parametrized with the bootstrap resamples. The vertical lines represent the mean entropies for the chords having one of the specified features. For example, all chords in Beethoven's string quartets with suspensions or added notes (e.g., V(64) or ii(+b9)). The average normalized conditional entropy of these chords in the major segments of the string quartets is $\mu_0 = 0.214$ and is shown as the blue vertical line in Figure 7.3a in the leftmost panel. Subsequently, for all chords in major segments of the string quartets $\mathbf{z}$, their normalized conditional entropies were transformed by subtracting their mean $\mu(\mathbf{z})$ and adding the null-hypothesis mean μ_0. From this translated sample $\tilde{\mathbf{z}}$, $B = 100{,}000$ bootstrap resamples were generated, each of the same size as the number of chords with suspensions and added notes ($N = 400$). For each of theses samples, its mean μ_b^* was calculated. The distribution of these means is shown by the blue density in Figure 7.3a. Since in this case none of the μ_b^* is more extreme than μ_0, $\mu_b^* < \mu_0, \forall b = 1, \ldots, B$, one can reject the null hypothesis and conclude that the normalized conditional entropies of chords with suspensions or added notes in the major mode segments of Beethoven's string quartets are on average significantly lower than a random sample. Recall that we only can assert a significant difference in the effect of chord features on predictability if the vertical lines are more extreme than the chosen significance level. For almost all sub-corpora, this is the case for chord suspensions ("suspended") and chords on top of pedal tones ("over_pedal"). Chords with suspensions have on average a much lower entropy than non-suspended chords, which indicates that the implied voice-leading strongly increases predictability of the subsequent event. Although inversions can have strong implications (e.g., V2 → I6), they do also occur in contexts of chord prolongation, e.g. I → I6 → I64 → I. Hence, chord inversion as a categorical feature does not significantly affect predictability of the subsequent event. From a musicological perspective, the most surprising finding is that chords over pedal notes are much less predictable than the average. It suggests that the pedal note is much more important for the prediction of the next event than the chord itself. An exception is Debussy's *Suite bergamasque* where the annotations do not contain any chords over pedal tones in either mode. Applied chords ("applied") significantly increase chord predictability, a tendency that can also be observed in the other corpora. Note also that Figure 7.3h does not show a density for the major mode for the chords with alterated root ("altered") because the only three chords having this feature all have zero entropy.

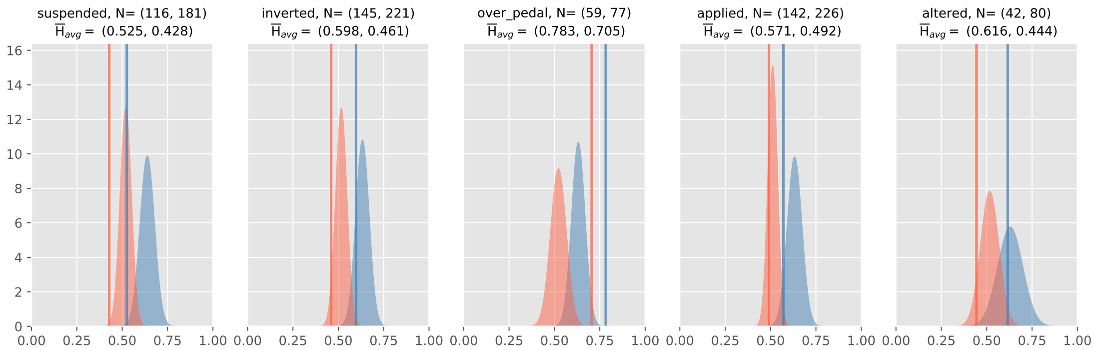
The methodology presented here (following Moss et al., 2019b) has used the average normalized entropies of chord bigrams to study the affect of the presence of certain chord features on the predictability of the next chord event using the bootstrap method. It is obviously not limited to the set of features considered here but may in principle be applied to any situation where the prefixes of bigrams in a corpus are distinguished by certain properties and hence constitutes a very general and useful technique for corpus studies, not only in the musical but in any domain that might find broader applicability within the digital humanities and beyond.



(a) Beethoven: String quartets.

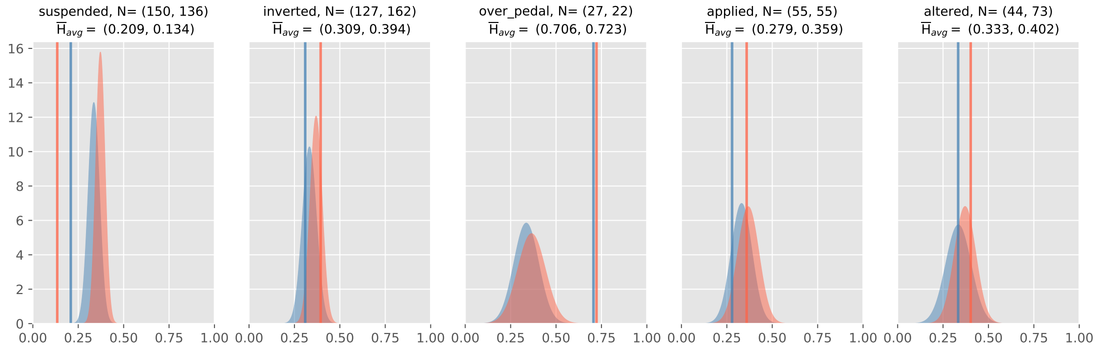


(b) Schubert: *Winterreise*.

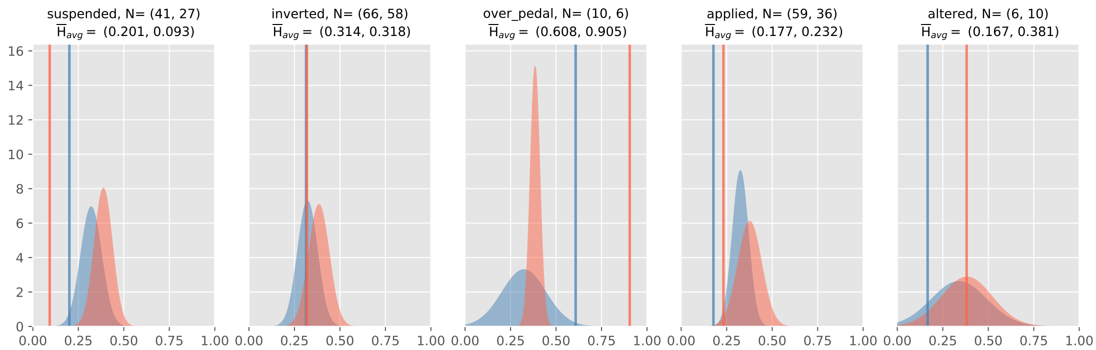


(c) Chopin: *Mazurkas*.

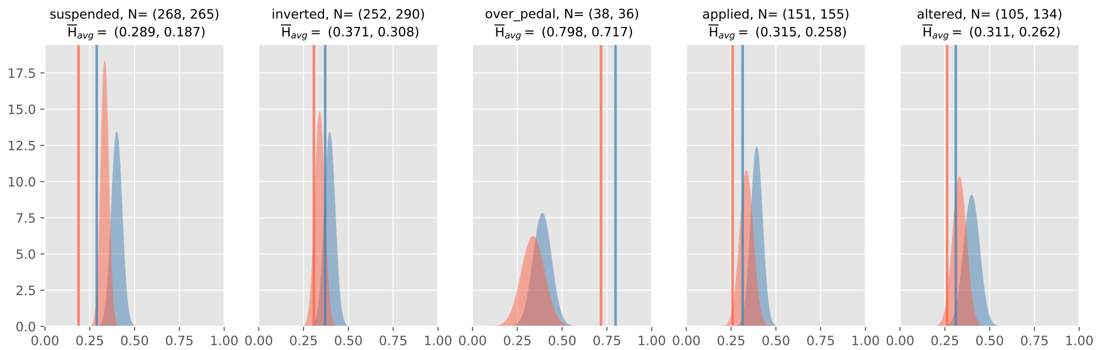
Figure 7.3 – Average normalized conditional entropies $\overline{\mathrm{H}}_{avg}$ of chord types with a certain feature (vertical lines) for major (blue) and minor (red) compared to 100,000 bootstrap samples of the same size N (densities) under the null hypothesis. Subfigures display the different features suspensions and added notes ("suspended"), inversions ("inverted"), chord over pedals ("over pedal"), applied chords ("applied"), and chords with altered roots ("altered"). The first number in parentheses refers to the major mode, the second number to the minor mode.



(d) Liszt: *Années de pèlerinage.*

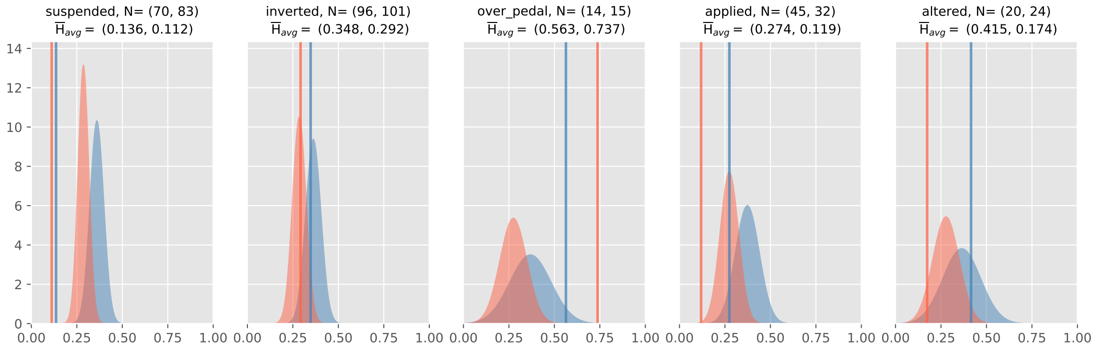


(e) Dvořák: *Silhouettes.*

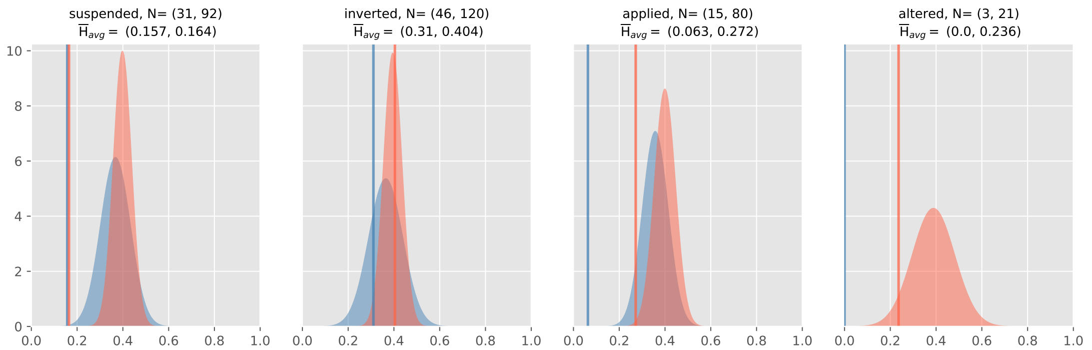


(f) Grieg: *Lyrical pieces.*

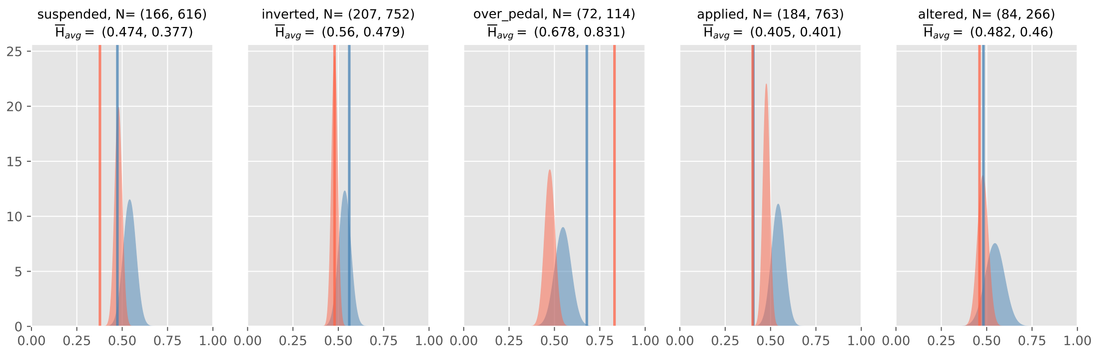
Figure 7.3 – Average normalized conditional entropies $\overline{\mathrm{H}}_{avg}$ of chord types with a certain feature (vertical lines) for major (blue) and minor (red) compared to 100,000 bootstrap samples of the same size N (densities) under the null hypothesis. Subfigures display the different features suspensions and added notes (“suspended”), inversions (“inverted”), chord over pedals (“over pedal”), applied chords (“applied”), and chords with altered roots (“altered”). The first number in parentheses refers to the major mode, the second number to the minor mode.



(g) Tchaikovsky: *Seasons.*



(h) Debussy: *Suite bergamasque.*



(i) Medtner: *Fairy tales.*

Figure 7.3 – Average normalized conditional entropies $\overline{H}_{avg}$ of chord types with a certain feature (vertical lines) for major (blue) and minor (red) compared 100,000 to bootstrap samples of the same size N (densities) under the null hypothesis. Subfigures display the different features suspensions and added notes ("suspended"), inversions ("inverted"), chord over pedals ("over pedal"), applied chords ("applied"), and chords with altered roots ("altered"). The first number in parentheses refers to the major mode, the second number to the minor mode.

7.4 Authentic and plagal chord progressions

In the previous sections we have treated each chord type in the chord vocabulary as equally distant from all other chords according to the discrete metric

$$d(\mathsf{a},\mathsf{b}) = \begin{cases} 0 & \text{if } \mathsf{a} = \mathsf{b} \\ 1 & \text{if } \mathsf{a} \neq \mathsf{b} \end{cases} \tag{7.11}$$

under which, for example, V is as different from I as from V7 or bVII7(b9#11). This might seem somewhat counterintuitive from a music theoretical point of view. It would be more appropriate to consider the chords V and V7 as more similar than, say, V and bVII7(b9#11) because they share many common tones, in particular the same root. A number of proposals exist for more fine-grained distances between chord symbols (e.g., de Haas et al., 2011; Rohrmeier and Graepel, 2012; Moss et al., submitted) but all of them rely on certain specific assumptions and there is no canonical choice for a metric between for the distance between chord symbols. In this section we investigate chord progressions by grouping all chord symbols together that have the same Roman numeral as root. Recall that the root is the only mandatory part of the regular expression CHORD so that the root is defined for every chord symbol in the entire corpus. Consequently, the chord alphabet as well as the chord progressions are reduced and the progressions are transformed to larger and more general categories.

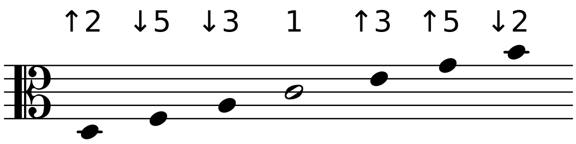
Harmonic analysis based on chord roots is widely used and has a long-standing historical tradition (e.g., Rameau, 1722; Sechter, 1853; Schenker, 1906; Schoenberg, 1969; Bárdos, 1979). This generalization by reduction is similar to a procedure in Natural Language Processing (NLP) called *stemming* (Manning and Schütze, 2003) where word tokens are stripped off their affixes and reduced to their word stems, effectively disregarding plural forms, inflections, conjugations, etc. Analogously we reduce all chords to their roots, removing all morphological features such as inversions, suspensions, alterations, etc. The biological metaphors of 'stems' and 'roots' attest to a certain kinship in the conceptualization of these reductive processes. As an example for such a reduction, the transitions V7 → i and III2 → vi6 are treated as equivalent because in both cases the progression describes a descending fifth. Moreover, the chord progression bV → I is also equivalent since this interval is a descending (diminished) fifth. Since the chord symbols do not specify the octave in which the chord lies this generalization reduces the chord vocabulary to only seven items, the scale degrees $\{\mathsf{I},\mathsf{II},\mathsf{III},\mathsf{IV},\mathsf{V},\mathsf{VI},\mathsf{VII}\}$.[3] Intervals between scale degrees that do not distinguish octaves and the interval quality, e.g. perfect, major, minor, diminished, or augmented, are called *directed generic interval classes* (Harasim et al., 2016). They are determined by their magnitude (difference between scale degrees modulo 7) and direction. These generic intervals are called 'unison', 'second', 'third', 'fourth', 'fifth', 'sixth', and 'seventh' and supplemented with a qualifier that determines

[3] The uppercase spelling of the scale degrees does not imply that a chord is a major triad as it did in the previous sections for chord symbols. The interpretation of uppercase Roman letters as either chord symbols or scale degrees becomes clear from the context.

Table 7.2 – Summary of interval classes, their complements, and corresponding authentic and plagal progressions.

interval class	complement	type	symbol
unison	octave	stationary	1
ascending second	descending seventh	authentic	↑ 2
descending second	ascending seventh	plagal	↓ 2
ascending third	descending sixth	plagal	↑ 3
descending third	ascending sixth	authentic	↓ 3
ascending fourth	descending fifth	authentic	↓ 5
descending fourth	ascending fifth	plagal	↑ 5

the direction ('ascending' or 'descending').[4] Since we assume octave equivalence, they form complementary pairs: unison and octave, second and seventh, third and sixth, and fourth and fifth. Generic intervals are traditionally classified into *authentic* and *plagal* ones (Bárdos, 1979). The systematization by Gárdonyi and Nordhoff (2002) is shown in Example 7.1. Note that their definition relies on generic intervals and thus departs from some other accounts that distinguish, for example, minor from major seconds. Additionally, we have added the category of stationary progressions (indicated by the 1 above the central note) that do not change the scale degree but might involve chord changes such as resolutions of suspensions.



Example 7.1 – Classification of generic intervals into authentic and plagal according to Gárdonyi and Nordhoff (2002). Left to the central C (white note) are the authentic intervals descending third (↓ 3), descending fifth (↓ 5), and descending seventh (↑ 2); right to it are the plagel intervals ascending third (↑ 3), ascending fifth (↑ 5), and ascending seventh (↓ 2).

A summary of the generic intervals, their complements, quality (authentic, plagal, or stationary), as well as the used symbols is given in Table 7.2. It shows that authentic progressions comprise descending odd intervals (thirds, fifths, sevenths) and ascending even intervals (sixths, fourths, seconds).[5] Authentic intervals are considered to be prevalent in tonal harmony (Gárdonyi and Nordhoff, 2002; Weiß et al., 2018). Plagal progressions, in contrast, reverse the direction of the authentic intervals and consist of ascending odd intervals (thirds, fifths, and sevenths) and descending even intervals (sixths, fourths, seconds).

A graphical representation of these relations is given in Figure 7.4. It shows the seven scale

[4] Note that musical and mathematical terminology deviate from each other. The mathematical distance of 0 scale degrees has the musical name 'prime' or 'unison', the mathematical distance of 1 has the musical name 'second', etc.

[5] Note that these are generic intervals that do not distinguish between specific qualities, such as perfect, major, minor, diminished, or augmented. For this reason, the tritone does not appear since the tritone is an (augmented) fourth.

degrees and all directed generic interval classes from starting at scale degree I. Intervals are represented as arrows between the nodes on the circle. Authentic progressions are drawn as solid arrows and plagal progressions are drawn as dashed arrows. The arrow labels specify the direction and magnitude of the interval class. For example, the unison 1 is the interval from I to I, which we will call a *stationary* progression, the interval from I to III is an ascending third ↑ 3, and so forth.

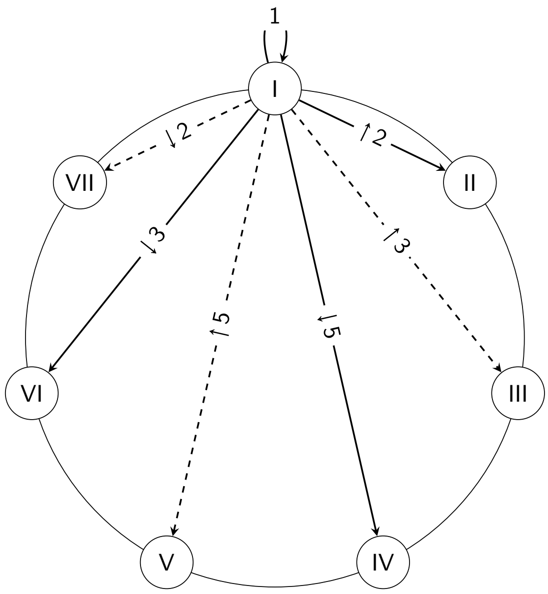
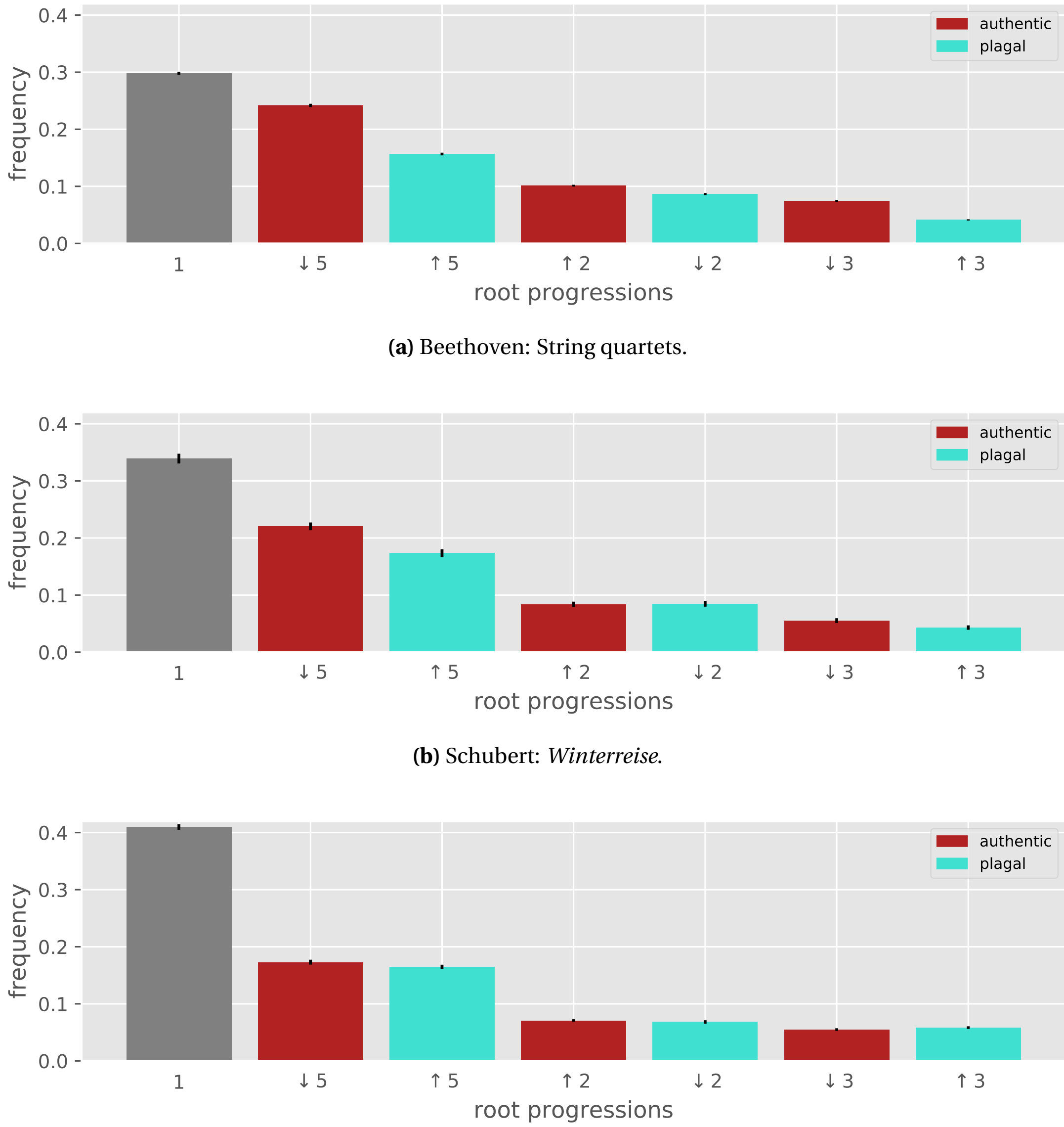


Figure 7.4 – Schematic depiction of diatonic scale-degrees as well as authentic and plagal generic intervals between them. Scale degrees are given by Roman numerals on the diatonic circle. Authentic and plagal progressions are indicated by solid and dashed arrows, respectively. Arrow labels show the complementary directed interval classes with the symbols from Table 7.2.

The classification of all chord progressions into authentic, plagal, and stationary ones allows us to calculate the ratios of these progression types in the sub-corpora, leaving aside progressions from and to the special @none chord symbol. The mean frequencies of these progression types in $B = 100,000$ bootstrap samples for each sub-corpus are shown in Figure 7.5. The error bars in Figure 7.5 represent the standard deviations of the bootstrap samples.

The values for the ratios of authentic, plagal, and stationary progressions are given in Table 7.3. Beethoven's string quartets contain 41.7% authentic progressions (descending fifths: 23.9%; descending thirds: 7.5%; ascending second: 10.2%) and 28.9% plagal progressions (ascending fifths: 15.6%; ascending thirds: 4.2%; descending seconds: 9.1%). Almost a third of all chord transitions (29.5%) are stationary ones and maintain the same root. These can, for instance, be attributed to chord resolutions of suspensions or chord arpeggiations where the bass note changes but the root is invariant. In Schubert's *Winterreise* the pattern is largely similar: Progressions by fifth are most common, while progressions by seconds are much less frequent. Third progressions are least common. A striking difference can be observed by comparing

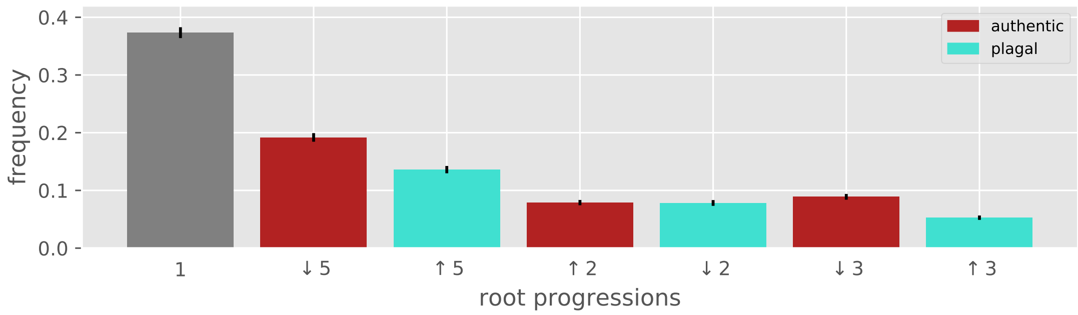


(a) Beethoven: String quartets.

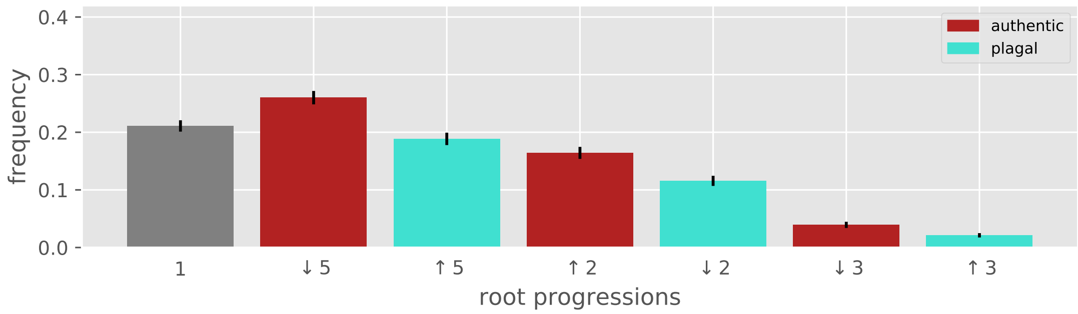
(b) Schubert: *Winterreise.*

(c) Chopin: *Mazurkas.*

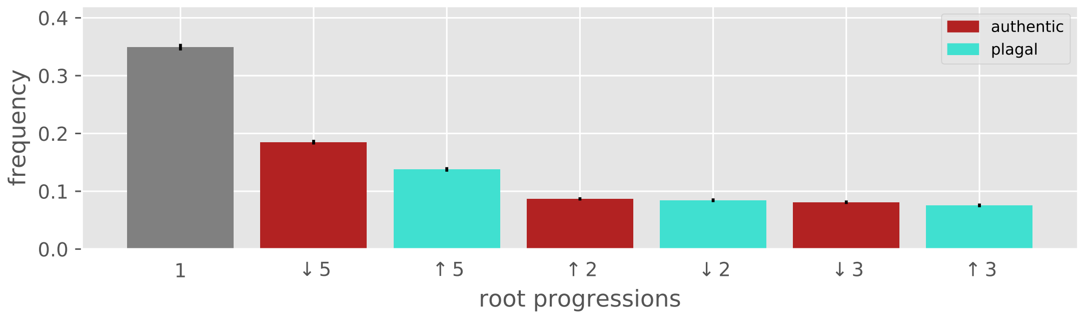
Figure 7.5 – Distribution of bootstrapped means of root progression frequencies between chords in tonal harmony. Error bars show the standard deviation from the mean. Authentic progressions are more common than plagal progressions, and the ranked interval sizes of root motions are fifths, seconds, and thirds.



(d) Liszt: *Années de pèlerinage.*

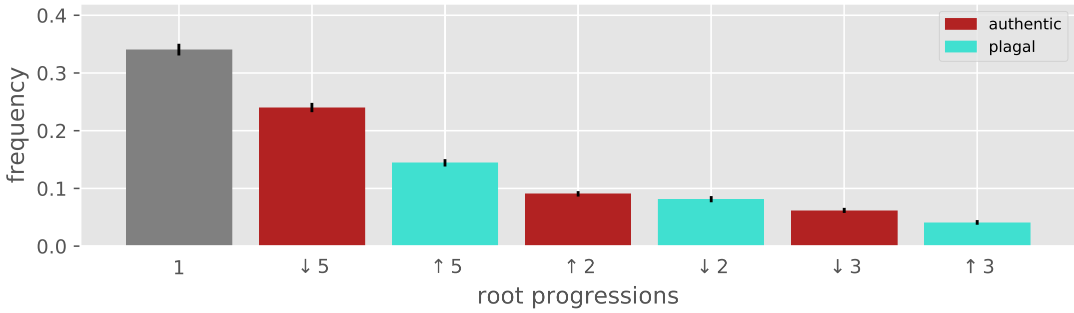


(e) Dvořák: *Silhouettes.*

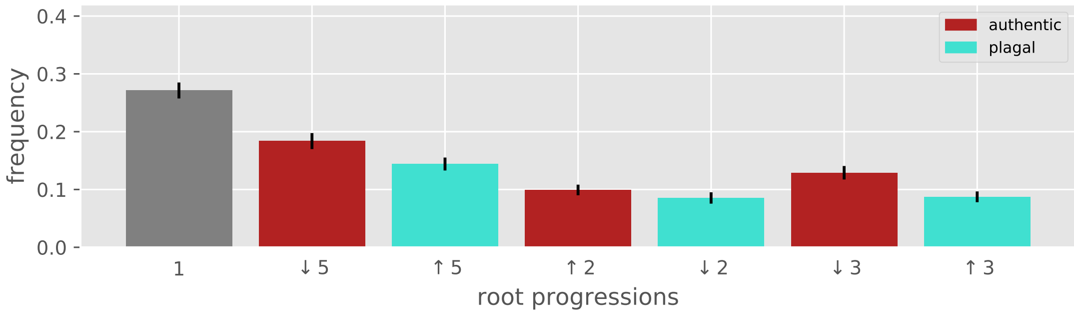


(f) Grieg: *Lyrical Pieces.*

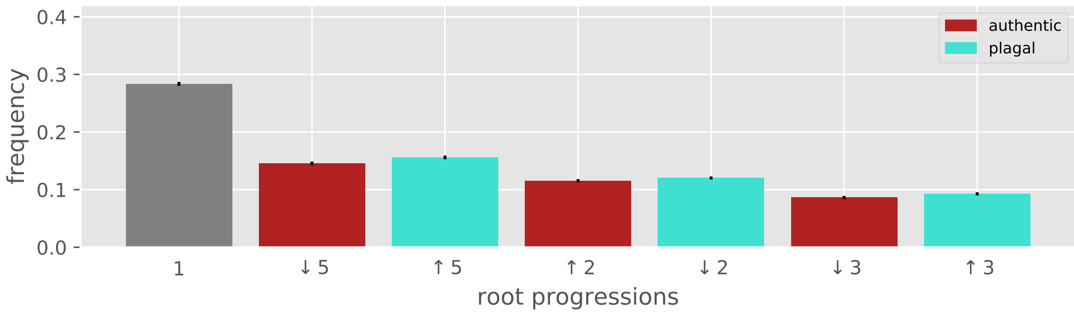
Figure 7.5 – Distribution of bootstrapped means of root progression frequencies between chords in tonal harmony. Error bars show the standard deviation from the mean. Authentic progressions are more common than plagal progressions, and the ranked interval sizes of root motions are fifths, seconds, and thirds (cont.).



(g) Tchaikovsky: *Seasons.*



(h) Debussy: *Suite bergamasque.*



(i) Medtner: *Fairy tales.*

Figure 7.5 – Distribution of bootstrapped means of root progression frequencies between chords in tonal harmony. Error bars show the standard deviation from the mean. Authentic progressions are more common than plagal progressions, and the ranked interval sizes of root motions are fifths, seconds, and thirds (cont.).

the patterns in the *Winterreise* to the ones in the string quartets. It seems that the direction of the interval classes are distributed almost equally. Authentic and plagal progressions (for the respective interval classes) differ much less than in Beethoven's string quartets. In Chopin's *Mazurkas*, this trend is repeated. Here, too, authentic and plagal progressions to not differ greatly. In fact, descending seconds are slightly more common than ascending ones, leading to a small prevalence of plagal vs. authentic seconds. Moreover, seconds and thirds do not differ much. Whereas in Beethoven and Schubert, one could clearly distinguish three generic intervals by their frequency, it seems that in Chopin's *Mazurkas* only fifths and "other" intervals can be clearly distinguished.

Liszt's *Années de pèlerinage* are surprisingly similar to the *Lyrical Pieces*, with the exception that Liszt's third most common progression is the descending (authentic) third which is the least frequent progression in Grieg's case. Tchaikovsky's *Seasons* and Grieg's *Lyrical Pieces* are much more similar to Beethoven. Authentic fifth progressions are more common than plagal ones. While the distributions of seconds and thirds (both authentic and plagal) are also very similar in Grieg's pieces, Tchaikovsky favors progressions by second over progressions by third. The *Suite bergamasque* by Debussy follows a similar trend as Beethoven and Schubert but, in contrast, puts more weight on the third progressions, in particular the authentic ones.

The most distinct corpus in terms of root progressions are Medtner's *Fairy Tales*. One can see that the differences between fifths, thirds, and seconds is smallest among all the corpora. Moreover, plagal fifth progressions dominate all other ones, if only slightly. In all corpora shown in Figure 7.5, the most frequent chord progression is the stationary one ("1", gray). This is a result of the reduction of chords to their roots. The last column of Table 7.3 contains the ratio of authentic-to-plagal progressions in the respective sub-corpora. Values larger than 1 mean that authentic progressions are more common. For some of the sub-corpora this ratio is close to 1 which seemingly contradicts the finding of asymmetry in Section 7.2. But recall that the mode symmetries were calculated on the basis of the full chord representation while we did only consider the scale degrees here. Despite the finding that chord progressions in both the major and minor movements in all sub-corpora are considerably asymmetrical (see Table 7.1) it seems to be the case that the 19th-century corpora are much more symmetric in terms of the ratio between authentic and plagal progressions. In reverse, we can conclude that this asymmetry in the corpora can not be explained by scale-degree relations between the chords but rather mostly by other chord features, such as inversions, suspensions, alterations, etc. The most common progressions in all sub-corpora, apart from the stationary ones, are by fifths, either ascending or descending. The tendency of tonal music for fifth-related progressions and is also predicted by several state-of-the art syntactic models of tonal harmony (Rohrmeier, 2011; Rohrmeier and Neuwirth, 2015). This preference for fifth-related progressions presumably emerged over the course of the 16th and 17th centuries (Tymoczko, 2011). This section has shown that it is relevant until far into the 19th century. The relation of chords by fifths—either ascending or descending, either perfect, diminished, or augmented—can thus arguably be taken as a central property of Western classical music throughout its history.

Table 7.3 – Bootstrapped mean frequencies of stationary, authentic, and plagal progressions.

corpus	stationary	authentic	plagal	ratio
Beethoven	.298	.417	.285	1.463
Schubert	.339	.359	.301	1.193
Chopin	.409	.299	.292	1.024
Liszt	.374	.360	.267	1.348
Dvořák	.211	.465	.323	1.440
Grieg	.349	.353	.298	1.185
Tchaikovsky	.339	.393	.268	1.466
Debussy	.272	.415	.313	1.326
Medtner	.283	.348	.369	0.943

8 Summary and discussion

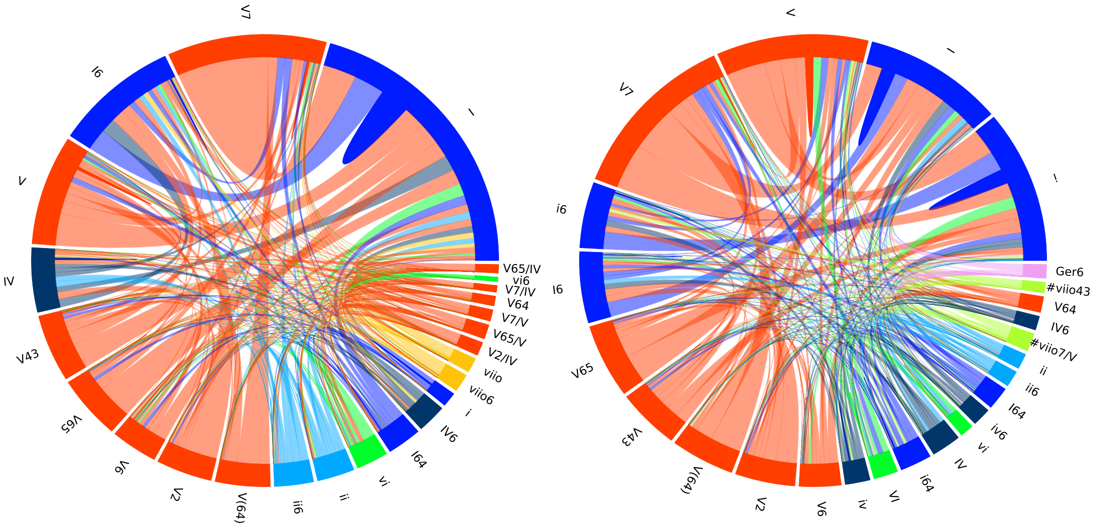
In this part, we have studied a corpus of harmonic labels of nine 19th-century composers, building on a model for a comprehensive chord morphology (Chapter 5). We analyzed the relation of the chord types to the overall chord vocabulary and to the core, i.e. the small set of chord symbols that are used by all composers in our sample (Section 6.1), and compared the vocabularies of the different sub-corpora with each other (Section 6.2). Differentiating between segments in the major and the minor mode, we have seen that the distribution of chords in the sub-corpora and the pieces contained in them can be modeled by Heaps' law (Section 6.3). We have studied the frequencies of chord tokens (Section 6.4) and seen that the relation between the frequency of chord tokens and their rank in a corpus can be modeled by Zipf's law (Section 6.5). The distribution of chord roots (Section 6.6) has revealed that most of the chords in all sub-corpora have the root V, signaling its importance for tonality. In Chapter 7, we analyzed the distributions of chord progressions (Section 7.1), finding that they tend to be asymmetrical for each of the two modes (Section 7.2). We have used the entropy of conditional probability distributions to study the effect of the presence of certain chord features on predicting the next chord in a bigram model and found that, generally speaking, chord suspensions increase predictability whereas a pedal note decreases it (Section 7.3). Finally, we have analyzed authentic and plagal progressions (Section 7.4). The findings show that they are approximately equally frequent in the corpora with few exceptions, and that ascending and descending fifth progressions constitute the majority of all non-stationary chord progressions in this dataset, emphasizing the eminent role of this interval for tonality in the 19th century.

Many of the results presented in this chapter are summarized in the visualizations in Figure 8.1. The frequencies of the 25 most common chords in the respective sub-corpora are shown by the length of the circular arcs next to the chord symbols. The arcs are colored according to the root of the chord symbols. For example, chords with roots I and V are colored in blue and red, respectively. Note that, in the case of applied chords such as V7/IV or #viio7/V, the coloring was chosen according to the chord before the slash symbol, i.e. all applied dominants have the same color (red) as chords on scale degree V. The transitions between these chords is shown by the bows connecting these arcs. Their color corresponds to the color of the first chord of the more frequent bigram. For example, in the plot showing the chords in the major mode in Beethoven's string quartets (left plot of Figure 8.1a), a substantial proportion of the red V7 arc goes into the blue I arc, corresponding to the 53.3% of V7 chords that are followed by I, as was shown in Figure 7.1a. Since the frequency of this transition is much larger than its reversal, $p(\mathsf{I} \rightarrow \mathsf{V7}) = .103$, its red color matches the one for V chords. Bows leading an arc back to itself occur as a result of chord repetition. The annotators were generally encouraged not to repeat chord symbols when the harmonies do not change but repetitions can nonetheless happen, for instance at phrase boundaries. This is why these transitions occur mostly on scale degrees I and V which are the roots of the two harmonies that most typically initiate or conclude phrases (Caplin, 1998; Diergarten and Neuwirth, 2019). Other reasons for a repeated scale degree can be the resolution of chord suspensions or chord inversions.

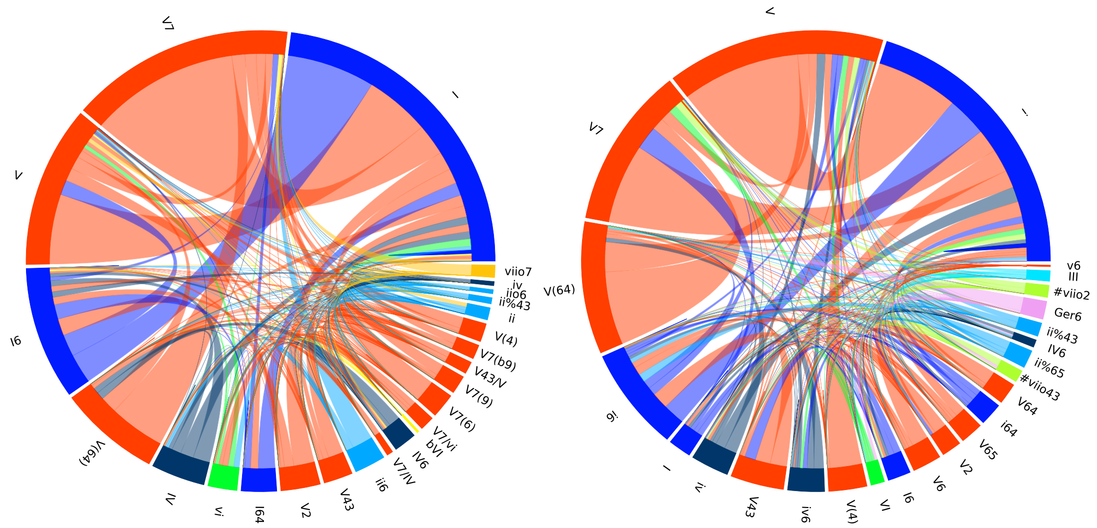
A comparison of these visualization between the respective sub-corpora reveals several differences. First, in the earlier corpora, especially Beethoven and Schubert, we see in both modes an overabundance of chords with roots V and I. This changes and the variety of chords and chord roots among the 25 top chord symbols increases. Second, the minor mode is generally more varied than the major mode. This is partially due to the fact that melodic and voice-leading constraints require to use non-scale chords, such as the altered seventh scale degree #vii (shown in yellow) or the flattened second scale degree bII (shown in light green) and their variants. In principle this kind of visualization is possible not only for the top 25 but for any number or all of the chord symbols in a corpus but increasing this number quickly renders the depiction illegible due to the Zipfian distribution of chord frequencies and transitions.

The results presented in this part largely corroborate traditional musicological and music theoretical knowledge about tonality but put it on firmer empirical grounds by computationally studying a sample much larger than any set of manually analyzed examples could be. The preeminence of tonic and dominant chords are surely no surprise to any scholar of Western classical music. What is new is the fact that, instead of collecting evidence of a handful of singular cases, the methods are much more general. We have not only asked "How frequent are tonics in this repertoire?" but also "How can we model the distribution of chords?" and thus generalized the scope of the question. Similarly familiar should the finding be that chord suspensions facilitate the prediction of subsequent harmonic events. The methodology introduced here[1] has broadened the horizon by more generally studying a larger variety chord features in terms of a single measure, their average normalized conditional entropy, using the bootstrap method. We hope that this change of perspective will inspire future work in similar directions, e.g. by studying different sets of features and by including different corpora. This part also has yielded some new insights. For instance did we see that a vast proportion of chords in the corpus consists of a surprisingly small set of only 43 chord symbols. We have also seen that chords on the third scale degree, in particular in the major mode, occur much more

[1] As an extension of the work in Moss et al. (2019b).

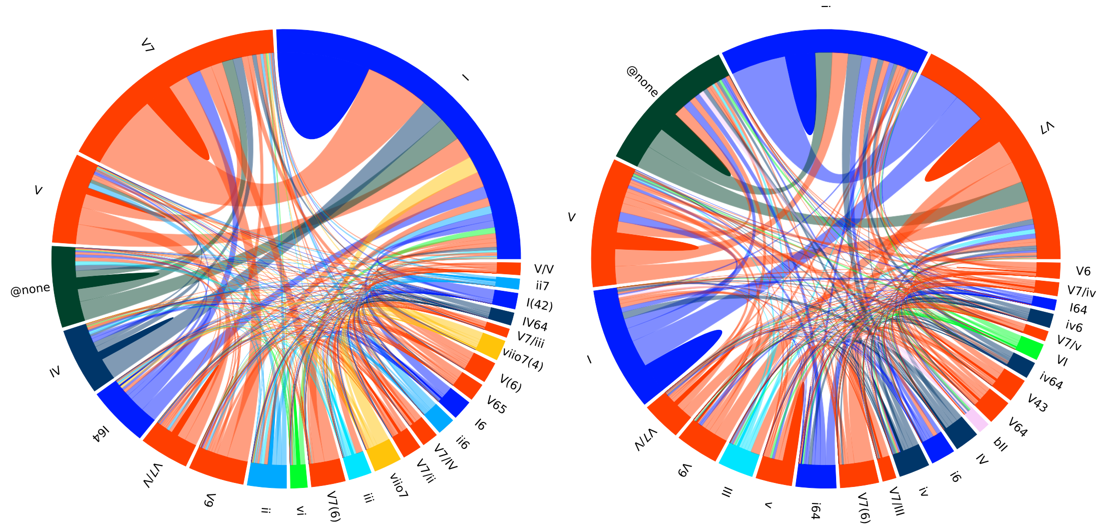


(a) Beethoven: String quartets.

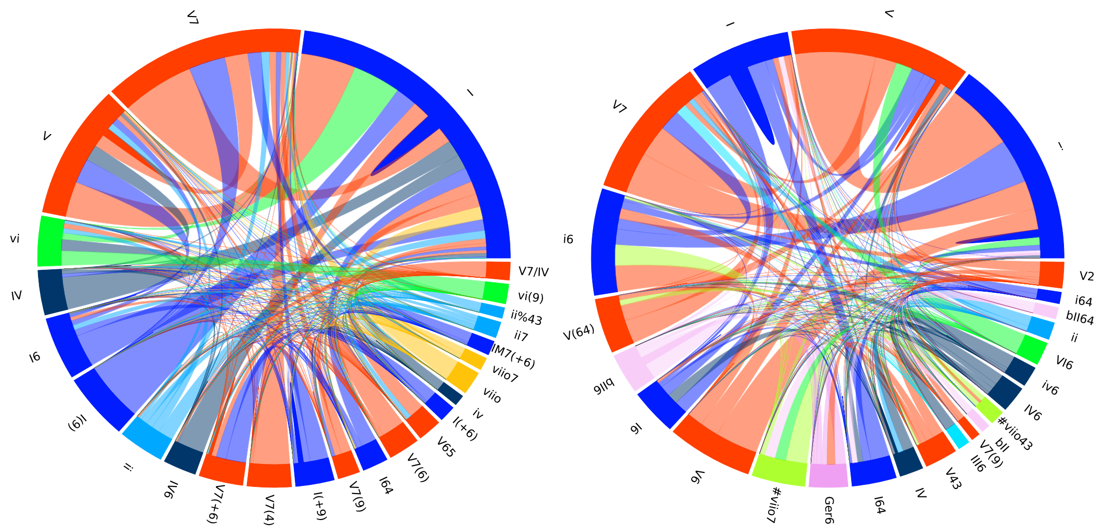


(b) Schubert: *Winterreise*.

Figure 8.1 – Distribution of the first 25 bigrams in major (left) and minor (right). Colors of the circle segments represent scale degrees. The lengths of the segments is proportional to the unigram frequency of the scale degree. The width of the bows between scale degrees is proportional to the bigram frequencies between the scale degrees that they connect and they are colored according to the direction of the more frequent bigram.

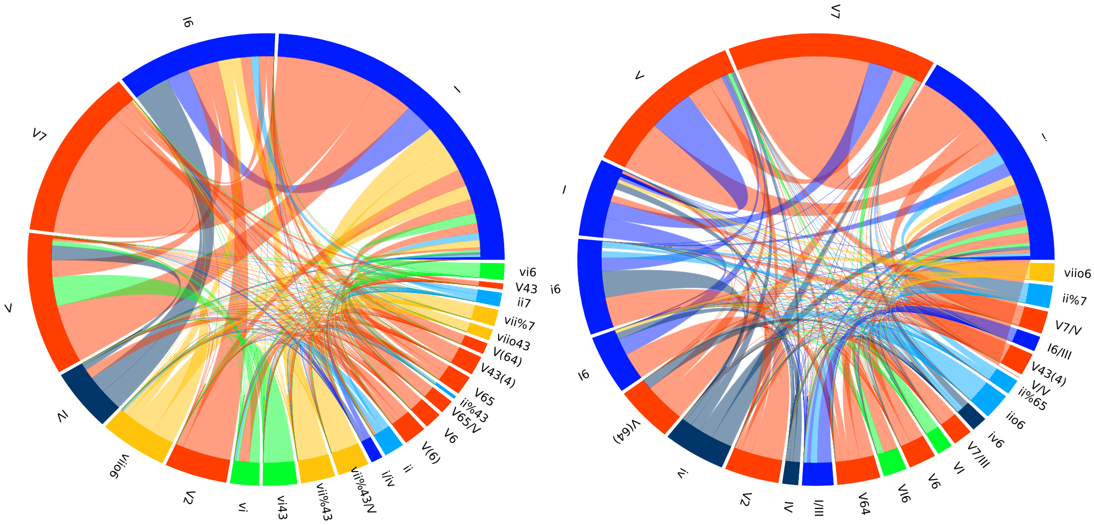


(c) Chopin: *Mazurkas*.

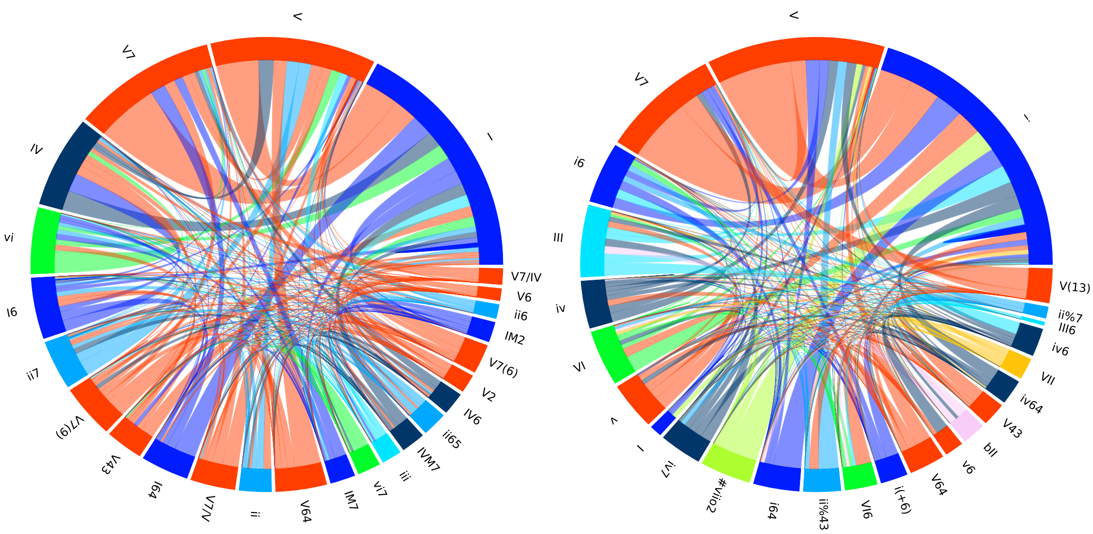


(d) Liszt: *Années de pèlerinage*.

Figure 8.1 – Distribution of the first 25 bigrams in major (left) and minor (right). Colors of the circle segments represent scale degrees. The lengths of the segments is proportional to the unigram frequency of the scale degree. The width of the bows between scale degrees is proportional to the bigram frequencies between the scale degrees that they connect and they are colored according to the direction of the more frequent bigram (cont.).

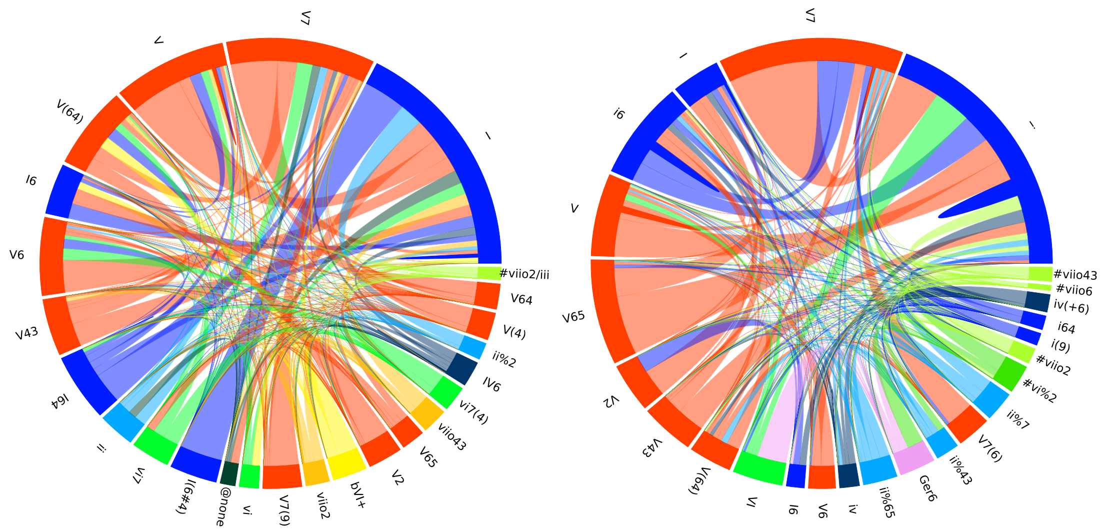


(e) Dvořák: *Silhouettes.*

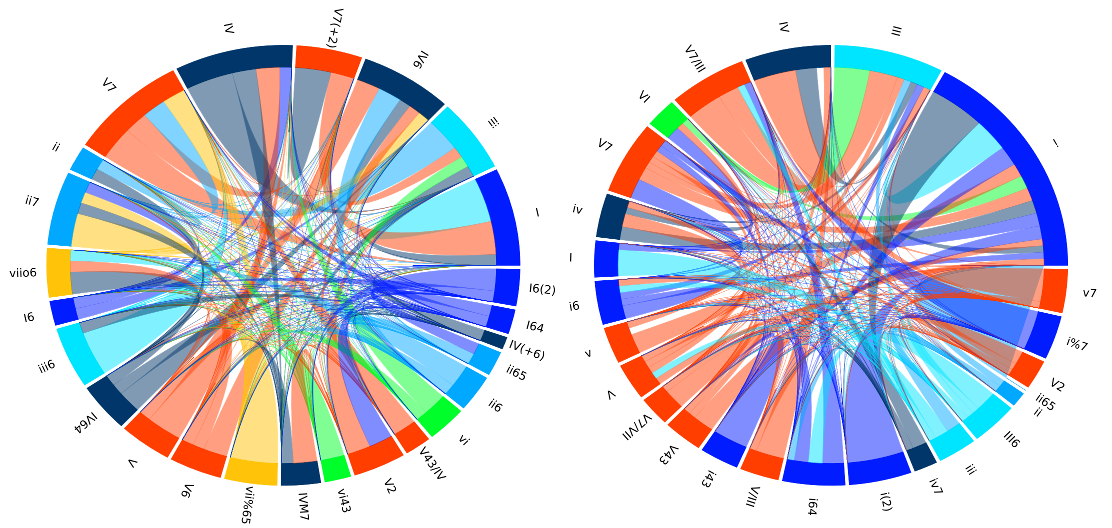


(f) Grieg: *Lyrical pieces.*

Figure 8.1 – Distribution of the first 25 bigrams in major (left) and minor (right). Colors of the circle segments represent scale degrees. The lengths of the segments is proportional to the unigram frequency of the scale degree. The width of the bows between scale degrees is proportional to the bigram frequencies between the scale degrees that they connect and they are colored according to the direction of the more frequent bigram (cont.).

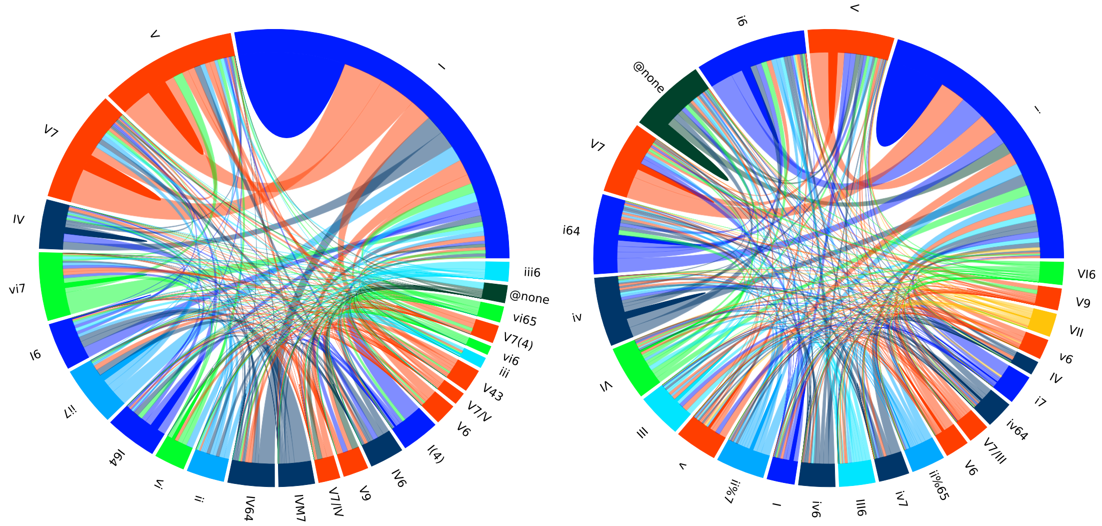


(g) Tchaikovsky: *Seasons*.



(h) Debussy: *Suite bergamasque*.

Figure 8.1 – Distribution of the first 25 bigrams in major (left) and minor (right). Colors of the circle segments represent scale degrees. The lengths of the segments is proportional to the unigram frequency of the scale degree. The width of the bows between scale degrees is proportional to the bigram frequencies between the scale degrees that they connect and they are colored according to the direction of the more frequent bigram (cont.).



(i) Medtner: *Fairy tales.*

Figure 8.1 – Distribution of the first 25 bigrams in major (left) and minor (right). Colors of the circle segments represent scale degrees. The lengths of the segments is proportional to the unigram frequency of the scale degree. The width of the bows between scale degrees is proportional to the bigram frequencies between the scale degrees that they connect and they are colored according to the direction of the more frequent bigram (cont.).

often in the 19th-century corpora than one would assume based on music theory textbooks which are often biased towards the Baroque and Classical periods.

To conclude, the Mesoanalysis part of this thesis has not only shown a variety of results with respect to chords and chord progressions in the works of nine 19th-century composers but also demonstrated a general, model-based methodology for the analysis and the computational study of musical datasets with harmonic labels. The methods presented here are not restricted to 19th-century composers but can be applied to any corpus of harmonic annotations that are formalized according to a regular expression. Therefore, this study can easily be expanded to a broader historical scope and a wider range of composers. The methods can, in principle, also be used to compare different musical styles, such as Western classical music, Jazz, Rock, Pop, or non-Western styles, with the only condition being that can be analyzed within a single music analytical framework.

Macroanalysis Part IV

9 Modeling tonality in musical pieces

> When we think about harmony, we automatically think about chords. In fact, we are so fixated on chords that we sometimes forget they tell only part of the story.
>
> Dmitri Tymoczko, *A geometry of music*

In the previous part (Mesoanalysis) we have studied tonality through the chords and chord transitions in a corpus of pieces by nine 19th-century composers. In the third and concluding part of the present study, we change the perspective and investigate historical changes of tonality by focussing on *tonal spaces*. Broadly speaking, tonal spaces describe the relations between notes in musical pieces. The main question guiding the analysis in this part is: Which of these relations can be inferred from a large historical corpus? While tonal spaces can be formalized mathematically (e.g. Lewin, 1987; Mazzola, 2002; Tymoczko, 2011; Chew, 2014; Bigo and Andreatta, 2016), we specifically acknowledge that the relations between notes and consequently the conceptions and models of tonal spaces are subject to historical change (Brower, 2008; Damschroder, 2008). Moreover, we do not interpret the relations between musical notes as expressing physically or logically necessary relations (as Fétis did, see Section 1.1), but rather as reflecting the underlying cognitive models of composers that use them to write music. Consider a composition that contains an enharmonic reinterpretation of an augmented sixth chord as a dominant seventh chord. The mere occurrence of the reinterpretation attests that the composer's mental model of tonality contains the option for enharmonic equivalence. As another example, the appearance of a chromatic passing tone in an otherwise diatonic composition reveals that—at least in principle—chromaticism is a device that the composer can choose to employ. We will, however, not present one of these models as *the* canonical description of tonal relations. There is no reason to assume that a composer has only one unique conception of how to combine notes to write music. Instead, a piece of music can invoke any number of these mental models, two of which we will consider in particular, the line of fifths (Chapters 10 and 11) and the Tonnetz (Chapter 12).

9.1 A large historical corpus

In this part, we broaden the scope of inquiry in several respects. First, the representation of musical pieces switches from harmonic annotations of scores to the one based on the actual note content of a composition. Moreover, the historical range of the corpus is extended to cover not only a single but almost six centuries of music. The corpus studied in this part consists of 2,012 musical pieces in MusicXML format, gathered from a variety of resources (see Chapter 3 as well as Table B.3 in the appendix). Each piece in the corpus must have a designated date to be useful for a historical study, but assigning a date to a composition can be difficult. Sometimes there lie many years between the finalization of a composition and its publication which makes it difficult to decide which year to choose to represent the piece, sometimes one of them or both are uncertain or unknown. To tackle this problem, the following procedure was applied. For each piece in the corpus, the dates of composition or publication were collected along with the scores as given in the respective sources. In questionable cases they were manually cross-checked with the metadata given by the International Music Score Library Project (IMSLP). If the year of composition of a piece was given, this year determined the date of the piece; otherwise, the publication date was adopted. In the rare cases where both were unknown, the mean of the composer's life was taken as an estimate of the year. This procedure provides a date value for each piece in the corpus but this assignment is not unique. Many pieces have exactly the same date attribute, for instance because they are published in a collection, e.g. almost all pieces by Dufay (1474) or all pieces in J. S. Bach's *Well-Tempered Clavier* (1722 and 1740). Following the procedure explained above leads to only 173 unique years for all 2,012 pieces in the corpus for the whole range of 582 years from 1361 to 1942. The distribution of the pieces over time is shown in Figure 9.1.

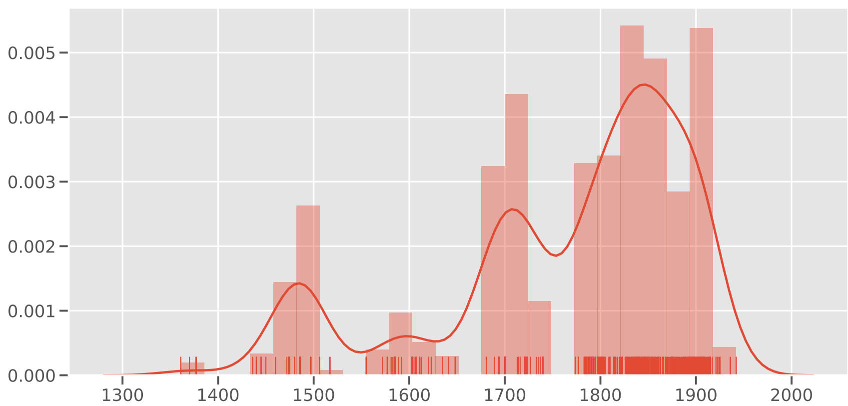


Figure 9.1 – Chronological distribution of pieces in the XML corpus.

It can be seen that there are some gaps in the historical timeline for which there are no pieces in the corpus and that some epochs are represented more than others, in particular the Renaissance (ca. 1450–1550), the Baroque (ca. 1680-1730), and the Romantic periods (ca. 1800–1900). An overview of the life dates of the composers in the corpus is shown in Figure 9.2.

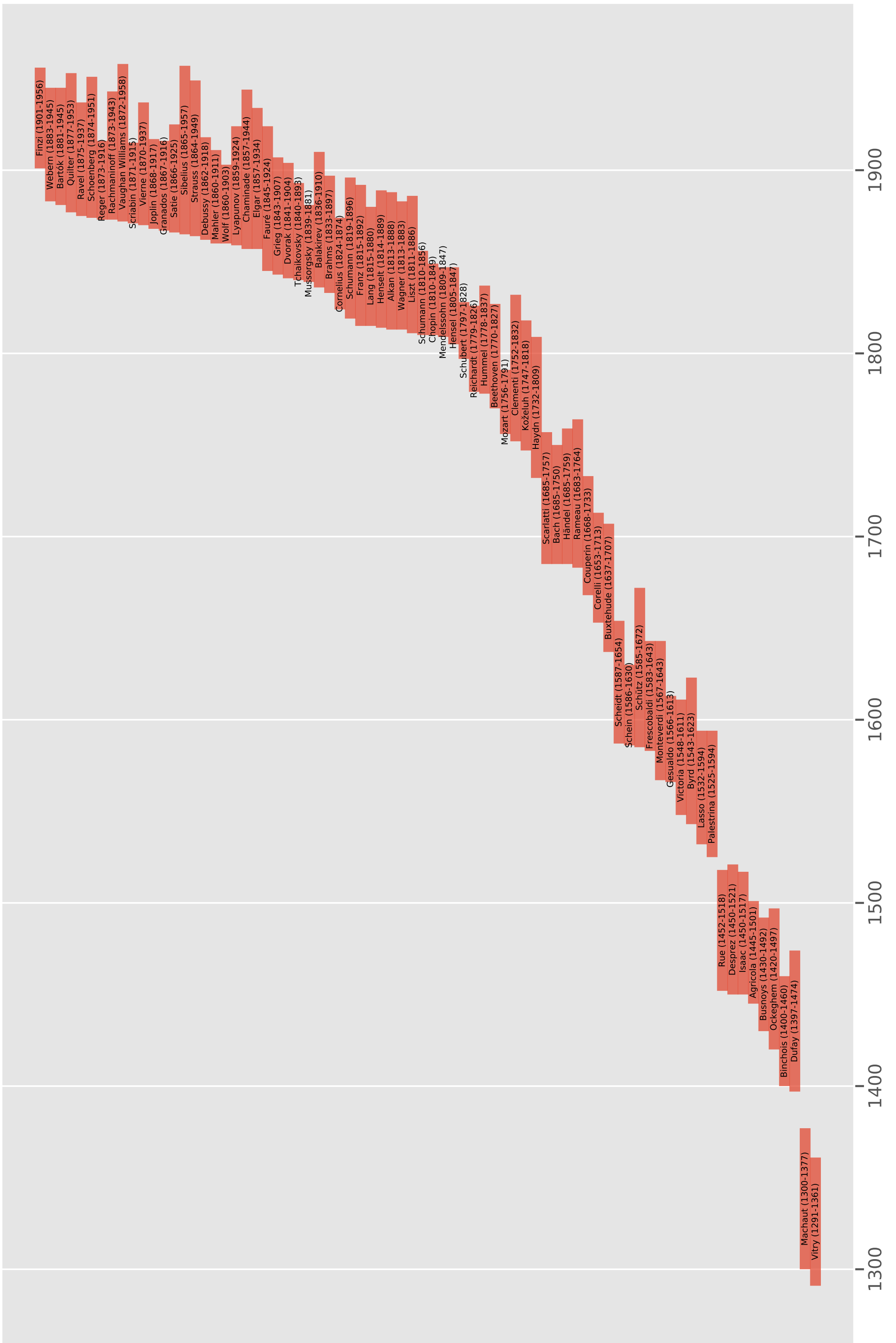


Figure 9.2 – Life dates of composers in the XML corpus. The vertical arrangement of the blocks reflects the order according to the composers' birth dates.

9.2 Musical pieces as tonal pitch-class distributions

Musical pieces are composed of notes which have a certain pitch, a certain duration, and are located in a specific position in a piece. Here, we disregard both the location and the duration of notes and consider only the pitch dimension of musical notes when counting them. It is moreover common to consider notes to be equivalent if their respective pitches are related by one or multiple octaves, and thus to speak of *pitch classes*. Pitch classes come in two varieties. The first, most commonly used representation in computational musicology and music information retrieval, distinguishes twelve different pitch classes. This representation system also assumes the enharmonic equivalence of certain notes, e.g. F♯ and G♭, C♭ and B, etc. The assumption of enharmonic equivalence is usually a consequence of music encoding formats that assume twelve-tone equal temperament, e.g. MIDI, in which enharmonically equivalent notes are indistinguishable. The assumptions of octave and enharmonic equivalence allow to represent pitch classes as residuals in $\mathbb{Z}_{12}$ and to arrange them on a circle. This arrangement of pitch classes is shown in Figure 9.3, called the *circle of fifths*. The numbers correspond to pitch classes in the order of fifths that can be transformed to chromatic ordering, the *chromatic circle*, by the mapping $t \mapsto 7t \mod 12$. Pitch classes that correspond to white keys on the piano are shown in white and pitch classes that correspond to black keys on the piano are shown in black. The second variant of pitch classes does not assume enharmonic equivalence but only that octave-related notes are equivalent, and is hence more general. This representation allows to arrange pitch classes on the *line of fifths* (Temperley, 2000). This linear of ordering tonal pitch-classes has been used by a number of music theorists, e.g. Weber (1851), Riemann (1900), Handschin (1948), and Bárdos (1956). It is shown in Figure 9.4.

Following Temperley (2000), we call the first representation *neutral pitch-classes* and the second one *tonal pitch-classes*. The linear structure of the line of fifths allows to associate each tonal pitch-class with an integer $k \in \mathbb{Z}$ such that this integer represents the number of perfect fifths that lie between this pitch-class and C (Gárdonyi and Nordhoff, 2002). In other words, the integer $k \in \mathbb{Z}$ corresponds to the number of flats (negative integers) or sharps (positive integers) that the major key has in which this tonal pitch-class is the root. For example, D is mapped to $+2$ because D major has a key signature with two sharps, A♭ is mapped to -4 because the key signature for A♭ major has four flats, and C is mapped to 0 because its key signature does not have any accidentals. Yet another benefit of this representation can be seen in the way in which we associate each tonal pitch-class with a color. Positive integers ('sharpened' tonal pitch-classes) are associated with increasingly darker shades of red, negative integers are associated with increasingly darker shades of blue ('flattened' tonal pitch-classes), and C is associated with white as the neutral origin of the line of fifths. This color mapping is used throughout this part. The line of fifths does not only contain all tonal pitch-classes but also a number of central musical scales. For example, pentatonic scales are segments of length 4 (containing five pitch classes) on the line of fifths, e.g. from G♭ to B♭, diatonic scales are defined by segments of lenght 6, e.g. from F to B, the early extensions of the natural diatonic scale by B♭ and F♯ correspond to the segment spanning the eight fifths on

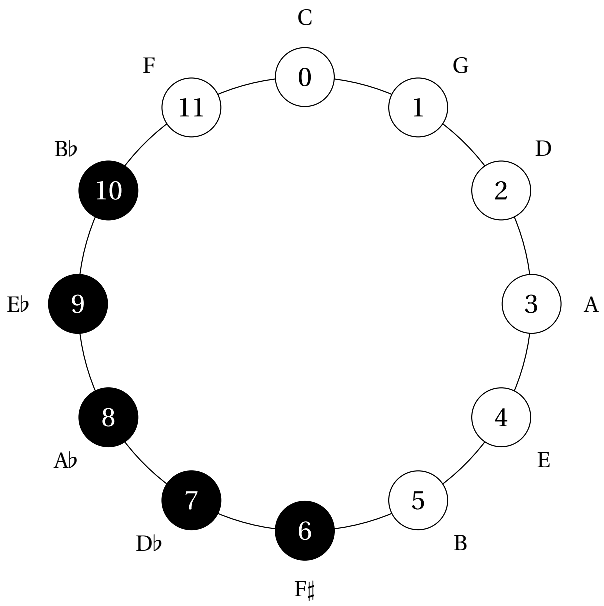


Figure 9.3 – Schematic depiction of the twelve neutral pitch-classes in $\mathbb{Z}_{12}$ on the circle of fifths. One representative of each neutral pitch-class is shown as a tonal pitch-class label next to the node. The coloring of the nodes corresponds to the colors of the keys on the piano.

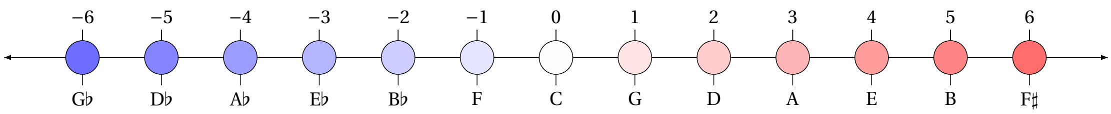


Figure 9.4 – Schematic depiction of the tonal pitch-classes on the line of fifths mapped to integers in $\mathbb{Z}$.

the line between these two tonal pitch-classes, and the two whole-tone scales correspond to the odd and even numbers, respectively. Theoretically, the line of fifths extends to infinity in both directions but in actual compositions only a small segment of it is used. In the corpus that is used here, we consider only the segment from F♭♭ to B♯♯ because no piece in the corpus contains tonal pitch-classes outside this range. The vocabulary size of the corpus is thus $V = 35$, consisting of the seven natural pitch-classes F, C, G, D, A, E, B with two, one, or zero sharps or flats, respectively. While the transformation of tonal into neutral pitch-classes is achieved by deterministically mapping a tonal pitch-class $t \in \mathbb{Z}$ to a neutral pitch-class $t \mapsto t \mod 12 \in \mathbb{Z}_{12}$ (in fifths ordering), the reverse direction involves some kind of inference and is called the problem of *pitch spelling* (Temperley, 2001; Cambouropoulos, 2003; Stoddard et al., 2004; Chew and Chen, 2005; Meredith, 2006).

The distinction between the two types of pitch-classes is not only relevant for music encoding and representation but also relates to the conceptualization of tonal relations in music. Concretely, neutral pitch-classes assume enharmonic equivalence which is the basis for many musical styles, for instance the compositions based on twelve-tone rows by atonal composers such as Arnold Schoenberg, Alban Berg, and Anton Webern (Schoenberg, 1975; Straus, 2005). It is also largely employed in Jazz, giving rise to phenomena such as the *tritone substitution* (Biamonte, 2008; Levine, 2011). Moreover, enharmonic equivalence lies at the heart of compositions by tonal composers who use scales that are based on symmetric divisions of the octave, e.g. Stravinsky (Tymoczko, 2002; Van Den Toorn, 2003), Debussy (Forte, 1991), and Messiaen (Messiaen, 1944). The representation of tonal pitch-classes, on the other hand, with its differentiation between enharmonically equivalent notes is more closely related to a diatonic way of thinking and might potentially entail even different tuning systems other than equal temperament (Sethares, 2005). The issue of tonal versus neutral pitch-classes also relates to the question of music notation and its *orthography*. The Western musical notation system was developed in order to accomodate largely diatonic music in which the tonal material is confined to small regions on the line of fifths. Music that is highly chromatic or is built upon symmetrical scales can not always be notated correctly—within the Western notation system, a composer has to chose a tonal spelling, e.g. opt for either F♯ or G♭, even if the composition is meant to be atonal. It is telling that the question of orthography and its relation to tonality is largely discussed in the context of 19th-century composers, e.g. Mendelssohn, Schumann, Chopin, or Liszt (Atlas, 1990), Mussorgsky (Perry, 1995, 1998), Schubert (Cohn, 1999; Noll, 2009), or composers at the turn to the 20th century, such as Scriabin (Perle, 1984; Wai-Ling, 1993) and Bartók (Gillies, 1983). We will see throughout the following chapters that, over the course of the 19th century, crucial changes take place.

Due to the encoding of the corpus, we adopt the representation of pieces as bags of notes (see Section 3.2) and represent each one as a distribution over the $V = 35$ tonal pitch-classes. That is, in the corpus used here with D =2,012 pieces, the tonal pitch-class distribution of a musical piece x_d is given by the relative frequencies of the tonal pitch class in that piece, for $d \in \{1, \ldots, D\}$. Each piece is thus described by a V-dimensional vector, $x_d \in \mathbb{R}^V, \forall d \in \{1, \ldots, D\}$, containing positive real numbers that sum to 1. In this chapter, we do not make any further assumptions about the process that generated this distribution and will postpone these considerations until Chapter 11. In this view, pieces simply correspond to points in the $V-1$-simplex

$$\Delta^{V-1} = \left\{ x_d \in \mathbb{R}^V \;\middle|\; \sum_{i=1}^{V} x_{d,i} = 1;\; x_{d,i} \geq 0 \right\}. \tag{9.1}$$

In this space, those pieces with very different tonal pitch-class distributions will be very distant, whereas pieces that have similar tonal pitch-class distributions will be closer to one another and form clusters in the $V-1$-simplex. It is important to note that the bag-of-notes representation relies on the assumption that the V dimensions are independent, meaning that this model does not *a priori* assume any particular order between the tonal pitch-classes.

The average tonal pitch-class distribution of all pieces in the corpus is shown in Figure 9.5. Figure A.1 in the appendix shows this distribution separately for each century from 1361 to 1943. Although we know that one can in principle order all tonal pitch-classes on the line of fifths (Figure 9.4), we do not incorporate this assumption into the model but will show instead that it can be inferred from the data.

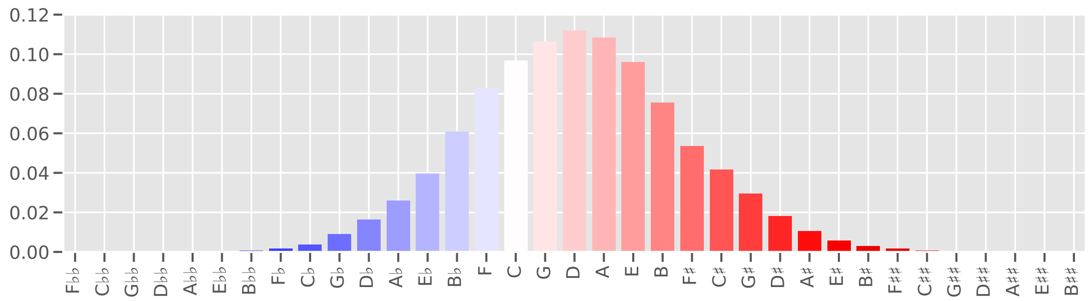


Figure 9.5 – Average tonal pitch-class vector representing the distribution of pitch classes in the corpus.

The frequencies in Figures 9.5 and A.1 have been sorted and colored along the line of fifths, revealing that they are almost normally distributed in this tonal space, centered on D. This shows that, on average, the natural tonal pitch-classes (from F to B) are most common, while altered tonal pitch-classes (with sharps or flats) are much less common in the corpus. Note that this average distribution is based on the relative frequencies of untransposed tonal pitch-classes in the pieces of the corpus *without* taking into account the key or the mode of the piece. The distribution in Figure 9.5 shows that the tonal pitch-class distributions of the pieces in the corpus are concentrated towards the natural tonal pitch-classes. Moreover, they are centered not on C but on D. While C is the center on the line of fifths under the interpretation that the association with integers depends on the number of accidentals (see Figure 9.4), D is the center within the diatonic scale, having three natural tonal pitch classes to its left (F, C, G) and to its right (A, E, B). This means that, on average, most pieces contain no or only a few tonal pitch-classes with accidentals. While this does not tell us, for instance, whether the distributions of pieces in different keys are very similar or not, it certainly shows that not all keys are used in the same way. Rather, compositions are in general located at the center of the line of fifths which may be simply related to the fact that this facilitates the notation of the music. The next section studies relations between the pitch-class distributions of the pieces.

10 The line of fifths

10.1 Discovering the line of fifths in the corpus

A number of recent studies has used the geometrical interpretation of note distributions in pieces as points in a high-dimensional space (Huang et al., 2017; Weiß et al., 2018; Harasim et al., submitted), based on large corpora of MIDI-encoded data. As mentioned above, this encoding entails the assumption of enharmonic equivalence and thus limits the extent to which tonal relationships can be extracted from the data, since musical pieces are represented by only twelve neutral pitch-classes. This is not the case here, where the XML corpus contains the exact spelling of the pitches and we only make the assumption that octave-related notes are equivalent. Since the space we consider here has 35 dimensions, it is impossible to visualize this space and pieces in it to inspect whether their arrangement contains any meaningful information. We address this problem by using a method for *dimensionality reduction* called Principal Component Analysis (PCA; Jolliffe, 2002) that projects the data into a lower-dimensional space of dimension M while at the same time maintaining characteristic properties of the original space. PCA thus can aid to achieve a better understanding of the global structure of the space.

To perform a PCA, the data is represented as a matrix

$$X = \begin{bmatrix} x_1^\top \\ \vdots \\ x_D^\top \end{bmatrix} \in [0,1]^{D\times V},\ x_d \in \Delta^{V-1}, \tag{10.1}$$

where the rows are given by the D data points (the pieces in the corpus), and the columns are given by the V features (the number of distinct tonal pitch-classes in the vocabulary). All entries in X range from 0 to 1. PCA determines the $M \leq V$ largest directions and magnitudes of the variance in the data in X by first calculating the *covariance matrix*

$$K_X = \mathrm{cov}[X, X] = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^\top] \in \mathbb{R}^{D\times D}, \tag{10.2}$$

where $\mathbb{E}$ denotes the expected value. The main directions of the variance in the data and their magnitude is given by the eigenvectors w_i and eigenvalues λ_i of K_X which are calculated by solving the equation

$$K_X \cdot w_i = \lambda \cdot w_i. \tag{10.3}$$

The projection into the lower-dimensional space is then achieved by selecting the M largest eigenvalues and their corresponding eigenvectors, and transforming the data to X', the dimensionality reduction of X, by

$$X' = X \cdot [w_1, \ldots, w_M] \in \mathbb{R}^{D \times M}, \tag{10.4}$$

where each row of $X' \in \mathbb{R}^M$ corresponds to the datapoints in the reduced space. The sum of all eigenvalues λ_i is the total amount of variance in the data and the variance explained by each principal component is given by $\lambda_i / \sum_j \lambda_j$. In the present context, X was transformed to have zero mean before applying PCA, but the variance was not standardized to 1. This is justified by the fact that all features are on the same scale, and because the differences in the variance between the respective tonal pitch-classes is of particular interest here.

Figure 10.1 shows the data reduced to the two-dimensional Euclidean plane $\mathbb{R}^2$ ($M = 2$). The data was whitened (Kessy et al., 2018) so that the unit of both axes is the standard deviation. Each dot represents a piece and the color of each piece corresponds to the line-of-fifths position of its *tonal center* which we operationalize as the most frequent tonal pitch-class in that piece. This definition follows Tymoczko (2011, p. 4) who defines "centricity"—which can be established by the most frequent note—as one of the core components of tonality. The dimensionality reduction shows that pieces with similar coloring are close together and additionally shows that the colors are largely ordered along the line of fifths. The arrangement of pieces in this reduced space implies that musical pieces that have tonal centers which are close on the line of fifths largely also have similar tonal pitch-class distributions.

While PCA is one of the most commonly used methods for dimensionality reduction, there are many others (sometimes relying on particular assumptions about the distribution of the data). In a qualitative comparison, several dimensionality reduction methods were contrasted: Independent Component Analysis (ICA; Hyvärinen and Oja, 2000), Isomap (Tenenbaum et al., 2000), Spectral Embedding (SE; Belkin and Niyogi, 2003), Multidimensional Scaling (MDS; Kruskal, 1964), and t-distributed Stochastic Neighbor Embedding (t-SNE; Van Der Maaten and Hinton, 2008). The latter, which emphasizes the local over the global structure of the data, found multiple sub-clusters that were relatively homogenuous with respect to the tonal centers, thus emphasizing more strongly the local similarities between pieces than the global structure of the space. All other methods achieved a similar picture to PCA by also largely arranging the data along the line of fifths. Figure A.2 in the appendix shows the clusterings of the pieces in two-dimensional space according to these methods.

In the present context, it was opted for PCA because it preserves most of the global structure,

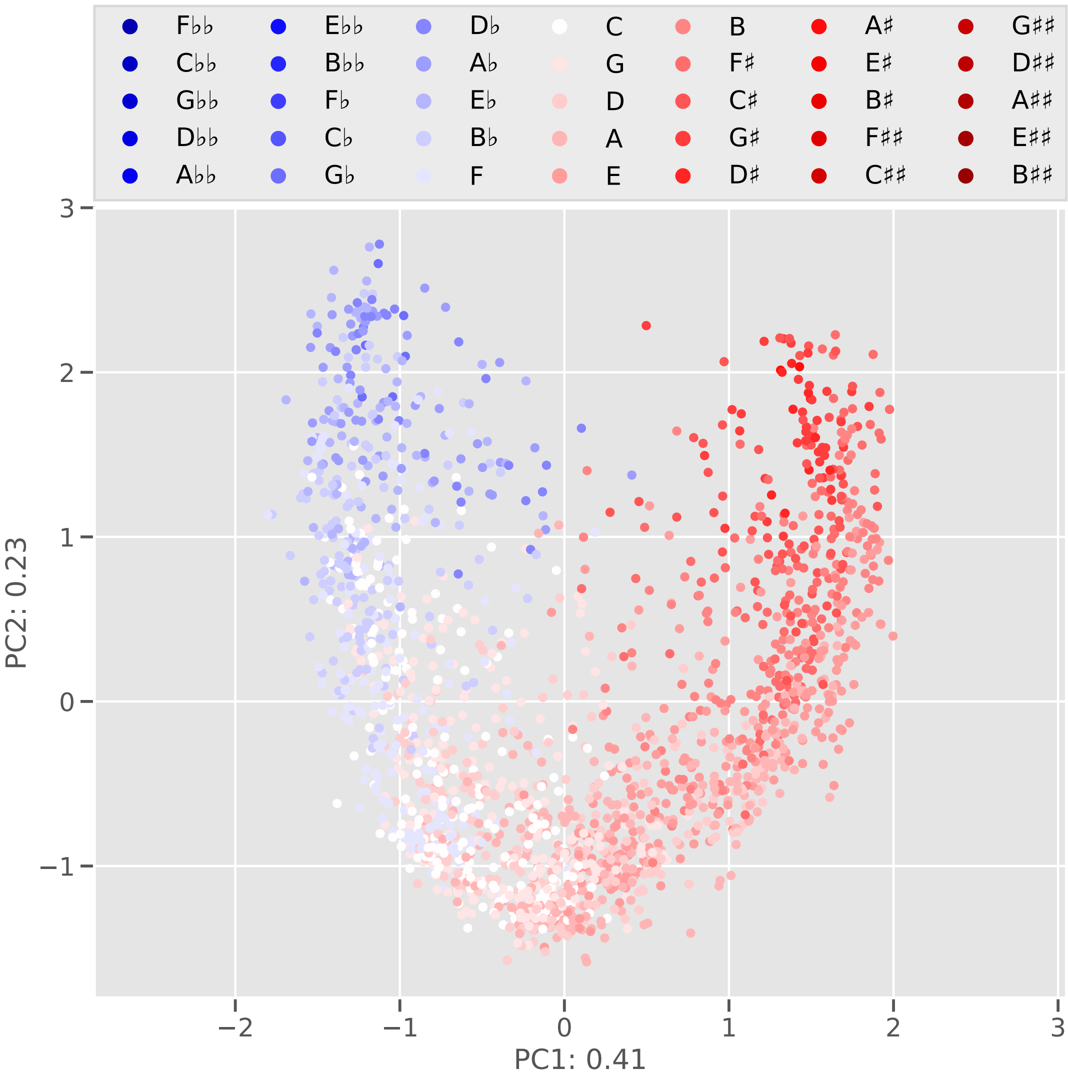


Figure 10.1 – Dimensionality reduction via PCA. Points represent pieces and the coloring corresponds to the position of the tonal center on the line of fifths.

and the interpretation of the results is straight-forward. Although we cannot visualize the corpus in the V-dimensional space Δ^{V-1}, we can calculate some statistics of this empirical distribution such as its mean and its variance. The mean distribution of all pieces was already shown in Figure 9.5. One of the advantages of the dimensionality reduction method PCA is that the axes in the reduced space, called the *principal components*, can be well interpreted since they express how much of the data variance in the original space Δ^{V-1} is retained in the reduced space $\mathbb{R}^2$. The first two components are shown separately in Figure 10.2.

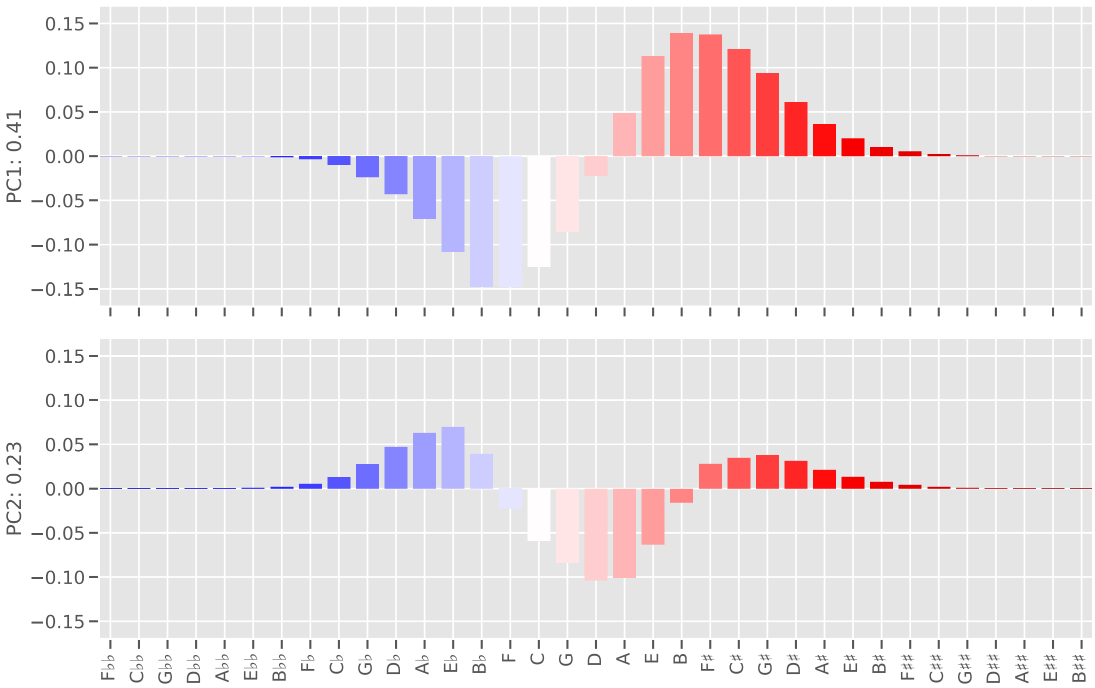


Figure 10.2 – First two principal components for the tonal pitch-class distributions.

The first principal component (PC1, top panel) distinguishes the two directions of the line of fifths—in ascending vs. in descending direction on the line of fifths (red vs. blue coloring)—as seen from D, the most frequent tonal pitch-class in the corpus. This dimension accounts for 41 percent of the data variance. It is related to the tonality of the pieces in that keys that are close on the line of fifths have a larger intersection of tonal pitch-classes that they can use. For example, the key of F major shares all but one tonal pitch-class with the key of B♭ major, which is its direct neighbor on the line of fifths, but shares no tonal pitch-classes with the key of F♯ major, which is seven fifths away from F major. Of course, pieces do not only use the tonal pitch-classes of their main key. They can employ many out-of-key notes, e.g. in modulations to more distant keys, in chromatic passages, or chord alterations. The first principle component is thus not strictly related to keys but to a more general conception of the 'global' tonality of a piece and the position of its tonal center on the line of fifths. The second principal component (PC2, bottom panel) represents the distance to the center of the line of fifths —regardless of the direction—and distinguishes pieces with natural tonal centers (from

F to B; white or very light colors) from more alterated tonal pitch-classes (darker shades of blue and red). This distinction accounts for 23 percent of the variance in the data. The first two principal components represent the direction and the distance on the line of fifths. They together account for a total of 64 percent of the variance in the data but simplify the space from $V = 35$ dimensions to just two. The dimensionality reduction of tonal pitch-class distributions thus captures an essential aspect of tonality, namely the importance of the line of fifths for the organization of notes in compositions.

10.2 Historical expansion on the line of fifths

Based on the importance of the line of fifths for the organization of tonal material, we investigate in this section how broadly the pieces in the corpus are distributed on it. Each piece does not only determine a position on the line of fifths through its tonal center, it also defines the smallest line-of-fifths segment that contains all of its tonal pitch-classes. The length of this segment, in turn, determines which intervals can potentially be used in this piece. In other words, each tonal pitch-class distribution also defines which intervals are potentially used in this piece. For example, a piece containing only the pitch classes D and A can only contain the intervals of the unison and the fifth as well as their complements, the octave and the fourth. A piece that contains only natural tonal pitch-classes cannot contain any chromatic notes since it contains the diatonic but not the chromatic semitone. The length of a line-of-fifth segment containing all tonal pitch-classes in a piece is called the *fifth range* of that piece (Gárdonyi and Nordhoff, 2002). The fifth range of strictly diatonic pieces is 6, while pieces with larger fifth ranges can contain chromatic notes, and pieces with a fifth range larger than 12 potentially contain enharmonic notes. It is thus reasonable to expect that the fifth range of pieces increases over time since, for example, Renaissance pieces rarely contain chromaticism which is more common in Classical and in particular in Romantic compositions. This simple measure of the fifths range enables us to compare all pieces in the corpus and to trace the historical changes regarding the spread of musical pieces on the line of fifths. The fifth range of all pieces in the corpus is shown in Figure 10.3. The two horizontal lines (gray, dashed) separate the diatonic (top) from the chromatic (middle) and enharmonic (top) segment.

We use Locally Weighted Scatterplot Smoothing (LOWESS; Cleveland and Devlin, 1988) to show historical trends in the distribution of fifths ranges. LOWESS fits a local polynomial regression not to the entire dataset but to a neighborhood of each datapoint that is determined by a fraction parameter δ defining what percentage of the whole data is taken into account when calculating the regressions. The larger this fraction is the smoother the resulting lowess curve will be. Here, this parameter was set to $\delta = .15$. Note that the range of years covered by the neighborhood can vary, depending on how the data is distributed over time. In periods with fewer pieces, a larger time range will be taken into account and *vice versa.* This is why the lines are much smoother before ca. 1700 and show much more variability in later decades and centuries. The weights for this regression are chosen so that they give less weight to datapoints further away from x_0. A commonly used weighting function is the so-called *tricube function*,

$$w(x_i) = (1 - |x_i - x_0|^3)^3. \tag{10.5}$$

The bundle of black lines in Figure 10.3 shows 500 boostrapped LOWESS regression lines.

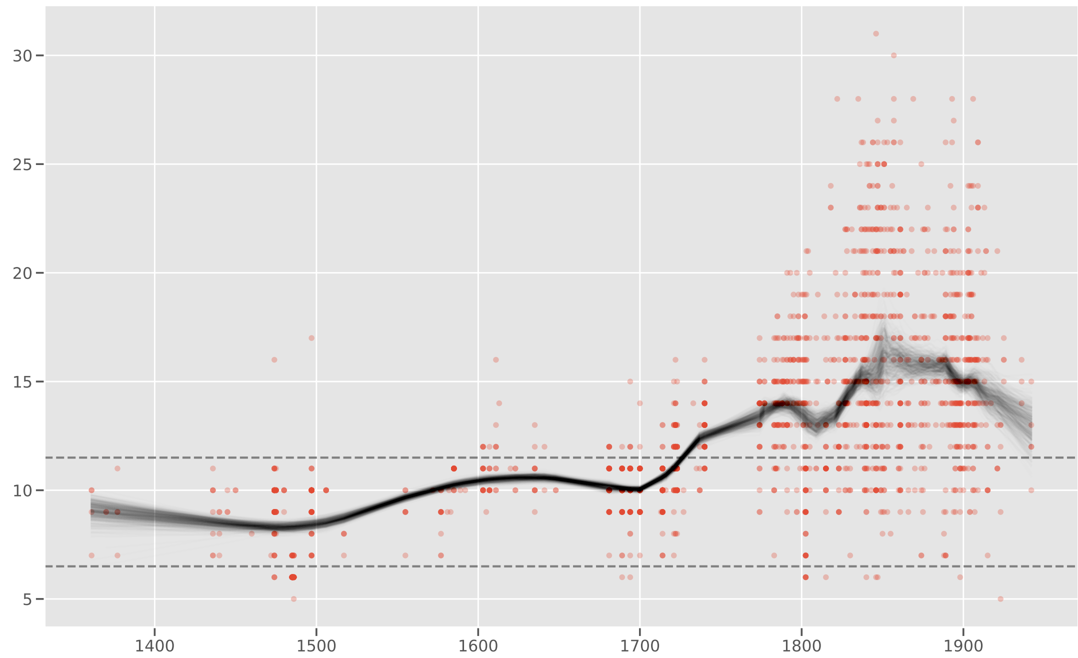


Figure 10.3 – Diachronic changes in the fifth-range of tonal pitch-class distributions of musical pieces. The black lines show 500 bootstrapped LOWESS curves with fraction parameter $\delta = .15$.

It can be seen that, over the course of the historical timespan under consideration, there is a substantial increase in the fifth range of musical pieces. While there are very few entirely diatonic pieces, the fifth range grows continuously, supporting a history of tonality that proceeds from diatonicism through chromaticism to enharmonicism (Fétis, 1844; Gárdonyi and Nordhoff, 2002). The piece with the largest fifth range is Frédéric Chopin's *Polonaise*, op. 61 (1846), spanning 31 fifths. The variance in the bootstrapped trendlines is relatively large before 1400, a consequence of the sparsity of data in that period. However, the variance is largest in the 19th century where the data is least sparse. Hence, the variance can be mainly attributed to the actual variability in the compositions. While the corpus does contain composers such as Chopin, Alkan, and Liszt who are known for their highly chromatic style and usage of enharmonicism, leading to a generally higher fifth range of their compositions, the corpus does also contain composers such as Henselt, Lang, Franz, and Cornelius, whose musical language exhibits a different, more traditional style. The variance in the boostrapped trendlines points to the fact that, while there is a generally growing trend towards chromaticism and enharmonicism, the diversity in the usage of musical styles—as manifested in the fifth range of compositions—increases, too.

10.3 Tonal pitch-class evolution

The line of fifths appeared as the central organizing space for tonal pitch-class distributions in Section 10.1 without taking the temporal distributions of pieces into account. We now turn to the study of historical developments in the usage of tonal pitch classes. What can we infer about tonality from the changing use of tonal pitch-classes? Recall that we model a piece x_d as a distribution over the $V = 35$ tonal pitch-classes. Moreover, each piece is associated with a year of composition or publication. As Figure 9.1 has shown, the dataset is not uniformly distributed over time. On one hand, there are some large gaps between periods, whereas on the other hand some years contain many pieces at the same time. In order to trace the historical usage of tonal pitch-classes, the *tonal pitch-class evolution*, we calculate the average tonal pitch-class distribution for a given year, analogously to the overall average tonal pitch-class distribution in the corpus (Figure 9.5). For years without data, we assume no change in the evolution of tonal pitch-classes and replicate the values from the previous year. This results in a tonal pitch-class distribution for each year in the historical range of 582 years from 1361 to 1942.

What is the interpretation of a per-year tonal pitch-class distribution? If a corpus of musical pieces under the bag-of-notes model represents an approximation of the entire musical material within the historical range, then the per-year distributions can be interpreted as temporal slices of the entire range. In a way, they approximate the music that was present in a given historical year and show the average relative frequency of each tonal pitch-class for each year and piece in the corpus. Moreover, we use a *moving average* to smoothen these distributions such that, for each year in the range under consideration, an average value is calculated based on the previous 50 years. The resulting tonal pitch-class evolution plot is shown in Figure 10.4. The Figure legend shows the mapping of tonal pitch classes to colorsand the tonal pitch-classes are ordered along the line of fifths.

The solid black curve line shows the normalized entropy of the pitch-class distributions at any given point in time. This line is smoothed by the same procedure as the individual per-year pitch-class distributions and is thus an adequate measure for the randomness in these distributions for a given year. The value of this line is independent of the number of non-zero tonal pitch-classes in a given year, since it is normalized by its maximal value which is given by $\log_2(n)$ where n is the number of non-zero tonal pitch classes in that year. If the tonal pitch-class distribution for some year were uniform, the normalized entropy would be maximal at 1. The relatively stable entropy line expresses the fact that, although the number of used tonal pitch-classes increases over time, their distribution in at a given historical moment has a similar normalized entropy that rises only slightly, i.e. the randomness in these distributions remains largely similar. Comparing the entropy curve with its average (shown by the dashed horizontal line in Figure 10.4) emphasizes this increasing trend and also shows that there is a turning point around 1700. Prior to that point, the normalized entropy is lower than the average, and after 1700 it is larger.

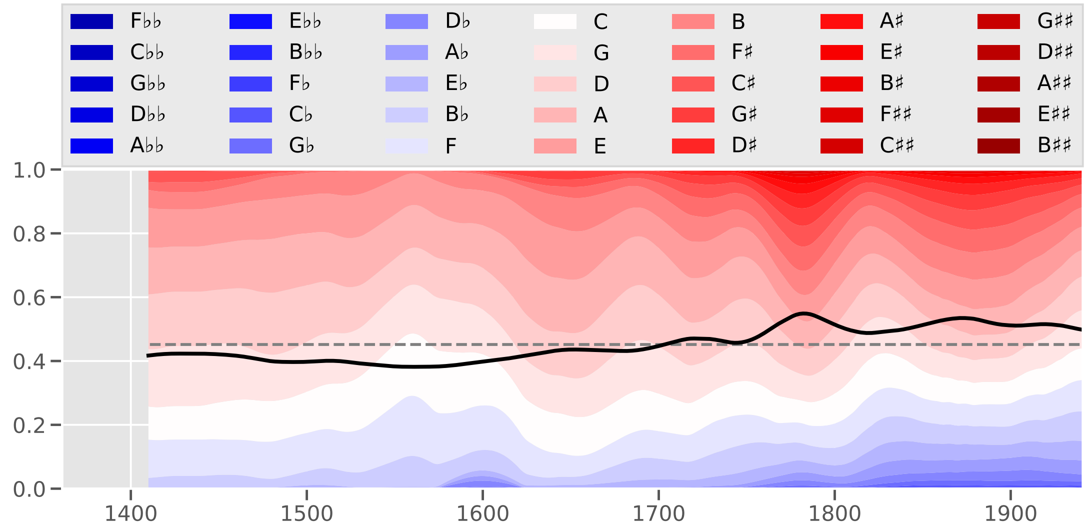


Figure 10.4 – The evolution of tonal pitch classes taking into account a 50-year window. The solid black curve shows the smoothed normalized entropy over the pitch-class distributions in each year and the dashed horizontal line shows the average normalized entropy across the historical timespan.

The smoothed trends show that sharpward (red) tonal pitch-classes are generally much more common. This is due to the fact that all natural tonal pitch-classes from G to B have one or multiple sharps in their major key signatures. Until into the 16th century, pieces in the corpus consist almost exlusively of natural tonal pitch-classes plus B♭, F♯, and C♯. The increasing fifth range (Figure 10.3) is reflected in a larger number of pitch classes. This number increases in particular after 1700 where composers begin to use more flat as well as more sharp tonal pitch-classes. For some tonal pitch-classes, it seems to be the case that their evolution curves are almost parallel, at least within some periods.

How are the evolution curves for the individual tonal pitch-classes related to each other? In other words, what can be inferred from the *co-evolution* of tonal pitch classes? Every tonal pitch-class is associated with a vector that contains the probability of this pitch class for each year. We define the co-evolution of two tonal pitch classes as the pairwise correlation ρ of their evolution vectors p and q, given by

$$\rho_{p,q} = \frac{\mathrm{cov}(p,q)}{\sigma_p \sigma_q}, \tag{10.6}$$

where $\mathrm{cov}(p,q)$ is the covariance and σ the standard deviation. The correlations between all tonal pitch-classes are shown in Figure 10.5.

This correlation matrix exhibits a number of interesting regularities. First, its block structure almost perfectly coincides with segments on the line of fifths that are determined by the number of accidentals. These segments are here emphasized by the white lines. The three blocks

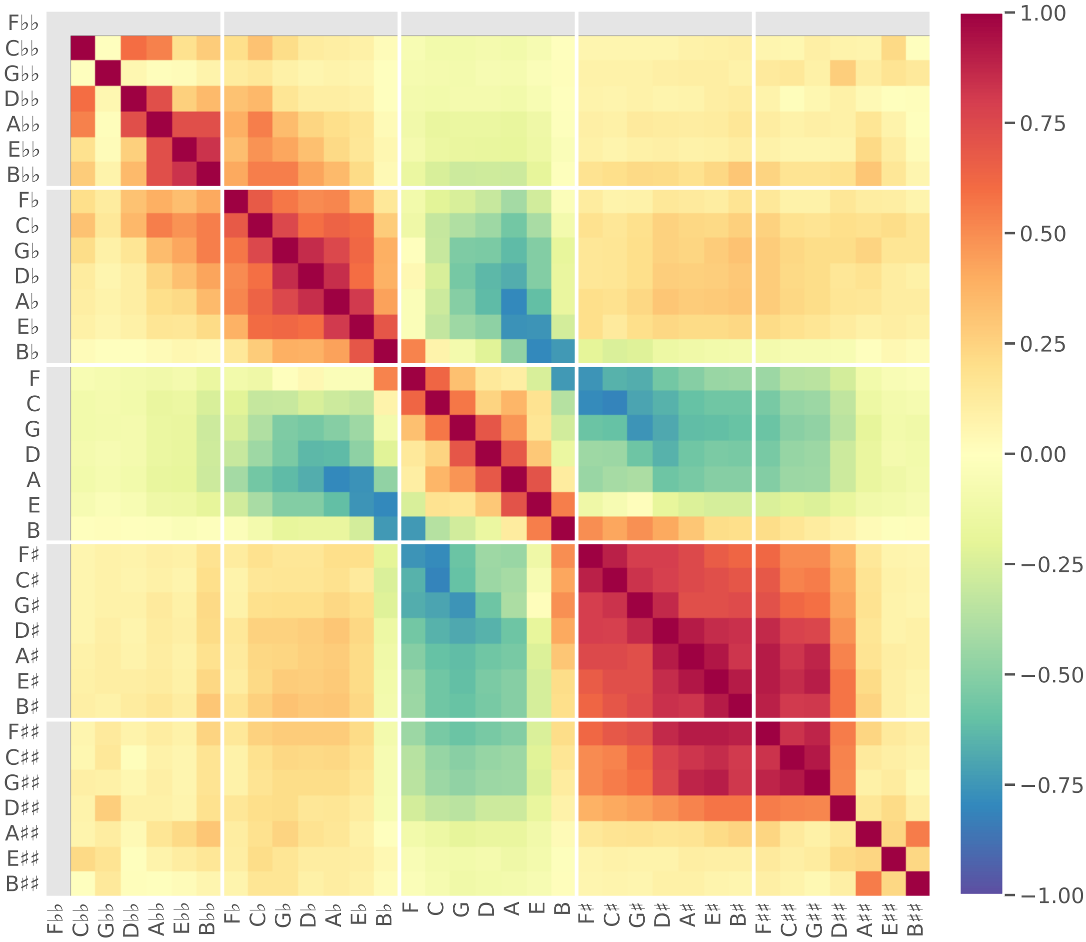


Figure 10.5 – The co-evolution of tonal pitch classes.

along the main diagonal of the matrix with relatively strong correlation values correspond to the co-evolution of tonal pitch-classes with flat, sharp, and no accidentals, respectively (top left to bottom right). The two blocks in the lower left and upper right corners of the matrix with moderate but positive tonal pitch-class co-evolution values correspond to the correlations of flat with sharp and, per symmetry sharp with flat tonal pitch-classes. Note that the diagonals in this matrix describe intervals between tonal pitch-classes. The main diagonal describes the unison, the diagonal above the perfect fifth, the one above that the major second, etc. Since the strongest correlations are found on intervals close to the diagonal, we can conclude that tonal pitch-classes that are close on the line of fifths also correlate highly in their historical evolution. The blocks with negative correlations are equally interesting. The weakest correlations overall can be seen in the parallel diagonals depicting the invervals of the chromatic semitone and the tritone, e.g. between A and A♭ and E♭, between E and E♭ and B♭, etc., but not through the entire space but rather for the central segment of the line of fifths from F to A♯. One can deduce that the role of these intervals, the chromatic semitone and the tritone, is very distinct in white-key music based on the natural tonal pitch-classes but that they are less pronounced in keys that are further apart from C on the line of fifth. It seems to be the case that pieces in keys with more accidentals as key signatures are also more chromatic in general.

Recall that we can find the directions of the largest variance in a dataset via PCA that involves the calculation of a covariance matrix. Since the values of the heatmaps in Figure 10.5 are given by a correlation matrix, we can apply the same methodology here. The difference in using the correlation instead of the covariance matrix is that the latter is expressed in terms of the data that has been standardized to have unit variance. Accordingly we can apply PCA to the correlation matrix in Figure 10.5 to find the principal components that account for most of the variation in the tonal pitch-class co-evolution and to quantify some of the earlier observations. Figure 10.6 shows the dimensionality reduction of the co-evolution matrix. Tonal pitch-classes that have high correlations appear close together (e.g. C and G), while those having low correlations are more distant (e.g. C and F♯). The black line connecting the tonal pitch-classes was added to emphasize that the co-evolution of tonal pitch-classes also reveals the line of fifths, at least for the tonal pitch-classes on its center.

The first two principal components for the tonal pitch-class co-evolution are shown separately in Figure 10.7. The variance explained by each of the components can be interpreted as the importance of these dimensions for the data. The first principal component (PC1) accounts for 66 percent of the variance in the tonal pitch-class co-evolution and confirms the observation that tonal pitch-class co-evolution is largely determined by regions of the line of fifths where the tonal pitch-class have the same number of accidentals. The wave-like pattern in the first principal component switches from positive to negative values and back almost exactly at the boundaries between tonal pitch-classes with two flats (♭♭), one flat (♭), no accidentals, one sharp (♯), and two sharps (♯♯), although this is not as clear between the single and double sharps. The second principal component (PC2) accounts for 20 percent of the variance in the data and represents the distinction between 'flat' and 'sharp' tonal pitch-classes. Recall that

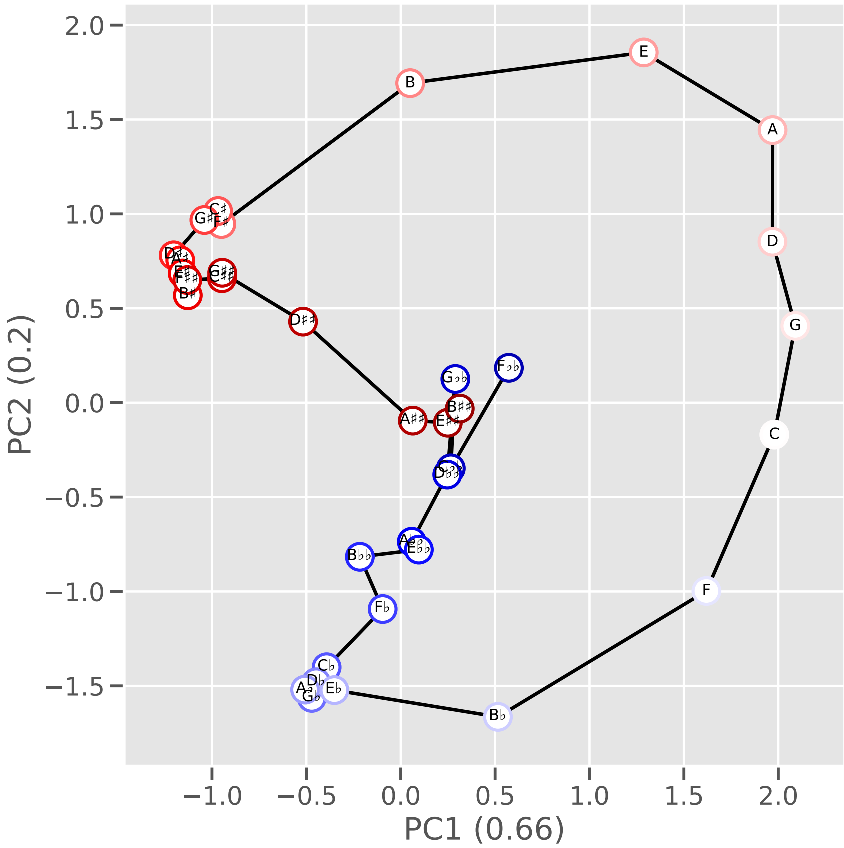


Figure 10.6 – Two-dimensional PCA reduction of tonal pitch-class co-evolution.

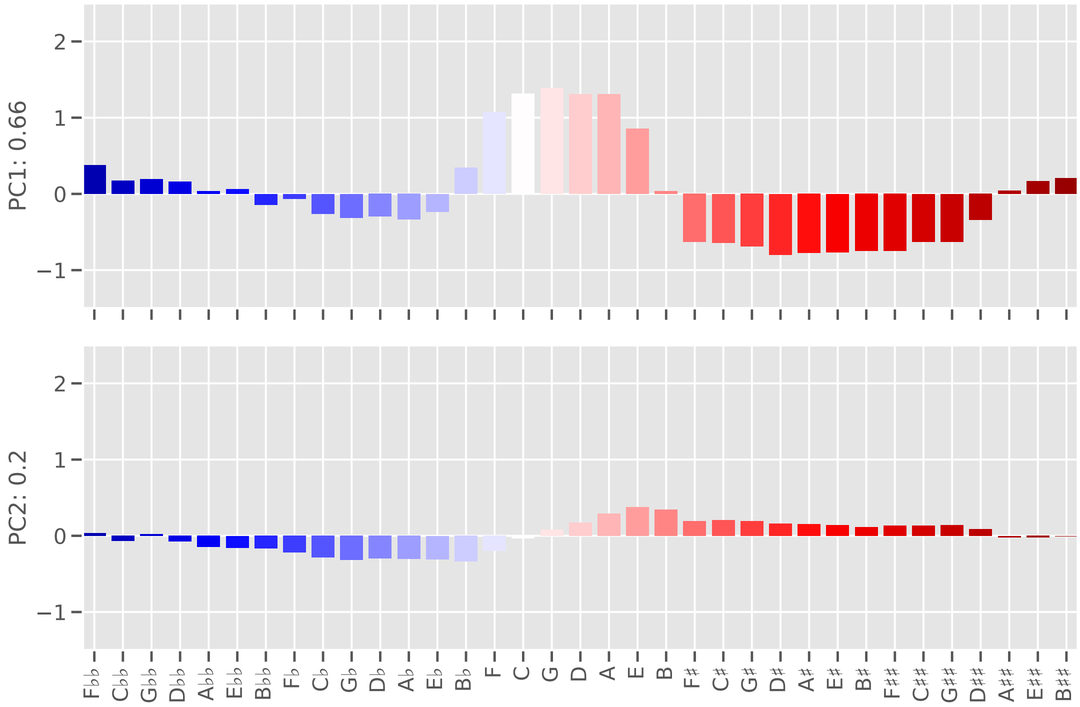


Figure 10.7 – Separate plots of the first two principal components jointly accounting for 86% of the variance in the data.

the ‘sharp’ tonal pitch classes also include the natural pitch classes G, D, A, E, and B, since the major keys in which they are the tonics all have one or several sharps as key signature. While more extreme sharp pitch-classes, such as F♯♯ and flat pitch-classes such as B♭♭ are positively correlated, the low and negative correlations between more central tonal pitch-classes such as B and B♭ pulls them apart in the reduced space. The principal components of the dimensionality reduction for the tonal pitch-class co-evolution matrix (Figure 10.7) is astonishingly similar to the principal components of the reduction of the distributions of tonal pitch-classes (Figure 10.2). In both cases, the distance on the line of fifths, as well as the distinction between flat, natural, and sharp tonal pitch-classes accounts for the largest proportions in the variance of the data.

Tonal pitch-class co-evolution per historical period. As the last sections have shown, the fundamental structure underlying tonal pitch-class co-occurrence as well as the co-evolution of tonal pitch-classes is the line of fifth. In both cases we have taken the entire corpus into account. We conclude this chapter by studying in more detail the tonal pitch-class co-evolution in separate historical periods. To that end, we divide the corpus into centuries within the historical range and calculate the tonal pitch-class co-evolution values, the correlations between the evolution curves, for each period separately. Since the whole corpus covers a range of approximately 600 years, we study here seven periods with a duration of 100 years each, except for the first and last period, which are somewhat shorter since the years of the earliest and the latest piece in the corpus do not coincide with the boundaries of the centuries. Figure 10.8 shows the heatmaps for these periods. Each of these heatmaps shows the tonal pitch-class co-evolution values for all pieces of the corpus in that century. The titles of the subplots also indicate the number of pieces and the number of distinct years per period. As we have shown before, musical compositions historically spread out across the line of fifths. This can be seen in the decrease of gray areas in the heatmaps. Since in some periods certain tonal pitch-classes are not contained in any piece, the correlation with other pitch classes is not defined and no correlation value can be shown. Recall that the tonal pitch-class F♭♭ does not occur in any piece in the whole corpus. For this reason, the topmost row and the leftmost column in all heatmaps is gray.

The period before 1400 contains only a very small subset of ten pieces in three distinct years. We will therefore not consider it in the following. While the 15th and the 16th century contain 218 and 60 pieces, respectively, there are only twelve and eleven unique years within these ranges. Recall that the tonal pitch-class distributions underlying the calculation of the tonal pitch-class co-evolution consist of aggregates per year. The number of unique years is particularly small for the earlier periods because most pieces in the corpus from these periods are published in collections and their date of composition is not known. Keeping these caveats in mind, one can observe that, over the the course of the 16th and 17th century, the blocks dividing the natural from the flattend and sharpened tonal pitch-classes stabilize.

The situation is quite different for the later periods. Not only is the number of pieces in these periods larger than in most of the earlier ones, but the number of unique years of composition or publication increases as well. The largest difference between the respective heatmaps can be seen between the 18th and the 19th century, containing 416 and 873 pieces in 26 and 86 unique years, respectively. The 18th century most closely resembles the overall co-evolution matrix (Figure 10.5). We see a very clear partition into regions on the line of fifths resulting in very pronounced blocks of high positive and negative correlation. The flattened, the natural, and the sharpened tonal pitch-classes correlate strongly and positively with themselves but strongly and negatively with each other. This is quite different in the 19th century, where the correlations between the flats and sharps become much weaker. While in the centuries before, tonal pitch-classes with accidentals do occur in approximately the same amount (including not at all), they gain more independence in the 19th century where they still tend to co-occur but their evolution curves are not as strongly correlated as before. The tonal pitch-class co-evolution thus provides data-based evidence for a stark shift in compositional practice between the 18th and the 19th century.

Another difference is that the block structure of the heatmap in the 18th century is replaced by a broad band along the main diagonal in the 19th century. Apparently, the usage of tonal pitch-classes in the 18th century, and consequently the intervals between them, depend more on their absolute position on the line of fifths than in the 19th century, where the co-occurrence of certain tonal pitch-classes is somewhat independent from the position on

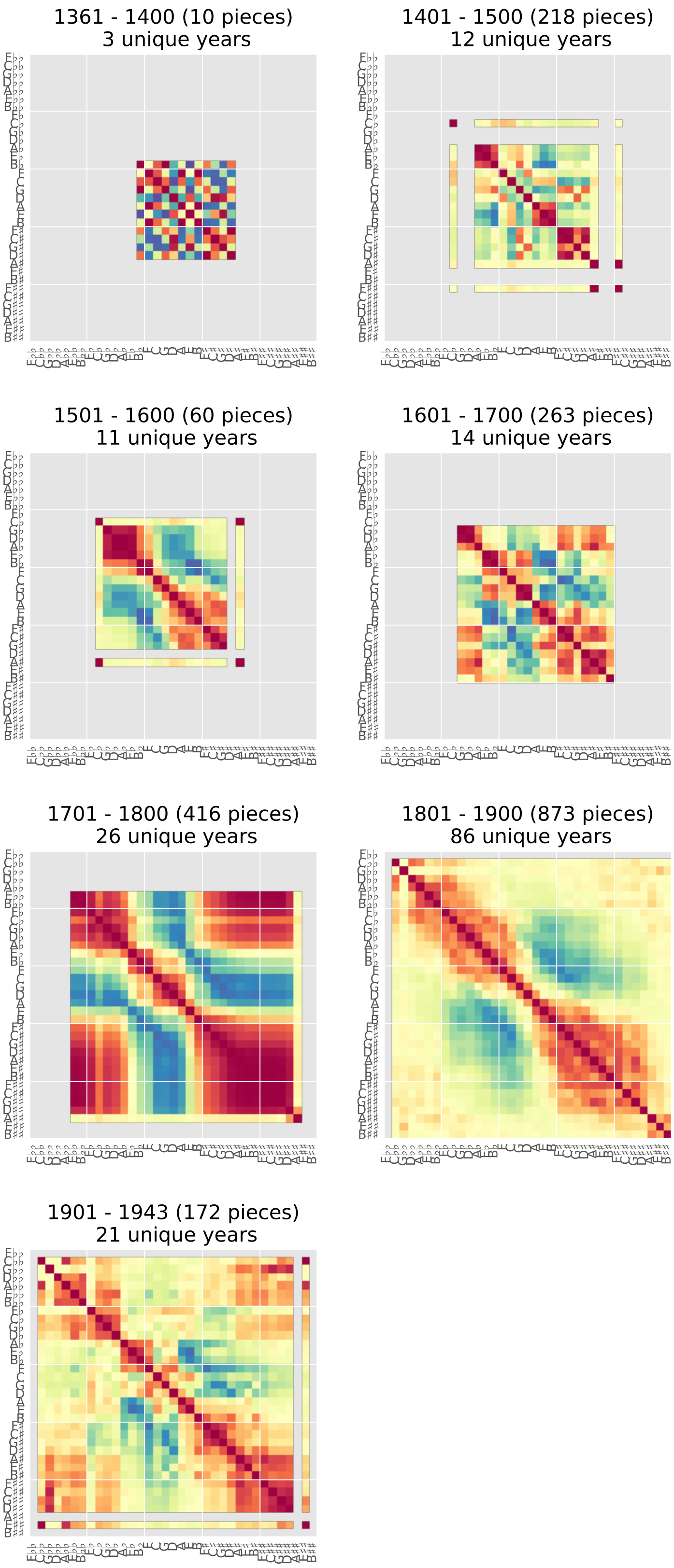


Figure 10.8 – The co-evolution of tonal pitch classes in different historical periods.

the line of fifths. This entails also that the question of transpositional invariance between keys—whether the distribution of tonal pitch-classes in a piece is independent of the key—can neither be answered universally, since it is dependent on the historical context, nor can it be conclusively answered based on the assumption of enharmonic equivalence across the historical timeline (see also Quinn and White, 2017). The tonal pitch-class co-evolution in the early 20th century shows resemblance to both the block structure of the 18th century as well as to the broad band along the main diagonal in the 19th century and thus somewhat represents a consolidation of these two extremes.

11 Tonal profiles

We have used the tonal pitch-class distributions of musical pieces to draw inferences about the underlying tonal space. In fact, as a number of recent musical corpus studies demonstrate, many characteristics of tonality in musical pieces can be deduced from such distributions (Temperley, 2009; Albrecht and Shanahan, 2013; White, 2014; Albrecht and Shanahan, 2016; Quinn and White, 2017; Weiß et al., 2018). One of the central findings in computational musicology is that these distributions of notes in corpora strongly correlate with listener ratings of tonal stability in psychological experiments (Krumhansl, 1990; Aarden, 2003; Huron, 2006; Rohrmeier, 2007; Temperley and VanHandel, 2013). This finding is assumed to be caused by statistical learning mechanisms (Rohrmeier and Rebuschat, 2012; Koelsch et al., 2016). In most empirical studies on Western music, the distributions over notes or the stability ratings of tones are based on the assumption that there are only twelve distinct pitch-classes, the *neutral pitch-classes* in twelve-tone equal temperament (Temperley, 2000) that correspond to the twelve keys within one octave on the piano (see Section 9.2). Either the chromatic circle or the circle of fifths is consequently assumed to be the relevant underlying model of tonal space. This assumption is not always made explicit and is often a pragmatic choice due to the encoding of data (mostly the MIDI format) rather than an explicit modeling decision. These distributions of the twelve neutral pitch-classes are commonly called 'key profiles' or 'tone profiles'. A second common modeling assumption is that musical keys are transpositionally invariant, i.e. that musical pieces do not change their essential characteristics when transposed to a different key. Accordingly, the distributions of neutral pitch-classes do in fact not represent key profiles but rather 'mode profiles', since modes can be understood as equivalence classes of keys (Harasim et al., submitted). Although transpositional invariance is a widely accepted assumption, it is not undisputed (Rom, 2011; Quinn and White, 2017), in particular regarding its historical adequacy.

The tone profiles from four different sources are juxtaposed in Figure 11.1 for both the major mode (top) and the minor mode (bottom). Krumhansl and Kessler (1982, red bars) let listeners rate how well these pitch classes fit to previously sounding tonal contexts, Temperley (2001, blue bars) provides a music-theoretically motivated adjustment of these values, Albrecht and

Shanahan (2013) and Harasim et al. (submitted) provide statistics inferred from large datasets in MIDI format (violet and gray bars, respectively). While the tone profiles are largely similar between the four sources, it is noteworthy that the corpus-derived tone profiles (violet and gray bars) attribute much less weight to out-of-scale neutral pitch-classes (1, 3, 6, 8, and 10 in the major mode, top panel; 1, 4, 6, and 9 in the minor mode, bottom panel) than the studies based on listener ratings (red and blue bars). In the minor mode, both of the pitch classes 10 and 11 are relatively strong because the pitch class 10 is in-scale but the out-of-scale pitch-class 11 is the leading-tone to the tonic in both modes.

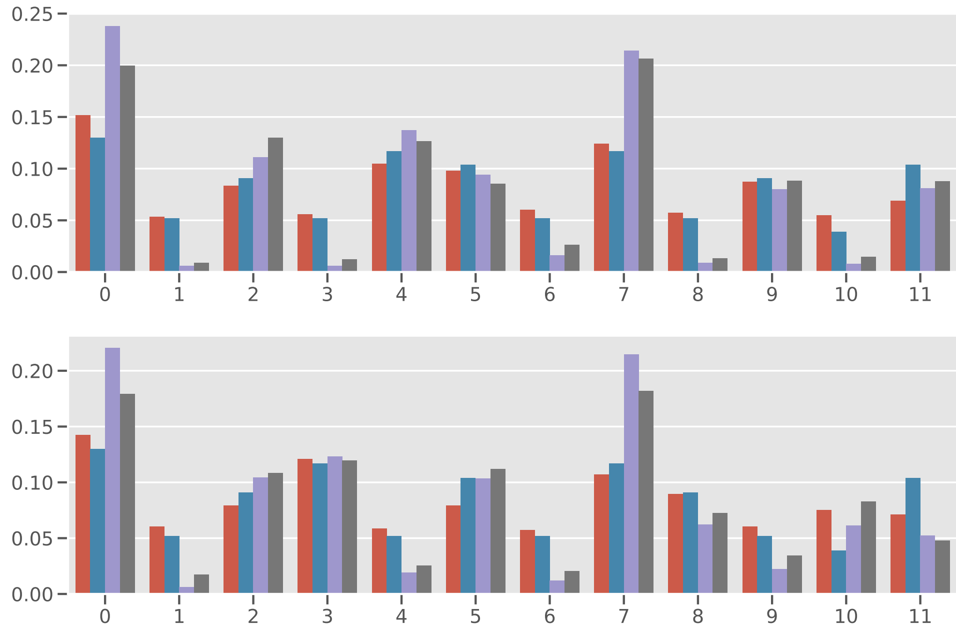


Figure 11.1 – Tone profiles for the major mode (top) and the minor mode (bottom) from Krumhansl and Kessler (1982, red bars), Temperley (2001, blue bars), Albrecht and Shanahan (2013, violet bars), and Harasim et al. (submitted, gray bars).

The corpus used in this part of the thesis contains musical pieces encoded in MusicXML, which allows to represent notes as tonal, as opposed to neutral, pitch-classes. Using the chromatic circle or the circle of fifths would thus be an inadequate representation because we would lose important information, for instance enharmonic differences between notes (e.g., between C and B♯). As the previous chapter has demonstrated, the line of fifth is much more suitable in this case. We consequently generalize the notion of tone profiles and define *tonal profiles* to be distributions of tonal pitch-classes on the line of fifths and thus on a potentially infinite number of distinct notes. For practical reasons, we restrict the vocabulary of tonal pitch-classes to those contained in the line-of-fifths segment from F♭♭ to B♯♯ since no piece in the corpus contains a note outside of this range. Recall the distribution of tonal pitch-classes from the first movement of Alkan's *Concerto for Solo Piano*, op. 39, no. 8 (Figure 3.1, bottom panel). Since there we did not assume any structure underlying the distribution of tonal pitch-classes, the frequencies were ordered by rank. Based on the results from the previous chapter, we can acknowlede the line of fifths as central structure for tonal space and represent the same distribution differently. The distribution of tonal pitch-classes in this piece on the line of fifths is shown in Figure 11.2.

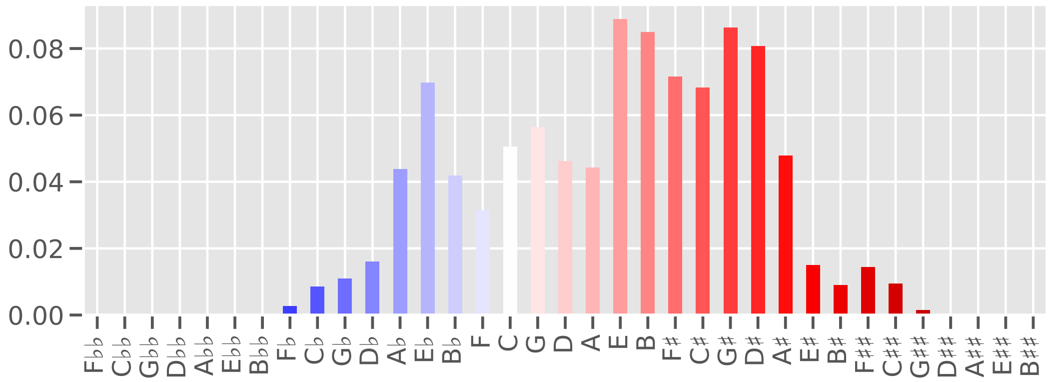


Figure 11.2 – Distribution of tonal pitch-classes of the first movement of Alkan's *Concerto for Solo Piano*, op. 39, no. 8.

The colors emphasize the linear structure of the line of fifth as well as the distance from the central C by the intensity of the colors and the direction towards more flat (blue) or sharp (red) tonal pitch-classes. Note that this piece contains more than twelve different tonal pitch-classes. Hence, representing it on the chromatic circle would unnecessarily discard valuable information and obliterate enharmonic differences.

11.1 Topic modeling with Latent Dirichlet Allocation

Where does this distribution of tonal pitch-classes come from? The multimodal shape of the distribution in Figure 11.2 suggests to model it as a weighted mixture of a small number of simpler distributions. This modeling assumption entails understanding the overall distribution of tonal pitch-classes in a piece as a mixture of different tonal profiles. These might, for example, correspond to the tonal pitch-class distributions of sections of a piece which are in different keys and which may contain chromaticism and enharmonicism. Instead of representing a piece as a single bag of notes as in the previous chapter, a piece is then represented as a collection of several of such bags of notes with potentially different sizes. This approach is commonly subsumed under the term *topic modeling* (Steyvers and Griffiths, 2007). One of its most prominent variant is Latent Dirichlet Allocation (LDA; Blei et al., 2003). Topic models define a composite model describing documents as mixtures of several topics which, in turn, are determined by distributions over items, e.g. words. Let us consider a textual example. One would expect that a document about 'politics' contains many names of politicians, institutions, states, political events like elections, wars, and so forth. It is rather unlikely that such a text would contain the names of composers, musical pieces, or music-theoretical terms like "augmented sixth chord", "symphony", "Passacaglia", and the like. Topic models turn this argument around and assume that the hypothetical text is about politics precisely *because* it contains many words from political topics. That is, topics are defined by the frequency of co-occurrence of certain words. This is sometimes called the *distributional hypothesis* (Harris, 1954).

This notion of a topic can directly be translated to the musical case if one considers pieces to be the documents and tonal pitch-classes to play the role of words in these documents. The vocabulary is then the set of all tonal pitch-classes that appear in any document in the corpus and topics correspond to profiles of tonal pitch-classes that statistically tend to co-occur. In this section, we treat these topics as an operationalization of tonality. Recall that tonality and the concept of a musical work are tightly intertwined (Dahlhaus, 1978; Polth, 2001; Schild, 2010). A piece containing only one or a few topics can thus be considered to be tonally more coherent than one with a multitude of different topics. Based on the historical accounts of the development of tonality (see Chapter 1), one would thus expect that pieces in the 19th century are more diverse with respect to the number of topics they contain. This conception of tonality as a topic[1] also resonates with one of the definitions of tonality given by Hyer (2001) in *The new Grove dictionary of music and musicians*, stating that

> [t]onal phenomena are musical phenomena (harmonies such as the tonic, dominant and subdominant, cadential formulae, harmonic progressions, melodic gestures, formal categories) arranged or understood in relation to a referential tonic, which imbues the music—in the case of C major—with 'C-ness'.
> (Hyer, 2001, p. 583)

or as Huron (2006) expresses it:

> One simple definition of tonality is a system for interpreting pitches or chords through their relationship to a reference pitch, dubbed the *tonic*.
> (Huron, 2006, p. 143)

The crucial point in these definitions is that "tonal" is synonymous with "relation to a referential tonic" that, in turn, determines the 'tonic-ness' of the music. Representing topics as bags of notes enables us to quantify the 't-ness' for this topic not only for a single distinguished note but for every tonal pitch-class t in the vocabulary.

Topic models, in particular LDA, have found wide applications in the Digital Humanities, mostly in the context of text-based studies (e.g., Blei, 2012b; Goldstone and Underwood, 2012; Rhody, 2012; Jockers and Mimno, 2013; Goldstone and Underwood, 2014). In contrast, there is only one application of LDA to music that is not based on textual data, such as metadata

[1] Note that this definition of 'tonality as a topic' is unrelated to other musicological definitions of a 'topic', for instance in Dickensheets (2012) or Johnson (2017) which are both situated in the context of Topic Theory (Ratner, 1980; Mirka, 2014b) that considers topics like 'Minuet' or 'March' (called 'types'), and topics such as 'military' or 'hunting' (called 'styles'; Mirka, 2014a).

or lyrics, but takes the tonal content of musical pieces into account. Hu and Saul (2009a,b) interpret musical keys as topics and use LDA to infer tone profiles (in the sense of distributions over the twelve neutral pitch-classes) from pieces. They evaluate the inferred topics against the tone profiles provided by Krumhansl and Kessler (1982, red bars in Figure 11.1) and moreover use their topics to trace key changes over the temporal course of a musical piece. Unfortunately, they do not report the numerical values of their profiles so that a direct comparison with the other profiles is impossible. The present approach differs from theirs in several important aspects. They use a relatively small sample of 235 manually selected pieces[2] by six composers, namely Bach (1685–1750), Vivaldi (1678–1741), Mozart (1756–1791), Beethoven (1770–1827), Chopin (1810–1849), and Rachmaninoff (1873–1943). Their dataset thus approximately spans a historical range of 260 years, whereas the number of pieces and composers as well as the extent of the historical range of the corpus supporting the present study is much larger (see Section 3.1). In Hu and Saul's approach, the basic units of the model are not individual notes but short segments of music. They thus incorporate a certain amount of temporal information, which is not the case here. The two topics that they find resemble the major and minor profiles in Krumhansl and Kessler (1982), but their major profile in particular seems rather to resemble a major triad than a major key profile—possibly an artifact of the segment level that they introduce—with the other in-scale pitch-classes having very low weights and the minor seventh (pitch class 10) being stronger than the major seventh (pitch class 11). Most importantly, their data is encoded in MIDI format, which, as we have seen above, implicitly models notes as neutral pitch-classes, whereas the representation of the data in our corpus models notes as tonal pitch-classes, and thus is based on a larger vocabulary and able to express enharmonic differences. Our musical interpretation of the LDA model is well-supported by empirical findings on hierarchical accounts of tonality (Krumhansl, 2004; Koelsch et al., 2013). The model architecture moreover dovetails nicely with music theoretical conceptions of tonality that emphasize its nested nature (e.g., Hauptmann, 1853; Schenker, 1935; Salzer, 1952; Lerdahl and Jackendoff, 1983; Lerdahl, 2001; Rohrmeier, 2011; Moss, 2014).

Generating pieces. The LDA model establishes probabilistic relations between topics and documents[3] in a corpus and specifies a generative process that is assumed to underlie the distribution of notes in pieces (Blei, 2012a). This generative model can thus also be used to create new documents, given a certain setting of the parameters of the model. It is important to note that 'generating' does not mean that LDA attempts to simulate the process of the composition of a piece. We will first illustrate the model by artificially generating an artificial corpus with $D = 20$ documents and $K = 3$ topics. Subsequently, we will use the LDA model to infer $K = 7$ topics from the data in the XML corpus and also compare the results for different values of K. In general, the more topics one assumes to exist in a corpus, the more specific they will be, whereas a small value of K leads to broader, more general topics. The vocabulary size is $V = 35$ since we consider tonal pitch-classes from F♭♭ to B♯♯ on the line of fifths. We then

[2] The data was gathered from http://www.classicalmusicmidipage.com.

[3] Since most applications of LDA work with textual data, we will use the terms 'document', 'piece', and 'composition' interchangeably.

create a distribution of topic weights for each document in the corpus. These distributions $\theta_{1:D}$ determine how prominent the K topics are in each of the the D documents.[4] The probability of topic $k \in \{1, \dots, K\}$ in document d in the artificial corpus is given by $\theta_{d,k}$. The distribution θ_d over all topics for document d is modeled as a sample from a Dirichlet distribution with a K-dimensional hyperparameter α,

$$\theta_d \sim \mathrm{Dir}(\alpha). \tag{11.1}$$

We set $\alpha = \mathbf{1}_K$ for all D documents in the artificial corpus, where $\mathbf{1}_K$ is the K-dimensional vector containing only 1's. The sampled probabilities of the three topics in all 20 documents are shown in Figure 11.3. Topic 1 is shown in red, topic 2 in blue, and topic 3 in violet.

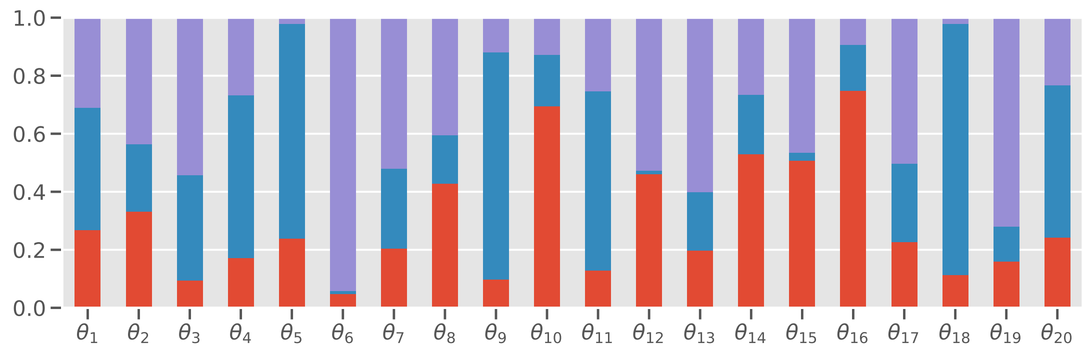


Figure 11.3 – Topic weights for $K = 3$ topics and a corpus of $D = 20$ artificial documents sampled from a Dirichlet distribution with hyperparameter $\alpha = (1, 1, 1)$.

For example, the topic distribution in document 1 is relatively balanced with $\theta_1 = (.27, .42, .31)$, and the one for document 15 is $\theta_{15} = (.51, .03, .46)$, consisting almost exclusively of topics 1 and 3. The topic distributions vary considerably due to the setting of the α that specifies a uniform distribution over the K-simplex to the effect that all configurations for θ_d, $d \in \{1, \dots, D\}$, are equally probable. Note that, at this stage in the generative process, we do not know yet how the K topics are composed, i.e. what the tonal pitch-class distributions look like that they represent.

In the next step, each of the K topics is associated with a distribution over the V tonal pitch classes that expresses how important each tonal pitch-class is for this topic. Accordingly, the topics $\phi_{1:K}$ are sampled from a Dirichlet distribution with V-dimensional hyperparameter β,

$$\phi_k \sim \mathrm{Dir}(\beta). \tag{11.2}$$

For all K topics, we set $\beta = \mathbf{1}_V$, where $\mathbf{1}_V$ is the V-dimensional vector containing only 1's. Both α and β are fixed hyperparameters for the whole corpus. The distribution of tonal pitch-classes for the $K = 3$ topics is shown in Figure 11.4.

[4] The notation follows Blei (2012a) where $x_{1:N}$ is a shorthand notation for $\{x_1, \dots, x_N\}$.

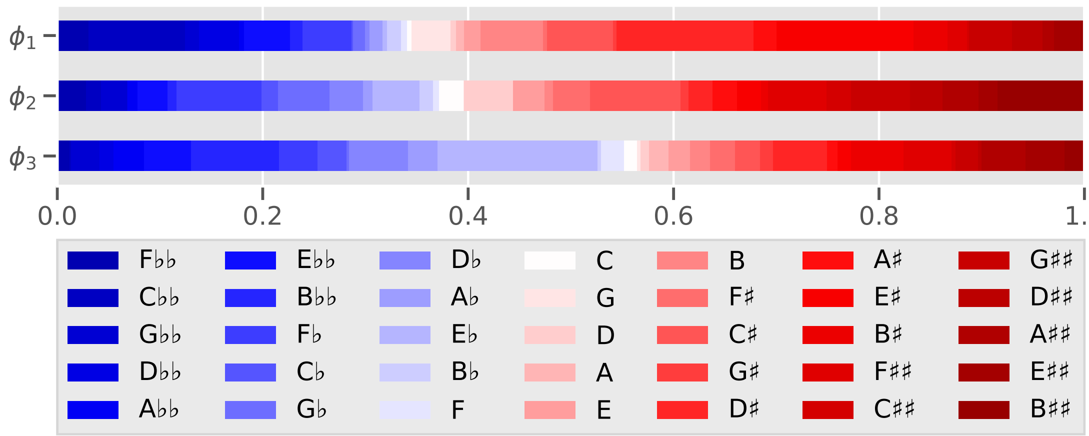


Figure 11.4 – Tonal pitch-class distributions for $K = 3$ artificial topics.

As mentioned before, a small number of topics K leads to broader topics that are not very specific. It can be seen that all three artificial topics cover the whole range of tonal pitch-classes and thus each of them contains the entire vocabulary. Yet comparing the three topics also reveals differences. For example, the probability for tonal pitch-class C to occur is .0004 for topic 1, .023 for topic 2, and .012 for topic 3. These are represented by the white parts of the bars in Figure 11.4. The probabilities of some of the other tonal pitch-classes vary even more. So far, we have sampled distributions over the K topics $\theta_{1:D}$ for each of the D documents (Figure 11.3) and distributions over the V tonal pitch-classes $\phi_{1:K}$ for each of the K topics (Figure 11.4).

The final step in the generative process of the LDA model consists in sampling the actual tonal pitch-classes for the D documents, given the distributions over topics $\theta_{1:D}$ and the distributions of tonal pitch-classes for all topics $\phi_{1:K}$. For each of the D documents in the corpus we sample its length, the total number of notes N_d in the dth document, from a Poisson distribution with hyperparameter $\lambda = 100$. This step is only necessary for the generation of new documents since the number of notes is deterministically given in actual documents in a corpus. Next, we use the topic weights θ_d to sample a topic assignment $z_{d,n}$ for the nth note in the dth document from a Categorical distribution using the corresponding parameter θ_d,

$$z_{d,n} \sim \mathrm{Cat}(\theta_d). \tag{11.3}$$

Given this topic assignment $z_{d,n}$, the actual nth tonal pitch-class $x_{d,n}$ in document d is sampled from a Categorical distribution with parameter $\phi_{z_{d,n}}$,

$$x_{d,n} \sim \mathrm{Cat}(\phi_{z_{d,n}}). \tag{11.4}$$

The tonal pitch-class distribution for the dth document is then given by x_d, the collection of all notes sampled by the generative LDA process. The model is summarized by the joint probability[5]

$$p(\phi_{1:K}, \theta_{1:D}, z_{1:D}, x_{1:D} \mid \alpha, \beta) = \prod_{d=1}^{D} \left(p(\theta_d \mid \alpha) \prod_{n=1}^{N} p(x_{d,n} \mid z_{d,n}, \phi_{z_{d,n}}) p(z_{d,n} \mid \theta_d) \right) \cdot \prod_{k=1}^{K} p(\phi_k \mid \beta) \tag{11.5}$$

over all the random variables specified in Equations 11.1–11.4.

A graphical representation of the LDA model in so-called *plate notation* (Bishop, 2006; Koller and Friedman, 2009) is shown in Figure 11.5. Circular nodes represent the observed (shaded) and latent (white) random variables of the model. The two hyperparameters α and β are shown without circles. Arrows between the variables represent the probabilistic dependencies between them. The boxes surrounding sets of variables are called plates. They stand for repeated sampling procedures of the random variables they contain.

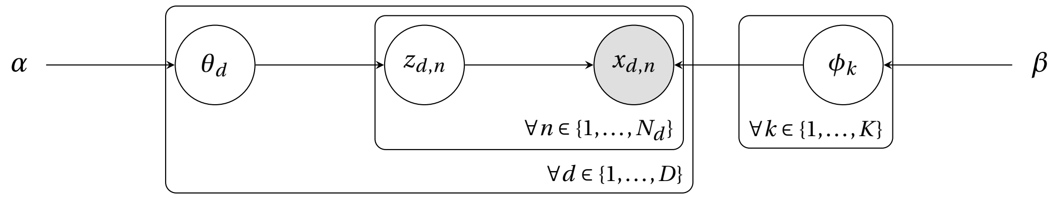


Figure 11.5 – Graphical model for Latent Dirichlet Allocation (LDA) describing the relations between random variables $x_{d,n}$ (nth note in document d; observed), $z_{d,n}$ (topic assignment of nth note in document d; latent), θ_d (topic distribution in document d; latent), ϕ_k (tonal pitch-class distribution of topic k; latent), and Dirichlet hyperparameters α and β, where N_d is the number of notes in document d, for $d \in \{1,\ldots,D\}$ and $k \in \{1,\ldots,K\}$.

The artificial corpus of all $D = 20$ documents generated with the hyperparameters α and β as above is shown in Figure 11.6. Because we set the parameter for the Poisson distribution to sample the lengths of the pieces in the artificial corpus to $\lambda = 100$, the average piece length is 100 notes. Since all $K = 3$ topics have non-zero probabilities for each of the $V = 35$ tonal pitch-classes in the vocabulary (the line-of-fifths segment from F♭♭ to B♯♯; see Figure 11.4), it is not surprising that each piece in the articifial corpus contains every tonal pitch-class at least once. In order to obtain a distribution of tonal pitch-classes in document d, we can normalize x_d, the count of tonal pitch-classes in the dth document, by dividing it by N_d, the total number of notes in this document.

Compare the distribution of tonal pitch-classes in the artificial corpus to the tonal pitch-class distribution in the entire actual corpus that is studied in this part (Figure 11.7). The distribution of tonal pitch-classes in the artificial corpus (Figure 11.7a) is spread out across the whole line of fifths, reflecting the arbitrary parameter settings used in the illustration of

[5]See Equation 1 in Blei (2012a).

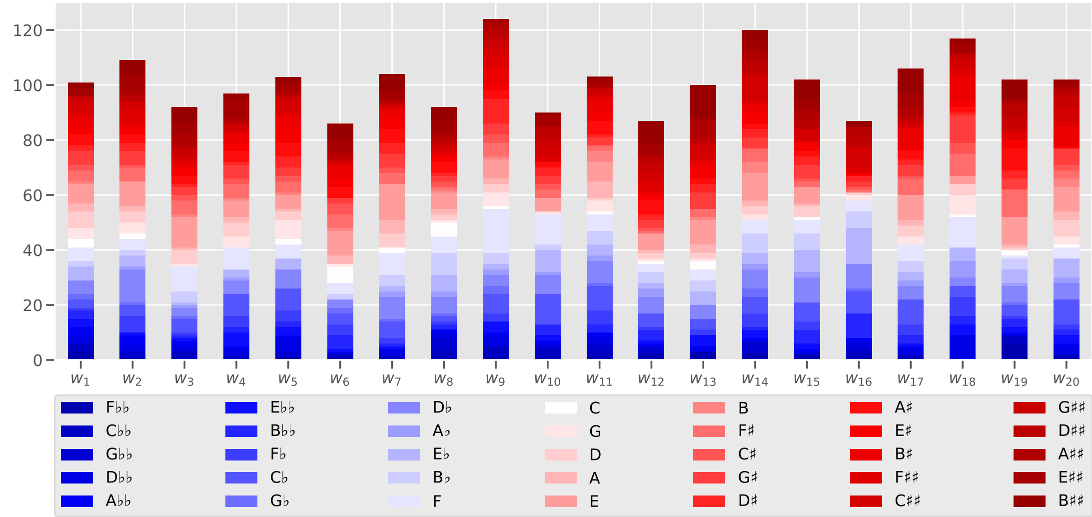
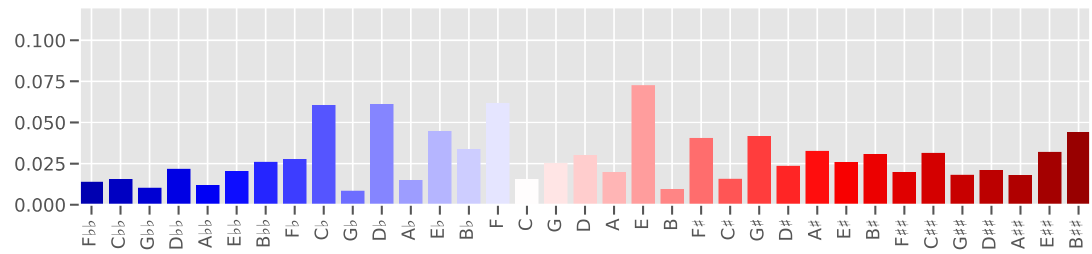
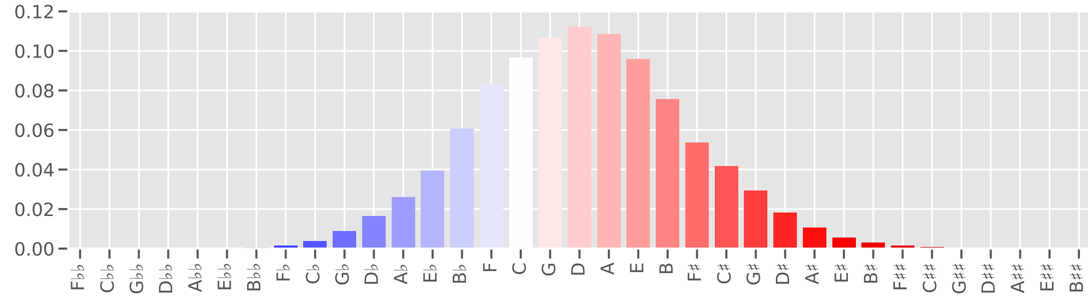


Figure 11.6 – Distribution of tonal pitch-classes in $D = 20$ documents in a randomly generated corpus with $K = 3$ artificial topics.



(a) Overall tonal pitch-class distribution in artificially generated corpus with $K = 3$ topics and $D = 20$ documents.



(b) Overall tonal pitch-class distribution in the XML corpus with $D = 2012$ documents; reproduced from Figure 9.5.

Figure 11.7 – Tonal pitch-class distribution in a corpus of artificially sampled pieces (top) and the XML corpus (bottom).

the generative process. The two most common tonal pitch-classes in the artificial corpus are E and D♭. It is evident that the overal tonal pitch-class distribution as well as the tonal pitch-class distributions of individual documents (Figure 11.6) have little to do with the actual tonal pitch-class distributions in musical pieces and the XML corpus (Figure 11.7b). In contrast to this almost normal distribution centered on the mid-range of the line of fifths, the tonal pitch-classes in the artificial corpus follow no recognizable pattern. This is not surprising since the weights of the topics in the corpus, the distribution of tonal pitch-classes per topic, and, consequently, the distribution of tonal pitch-classes in the $D = 20$ pieces of the corpus are all based on arbitrary choices for the parameters.

Inferring topics. While generating pieces corresponds to sampling from the joint distribution given in Equation 11.5, finding topics in the pieces corresponds to computing the conditional distribution of all variables given a corpus wich is given by

$$p(\phi_{1:K}, \theta_{1:D}, z_{1:D} \mid x_{1:D}, \alpha, \beta) = \frac{p(\phi_{1:K}, \theta_{1:D}, z_{1:D}, x_{1:D} \mid \alpha, \beta)}{p(x_{1:D} \mid \alpha, \beta)}, \tag{11.6}$$

according to the product rule of probability. Unfortunately, this is often difficult or impossible to calculate. For this reason, the topics have to be estimated using approximative methods. *Gibbs sampling* (Griffiths, 2002; Steyvers and Griffiths, 2007) is such a method and a popular algorithm from a larger class of so-called Markov Chain Monte Carlo (MCMC) sampling algorithms (MacKay, 2003; Bishop, 2006). A Gibbs sampler first initializes the variables randomly and then iteratively updates each of them in turn, conditioned on the values of the other variables from the previous iteration step. The rationale behind this procedure is that if, after many iterations, a tonal pitch-class $x_{d,n}$ has been repeatedly assigned to a certain topic $z_{d,n}$, the probability of assigning another instance of that tonal pitch-class in the same or in other pieces to that topic increases. Analoguously, if a certain topic is repeatedly used in a certain piece, it increases the probability of other tonal pitch-classes in this piece to be assigned to said topic. The topic assignment of tonal pitch-classes thus depends both on how likely this note is for a given topic and on how prevalent a topic is in a given document (Steyvers and Griffiths, 2007, p. 8).

Consider the tonal pitch-class counts in 20 randomly selected pieces from the XML corpus shown in Figure 11.8. Contrary to the 20 pieces in the artificial corpus (Figure 11.6), this set of pieces is much more diverse. None of the pieces spreads across the whole range of the line of fifths. The pitch-class distributions are also more structured. While some pieces contain only ‘sharp’ (red) tonal pitch-classes, e.g. x_{11} and x_{13}, others contain almost only ‘flat’ (blue) tonal pitch-classes, e.g. x_{14}. Interestingly, each of the 20 pieces in this sample spans a contiguous segment on the line of fifths, i.e. they do not contain any gaps. Another difference between the two samples is that their lengths, the absolute number of notes they contain, varies more in the actual corpus than in the artificial corpus.

The task of the Gibbs sampling procedure is now to find the most likely topics, i.e. tonal

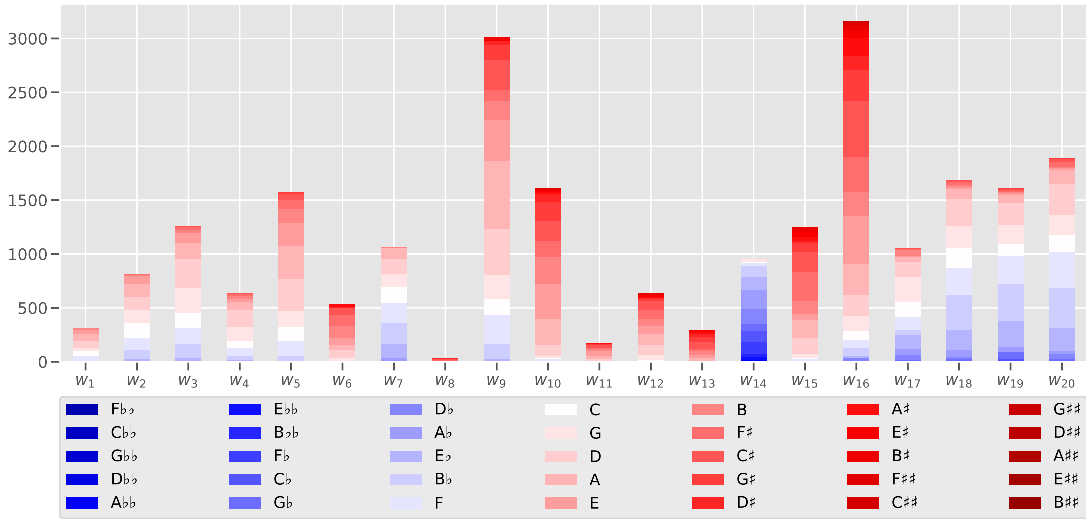


Figure 11.8 – Tonal pitch-class counts in a sample of 20 pieces from the XML corpus.

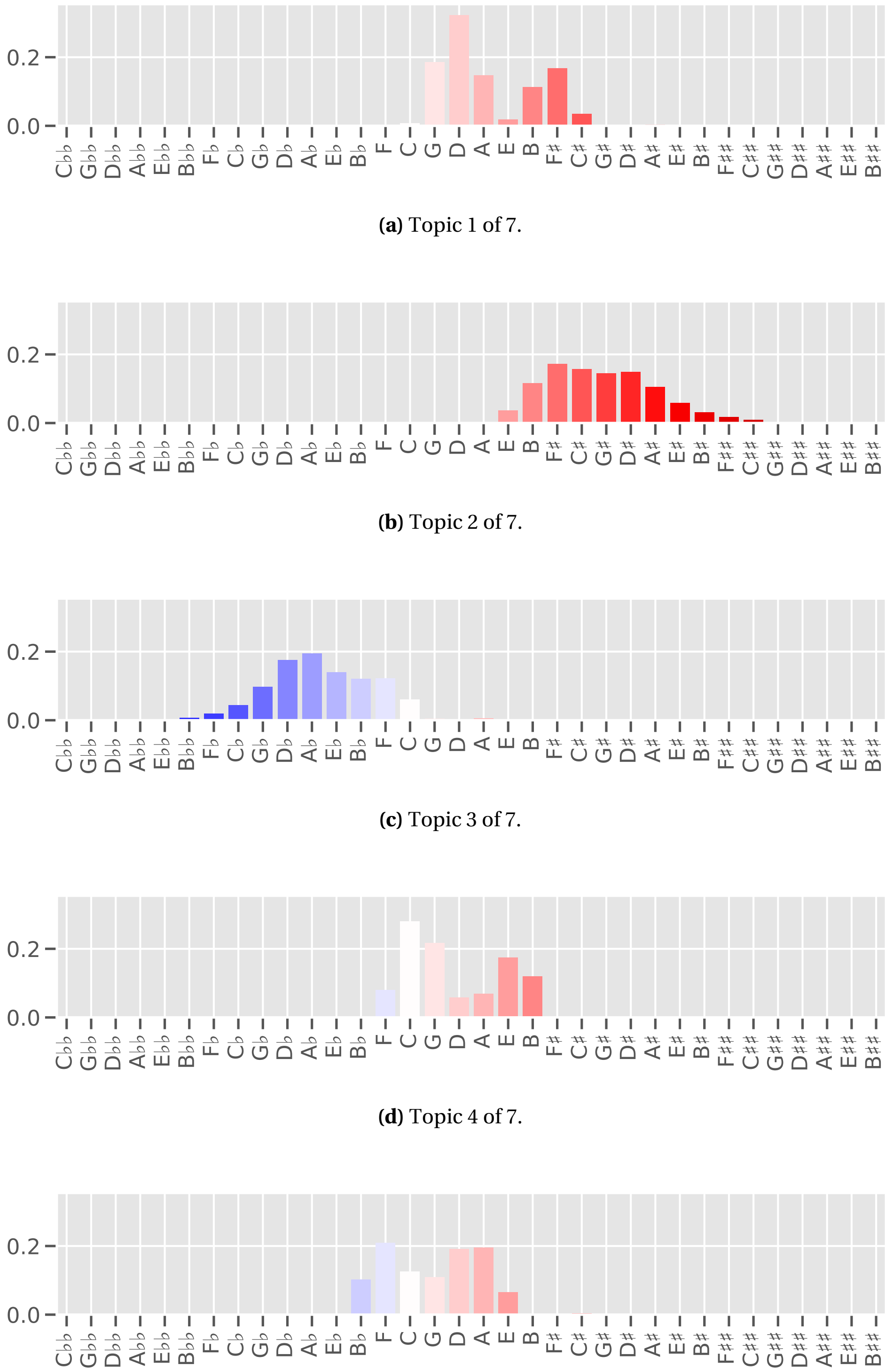
pitch-class distributions ϕ_k as well as their proportions for all documents θ_d, that gave rise to the tonal pitch-class distributions not only in this sample of 20 compositions but in the entire corpus, given a predefined number of topics K.

11.2 Finding topics in the corpus

Which topics are latent in the pieces of the XML corpus? As mentioned before, the LDA model takes a fixed parameter K, the number of topics assumed to exist in the corpus. Although there are various approaches to also infer the optimal number of topics from the data in order to make LDA fully unsupervised (e.g. Teh et al., 2006; Chang et al., 2009; Wallach et al., 2009), the interpretation of the found topics is highly domain-dependent and it is a matter of discussion whether purely data-driven methods should determine what is optimal or to what degree domain knowledge is to be taken into account (e.g. Schmidt, 2012; Mohr, 2013; Tang et al., 2014; Binder, 2016). Whereas the interpretation of textual topics is relatively straight-forward for humans, this is not necessarily so in the musical case because tonal pitch-classes *per se* bear no semantic information. For this reason, it was opted to run the model multiple times for different settings of the parameter $K \in \{2,3,5,7,10,12,24\}$. Since this study is the first application of topic models to tonal pitch-class distributions, we have no means to compare the results to established findings using the same method in the literature. Rather, we will largely draw on music theory and use the shape of the distributions on the line-of-fifths in order to interpret the inferred topics.

The inferred tonal pitch-class distributions for the seven topics are shown in Figure 11.9. Note that the order of the topics has no particular meaning and that some weights for tonal pitch-classes in the topics are too small to be displayed. The numerical values for the probabilities of the tonal pitch-classes in the respective topics $\phi_{1:K}$ are given in Tables B.4–B.10 in the appendix. It turns out that the tonal pitch-class distributions for $K = 7$ topics largely correspond to distributions over diatonic sets. Topic 1 ranges from G to C♯ (six fifths) and thus covers all tonal pitch-classes in the keys of D major and B minor. The two most prominent pitch-classes are D and F♯, the tonic and the major third in D major and the third and the fifth in B minor. Similarly, topic 7 covers all tonal pitch-classes from D to D♯. It spans seven fifths on the line of fifths and thus exceeds the diatonic range of six fifths, comprising all tonal pitch-classes of either A major and F♯ minor, or of E major and C♯ minor. Topics 2 and 3 somewhat stand out in comparison to the other topics but are to a certain degree similar to each other. Both of them cover a larger range of tonal pitch-classes on the line of fifths. Topic 2 covers the range from E to C♯♯ (ten fifths) and topic 3 covers the range from B♭♭ to C (nine fifths). Moreover, both of them do not contain tonal pitch-classes that particularly stand out as in the topics considered so far. While topic 2 covers only 'sharp' (red) tonal pitch-classes, topic 3 contains only 'flat' (blue) tonal pitch-classes and C with substantial probability mass (it can be seen that topic 3 also assigns non-zero probability mass to A). The three remaining topics 4, 5, and 6 lie more or less on the center of the line-of-fifths segment and all cover diatonic sets (six fifths). Topic 4 ranges from F to B and thus contains all tonal pitch-classes of C major and A minor. Its three most prominent tonal pitch-classes are C, G, and E, the constituents of the C-major triad. Topic 5 ranges from B♭ to E and hence consists of all notes in the keys of F major and D minor. Its most prominent notes are F, A, and D, the components of a D-minor triad. Finally, topic 6 ranges from A♭ to D and thus covers all tonal pitch-classes of E♭ major and C minor with the most prominent notes being B♭ and E♭. With the exception of topics 2 and 3 the most likely topics are given by diatonic sets. Recall that topics are operationalizations of tonality. This means that the best explanation for the tonal pitch-class distributions in the XML corpus consists of several diatonic sets in the middle range of the line of fifths plus two topics representing the two extremes, flats (topic 3) and sharps (topic 2).

In Chapter 10 we found that the line of fifths emerges both from the co-occurrence of tonal pitch-classes under the bag-of-notes model (Section 10.1) and from the (co-)evolution of tonal pitch-classes (Section 10.3). The dimensionality reduction via PCA revealed differences between the natural tonal pitch-classes on one hand, and the flat and sharp tonal pitch-classes on the other hand (see Figures 10.2 and 10.7). The composite architecture of the LDA model has provided a means to further sub-divide the distributions over the natural tonal pitch-classes into the distributions of topics 1, 4, 5, 6, and 7. The topics inferred by the LDA model can thus be interpreted as corroboration and refinement of the earlier findings. The topic distributions for other values of $K \in \{2, 3, 5, 10, 12, 24\}$ are shown in Figures A.3–A.8 in the appendix. All topics consist of gapless segments on the line of fifths. This is not trivial since, as before, the model is agnostic to the interval relations between the tonal pitch-classes. As mentioned before, increasing the number of topics also increases their specificity. Comparing how the topics change with larger values of K shows that the growing specificity is largely reflected in emphasizing fifth and third relations between the most prominent notes. A fact that is predicted by several hierarchical models of tonal space (Krumhansl, 1990; Lerdahl,



(a) Topic 1 of 7.

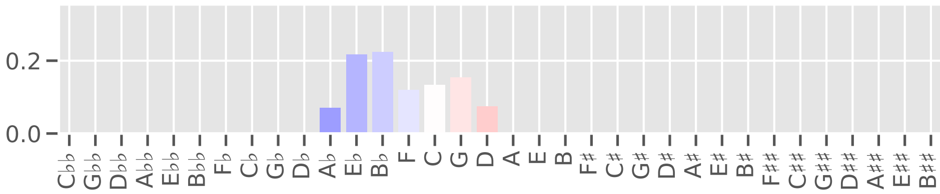
(b) Topic 2 of 7.

(c) Topic 3 of 7.

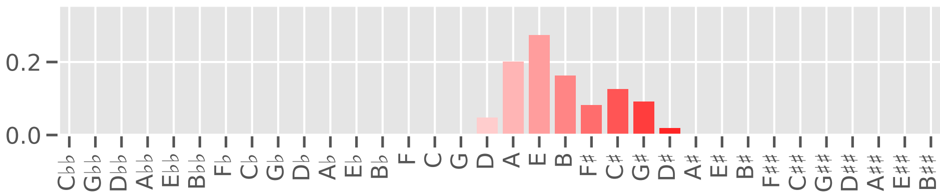
(d) Topic 4 of 7.

(e) Topic 5 of 7.

Figure 11.9 – The note distributions for the $K = 7$ topics.



(f) Topic 6 of 7.



(g) Topic 7 of 7.

Figure 11.9 – The note distributions for the $K = 7$ topics (cont.).

2001; Rohrmeier, 2011, see Figure 5.5).

Since topics are defined as distributions over tonal pitch-classes, one can define appropriate measures to assess the similarity between them. For two discrete probability distributions P and Q over a discrete support $\mathcal{X}$ the *Kullback-Leibler divergence* (Cover and Thomas, 2006) is defined as

$$\mathrm{KL}(P \parallel Q) = -\sum_{x \in \mathcal{X}} P(x) \log_2 \left(\frac{Q(x)}{P(x)} \right). \tag{11.7}$$

Here, $\mathcal{X}$ is the line-of-fifths segment from F♭♭ to B♯♯. The *Jensen-Shannon divergence* is a symmetricized version of the Kullback-Leibler divergence and is given by

$$\mathrm{JS}(P \parallel Q) = \frac{\mathrm{KL}(P \parallel M) + \mathrm{KL}(Q \parallel M)}{2}, \tag{11.8}$$

where $M = (P + Q)/2$ is the mean distribution of P and Q. Since $\mathrm{JS}(P \parallel Q)$ measures the average divergence from the mean of the two distributions, having zero values in the distributions is unproblematic because M is zero if and only if both $P(x)$ and $Q(x)$ are zero for $x \in \mathcal{X}$. The square root of the Jensen-Shannon divergence defines a distance metric

$$d_{\mathrm{JS}}(P, Q) = \mathrm{JS}(P \parallel Q)^{\frac{1}{2}}, \tag{11.9}$$

the *Jensen-Shannon distance* (Endres and Schindelin, 2003) that we can use to compare the topics with each other. Finally, because the distance d_{JS} is bounded between 0 and 1, we can express the *Jensen-Shannon similarity* between two topics ϕ_k and $\phi_{k'}$ as

$$s_{\mathrm{JS}}(\phi_k, \phi_{k'}) = 1 - d_{\mathrm{JS}}(\phi_k, \phi_{k'}), \tag{11.10}$$

where a similarity value of 1 implies that the two topics are identical.

The similarities between all pairs of the $K = 7$ topics are shown in Figure 11.10. Confirming our initial observations, the tonal pitch-class distributions that are most similar to each other are topics 4 and 5 with $s_{JS}(\phi_4, \phi_5) = .511$. These topics both lie in the center of the line of fifth and contain mostly the same notes. The main difference with respect to their support is that topic 4 contains B but not B♭ and *vice versa* for topic 5.

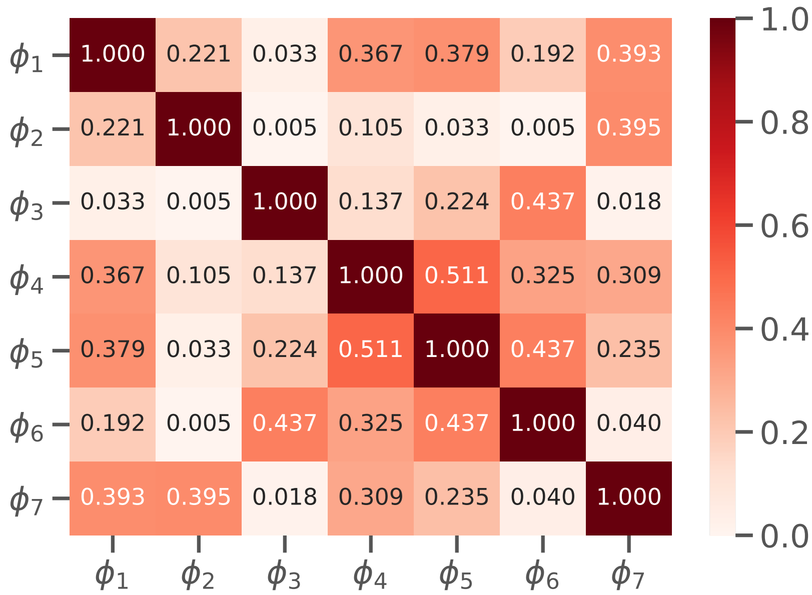


Figure 11.10 – Jensen-Shannon similarities for $K = 7$ topics.

Because the probabilities of tonal pitch-classes in these two topics are different (see, for instance the weights of tonal pitch-classes F, D, and A), their similarity is relatively low in absolute terms while still being the largest among all pairs of topics. The two most distinct topics are topic 2 and topic 3 with $s_{JS}(\phi_2, \phi_3) = .005$, a consequence of the fact that they have almost no tonal pitch-classes in common. Recall that some of the non-zero probabilities are too small to be displayed in Figures 11.9b and 11.9c. Table B.7 in the appendix shows that topics 2 and 3 both have non-zero probabilities for tonal pitch-class A, leading to a minimal but non-zero similarity between the two topics. The topic similarities for other values of $K \in \{2, 3, 5, 10, 12, 24\}$ are shown in Figures A.9–A.14 in the appendix.

11.3 The evolution of topics in the corpus

In the LDA model, each note $x_{d,n}$ in each of the d pieces is associated with a topic assignment $z_{d,n}$ that represents the most likely topic this note was generated from. Consequently, we can count these topic assignments for each document to obtain θ_d, a distribution over topics for each document. The average distribution over the $K = 7$ topics for all pieces in the corpus is shown in Figure 11.11. The black error bars represent the standard error of the mean. The most prominent topics in the corpus are topic 5 (.332) and topic 4 (.227) which are also the most similar topics to each other ($s_{JS}(\phi_4, \phi_5) = .511$) and which both are located at the very center of the line of fifths. The least frequent topics are topic 2 (.053) and topic 3 (.048) which are also the most dissimilar ones ($s_{JS}(\phi_2, \phi_3) = .005$).

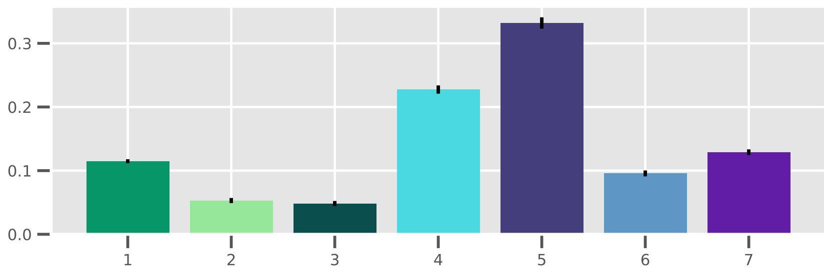


Figure 11.11 – Average distribution of topics for all documents in the XML corpus for $K = 7$ topics. Error bars show the standard error of the mean.

The average topic distributions for other values of $K \in \{2, 3, 5, 10, 12, 24\}$ are shown in Figures A.21–A.26 in the appendix. Recall that the distribution of tonal pitch-classes in the corpus was approximately normally distributed on the line of fifths and centered on pitch-class D (Figure 9.5). This distribution shows the relative frequencies of the observed notes in the pieces. The average topic distribution represents the overall distributions of the topics, the 'tonalities', that are latent in the corpus.

Using topic modeling in the context of historical studies entails certain assumptions. As mentioned before, LDA is based on the bag-of-notes model and does thus not know the order of notes within a piece. Beyond that, it also does not have a concept for the order of pieces in the corpus, although some recent variants of the model attempt to incorporate chronological information (Blei and Lafferty, 2006; Zhu et al., 2016; Beykikhoshk et al., 2018, e.g.). Under the basic LDA model, all pieces in the corpus are treated equally in order to infer the overall topics, regardless of the time of their composition. Since we do know the dates of composition or publication for the pieces in the corpus (see Table B.3 in the appendix), we are in a position to compare the topic distributions in the pieces diachronically in order to consider historical changes in these distributions. In order to trace the *topic evolution* in the corpus, we use the same methodology that we used in Section 10.3 to study tonal pitch-class evolution and calculate first the average topic distribution for each year for which we have pieces in the corpus. We moreover assume that the average topic distribution does not change if we do not have data for a year within the time range. Subsequently, we calculate a moving average over the distributions that returns smoothed values for each topic while at the same time ensuring that the distributions per year always sum to 1. As before, we use a windows size of 50 years as window size. Based on the previous chapter that has shown the historical changes in the usage of tonal pitch-classes, we can expect to see these changes also reflected in the historical development of the latent topics. In particular, one would expect that the topic evolution reflects historical changes in tonality, namely the increase in the usage of alterated notes as well as chromaticism and enharmonicism (see Chapter 10). The topic evolution over time for $K = 7$ topics is shown in Figure 11.12. The topic evolution plots for other values of $K \in \{2, 3, 5, 10, 12, 24\}$ are displayed in Figures A.15–A.20 in the appendix.

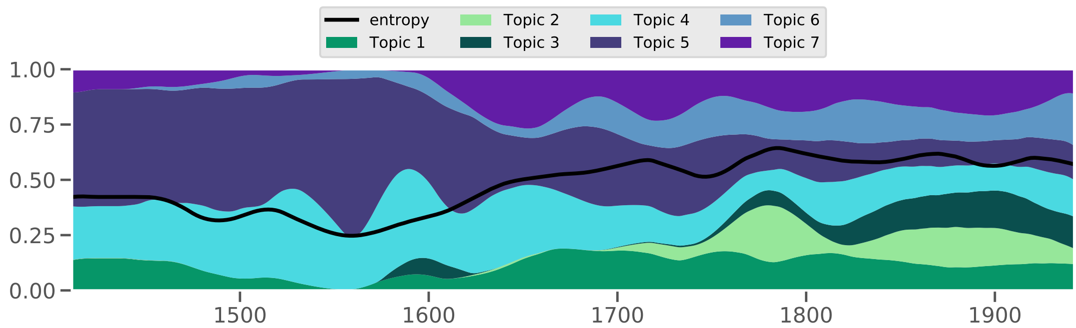


Figure 11.12 – Topic evolution for $K = 7$ topics.

It can be seen that, while the overall most common topics 4 and 5 are prevalent in the earlier centuries, the other topics—in particular topics 2 and 3 — gain traction over the historical timeline and seem to stabilize in the 19th century, bearing witness to the fact that the tonal pitch-class vocabulary spreads out on the line of fifths (see Chapter 10). The black line shows the smoothed normalized entropy of the topic distributions in each year. If in any year the topic distribution were uniform, the normalized entropy in this year would be equal to 1 before smoothing. It shows that the distributions over time tend to approach uniformity, i.e. that the diverse topics become, on average, more and more similar with respect to their frequencies of occurrence. Note that this is not a corrolary of the simple fact that more topics are used after ca. 1700 since the entropy is normalized for the number of topics in each year. The evolution of the latent topics does indeed corroborate the expectations for the historical development of tonality by showing that, over time, the number of used topics increases and is particularly strong in the 19th century.

As mentioned before, a larger number of topics increases their specificity. Indeed, a closer look at the tonal pitch-class distributions of the topics in Figures A.3–A.8 shows that the unimodal distributions for 2 and 3 topics become multimodal. This means that the probability of tonal pitch-classes in these topics is not anymore directly related to the line of fifths and that other intervals, in particular major and minor thirds, become more and more prevalent the larger the number of topics gets. For example, for topics 2–4 of 12 (Figure A.7) distribute most probability mass on tonal pitch-classes that are related by triadic intervals: perfect fifths as well as major and minor thirds (F♯, A, and D in topic 2, G, B, and D in topic 3, and C, E, and A in topic 4). The contributions of thirds in the constitution of tonality will be addressed in more detail in Chapter 12.

12 The Tonnetz

> Aussi dois-je avoüer que, nonobstant toute l'expérience que je pouvois m'être acquise dans la Musique, pour l'avoir pratiquée pendant une asseéz longue suite de temps, ce n'est cependant que par le secours des Mathématiques que mes idées se sont débroüillées, & que la lumière y a succedé à une certaine obscurité, dont je ne m'appercevois pas auparavant.[1]
>
> Jean-Philippe Rameau, *Traité de l'harmonie réduite à ses principes naturels*

The previous chapters have shown that distributions of tonal pitch-classes are largely organized around segments on the line of fifths. In particular, the chapter on the LDA model has shown that major and minor thirds do also play prominent roles for these distributions, a finding predicted both by historical as well as contemporary music theoretical models. The role of these intervals for the construction of tonal spaces was the topic of an active discussion in 19th-century music theory. This chapter reviews some approaches and shows how intervals between tonal pitch-classes can be conceived as concatenations of a few basic intervals. A new model, the Tonal Diffusion Model (TDM), is then introduced in order to derive the most likely interval weights for every piece in the corpus. The distributions of the model parameters are subsequently analyzed to reveal larger trends, and conclusions are drawn for the historical development of tonality.

[1] "Also I have to confess that—notwithstanding all the experience that I have been able to acquire in Music, having practiced it for quite a long time—it is only through the aid of Mathematics that my ideas got untangled & that the light succeeded in a certain darkness that I was not aware of before." (Rameau, 1722, préface); translation by the author.

12.1 Mapping tonal space

One of the earliest known graphical depictions of relations between tones dates back to Euler (1739) who conceived of a spatial diagram of musical intervals. It is shown in Figure 12.1; a later version is published in Euler (1774). Ascending perfect fifths are marked by 'V' and ascending major thirds are marked by 'III'. Euler clearly assumes enharmonic equivalence which is attested by the fact that only twelve tones are shown in the diagram and that the tone at the bottom is named B instead of A♯ which would be the tone a perfect fifth above D♯.[2] Note that this space topologically corresponds to a torus (Purwins et al., 2007) that is, going down a fifth from the top tone F leads one back down to the bottom tone B♭ ('B'), and moving left a major third from C leads to the right side of the diagram to G♯ ('Gs') which is enharmonically equivalent to A♭, the major third below C.

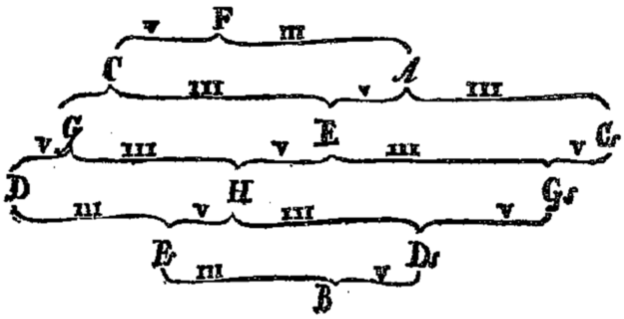


Figure 12.1 – The Tonnetz in *Tentamen novæ theoriæ musicæ ex certissimis harmoniæ principiis dilucide expositæ* (Euler, 1739, p. 147).

A number of 19th-century music theorists have proposed spatial representations for the relations between tones that do not assume enharmonic equivalence (e.g. Hauptmann, 1853; Weitzmann, 1860; von Oettingen, 1866; Hostinský, 1879; Riemann, 1896). These have become known as the *Tonnetz* (Cohn, 1997; Gollin, 2006). The Tonnetz is a graph where the nodes represent musical tones and the edges represent intervals between them. Most commonly, the edges describe the intervals of the perfect fifth, the major third, and, sometimes, also the minor third. This goes back to Hauptmann (1853) who conceives of the intervals of the octave, the perfect fifth, and the major third as "directly intellegible" and "unchangeable" (Hauptmann, 1888, p. 5)[3] and axiomatically defines them as the basic intervals underlying tonal space. The Czech music theorist Hostinský disregards the octave as merely an "Alterego" (Hostinský, 1879, p. 67), thus prefiguring the concept of tonal pitch-classes that we also adopt in this study. Figure 12.2 shows three examples of such graphical depictions of tonal space, namely the ones by von Oettingen (1866), Hostinský (1879), and Riemann (1896). The two representations by von Oettingen (1866) and Riemann (1896) (Figures 12.2a and 12.2c) are based on perfect fifths and major thirds. Note that these tables—in contrast to Euler's schema—theoretically continue infinitely in all directions and that the interval of the minor third appears here as

[2] Euler uses the German convention of spelling notes where H corresponds to B and B corresponds to B♭. Sharps are abbreviated with an 's' instead of '♯'.

[3] The German original is "direkt verständlich" and "unveränderlich" (Hauptmann, 1853, p. 21)

the difference of the fifth and the major third, visually corresponding to diagonal relations between tones. The map that Hostinský presents acknowledges the minor third as an interval on its on right, leading to the hexagonal representation shown in Figure 12.2b. Note that ascending major thirds are depicted downwards on Euler's and Hostinský's maps in contrast to the other two authors who depict them upwards. The latter way has come to be the standard of representing the Tonnetz in music theory.

12.2 Modeling tonal relations

Hostinský's conception of tonal space entails an interesting aspect regarding the combinatorial nature of intervals:

> We call two tones that are essential components of the same musical sonority *directly related*. The interval of two directly related tones is a *consonance*. [...] Two tones are indirectly related if they are directly or indirectly related not with each other but with a third [tone].[4] (Hostinský, 1879, p. 66)

Tones are "directly related" if they share an edge on his Tonnetz, i.e. if they are related by either a perfect fifth, a major third, or a minor third. All other interval relations are "indirectly related" by means of combinations of the direct intervals. These combinations correspond to paths that connect any two tones on the Tonnetz (see also Mazzola, 1985; Lewin, 1987; Longuet-Higgins, 1987a,b).

Adopting the Tonnetz as a representation for relations between tones, each tone t is related to its neighbors by one of six intervals, that we call *primary intervals*. These are the ascending perfect fifth (+P5), the descending perfect fifth (−P5), the ascending major third (+M3), the descending major third (−M3), the ascending minor third (+m3), and the descending minor third (−m3). As an example, Figure 12.3 shows the six tones that can be reached from C by means of these primary intervals. This corresponds directly to an inverted version of Hostinský's map, in which the ascending major thirds go upwards instead of downwards. Note that the primary intervals as defined here are not identical to Hostinský's direct intervals because we consider intervals in opposite directions to be different. For example, we differentiate between an ascending (+M3) and a descending major third (−M3).

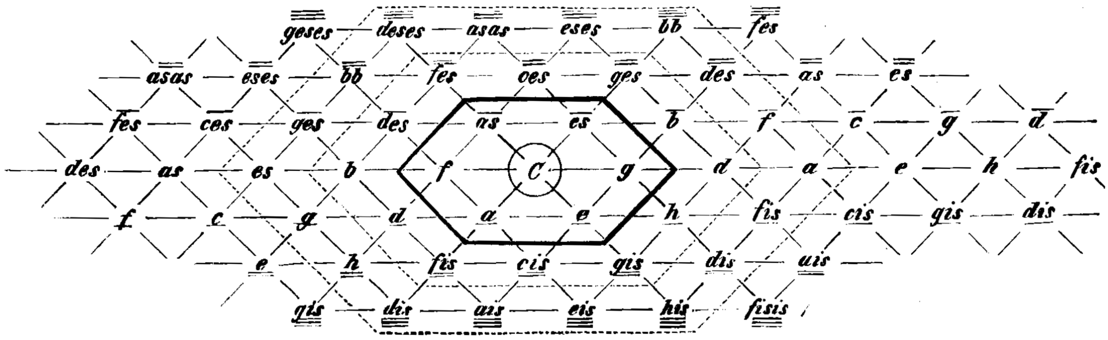
How can tonal relations other than by the primary intervals be expressed? Consider for instance the interval from C to B, an ascending major seventh. These two tonal pitch-classes are thus not directly related by one of the primary intervals. Consequently, they have to be related by concatenations of primary intervals. Some of the numerous paths on the Tonnetz from C to B using only the primary intervals are shown in Figure 12.4.

[4] The German original is "Zwei Töne, welche wesentliche Bestandtheile desselben musikalischen Klanges sind, nennen wir *unmittelbar verwandt*. Der Zusammenklang zweier unmittelbar verwandter Töne ist eine *Consonanz*. [...] Mittelbar verwandt sind zwei Töne, wenn sie zwar nicht untereinander, wohl aber beide mit einem dritten entweder unmittelbar oder mittelbar verwandt sind [...]"; translation by the author.

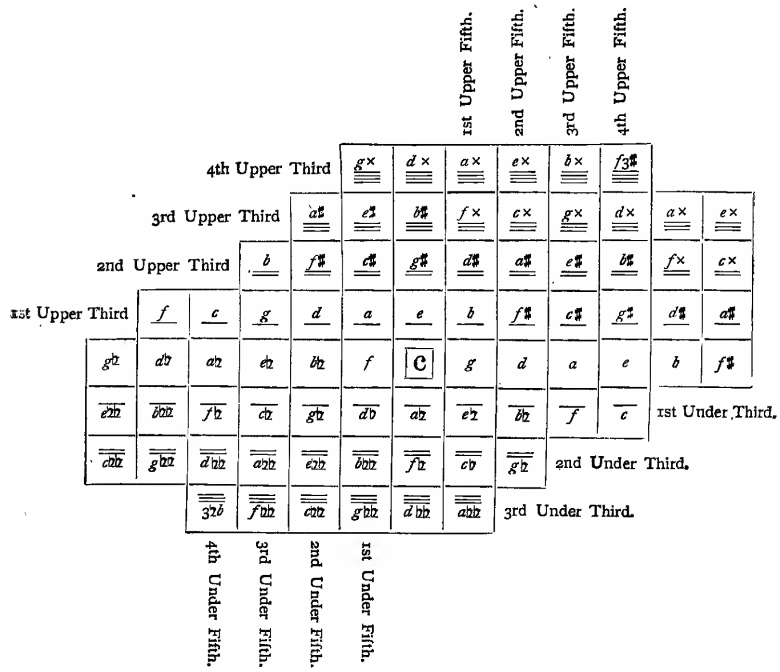
5^m 3^n.

n:	−8	−7	−6	−5	−4	−3	−2	−1	0	1	2	3	4	5	6	7	8
m 2	c	g	d	a	e	h	fis	cis	gis	dis	ais	eis	his	fisis	cisis	gisis	disis
1	as	es	b	f	c	g	d	a	e	h	fis	cis	gis	dis	ais	eis	his
0	fes	ces	ges	des	as	es	b	f	c	g	d	a	e	h	fis	cis	gis
−1	deses	asas	eses	bb	fes	ces	ges	des	as	es	b	f	c	g	d	a	e
−2	bbb	feses	ceses	geses	deses	asas	eses	bb	fes	ces	ges	des	as	es	b	f	c

(a) The Tonnetz in *Harmoniesystem in dualer Entwicklung* (von Oettingen, 1866, p. 15).



(b) The Tonnetz in *Die Lehre von den musikalischen Klängen: ein Beitrag zur aesthetischen Begründung der Harmonielehre* (Hostinský, 1879, p. 67).



(c) The Tonnetz in *Dictionary of music* (Riemann, 1896, p. 628).

Figure 12.2 – Schematic depictions of relations between tones in three historical sources.

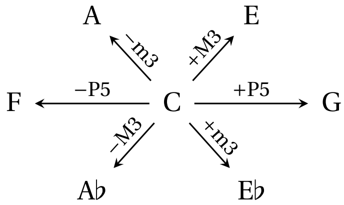


Figure 12.3 – The (directed) primary intervals perfect fifth (±P5), major third (±M3), and minor third (±m3) centered on C.

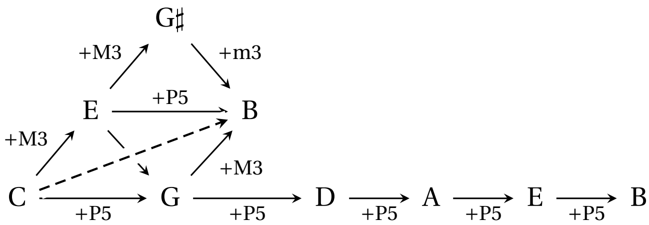


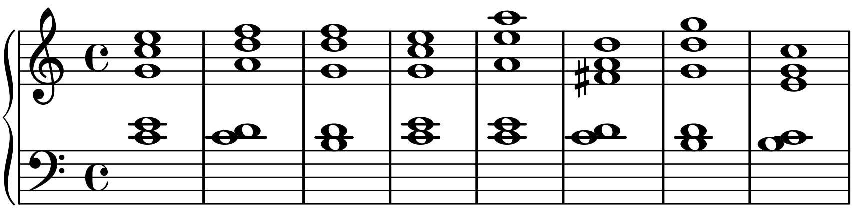
Figure 12.4 – Several paths on the Tonnetz describing the interval from C to B (dashed arrow).

For example, one could either first ascend from C by a perfect fifth (+P5) to G and then by a major third (+M3) in order to reach B, resulting in a path (+P5, +M3). Or, conversely, one could first ascend by a major third to E followed by a major fifth to B through the path (+M3, +P5). Another path is given by first ascending two major thirds to G♯ and then adding a minor third to B, resulting in the path (+M3, +M3, +m3). Using the line of fifths, one could reach B from C by ascending five perfect fifths (+P5, +P5, +P5, +P5, +P5). Yet another path would be to first ascend by a fifth to G, move back down to C, move back up to G, and then, finally, to B through the path (+P5, −P5, +P5, +M3). This last example illustrates that there are infinitely many paths connecting two tones in this graph since it is in principle possible to go back and forth any number of times. Note that the representation of tones as tonal pitch-classes does not allow to differentiate e.g. the ascending perfect fifth from the descending perfect fourth. Since the tonal pitch-classes do not contain the specification of the octave, both of these intervals are represented by +P5. It does, however, permit to distinguish between enharmonically equivalent intervals such as the ascending augmented fourth—the tritone—and the ascending diminished fifth.

Figure 12.4 also shows that tonal pitch-classes are represented multiple times on the Tonnetz. In this example, B is shown five steps to the right of C as well as one step to the right and one step to the upper right. There are infinitely many representatives of this tonal pitch-class, all lying on the line through the two representatives shown in Figure 12.4. In the historical graphs (see Figure 12.2), these were differentiated by one or several horizontal bars above or below the names of the tonal pitch-classes to indicate slightly different tunings. With respect to tuning, the difference between the two B's is a *syntonic comma* and the two pitch classes are called *syntonic images* (Cohn, 2012).

Here, we do not distinguish these elements for two reasons. First, we adopt the interpretation that the position on the Tonnetz of different instances of the same tonal pitch-class does not express different tunings but different *harmonic functions* of this tonal pitch-class with respect to some reference pitch class. This rather cognitive interpretation is not new and was already promoted by Riemann (Riemann, 1916; Wason and Marvin, 1992). More recently, Temperley (2000, 289f.) considers the context within which the tones occur in a piece more important than their tuning with respect to the cognition of the harmonic functions of a tone. We follow this interpretation but are confronted with the difficulty that, while the representation of the pieces in the corpus allows for the distinction of enharmonically equivalent tones, the representation of pieces as bags of notes eliminates their context, which renders the assignment of a harmonic function to each pitch class difficult. Concretely, in our representation of pieces as bags of notes and the representation of notes as tonal pitch-classes, the harmonic functions of the latter are ambiguous because all occurences of a tonal pitch-class, say B, are counted as tokens of the same type, regardless what their actual harmonic function in the piece is, i.e. whether B is the upper major third of G, the lower perfect fifth of F♯, or the upper major seventh of C, etc.

Consider as a concrete musical example the harmonic reduction of the first seven measures of J. S. Bach's Prelude in C major (Example 12.1). What is the relation between the notes C



Example 12.1 – Harmonic reduction of the first seven measures of J. S. Bach's Prelude in C major, BWV 846.

and F♯ in m. 6? The measure contains a D dominant seventh chord in third inversion, i.e. C is in the bass. Accordingly, the interval between C and F♯ is an ascending augmented fourth (an ascending tritone), which is the characteristic interval of the dominant seventh chord, entailing a resolution of the tritone into the minor sixth between B and G in the subsequent measure. The harmonic function of F♯ with respect to C is that it is two minor thirds below it within the D dominant seventh chord. The path from C to F♯ is thus given by (−m3, −m3).

But what is the harmonic relation between that same F♯ and the C in the very first measure of this piece? These two notes do not stand in an immediate relationship but rather in a mediated one. In the given context of Bach's *Prelude*, F♯ is the major third above the root D of the dominant seventh chord in m. 6 which, as the applied dominant to the G major chord in m. 7, stands in a fifth relation to the root of G major. This G is, in turn, the dominant of the root of the overall tonic C. In short, the harmonic function of F♯ in m. 6 in relation to the C in the first measure is that it is the major third of the secondary dominant of C, and the interval between C and F♯ is given by the path (+P5, +P5, +M3) that is shown by the black arrows in Figure 12.5. Yet another, less likely interpretation of the harmonic function of the F♯ in relation to the C in m. 1, could be that it is the lower fifth to C♯ which is the upper major third of A which is the lower mediant to C. This interval is given by the path (−m3, +M3, −P5) and shown by the dashed arrows in Figure 12.5.

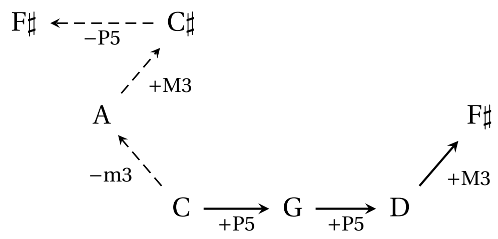


Figure 12.5 – Several paths on the Tonnetz describing different interpretations of the harmonic function of F♯ with respect to C.

To describe the relation between C and F♯, we have used the context, e.g. the local harmonic context in a measure or the context of the entire first six measures. This interpretation of paths on the Tonnetz also conforms to Lewin's (1987, 18ff.) conception of how listeners perceive ("intuit" in his words) musical intervals in tonal space. Rather than requiring that listeners are capable of intuiting any harmonic intervall directly, he argues for a cognitively more sparse explanation assuming that tonal relations are given by the recursive applications of the primary intervals, similar to Hostinský's reasoning. In this understanding, relating a tone *s* to a tone *t* implicitly requires to recursively apply a number of cognitive operations that each correspond to steps on the Tonnetz given by the primary intervals. The cognitive reality of hierarchical recursive interval relations has been investigated in a number of more recent studies (Krumhansl, 1990; Tillmann et al., 2000; Janata et al., 2002; Temperley and Marvin, 2008), although these usually rely on the representation of tones as neutral pitch-classes.

In the example above, we have said that not all paths are equally likely, i.e. we have implicitly associated some harmonic functions of a tone *s* with respect to tone *t* with a higher probability than others. As was illustrated before, there can be many different interpretations of the harmonic function of a tone, i.e. there are many paths on the Tonnetz to trace back a tone to another one corresponding to many different chains of cognitive operations. In the following section, we present a model that associates each of the primary intervals with a certain probability in a formalism that recursively traces each tonal pitch-class in a piece in the bag-of-notes representation back to a tone of reference, the tonal center.

12.3 The tonal diffusion model

We now present the Tonal Diffusion Model (TDM).[5] This novel model implements the fundamental idea that the occurrence of any tone in a musical piece is related, directly or indirectly, to the tonal center of that piece. This directly corresponds to historical models of musical

[5] This section was partially done collaboratively. The conceptualization and initial implementation of the Tonal Diffusion Model (TDM) was done by Robert Lieck. Expanding the model, fitting it to the corpus, analyzing and interpreting the results, as well as contextualizing it within the literature was done by Fabian C. Moss. A manuscript is in preparation (Lieck et al., in preparation).

intervals (Hostinský, 1879; Lewin, 1987) as well as contemporary hierarchical models of tonal space (Lerdahl, 2001). We assume that, in general, the relation of a tone s to a tone t is given by a number of steps on the Tonnetz which we interpret as cognitive operations that are necessary in order to arrive at t from s by means of the primary intervals. The probability $\pi(t \mid s)$ of performing such an operation is defined as

$$\pi(t \mid s) = \begin{cases} p_i & \text{if } \mathrm{int}(s,t) = i \text{ and } i \text{ is a primary interval} \\ 0 & \text{else ,} \end{cases} \tag{12.1}$$

where t and s are tonal pitch-classes and the function int returns the interval from s to t. The p_i are the probabilities of relating t to s by i.

This means that the probability of any interval other than the primary intervals has zero probability and hence can only be derived by combinations of the primary intervals, as in the example above where the seventh B was related to C by a number of different paths on the Tonnetz. Further, we assume that several of these basic operations can be applied in succession. Recall that all other intervals, such as ascending or descending seconds, tritones, or augmented sixths, and so forth, are expressed as combinations of these primary intervals, corresponding to paths on the Tonnetz. We also assume that $0 \leq \gamma < 1$ is the probability of applying one more operations, i.e. $1-\gamma$ is the probability of stopping. The parameter γ thus determines the depth of the tail-recursion with which a tone is related to another one. Smaller values of γ favor more immediate tonal relations, resulting in an overall more concentrated distribution on the Tonnetz, while larger values permit tones to be related through longer chains of intermediate relations and allow for more widely diffused distributions on the Tonnetz, hence the model's name.

A particular tone t has a *direct* probability of occurrence $\hat{p}(t)$ and also an *effective* probability of occurrence $p(t)$, which it gains by relating it to other tones. For tonal music we assume only the tonal center to have a non-zero direct probability of occurrence and hence define $\hat{p}(t)$ as

$$\hat{p}(t) = \begin{cases} 1 & \text{if } t \text{ is tonal center} \\ 0 & \text{else .} \end{cases} \tag{12.2}$$

As before (see Chapter 10), the tonal center of a piece is operationalized by its most frequent tonal pitch-class. The model is fully specified by Equation 12.3 that defines the probability of each tonal pitch-class t in terms of its direct probability and its effective probability. In other words, the probability of occurrence of a tonal pitch-class in a piece combines the probability of a tone to occur directly (which can only happen in the case that t is the tonal center) or to be recursively related to other tones:

$$p(t) = (1-\gamma)\,\hat{p}(t) + \gamma \sum_s \pi(t \mid s)\, p(s) \,, \tag{12.3}$$

where π and γ are defined as above.

Expanding this recursive definition reveals that $p(t)$ corresponds to the probability mass it receives over all possible ways of arriving at t from the tonal center, weighted with the probability γ of the corresponding operation to actually be applied

$$p(t) = (1-\gamma)\,\hat{p}(t) + \ldots \tag{12.4}$$
$$\ldots + \gamma \sum_{s_1} (1-\gamma)\,\hat{p}(s_1)\,\pi(t \mid s_1) \tag{12.5}$$
$$\ldots + \gamma^2 \sum_{s_1,s_2} (1-\gamma)\,\hat{p}(s_2)\,\pi(t \mid s_1)\,\pi(s_1 \mid s_2) \tag{12.6}$$
$$\ldots + \ldots$$
$$\ldots + \gamma^n \sum_{s_1,\ldots,s_n} (1-\gamma)\,\hat{p}(s_n)\,\pi(t \mid s_1)\cdot\ldots\cdot\pi(s_{n-1} \mid s_n), \tag{12.7}$$

where (12.4) is the probability mass gained through direct occurrence, (12.5) is the mass gained by applying one operation, (12.6) the mass gained by applying two operations, (12.7) the mass gained by applying n operations. The diffusion parameter γ ensures that the sum in Equation 12.3 converges. This recursive definition of interval relations in a piece directly incorporates the fact that tonal music is hierarchically organized and all notes are directly or indirectly linked to the tonal center. To account for more than one tonal center, the definition of $\hat{p}$ can in principle be generalized accordingly.

Mathematically, equation (12.3) defines a system of linear equations for p, which can either be solved directly or by dynamic programming using the following update equation

$$p^{(k+1)}(t) = (1-\gamma)\,\hat{p}(t) + \gamma \sum_{s} \pi(t \mid s)\,p^{(k)}(s), \tag{12.8}$$

where the initialization with $\hat{p}(t)$ guarantees a properly normalized distribution in each step k. The reasons for these updates to converge are the same as in other applications.[6] The advantage of the dynamic programming solution over directly solving the system of linear equations is that it is more flexible and allows for modifications that make Equation (12.3) non-linear or let γ or π depend on the update step (e.g. to restrict the maximum number of steps that are permissible).

12.4 Evaluating the model

In order to evaluate the model, it was fitted to each piece d in the corpus using *Sequential Least Squares Programming* (SLSQP; Nocedal and Wright, 2006). The optimal parameter setting

$$\theta_d = \{p_{+\mathrm{P5}}, p_{-\mathrm{P5}}, p_{+\mathrm{M3}}, p_{-\mathrm{M3}}, p_{+\mathrm{m3}}, p_{-\mathrm{M3}}, \gamma\} \tag{12.9}$$

for that piece was obtained by minimizing the *Kullback-Leibler divergence* (Cover and Thomas, 2006, see Section 11.2) between the tonal pitch-class distribution of the piece being fitted,

[6] See e.g. Equation (4.4) in Sutton and Barto (2018).

and the distribution of tonal pitch classes generated by the model with parameters θ_d. The Kullback-Leibler divergence is well-defined because the distribution of tonal pitch-classes estimated by the TDM contains no zero values as long as both fifth components and the diffusion parameter are strictly larger than zero. Figure 12.6 shows the distribution of minimal Kullback-Leibler divergences for all pieces in the corpus; its mean value is .228. Note that the axes are given in different units. The Kullback-Leibler divergences on the x-axis are given in bits while the values on the y-axis are a result of the constraint that the area under the estimated density curve has to sum to 1. Values on the y-axis can thus be much larger than 1 because the range of values on the x-axis is relatively narrow.

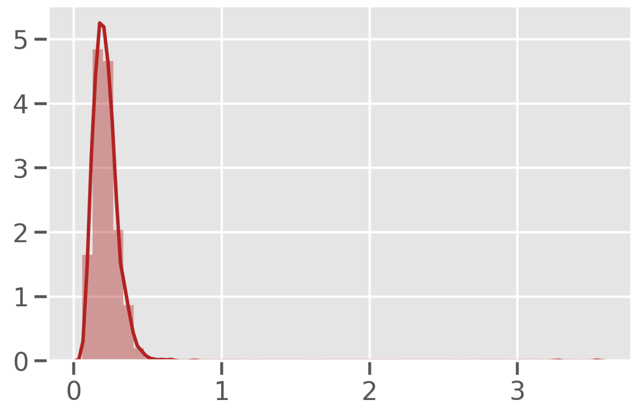


Figure 12.6 – Distribution of Kullback-Leibler divergences between the tonal pitch-class distributions of all pieces in the corpus and their respective model estimates.

The Kullback-Leibler divergences for all pieces grouped by composer is displayed in Figure 12.7 as boxplots. The line in the boxes shows the median Kullback-Leibler divergence of all pieces by a particular composer.

It can be seen that—with the exception of very few outliers—the model performs more or less equally well, regardless of composer and historical time. We conclude that the historical time of composition does not affect the model's performance on a particular piece and that we can compare the optimal parameters between pieces. The three pieces for which the model performes worst, the outliers shown by the diamonds in Figure 12.7 that have values larger than .8, are Charles-Valentin Alkan's *Esquisse*, op. 63, no. 23 "L'homme aux Sabbots" (1861) with a Kullback-Leibler divergence of 3.54; Hugo Wolf's *Eichendorff-Lieder*, no. 1 "Der Freund" (1889) with a Kullback-Leibler divergence of 3.27; and Franz Liszt's *Trauervorspiel und Trauermarsch*, S. 206 (1885) with a Kullback-Leibler divergence of .81. Although this does not contradict the previous statement that the model performance is independent of the historical time, it is telling that the largest outliers are all pieces by late 19th-century composers.

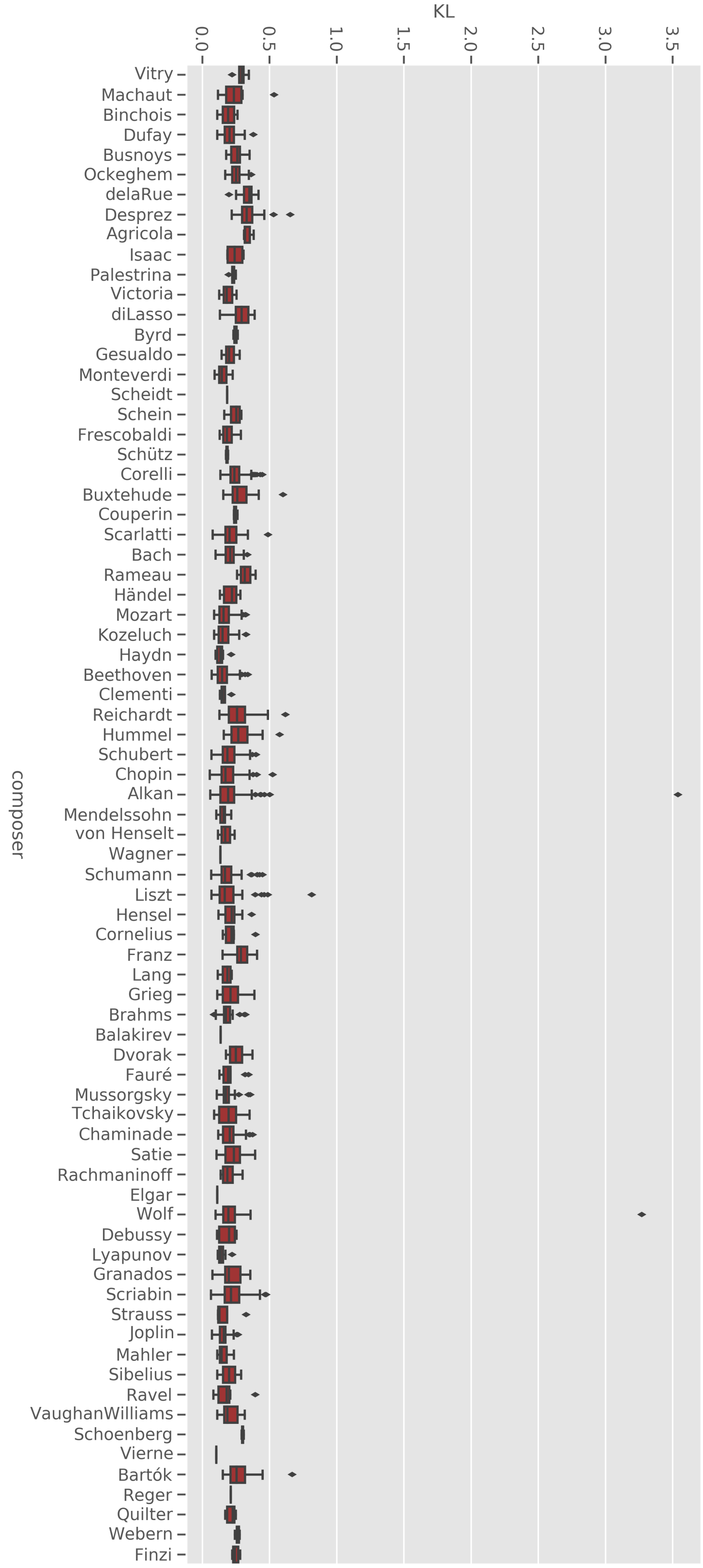


Figure 12.7 – Kullback-Leibler divergences grouped by composer.

12.5 Comparing pieces

For each piece in the corpus, the TDM infers an optimal set of parameters. In this section, we compare these parameters for four selected pieces from four different time periods, namely Orlando di Lasso's *Salve Regina* (1582), a six-part (SSATTB), largely homophonic vocal motet; Johann Sebastian Bach's Prelude in C major, BWV 846 (1722) from the first volume of the *Well-Tempered Clavier*; the first movement of Ludwig van Beethoven's Sonata in D minor (*Tempest*), op. 32, no. 2 (1802); and Franz Liszt's *Lugubre gondola*, S. 200/1 (1882).

Tonal pitch-class distributions on the Tonnetz. Since we represent pieces as distributions of tonal pitch-classes, we can make use of the theoretical models of tonal space, in particular the Tonnetz, and visualize the pieces as distributions on the Tonnetz. This is shown in Figures 12.8a–12.8d. Recall that the Tonnetz is an infinite space. Only the segment containing tonal pitch-classes in the respective pieces is shown. Each of the plots is centered at the most frequent note, the tonal center.

The distribution of tonal pitch-classes in the *Salve Regina* is shown in Figure 12.8a. The pitch classes D and A, the tonic and dominant,[7] are almost equally frequent. The frequency of the other diatonic notes F, C, G, E, and B decreases with distance to the tonal center D. The fact that B♭ is more frequent than its natural version B leads to the conclusion that this piece is in D minor rather than in D dorian. A comparison with the notes in the score shows that the chromatic notes F♯, C♯, and G♯ do appear as leading notes to G, D, and A, as one would expect in this music. F♯ is also used as the Picardy third at the end of the piece.[8]

The distribution in Bach's C major *Prelude* (Figure 12.8b) is not so different at first sight. The two most frequent tonal pitch-classes are C and G, the tonic and the dominant, respectively, also with very similar frequencies. In contrast to the *Salve Regina*, the major third above the tonal center is more frequent than the minor third above it, showing the differences between the two modes. Bach's piece spans the line-of-fifths segment from A♭ to C♯ (11 fifths), whereas Lasso's pieces ranges from B♭ to G♯ (10 fifths).

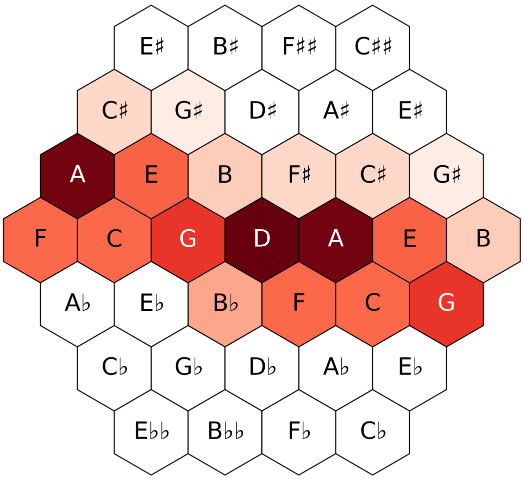
Beethoven's movement (Figure 12.8c) is much more distinct from these older pieces. The overall tonal pitch-class distribution spans the entire range from D♭ to B♯ (17 fifths). The most frequent note is not the tonic D but the dominant A. In fact, the constituent notes of the D-minor and A-major triads are particularly frequent. It is moreover evident that the tonal pitch-class distribution in this piece is much less concentrated than in the earlier pieces.

The most extreme distribution among the four examples is found in Liszt's piece (Figure 12.8d). It diffuses particularly widely on the Tonnetz, covering the range from F♭ to F♯♯ (21 fifths), thus potentially enabling the interval of an triple-augmented unison. Note, though, that the

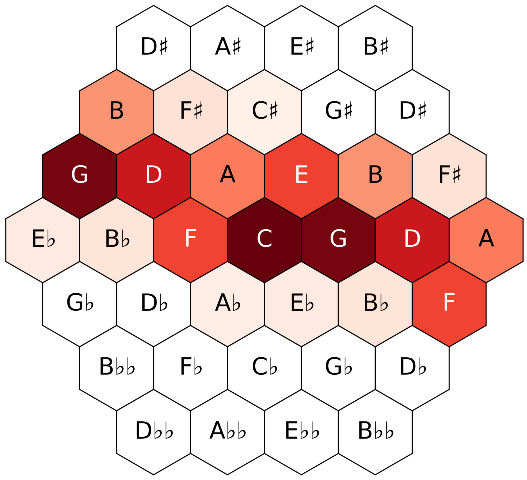
[7] The more appropriate terminology would be 'finalis' and 'repercussa' for this musical period.

[8] When a minor piece concludes with a major tonic triad, its third is called a Picardy third (Aldwell et al., 2010), named after the French region where this idiomatic ending supposedly originated.

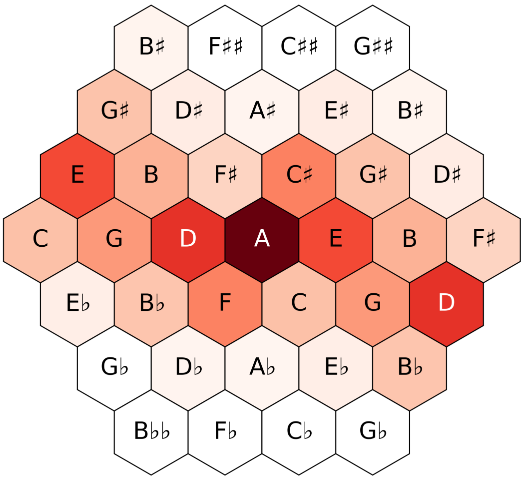
distribution of tonal pitch-classes on the line of fifths has 'gaps'; neither of C♭, E♯, or B♯ does appear, strongly indicating that the other Tonnetz dimensions are relevant as well.



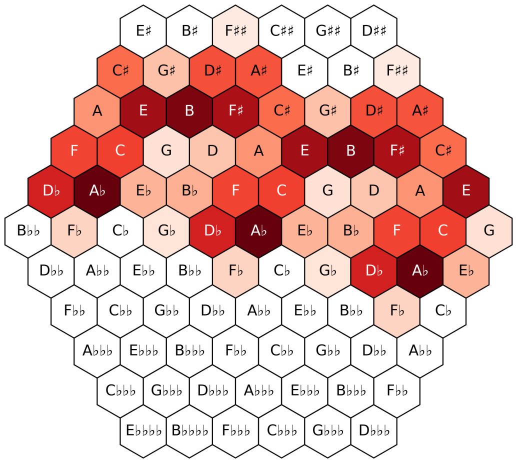
(a) Orlando di Lasso, *Salve Regina* a 6 (1582).



(b) J. S. Bach, *Prelude in C major*, BWV 846 (1722).



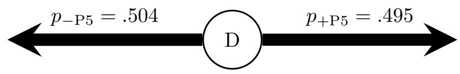
(c) Ludwig v. Beethoven, Sonata op. 31, no. 2 in D minor (*Tempest*) (1802).



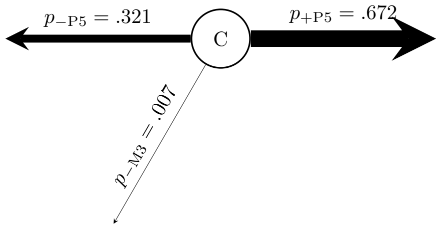
(d) Franz Liszt, *Lugubre gondola I*, S. 200/1 (1882).

Figure 12.8 – Tonal pitch-class distribution of four pieces on the Tonnetz.

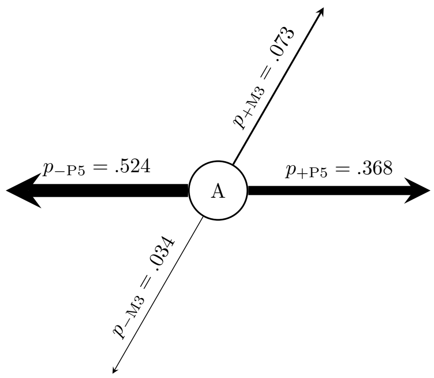
Comparing the model parameters. How does the TDM explain the distributions of tonal pitch-classes in these pieces? Recall that the model's explanation is captured by the optimal parameters θ_d for each piece. These are displayed in Figures 12.9a–12.9d. The parameters are plotted on the primary intervals (see Figure 12.3), where the widths of the arrows are proportional to the weights of the respective primary interval and the length of the arrows is proportional to the diffusion parameter γ. For visual purposes, the parameters are only shown if they are greater than a threshold $\varepsilon = 10^{-4}$.



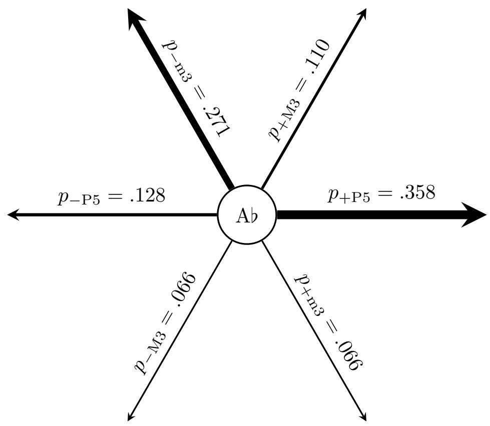
(a) Optimal parameters for Orlando di Lasso's *Salve Regina* a 6 (1582), $\gamma = .87$.



(b) Optimal parameters for J. S. Bach's *Prelude in C major*, BWV 846 (1722), $\gamma = .84$.



(c) Optimal parameters for Ludwig v. Beethoven's Sonata op. 31, no. 2 in D minor (*Tempest*) (1802), $\gamma = .82$.



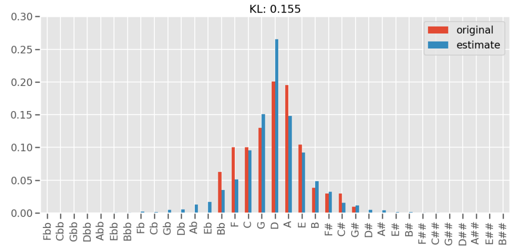
(d) Optimal parameters for Franz Liszt's *Lugubre gondola I*, S. 200/1 (1882), $\gamma = .93$.

Figure 12.9 – Optimal parameters of the tonal diffusion model (TDM) for four pieces. Arrow strengths are proportional to the primary interval weights and arrow lengths are proportional to the diffusion parameter γ.

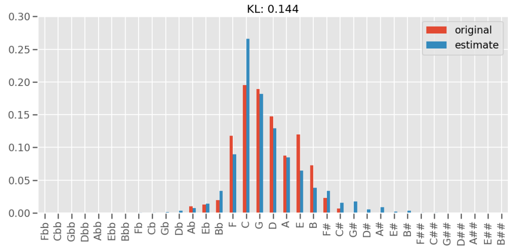
For Lasso's *Salve Regina* (Figure 12.9a), the model only assigns weights to the ascending and descending perfect fifths. Accordingly, each tonal pitch-class in this piece is related to the tonal center by some sequence of ascending or descending perfect fifths. Note also that the weights for the two intervals are almost identical, thus not favoring ascending over descending steps, or *vice versa.* The parameters found by the model for Bach's *Prelude* (Figure12.9b) exhibit two striking differences to Lasso's piece. While the two fifth components are also the strongest ones in this piece, they are not symmetric anymore, the ascending fifth dominates the descending one. For example, the model's interpretation of the relation of the tonal pitch-classes E and A to the center are not explained as the major third above and the minor third below the tonal center but as being related to it by a sequence of four and three ascending perfect fifths, respectively. Moreover, the model assigns a small weight to the descending major third, meaning that the occurrence of the tonal pitch-class A♭, for example, is partially related to the center by this interval. The larger spread of tonal pitch-classes in Beethoven's Sonata and Listz's piano piece leads to a weakening of both fifth components to the benefit of some third components that are stronger in these two pieces than in the other ones. Consider for example the pitch class F in Beethoven's piece, which is more frequent than its neighboring pitch class C. Since the TDM models each pitch class as diffusion from the tonal center, F could not have a higher probability if it were only derived as the fifth below C. Rather, the model has to assume that, at least some of the occurrences of F in this piece, are given by F as the lower major third of A, which is partially responsible for the non-zero weight of the +M3 component. The case of Liszt's piece is even more extreme, all primary intervals have relatively large probabilities. The high frequency of major third related pitch classes (e.g. D♭, F, and A; A♭, C, and E) is a result of the high frequency of augmented triads in the piece. However, inferred probabilities of the major third components are weaker than the major are weaker than the major third probabilities because the fifths are relatively strong so that the model explains these major thirds as the difference of perfect fifths and minor thirds. Compared to the extremly low probabilities of some components in the other pieces (e.g. $p_{-\mathrm{M3}} = .007$ in Bach's prelude) all components are relatively strong here.

Comparing the pieces to the model's approximations. As mentioned before, the Kullback-Leibler divergence can be used to compare the actual tonal pitch-class distributions of the pieces with the tonal pitch-class distributions generated by the model with the optimal parameters. Figures 12.10a–12.10d show both the actual distribution of tonal pitch-classes (red bars) and the ones generated by the model with the optimal parameters (blue bars). Note that the distributions shown by the red bars are exactly the same as the ones shown on the Tonnetz in Figures 12.8a–12.8d for the respective pieces. Representing both distributions on the line of fifths facilitates the visual comparison between the pitch-class distribution in the piece and its approximation by the TDM.

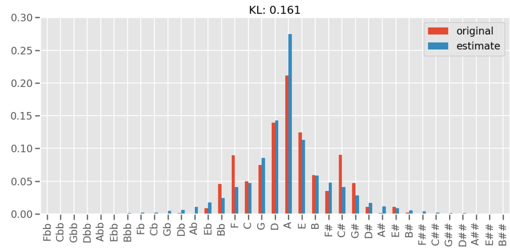
The best fit is achieved for Bach's *Prelude* with a Kullback-Leibler divergence of .144, the worst fit for Liszt's *Lugubre gondola* with a Kullback-Leibler divergence of .46 (recall that the average Kullback-Leibler divergence was .228). The juxtaposition of the two distributions for



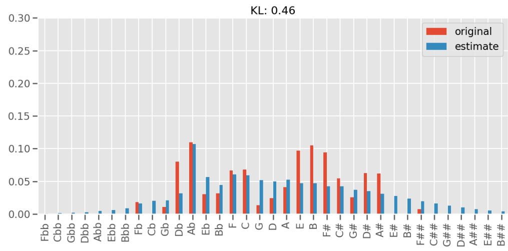
(a) Orlando di Lasso, *Salve Regina* a 6 (1582).



(b) J. S. Bach, *Prelude in C major*, BWV 846.



(c) Ludwig van Beethoven, Sonata op. 31, no. 2 in D minor (*Tempest*) (1802).



(d) Franz Liszt, *Lugubre gondola I*, S. 200/1 (1882).

Figure 12.10 – Comparison of the empirical (red bars) and modeled (blue bars) tonal pitch-class distribution for four pieces.

all pieces makes the quality of the fit evident. The *Salve Regina* is modeled quite well but the model overestimates how far tonal pitch-classes diffuse from the tonal center in descending fifths direction (to the left). It becomes also apparent that Liszt's piece is modeled poorly. It seems to be the case that the assumption of a single tonal center is hindering a better model fit. This is discussed further below and will be addressed in future adaptations of the model. Comparing the optimal parameters under the TDM for these four pieces shows that the parameter settings can vary considerably, although it is noticeable that the decay parameter γ is relatively similar for all of them, as shown by the lengths of the arrows. Certainly, each musical piece is unique in its own way, resulting in equally unique distributions of tonal pitch-classes. The next section studies whether the parameters found by the TDM for each piece do share commonalities when viewed under a historical perspective.

12.6 Diachronic relevance of primary intervals

For each piece in the corpus, the TDM finds an optimal parameter setting θ_d that best explains the distribution of tonal pitch-classes in this piece according to the model's assumptions. One can now compare these parameters for all pieces in the corpus in order in order to investigate larger historical trends.

Interval weights. Recall that the θ_d consist of six weight parameters for the primary intervals (see Figure 12.3) and one diffusion parameter γ for each piece. The distributions of the six parameter weights are studied in this section, the distribution of the diffusion parameter is analyzed in the next section. Figure 12.11 shows the six weight parameters for each piece over time for each interval separately. The left column shows ascending and the right column shows descending intervals. Perfect fifths are shown in the first row, and major and minor thirds are shown in second and third row, respectively. The ascending fifths (+P5) are shown in red, the descending fifths (−P5) in blue, the ascending minor thirds (+m3) in violet, the descending minor thirds (+m3) in gray, the ascending major thirds (+M3) in yellow, and the descending major thirds (−M3) in green. The black lines in each of the subplots are given by a a local linear regression using Locally Weighted Scatterplot Smoothing (LOWESS; Cleveland and Devlin, 1988, see Section 10.2).

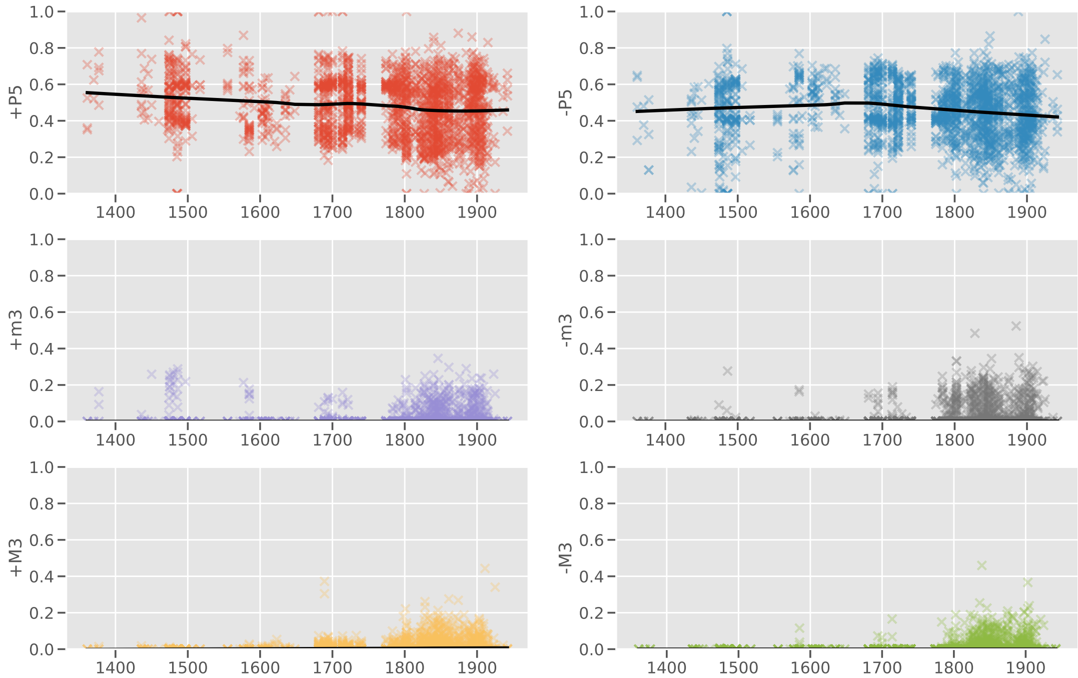


Figure 12.11 – Directed interval weights found by the tonal diffusion model across the historical timeline.

Note that the six parameters for each piece together form a distribution, i.e. for a given piece, the values of the weights for the respective primary parameters sum to 1. Observing the changing trendline, the first remark one can make is that, across the whole timeline from 1361–1942, the model assigns the largest weights by far to both the ascending and the descending perfect fifth (first row). While the ascending fifth direction slightly decreases over time, the descending fifth direction slightly rises until the end of the 17th century and then decreases again. For a large majority of pieces, the model does explain their tonal pitch-class distributions by means of ascending or descending fifths. This finding is in agreement with the result from previous sections that the line of fifths is the essential structure underlying the tonal pitch-class distributions in the corpus. In comparison to the fifths, the third components do not play a major role. While there are a number of non-zero weights for the third intervals, as can be seen by the colored crosses in the subplots in the second and third row of Figure 12.11, the trendline is pulled down by the overwhelming proportion of zero values. Recall the four pieces discussed above. For all of them at least one third component had zero weight. This does, of course, not entail that thirds are not important for tonal music. It rather means that the model explains these thirds in terms of fifths relations instead of assuming a separate third dimension. It can nonetheless be seen that the weights for major and minor thirds are particulary strong in the 19th century, especially for ascending major thirds as well as descending minor and major thirds. This does, however, not change the overall trend.

In order to see more subtle changes in the usage of the primary intervals, we have to zoom in and consider the relative changes over time in the trendlines given by the LOWESS curves. Moreover, we can apply the previously introduced bootstrap method in order estimate the variance in the interval weights. This is shown in Figure 12.12. Instead of displaying a single historical trendline derived from all pieces in the corpus, each line in the bundles of trendlines shown in the subplots corresponds to a LOWESS curve calculated from a bootstrap sample. The colors correspond to the colors assigned to the respective intervals shown before. The exact values of the pieces are not shown here to facilitate focussing on the trendlines. Note that the mean of the bootstrapped trendlines converges to the black trendlines shown in Figure 12.11 for a large number of bootstrap samples.

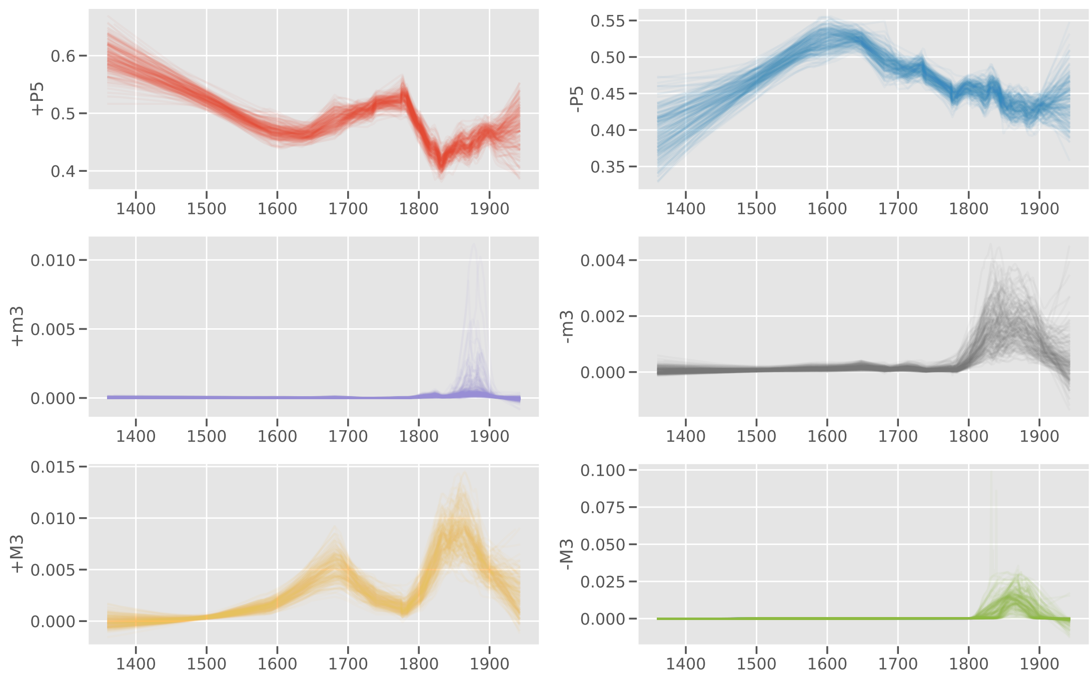


Figure 12.12 – LOWESS curves of 200 bootstrap samples for the directed interval weights.

While Figure 12.11 emphasized the differences between the primary intervals, the representation in Figure 12.12 renders a closer look on the historical changes for the intervals individually. It can be seen that the trends in the two fifth components are more complex than it was visible from the global point of view. For example, the peak of the ascending fifth during the 18th century could not be seen in Figure 12.11. The strongest difference between Figures 12.11 and 12.12 is evident in the third components. All four third components, ascending and descending major and minor thirds, show a relatively strong peak in the 19th century. Although the extent of these peaks is very different for each of the four third intervals, ranging from a scale of 10^{-4} in the case of the ascending minor third (+m3) to 10^{-2} in the case of the descending major third (−M3), they have in common that the 19th century leads to a drastic increase in weights of the thirds, a fact that was not visible in the black trendlines of Figure 12.11. Note also that the distribution of the weight parameters for the ascending major third (+M3; bottom left panel in Figure 12.12) has a peak in the late 17th century which aggrees with the findings of the previous results (see Figures 10.4 and 11.12). Recall that all notes in third direction on the Tonnetz can also be reached by fifths since the line of fifths contains all tonal pitch-classes. Only when it seems unlikely that thirds are explained by either of the two generally strong fifth components, the model will assign more weight to the third components. We have seen how this can happen in the example of Beethoven's *Tempest* Sonata above, where the relatively high frequency of the tonal pitch-class F could not only be explained by tracing it back to the tonal center A via ascending perfect fifths, but also necessitated to assign some weight to the ascending major third in order to directly backtrace F to A.

In conclusion, the tonal pitch-class distributions of the 19th-century compositions in the corpus are shaped in such a way that third-based explanations are much more likely than in earlier centuries. This is in strong correspondence with virtually all theoretical accounts of harmony and tonality in the 19th century (see Section 1.1). It was also shown that the absolute values of the parameter weights are weakest for the minor thirds, both ascending (+m3) and descending (−m3). This is in line with Hostinský's assessment of the relative importance of the primary intervals, concluding that "needless to say, the degree of relationship is strongest in the fifth direction, and weakest in the minor third direction"[9] (Hostinský, 1879, p. 67). While the results seem to support that assessment, it is important to note that one cannot rule out other explanations as well, for example the mere fact that minor thirds span a smaller range on the line of fifths (three fifths) than major thirds (four fifths) and hence are more likely to be derived by consecutive fifths rather on their own.

Diffusion parameter. We now inspect the distribution of the diffusion parameter γ of the TDM. It determines how far the notes in a musical piece diffuse from the tonal center in the six primary interval directions. In other words, it determines in how many steps a tonal pitch-class can be traced back to the tonal center. The smaller the value of γ the lower the probabilities for longer paths (see Equation 12.3). The diffusion parameter γ is independent of the six interval weight parameters and ranges from zero to one, $\gamma \in [0,1)$. If it were equal to

[9] The German original is "Selbstverständlich ist der Verwandtschaftsgrad in der Quintenrichtung am stärksten, in der Richtung der kleinen Terz am schwächsten" (Hostinský, 1879, p. 67); translation by the author.

zero, the piece would not diffuse at all and just put all probability mass on the tonal center, and if it were equal to 1 the probability mass would diffuse infinitely on the Tonnetz and the sum in Equation 12.3 would not converge. The distribution of the diffusion parameters γ for all pieces in the corpus is shown in Figure 12.13. As before, we use 200 bootstrap samples of the corpus to show LOWESS curves (black lines) in order to indicate historical trends in the parameter's distribution.

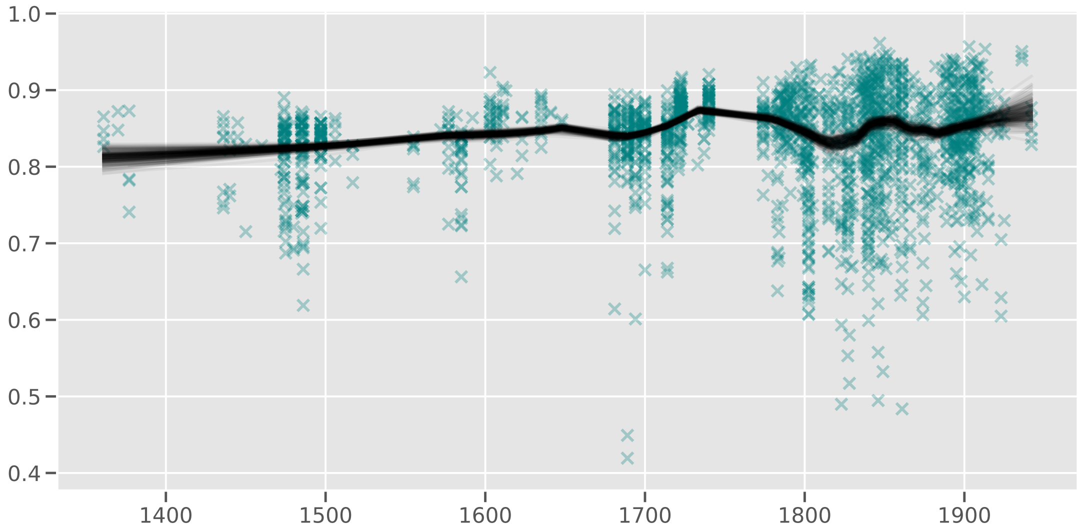


Figure 12.13 – Bootstrapped diffusion parameters of the TDM.

Overall, the variance exhibited by the bundle of LOWESS curves is relatively narrow with values well above .75 across the historical timeline. Moreover, while the curves exhibit several local extrema, there is a generally increasing trend in the distribution of the diffusion parameters. This means that the probability for tonal pitch-class distributions to diffuse more broadly on the Tonnetz historically increases. The exploration of larger ranges on the line of fifths has already been shown in Chapter 10. The model presented here has additionally provided a way to trace the larger distributions in other interval directions, namely major and minor thirds, both ascending and descending.

13 Summary and discussion

In this part we have studied a corpus of more than two thousand pieces in MusicXML format spanning a historical range of almost six hundred years. Specifically, we have represented these pieces as distributions of the tonal pitch-classes that are contained in these compositions. Without making any assumptions about the relations between these tonal pitch-classes we have shown that they are most strongly related by the line of fifths and that, over the course of time, composers explore ever wider regions in this tonal space. Further, we have introduced the concept of tonal pitch-class co-evolution and applied it to the entire corpus as well as to historical segments, and used it to confirm the primacy of the line of fifths also from a diachronic perspective. It is thus certainly not too far-fetched to state that the interval of the perfect fifth ties together the "Age of Tonality (ca. 1600-1918)" (Meyer, 1989, p. 20). The central role of the perfect fifth for the organization of tonal material confirms earlier empirical findings that have shown its importance for root progressions (Hedges and Rohrmeier, 2011) and for the historical development of Western tonality (Huang et al., 2017). However, these approaches found the circle of fifths instead of the line of fifths, simply due to the fact that their studies were based on MIDI files. This provides an excellent example of how implicit assumptions encoded in the data—in this case the assumption of enharmonic equivalence—can influence and bias the results. While in this case the divergence between the discovery of the circle of fifths on one hand and the line of fifths on the other hand is unproblematic since the two results rather reaffirm each other, the general lesson from this for empirical music research is how important it is to pay attention to unspoken premises and make them explicit whenever possible.

Using the more complex model of Latent Dirichlet Allocation, we have discovered that the latent topics assumed by this model correspond to contiguous segments on that line, regardless of the number of topics. This contiguity is interesting for two reasons. First, recall that the model is conditioned on pieces in the bag-of-notes representation and thus does not know the order in which these notes appear in the piece. The comparison to the topics in the artificial corpus generated with arbitrary parameter settings (Figure 11.4) demonstrates how essential the closeness of tonal pitch-classes on the line of fifths is for the distributions of notes in musical pieces. Thus, the tonal profiles of the topics found by LDA for several values of K support the findings in Chapter 10, reaffirming the primacy of the perfect fifth from yet another angle. Second, consider the great amount of 19th-century pieces in the corpus including, for instance, a large number of pieces by Frédéric Chopin, Charles-Valentin Alkan, Franz Liszt, Johannes Brahms, and Alexander Scriabin. Their compositions are known to contain highly chromatic passages, manifold non-diatonic key changes, abundant usage of enharmonic exchanges, and so forth (see Part II "Microanalysis"). Apparently this does not not outweigh the overall underlying diatonic structure, in the sense of relatively small contiguous segments on the line of fifths.

This point is more important than it might seem at first. If it were true that chromaticism and enharmonicism were to have replaced the more traditional diatonic system of writing music this would have led to topics that are more spread out on the line of fifths. The emphasis in musicology and music theory with respect to 19th-century compositions on new compositional techniques and innovative strategies leaves aside the fact that—with the exception of a small number of remarkable pieces such as Franz Liszt's *Nuages gris*, S. 199—the vast majority of the tonal material is still governed by the same tonal relations as earlier music when viewed from a global vantage point under the bag-of-notes model. This should by no means downplay the dramatic changes that have taken place in the way composers wrote music in the 19th century. What's more, the influence a composer and his or her pieces have on future artistic developments certainly do not solely depend on the number of notes they write, but it is rarely pointed out how much of the music is governed by similar distributions of notes. And we are not even taking into account composers that did not make it into the canon—that what Moretti (2000) calls the "slaughterhouse of literature". One should treat both positions, the emphasis on the particular and the emphasis on the general, as two sides of the same coin where each only makes sense in light of the other. Only in considering both the 'micro' and the 'macro' perspectives will serve to achieve a complete picure of the history of tonality and the impact of 19th-century composers.

The results presented the final chapter of the Macroanalysis part are based on the Tonal Diffusion Model (TDM), a novel model that is grounded in music theory and cognition. It models tonal pitch-class distributions in musical pieces, represented as bags of notes, by backtracing each tonal pitch-class to the tonal center of the piece. The results have shown differences between individual compositions as well as large-scale historical changes in the exploration of the Tonnetz. Modeling tonality with the tonal diffusion model has shown that, while the perfect fifth remains the most important primary interval throughout the considered range of music history, the importance of major and minor thirds changes drastically in the 19th century. One can conceive of several directions to build build upon these results and expand the current implementation of the model in the future. Tonality is a hierarchical phenomenon. While the current version of the TDM traces each tonal pitch-class back to the tonal center of the piece, one could also imagine a weighted mixture of TDMs, hence allowing for multiple tonal centers accordingly. This way, one could include fitting pieces that where one does not need to assume that sections are all tonally related, such in multimovement works such as sonatas, or even larger works such as oratorios or operas. In each of the primary interval directions, the model diffuses with an exponential decay given by the diffusion parameter γ which seems to be too strong a decay (compare the distributions in Figures 12.10a–12.10d). One step to achieve a better fit could be to exchange the constant diffusion parameter γ with more general diffusion functions that more accurately conform to the tonal pitch-class distributions in musical pieces. A further advancement of the model could be to go beyond the bag-of-notes representation for pieces and incorporate the temporal order in which tonal pitch-classes appear in a piece, for instance by relying on syntactical models of music. This will certainly lead to a more fine-grained view on the changes of tonality and is also likely to expose differences of the treatment of compositional strategies between composers. By relating temporal information of tonal pitch-classes to directions on the Tonnetz it will be also possible to see changes in the relation between harmony and musical form in the 19th century (and beyond). Based on the results of the current implementation of the TDM, these seem to be very promising directions for future research.

The findings in this chapter somewhat put into perspective the analyses in Part II and allow for more general reflections on example-based music analysis. Selecting musical pieces for analysis does not only entail the problem of misrepresenting the repertoire by focussing on the canon but also to be implicitly biased by salient phenomena. What is salient depends, naturally, on the contrast to some kind of more or less constant background against which a piece, or an event in the piece, stands out. Hence, the concentration of a small number of examples entails the risk of overgeneralization of the observations. The corpus-based approach taken here is an attempt circumvent this problem. Focussing only on particular examples and on historical changes may also steer the view away from the complementary perspective that what remains constant is interesting as well. Given a historical development of several hundreds of years, an artistic system like music in which composers explore, create, and develop new techniques, it is meaningful that one can find constant elements in the 'transitions of tonality'.

Conclusion Part V

Conclusion and prospects

> I hope that the relationships and connections developed in my sketch-history will seem interesting and illuminating, coherent and convincing. But they do not pretend to be definitive. They are hypotheses. Some may be downright wrong, others will require refinement. All need to be tested through application to genres and repertoires not considered here. It is a program of work to be done, of ideas and hypotheses to be evaluated and perhaps rejected, explored and perhaps extended.
>
> Leonard B. Meyer, *Style and Music*

This thesis has presented a corpus-based study on the history of tonality with a particular focus on the 19th century. The results document substantial historical changes: The tonal material employed by composers throughout approximately six hundred years of Western classical music is continuously growing. It has been shown that, within the 19th century, chord progressions by fifths become less prevalent in favor of progressions by thirds and seconds. We have seen that the co-evolution of tones changes drastically over the 18th, 19th, and 20th centuries. Latent tonal profiles, or topics, representing chromaticism manifest in the 18th and stabilize in the 19th century, the latter of which also witnesses a surge in musical compositions that explore the Tonnetz along the major and minor third axes. Although chromaticism is not an invention of the 18th century—one may think of Renaissance pieces, such as Carlo Gesualdo's *Moro lasso*—it becomes pervasive and statistically visible only over the course of the 18th century. In their totality, these results reveal fundamental aspects of transitions of tonality that reflect the different stages in its historical development as already described by Fétis (1844).

At the same time, it has been shown that there are many commonalities throughout the considered historical period: The relation between types and tokens as well as between the frequencies and ranks of chords in the annotated corpora of 19th century composers have been shown to conform to well-established models for these relations in corpus linguistics, Heaps' and Zipf's laws, respectively. We have seen that chord progressions are asymmetrical and interpreted this finding as another central factor of tonality, although composers differ in the degree to which their chord progressions are directed. As yet another stable component of tonality, the integral structure of the line of fifths has surfaced both in synchronic and diachronic observations. Small contiguous segments on this basic tonal space have been shown to act as latent topics, showing that this underlying structure ties together pieces as early as Philippe de Vitry's *Virtutibus laudabilis* (1361) and as late as Gerald Finzi's *Let us Garlands bring* (1942), regardless of their manifold differences. Despite the undeniable freedom of the composers' creative expressivity, these findings attest to the fact that music exhibits numerous regularities, possibly due to cognitive constraints both on the side of the composer and the audience, as well as cultural factors such as the tradition within which music subsists. Yet, the causes for these patterns remain elusive and are still not fully understood.

Studying tonality is not the same as studying music. Other important dimensions have been set aside, such as rhythm, meter, text, instrumentation, performance, and many more. With the exceptions of the analysis of Liszt's *Sonetto 47 del Petrarca* in the Microanalysis part and the study of chord progressions in the Mesoanalysis part, this study did also not take the temporal order of musical elements, such as chords or notes, into account. Undoubtedly, temporal order is an important variable for music, an art that unfolds in time. The integration of syntactic theories and models for the sequential ordering of musical elements on one hand and the distributional perspective advocated here on the other hand is certainly an important avenue for future research. Moreover, music does not exist in a vacuum. It is made by and for humans, thus the psychological dimension, that we have only touched upon in passing, needs to be considered more seriously in order attain a better understanding of music as both a cultural and a cognitive phenomenon. To what extent are the musical structures that were discovered in the corpora perceptually relevant? How do they shape our appreciation of music? And, conversely, how do the constraints given by our sensory and mental capacities shape the creation and stylistic development of music? On that matter, this conclusion leaves us with more questions than answers.

Against the backdrop of these results, this study does not fully subscribe to either of the two narratives on the history of tonality presented in Chapter 1: on one hand a teleological view in which tonality culminates in the classical period and dissolves in the 19th century, and on the other hand a view of paradigmatic changes that is shaped by several revolutionary turnovers leading to essentially new tonal systems; instead, it advocates for a less radical view in which both aspects play decisive roles. It is indeed the case that the music of the 19th century exhibits a number of extreme tendencies, but focusing only on those occludes the fact that they still have a lot in common with earlier music, in particular the importance of the fifth, both in the relation between tonic and dominant chords, as well as in the organization of tonal pitch-classes on the line of fifths. It might well be the case that the disagreement with respect to the historical processes results purely from implicit biases of the proponents of the two positions. Whether one considers tonality as having reached its pinnacle with Beethoven, or whether one understands extended tonality as a new tonal system that coexists with classical tonality, may solely depend on which aspects of tonality one considers to be relevant and which not. An evidence-based corpus study such as this one is not a universal remedy to avoid such biases since it also relies on assumptions that impact on the data representation, modeling choices, and interpretation of results, but ideally these are made explicit which eases comparison and critical discourse.

Many of the results of this doctoral thesis corroborate prior musicological knowledge and ground them in a larger empirical basis than approaches that only rely on a handful of carefully selected examples. Such examples are commonly chosen to represent a larger class of pieces, e.g. the oeuvre of a composer or even a musical style as a whole. As has been discussed at the beginning, representativity is not only an obstacle for such traditional approaches, it also constitutes a major challenge for corpus studies. It is extremely difficult to delineate *what exactly* a corpus should be representative of. Merely increasing the amount and sizes of corpora does not solve this fundamental epistemological problem but the augmented sample size as given by large corpora can increase the confidence in the consistency of research findings. Contrary to manual music analyses that are usually restricted to a few measures of music, corpus studies can be based on musical pieces as a whole, in an analogous way that literary corpus studies read large bodies of texts from a distance (Moretti, 2013).

We have seen that many music theoretical concepts can indeed be quantified to become applicable not only to a single examples but to much large datasets. By specifying the employed concepts formally, the present approach invites the adoption, evaluation, and critique by others, which is indispensable for advancing the state-of-the-art and for gaining deeper insights in the future. It is one of the main lessons to be drawn from this study that, if one wants to take music theoretical concepts seriously, it is worthwhile to put them to the test by working out how they can be operationalized. Despite being meticoulous and tiresome at times, this process is ultimately highly rewarding. Apart from concrete outcomes—the results that are reported in publications such as this one—there is an abunance of illuminating experiences, of insights from taking wrong turns and making mistakes, as well as of revisions of prior convictions that is highly gratifying.

But there is a drawback. The gain in generality as achieved by automated approaches comes with a loss in detail. In the current context, the entirety of a musical piece was oftentimes represented as just a single number—a stark contrast to the complexities of a composition that one can consider by focusing on a couple of measures. For this reason, manual analyses should not be replaced but be complemented by computational studies. The multi-level structure of this thesis, combining the close reading of a score in the Microanalysis part with algorithmic analyses in the Meso- and Macroanalysis parts, attempts at providing a showcase of what such a complementarity might look like. The three analytical levels have demonstrated the feasibility to address musicological and music theoretical issues from different perspectives, each laden with its own benefits and challenges. It remains to be seen whether more fine-grained or even continuously scalable methods will be developed in the future. The computational approach adopted here has moreover achieved some new findings that were beyond the scope of traditional music theory, simply due to the lack of appropriate models and datasets. This study has furthermore demonstrated a wide range of methods which are not only suitable for the questions asked here, but which are as well applicable to a broad spectrum of quantitative music studies.

It must also be acknowledged that corpus-based music research requires an immense amount of resources. Scores have to be located, digitized, transcribed, or annotated; assistants have to be trained in order to perform skillful music theoretical analyses; time has to be invested in an abysmal variety of tasks, e.g. the acquisition, preparation, transformation, and analysis of the obtained data as well as in the interpretation and contextualization of the results that are derived from it, in the mastery of statistical and data science methods, and in the setup of hard- and software—to name only a few. In addition, all of this has to be supported both financially and institutionally. Musical corpus research is—despite its advantages some of which this thesis has demonstrated—an expensive and precious endeavor. It should be apparent that this line of work cannot be done in isolation. The skill set required for musical corpus research is extremely broad and few can claim to possess it entirely. Collaboration within and across disciplinary boundaries is thus not an option but an obligation.

One major goal of this study was to build a bridge between traditional and computational music analysis. It is hoped that the multi-scalar approach provided here may serve as a case study of what a reconciliation between these two poles might look like. However, the methodological gap between musicological scholarship and empirical research remains a source of potential tension. In this regard, one has to acknowledge that bridges have to be accessible from both sides. In my view, it is imperative for musicologists to engage with quantitative practices and digital tools to be able to draw on the entire methodological repertoire of 21st-century musicology. It is equally important for empirically or formally working researchers to acknowledge the complexities that an object of study like music entails. Reducing it to 'just data' neglects its intricate embedding in human history, thought, perception, and aesthetic appreciation. The digital humanities are a promising arena where the traditional and the contemporary can meet, engage in communication and interaction, and explore new pathways.

Musical corpus research is to the date rapidly growing and gaining more and more momentum. New initiatives and projects as well as methods and datasets are published in high frequency. I hope that this work will be regarded as contributing a drop of water to this wave. Apart from very few exceptions, the field is still in a state where individuals or small groups develop their own formats, datasets, as well as analytical methods and tools. It will be an important task for the coming years and decades to join forces, develop standards, coordinate the creation of new corpora, and develop research paradigms that hopefully will transform musical corpus studies to a veritable discipline, benefitting the whole research community. The future of the field might depend on it.

A Figures

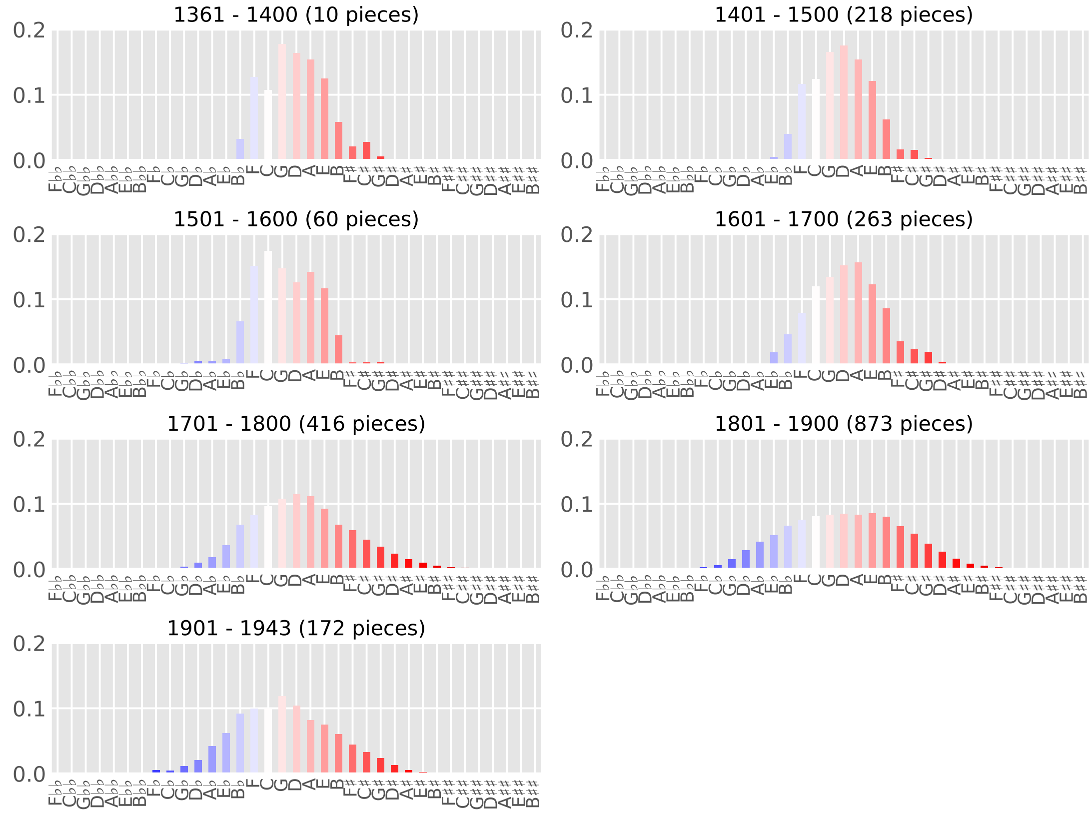


Figure A.1 – Average tonal pitch-class distribution for each century within the historical range of the corpus.

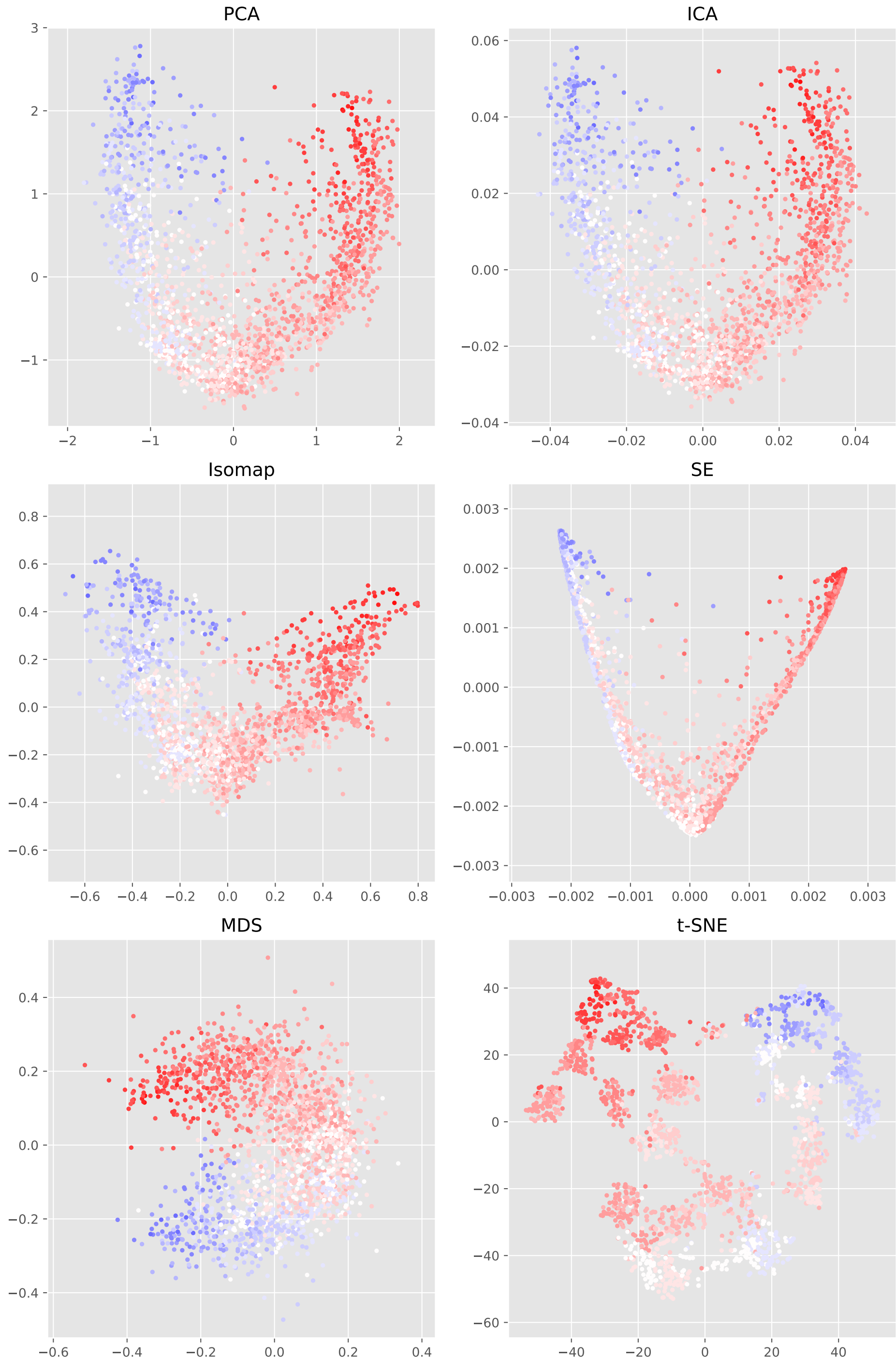
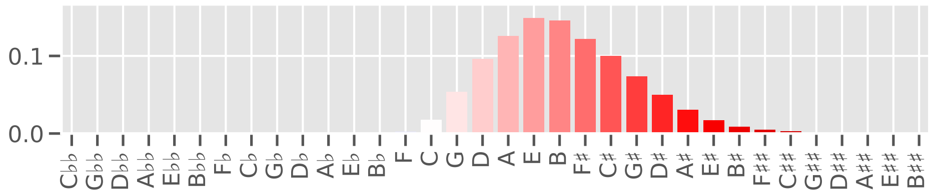
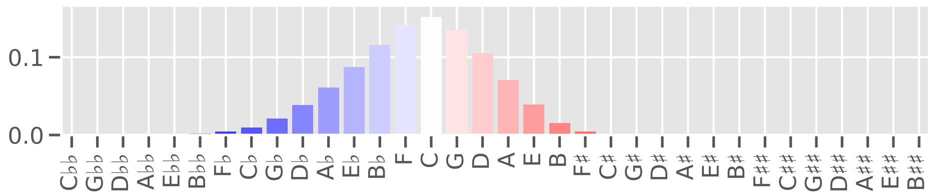


Figure A.2 – Comparison of several dimensionality reduction methods. *Principal Component Analysis* (PCA, top left; Jolliffe, 2002), *Independent Component Analysis* (ICA, top right; Hyvärinen and Oja, 2000), *Isomap* (center left; Tenenbaum et al., 2000), *Spectral Embedding* (SE, center right; Belkin and Niyogi, 2003), *Multidimensional Scaling* (MDS, bottom left; Kruskal, 1964), and *t-distributed Stochastic Neighbor Embedding* (t-SNE, bottom right; Van Der Maaten and Hinton, 2008). For PCA and ICA, the data was whitened (Kessy et al., 2018) to the effect that the unit of the axes in the corresponding subplots is the standard deviation.

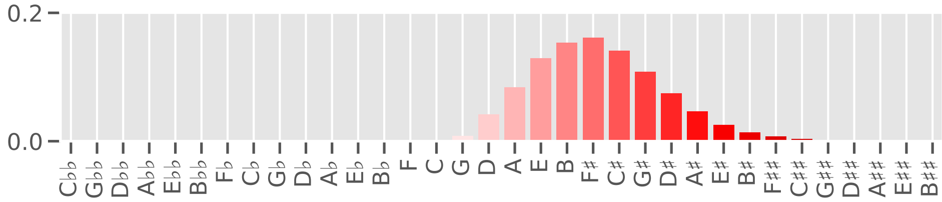


(a) Topic 1 of 2.

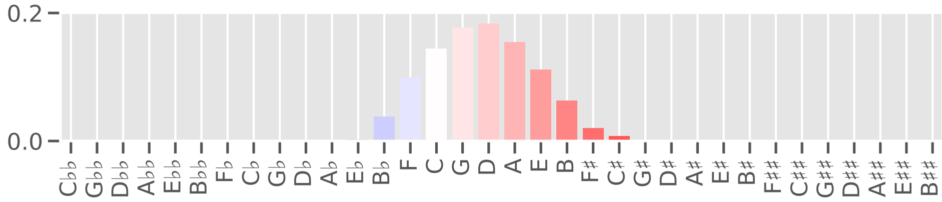


(b) Topic 2 of 2.

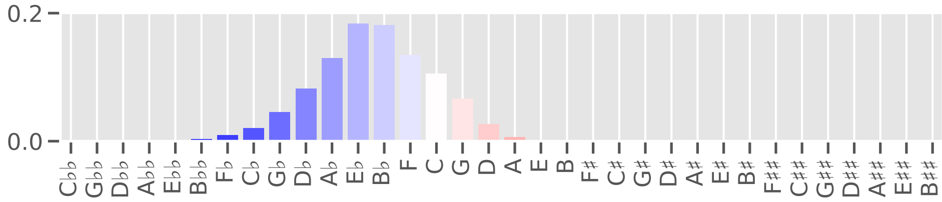
Figure A.3 – The note distributions for $K = 2$ topics.



(a) Topic 1 of 3.

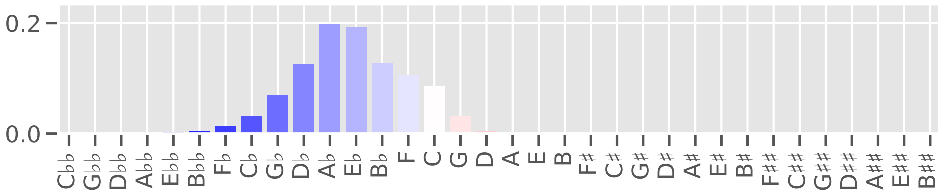


(b) Topic 2 of 3.

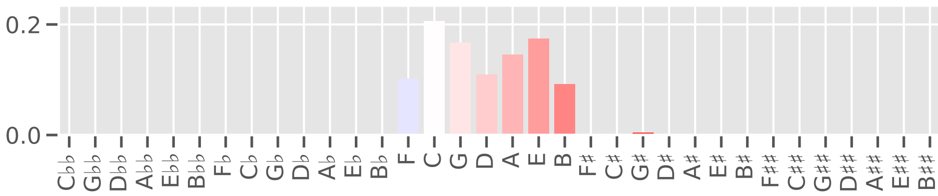


(c) Topic 3 of 3.

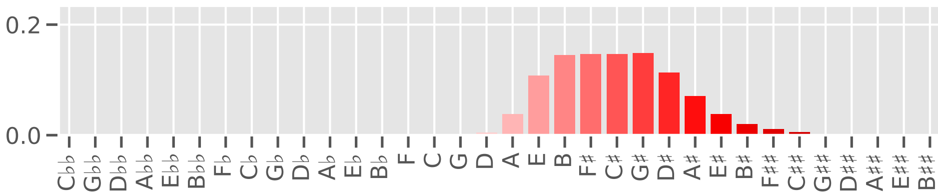
Figure A.4 – The note distributions for $K = 3$ topics.



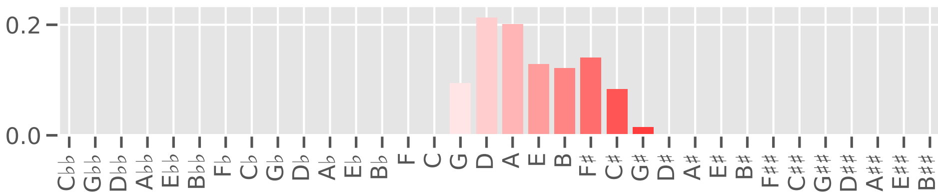
(a) Topic 1 of 5.



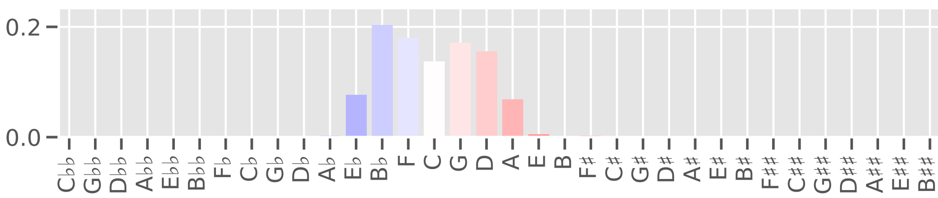
(b) Topic 2 of 5.



(c) Topic 3 of 5.

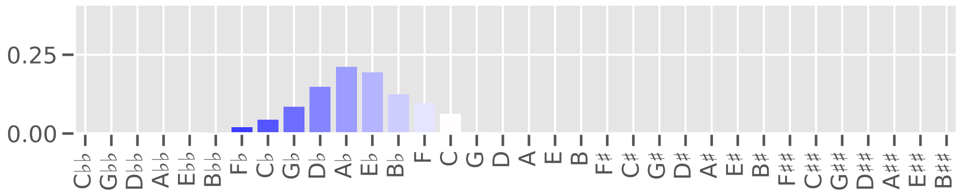


(d) Topic 4 of 5.

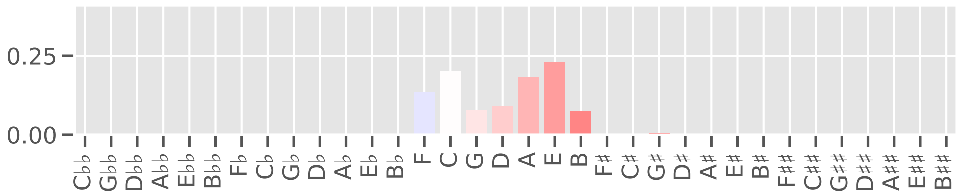


(e) Topic 5 of 5.

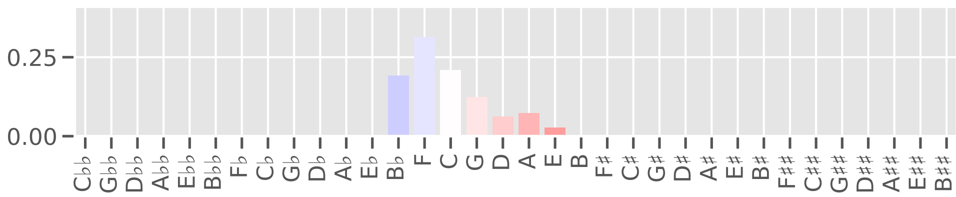
Figure A.5 – The note distributions for $K = 5$ topics.



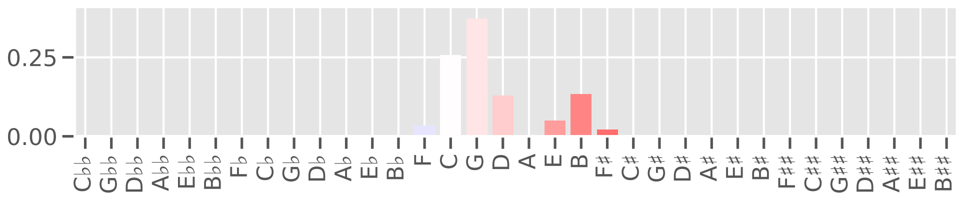
(a) Topic 1 of 10.



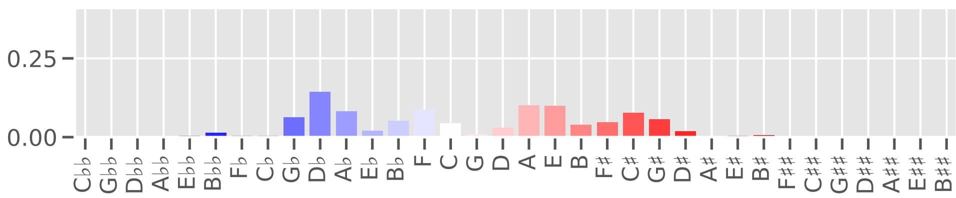
(b) Topic 2 of 10.



(c) Topic 3 of 10.

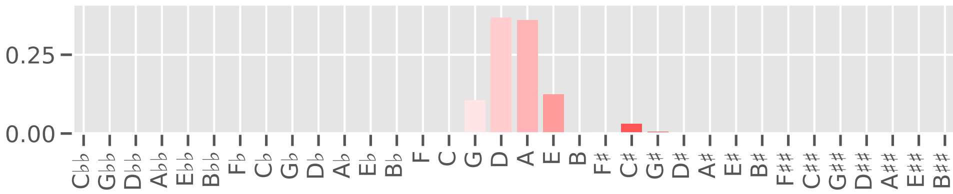


(d) Topic 4 of 10.

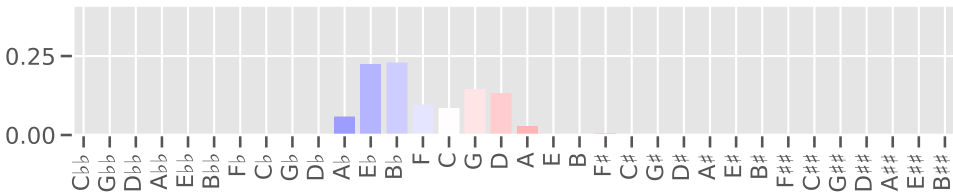


(e) Topic 5 of 10.

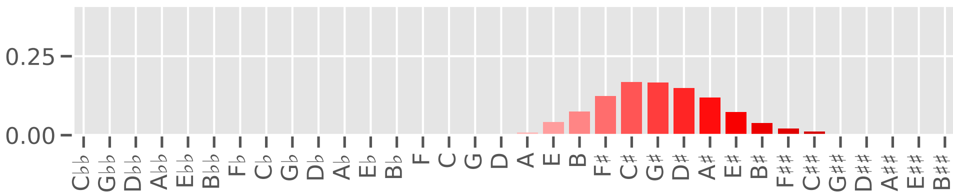
Figure A.6 – The note distributions for $K = 10$ topics.



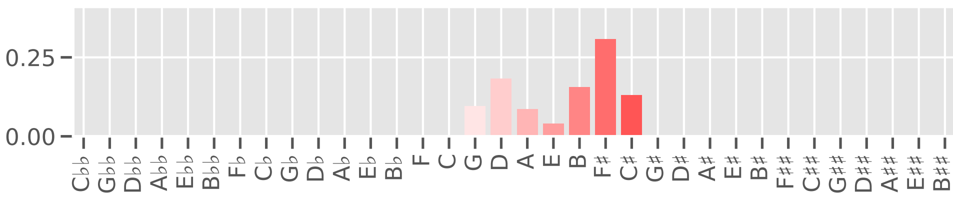
(f) Topic 6 of 10.



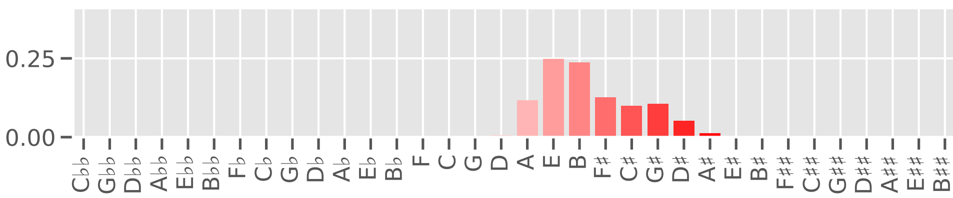
(g) Topic 7 of 10.



(h) Topic 8 of 10.

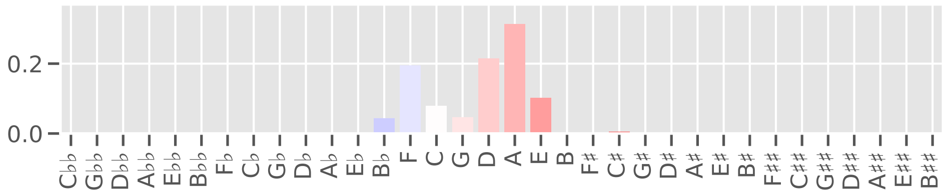


(i) Topic 9 of 10.

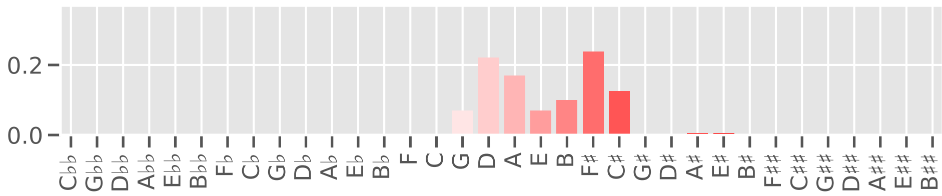


(j) Topic 10 of 10.

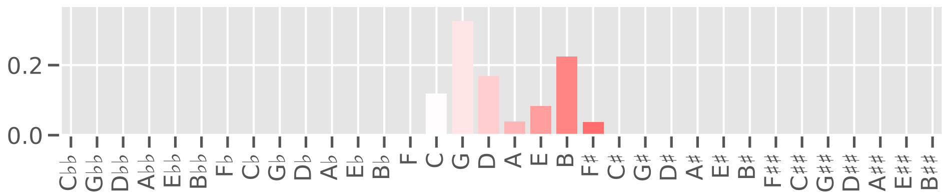
Figure A.6 – The note distributions for $K = 10$ topics (cont.).



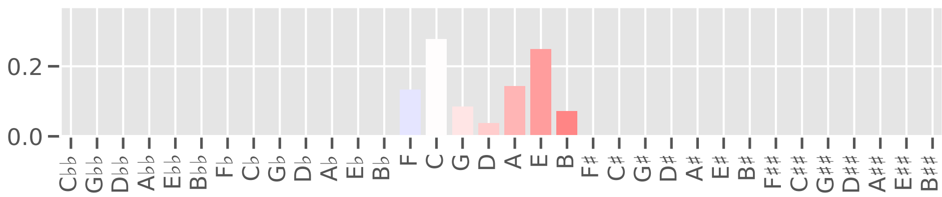
(a) Topic 1 of 12.



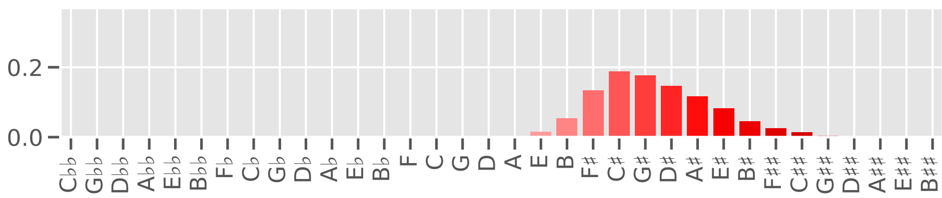
(b) Topic 2 of 12.



(c) Topic 3 of 12.

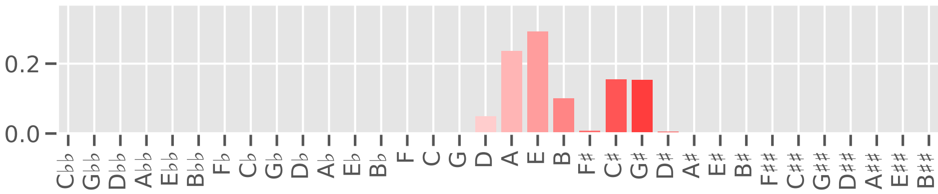


(d) Topic 4 of 12.

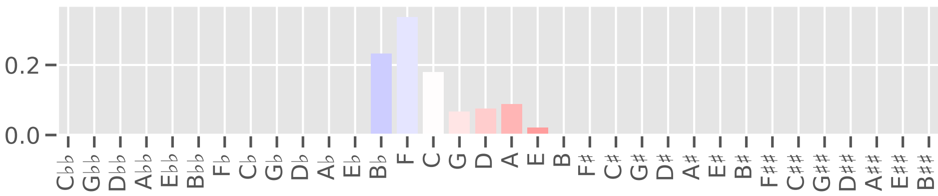


(e) Topic 5 of 12.

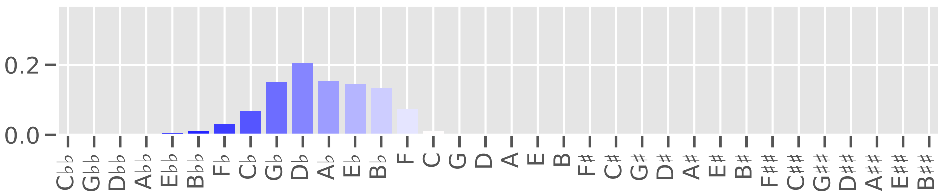
Figure A.7 – The note distributions for $K = 12$ topics.



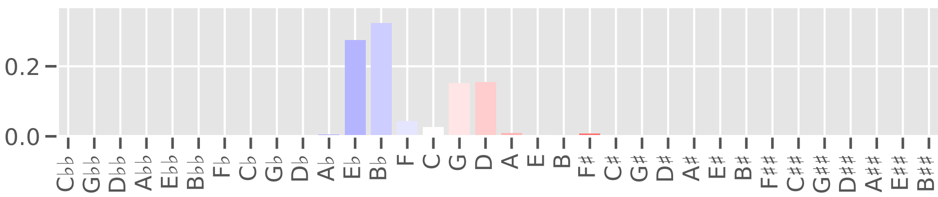
(f) Topic 6 of 12.



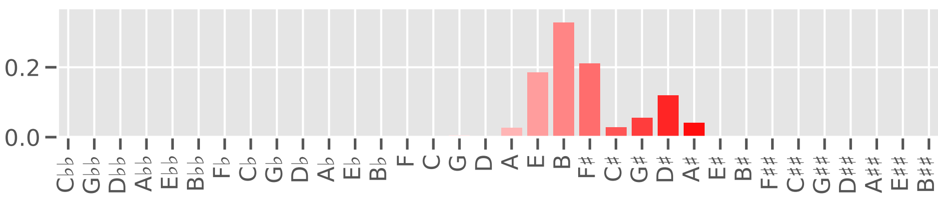
(g) Topic 7 of 12.



(h) Topic 8 of 12.

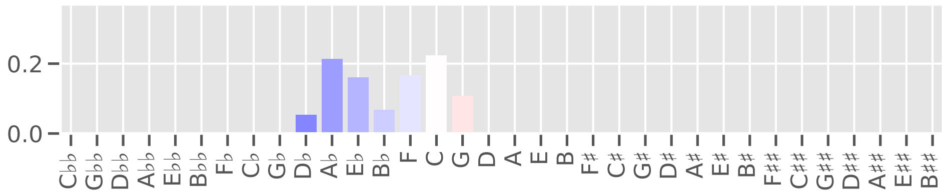


(i) Topic 9 of 12.

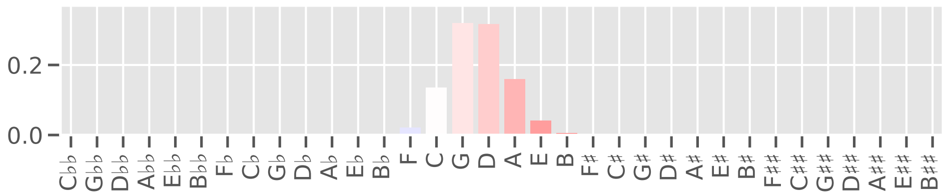


(j) Topic 10 of 12.

Figure A.7 – The note distributions for $K = 12$ topics (cont.).

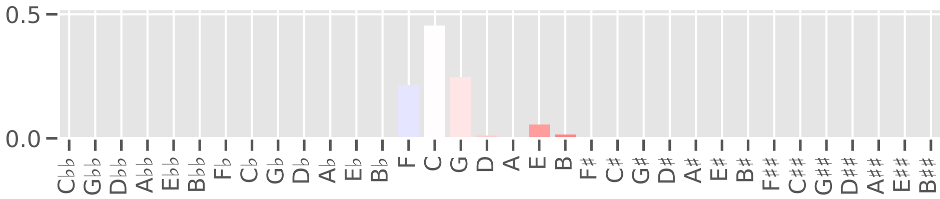


(k) Topic 11 of 12.

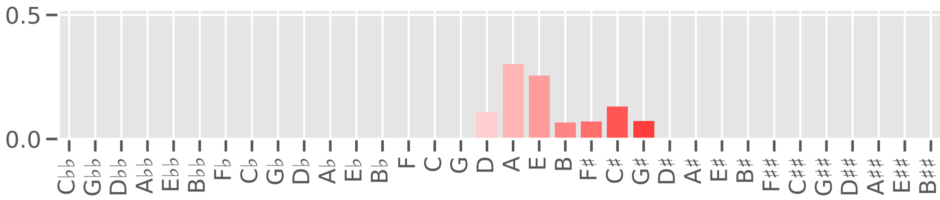


(l) Topic 12 of 12.

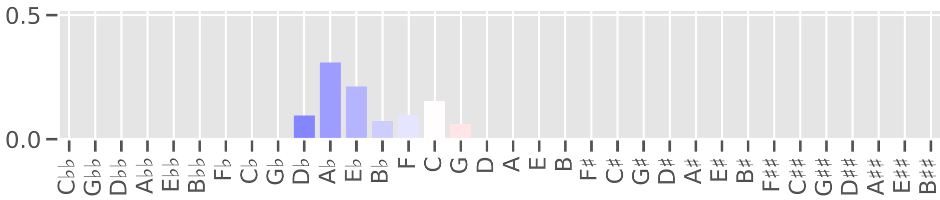
Figure A.7 – The note distributions for $K = 12$ topics (cont.).



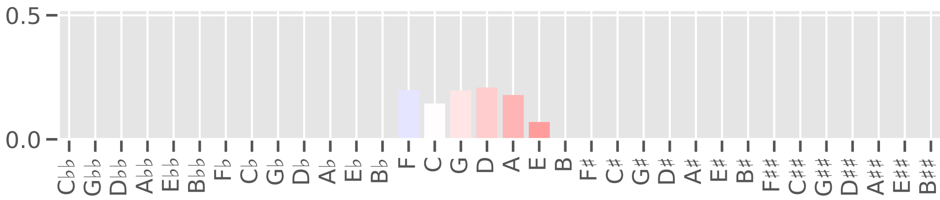
(a) Topic 1 of 24.



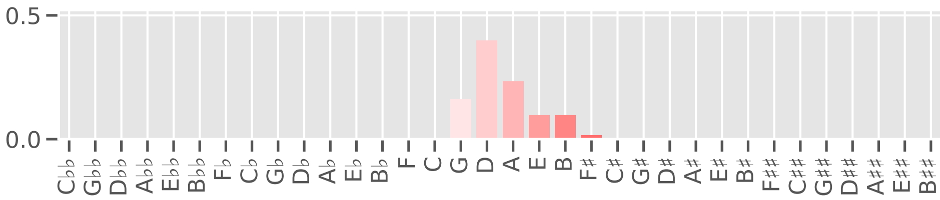
(b) Topic 2 of 24.



(c) Topic 3 of 24.

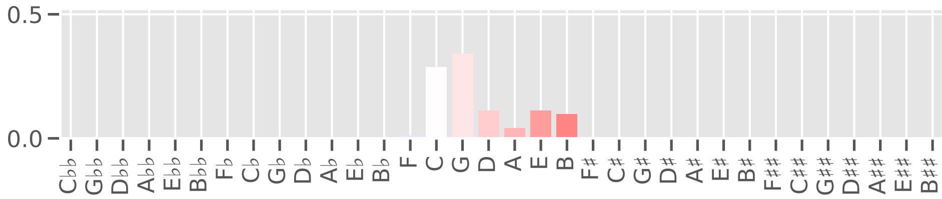


(d) Topic 4 of 24.

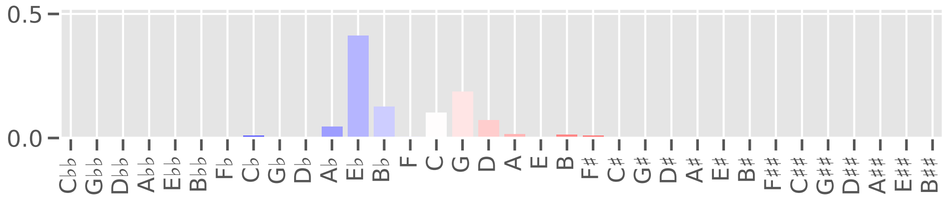


(e) Topic 5 of 24.

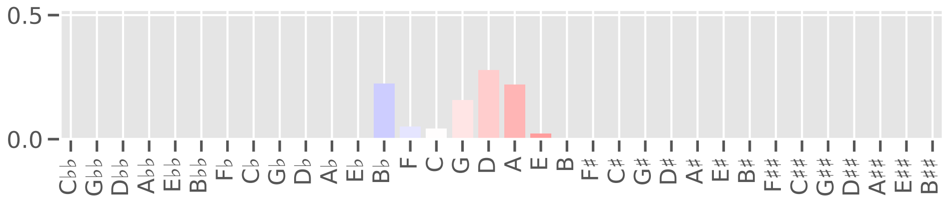
Figure A.8 – The note distributions for $K = 24$ topics.



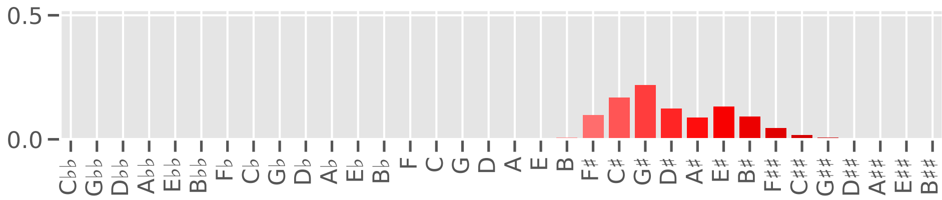
(f) Topic 6 of 24.



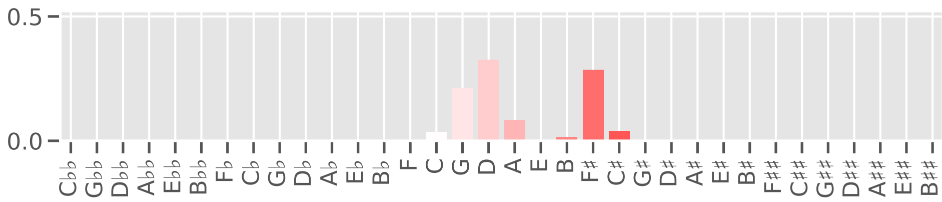
(g) Topic 7 of 24.



(h) Topic 8 of 24.

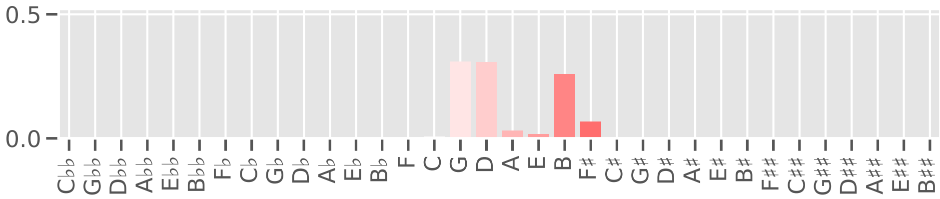


(i) Topic 9 of 24.

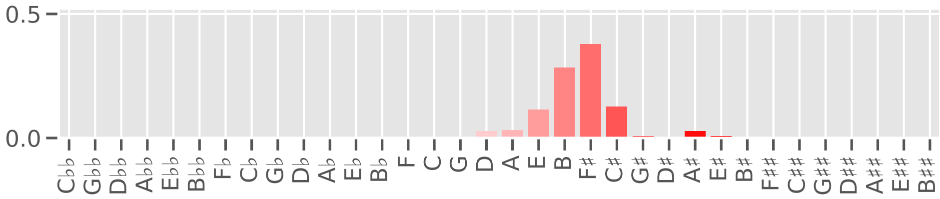


(j) Topic 10 of 24.

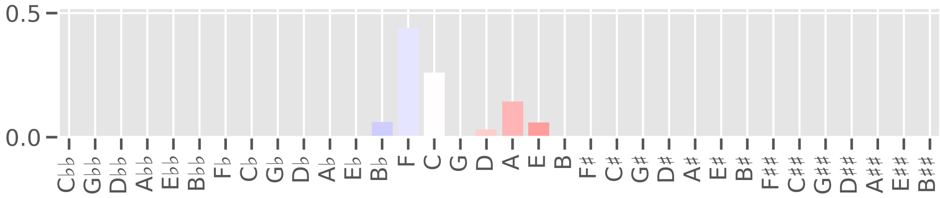
Figure A.8 – The note distributions for $K = 24$ topics (cont.).



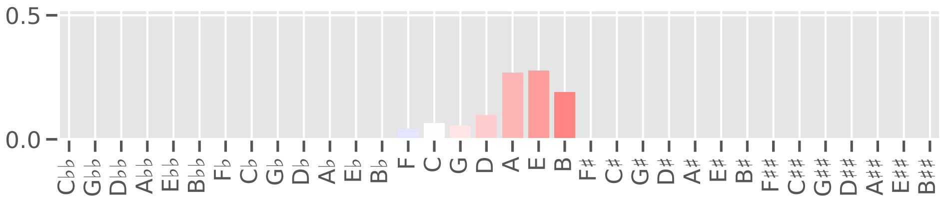
(k) Topic 11 of 24.



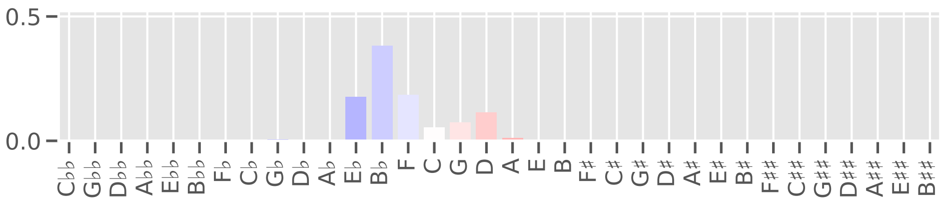
(l) Topic 12 of 24.



(m) Topic 13 of 24.

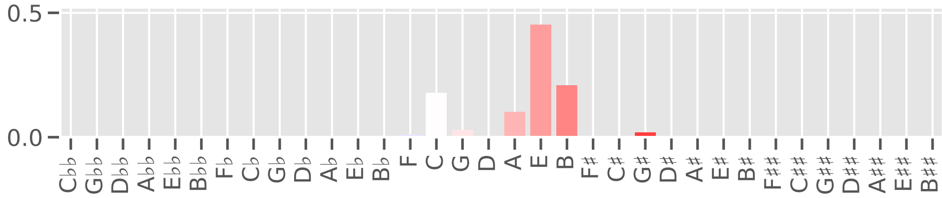


(n) Topic 14 of 24.

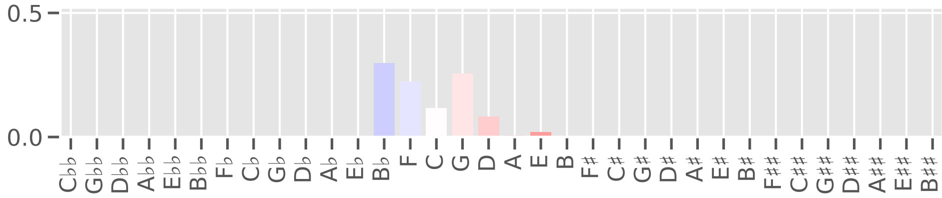


(o) Topic 15 of 24.

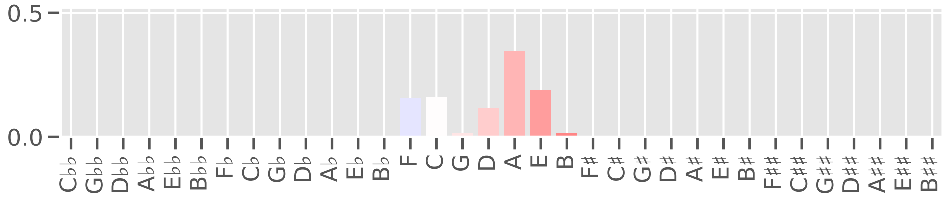
Figure A.8 – The note distributions for $K = 24$ topics (cont.).



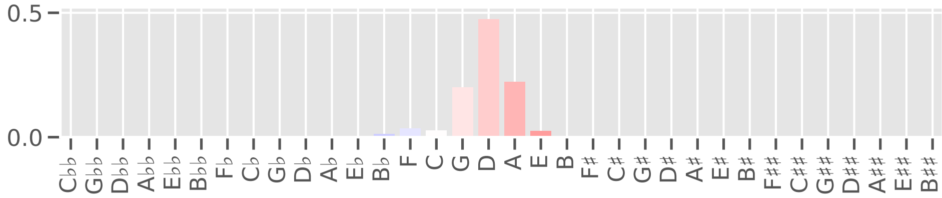
(p) Topic 16 of 24.



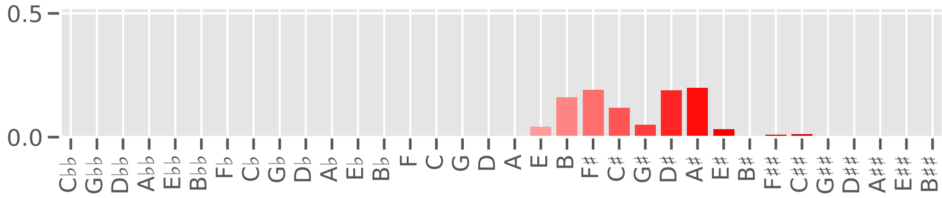
(q) Topic 17 of 24.



(r) Topic 18 of 24.

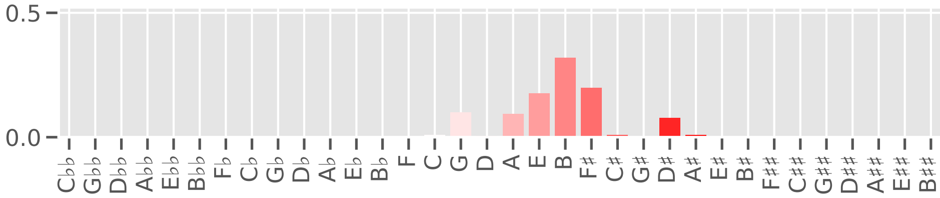


(s) Topic 19 of 24.

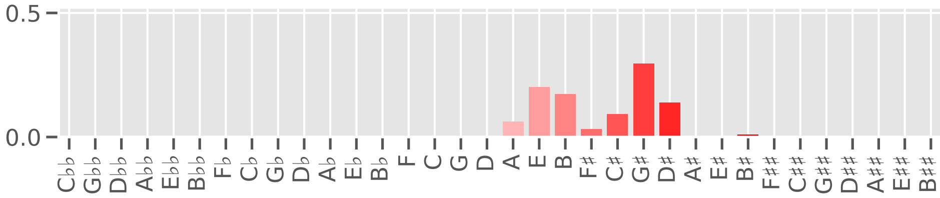


(t) Topic 20 of 24.

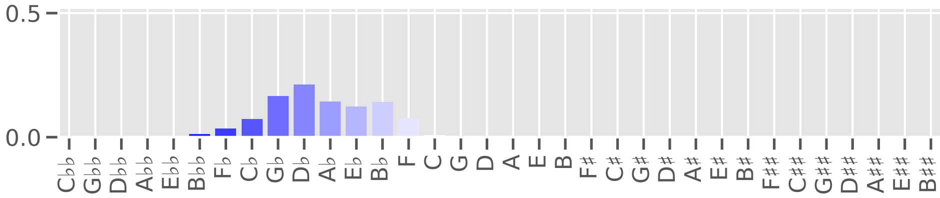
Figure A.8 – The note distributions for $K = 24$ topics (cont.).



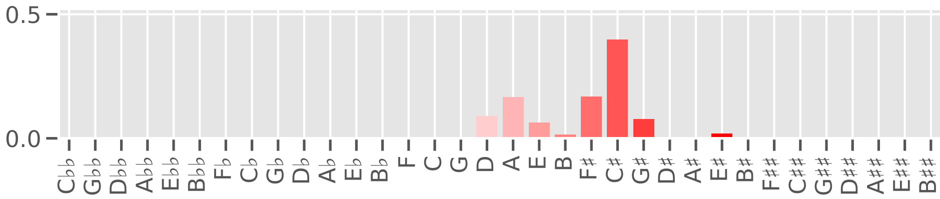
(u) Topic 21 of 24.



(v) Topic 22 of 24.



(w) Topic 23 of 24.



(x) Topic 24 of 24.

Figure A.8 – The note distributions for $K = 24$ topics (cont.).

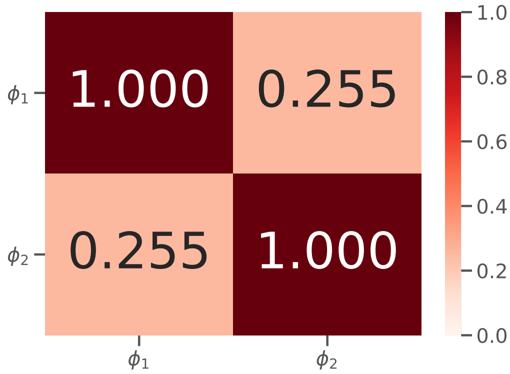


Figure A.9 – Jensen-Shannon similarities for $K = 2$ topics.

Figure A.10 – Jensen-Shannon similarities for $K = 3$ topics.

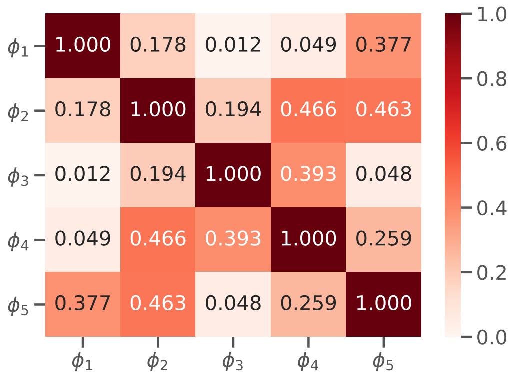


Figure A.11 – Jensen-Shannon similarities for $K = 5$ topics.

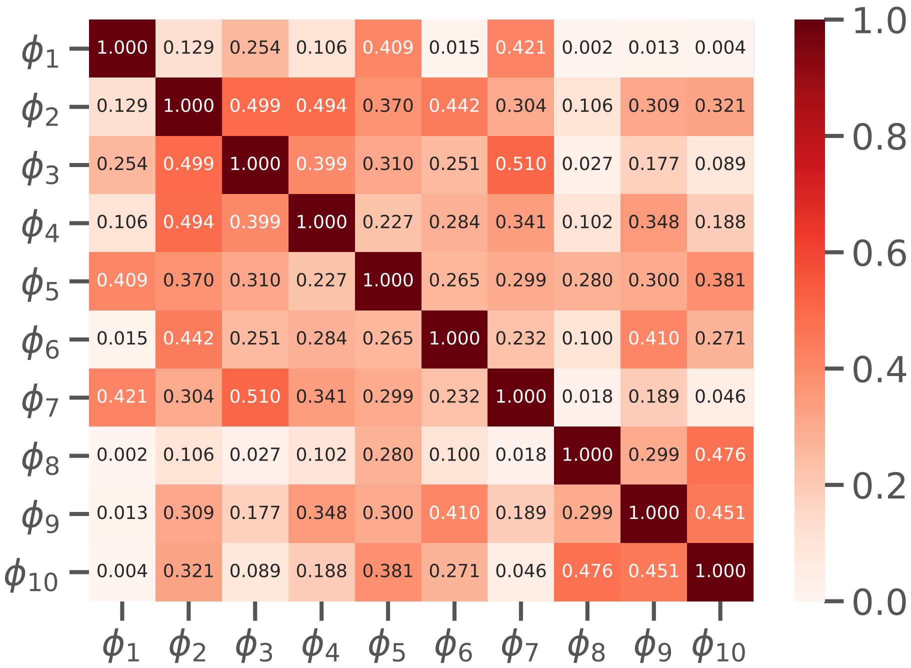


Figure A.12 – Jensen-Shannon similarities for $K = 10$ topics.

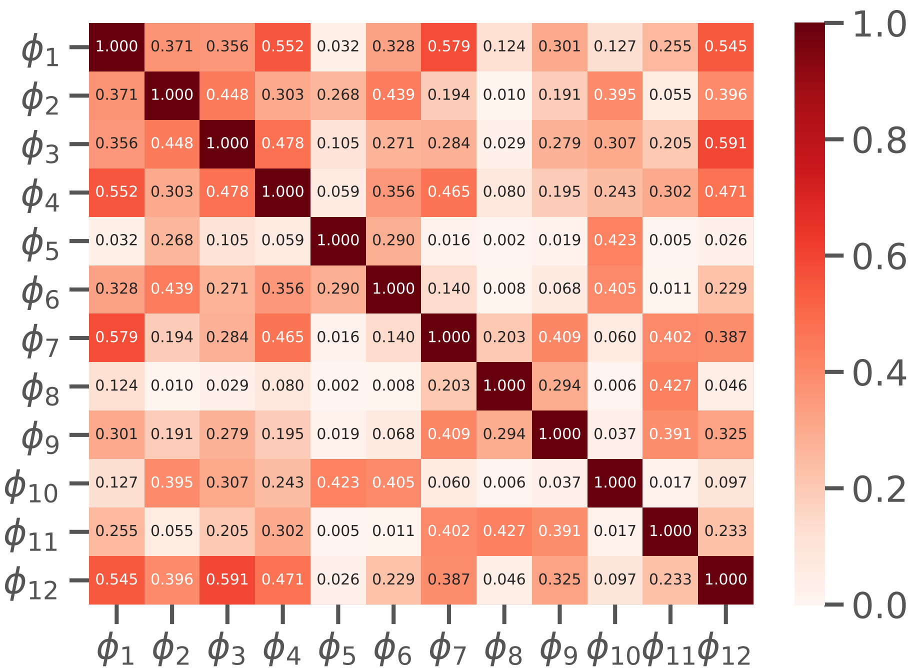


Figure A.13 – Jensen-Shannon similarities for $K = 12$ topics.

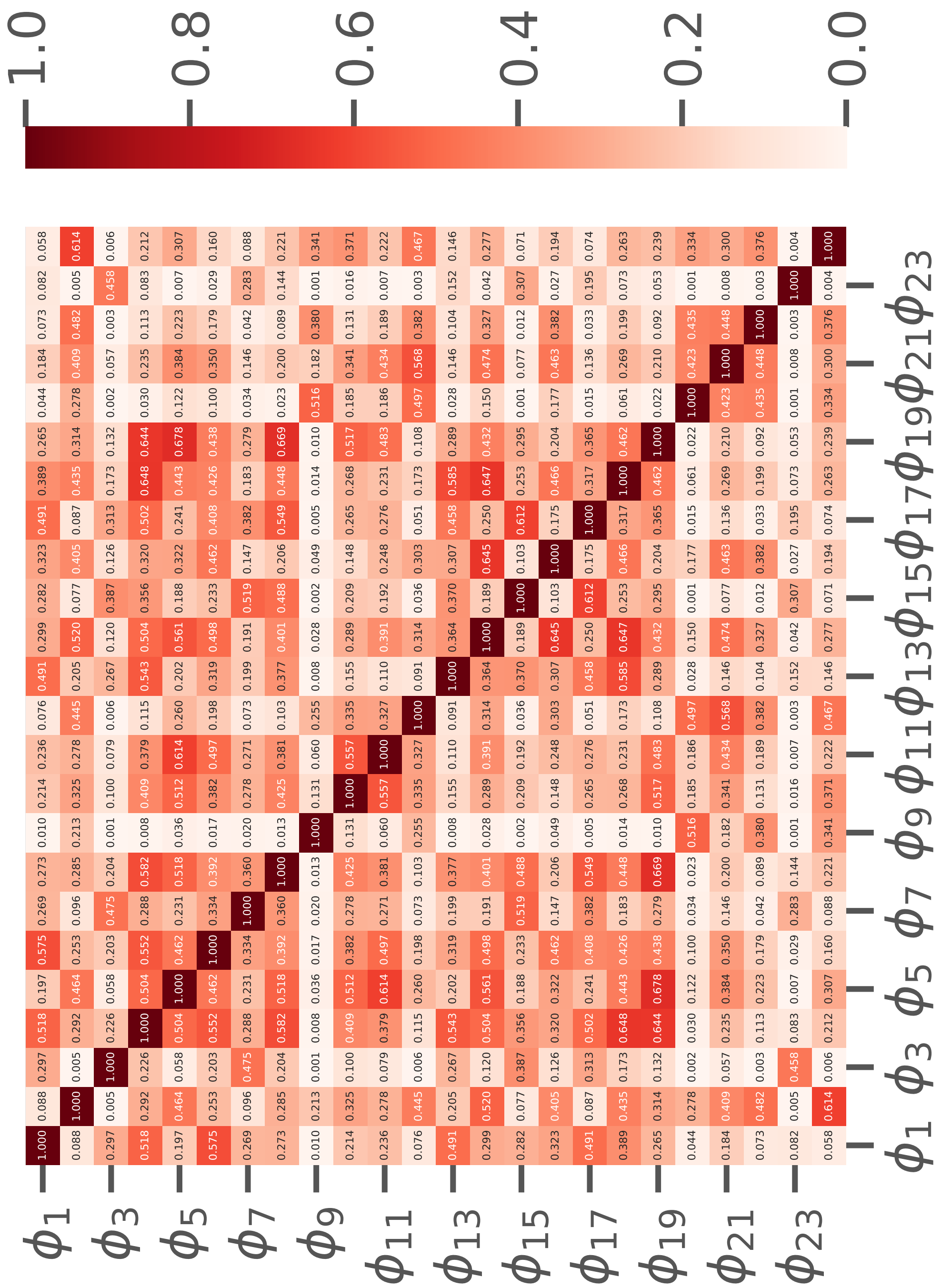


Figure A.14 – Jensen-Shannon similarities for $K = 24$ topics.

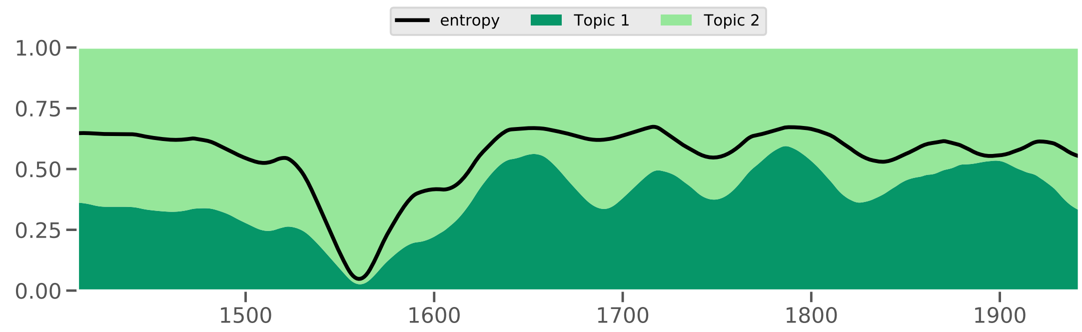


Figure A.15 – Topic evolution for $K = 2$ topics.

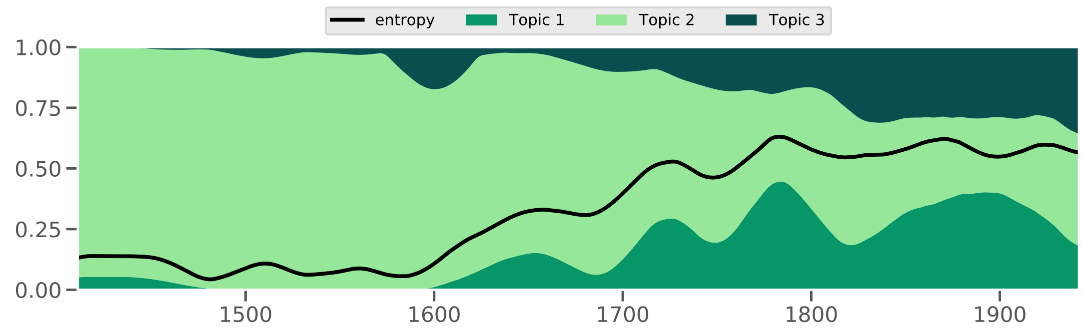


Figure A.16 – Topic evolution for $K = 3$ topics.

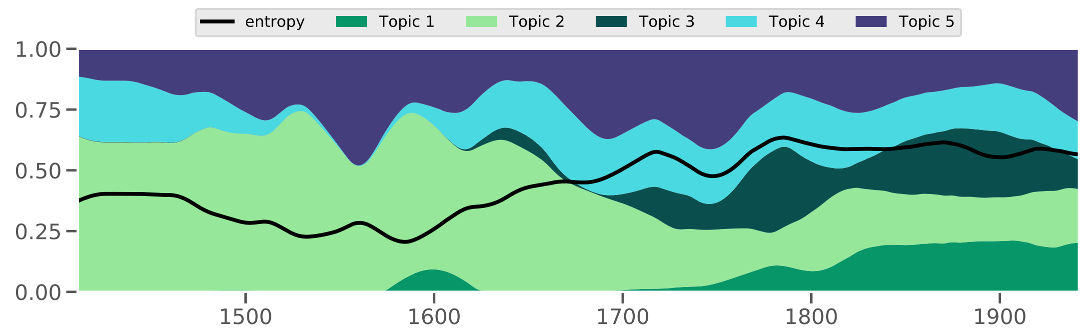


Figure A.17 – Topic evolution for $K = 5$ topics.

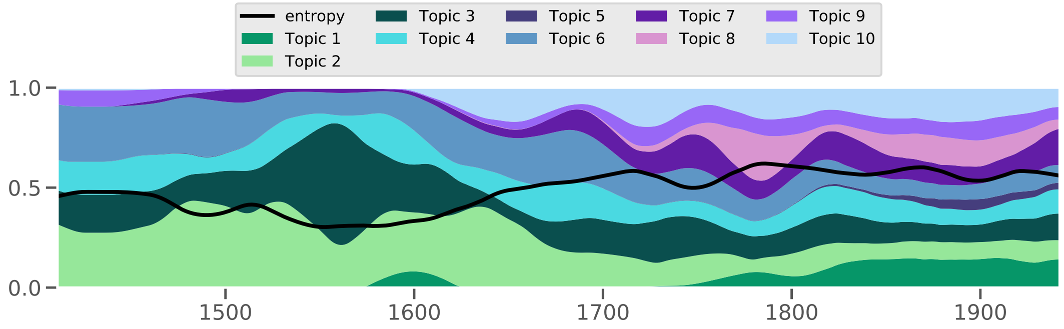


Figure A.18 – Topic evolution for $K = 10$ topics.

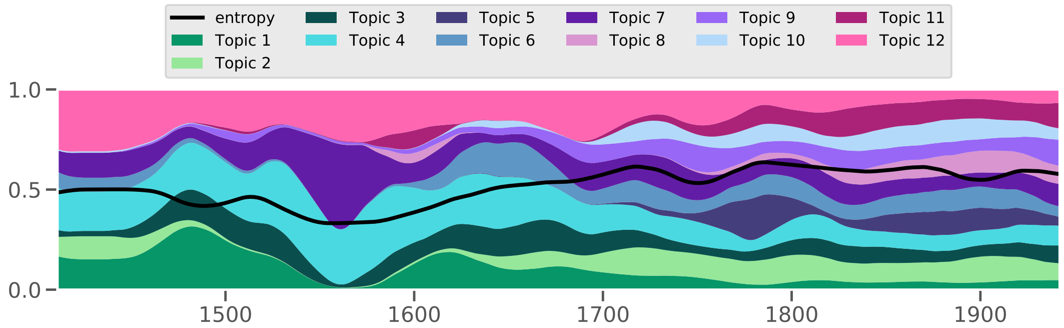


Figure A.19 – Topic evolution for $K = 12$ topics.

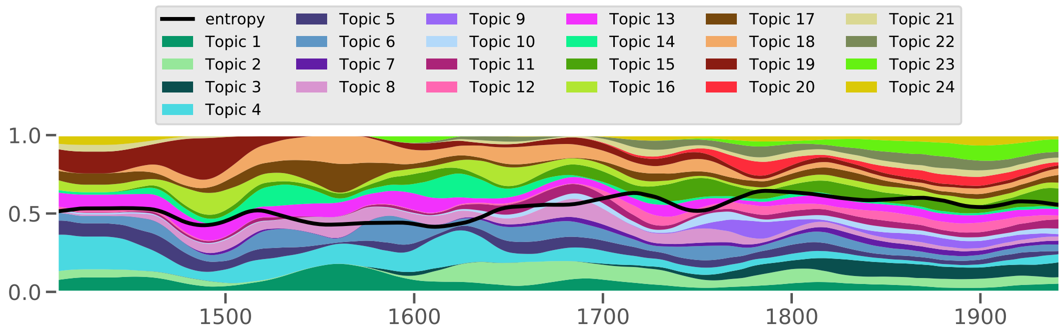


Figure A.20 – Topic evolution for $K = 24$ topics.

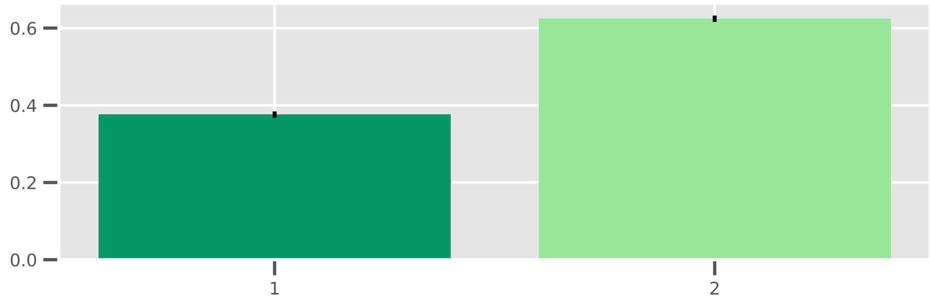


Figure A.21 – Average distribution of topics for all documents in the XML corpus for $K = 2$ topics. Error bars show the standard error of the mean.

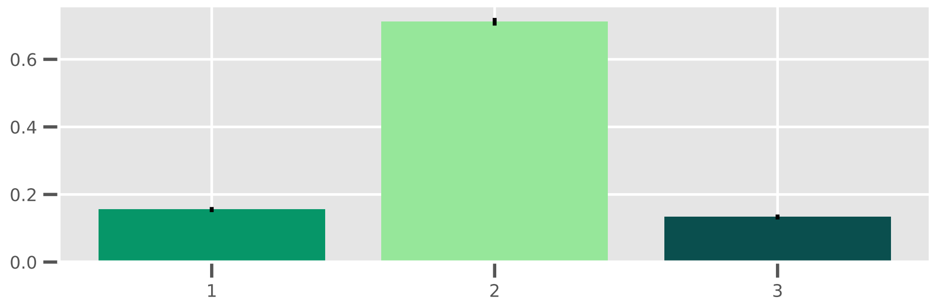


Figure A.22 – Average distribution of topics for all documents in the XML corpus for $K = 3$ topics. Error bars show the standard error of the mean.

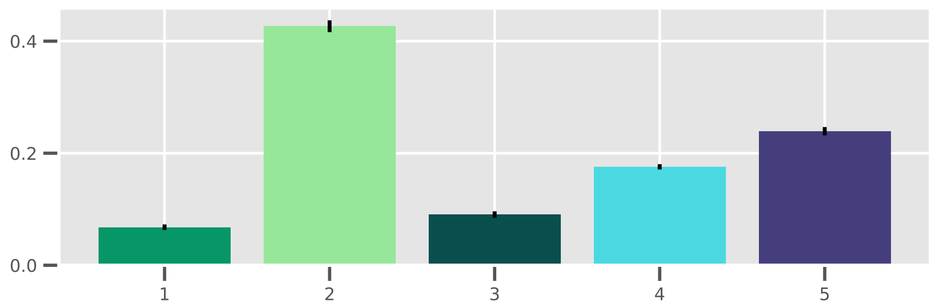


Figure A.23 – Average distribution of topics for all documents in the XML corpus for $K = 5$ topics. Error bars show the standard error of the mean.

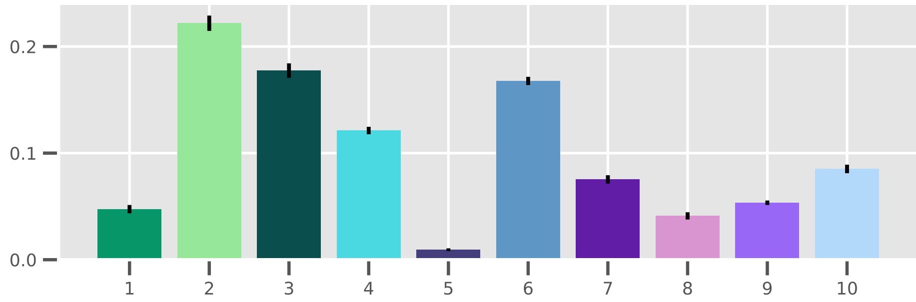


Figure A.24 – Average distribution of topics for all documents in the XML corpus for $K = 10$ topics. Error bars show the standard error of the mean.

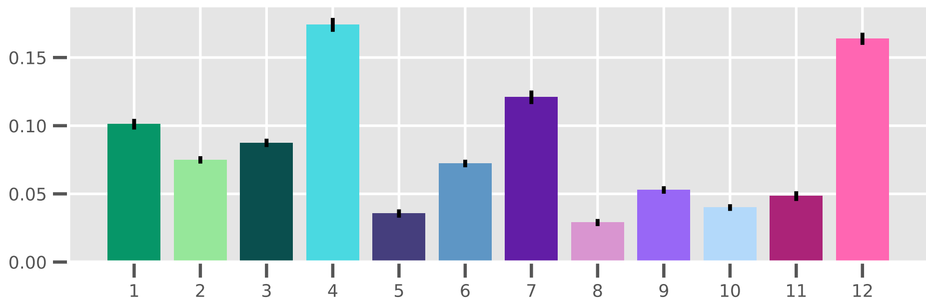


Figure A.25 – Average distribution of topics for all documents in the XML corpus for $K = 12$ topics. Error bars show the standard error of the mean.

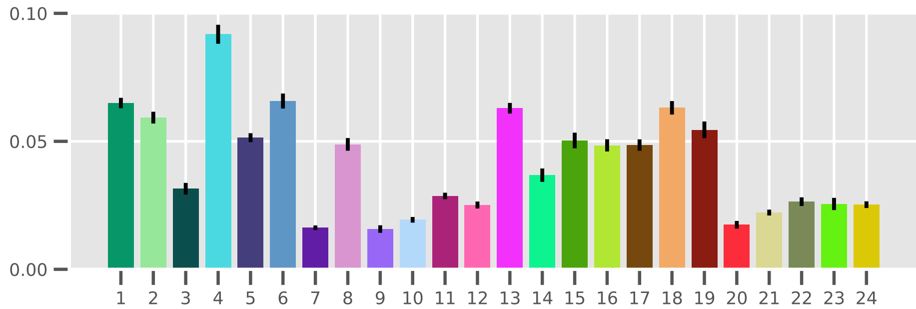


Figure A.26 – Average distribution of topics for all documents in the XML corpus for $K = 24$ topics. Error bars show the standard error of the mean.

B Tables

Table B.1 – Counts of mentioned works per composer in Meyer (1989), Harrison (1994), Kopp (2002), and Cohn (2012).

Composer	Meyer (1989)	Harrison (1994)	Kopp (2002)	Cohn (2012)
Arensky	1	–	–	–
Babbit	–	–	–	1
Bach, CPE	–	–	–	1
Bach, JS	4	2	–	3
Bartók	1	–	–	–
Beethoven	8	4	9	8
Benda	–	–	–	1
Berg	–	–	–	1
Berlioz	–	–	–	–
Brahms	6	2	4	7
Bruch	–	–	–	–
Bruckner	–	1	–	2
Busoni	–	4	–	–
Charpentier	–	–	–	–
Chausson	–	–	1	–
Chopin	7	1	11	10
Cornelius	–	–	–	1
Debussy	1	–	–	1
Distler	–	–	–	1
Donizetti	1	–	–	–
Dvořák	–	1	2	1
Elgar	–	–	–	1
Fauré	–	–	–	1
Franck	1	2	–	1
Geminiani	1	–	–	–
Gernsheim	–	–	–	–
Goldmark	–	–	–	–
Grieg	–	1	–	1
Händel	3	–	–	–
Haydn	9	–	–	2
Hoffmann	–	–	–	1
Karg-Elert	–	2	–	1
Kodály	–	–	–	1
Liszt	2	–	1	11
Mahler	3	1	–	1
Mendelssohn	1	–	–	1
Meyerbeer	–	–	–	–
Monteverdi	–	–	–	1
Mozart	18	1	–	3
Mussorgsky	–	–	–	1
Pfitzner	–	1	–	–
Prokofiev	–	–	–	1
Puccini	–	–	–	1
Raff	–	–	–	–
Reger	–	9	–	–
Reicha	1	–	–	–
Rimsky-Korsakov	1	–	–	2
Rossini	–	–	–	–
Sammartini	1	–	–	–
Schönberg	1	1	–	–
Schreker	–	1	–	1
Schubert	3	3	14	24
Schumann	3	–	3	3

Continued on next page

Table B.1 – Counts of mentioned works per composer in Meyer (1989), Harrison (1994), Kopp (2002), and Cohn (2012) (cont.).

Composer	Meyer (1989)	Harrison (1994)	Kopp (2002)	Cohn (2012)
Scriabin	–	1	–	1
Smetana	–	–	1	–
Strauss	1	9	–	2
Stravinski	3	–	–	–
Tallis	–	–	–	1
Tchaikovsky	1	–	–	1
Verdi	4	–	–	1
Wagner	–	2	3	6
Weber	–	–	–	–
Wolf	–	1	3	1

Table B.2 – Corpus used in Part III (“Mesoanalysis”).

#	Name	Year	Title	Source
0	Beethoven, L. van	1799	String quartets, op. 18, No. 1, mov. 1	ABC
1	Beethoven, L. van	1799	String quartets, op. 18, No. 1, mov. 2	ABC
2	Beethoven, L. van	1799	String quartets, op. 18, No. 1, mov. 3	ABC
3	Beethoven, L. van	1799	String quartets, op. 18, No. 1, mov. 4	ABC
4	Beethoven, L. van	1799	String quartets, op. 18, No. 2, mov. 1	ABC
5	Beethoven, L. van	1799	String quartets, op. 18, No. 2, mov. 2	ABC
6	Beethoven, L. van	1799	String quartets, op. 18, No. 2, mov. 3	ABC
7	Beethoven, L. van	1799	String quartets, op. 18, No. 2, mov. 4	ABC
8	Beethoven, L. van	1799	String quartets, op. 18, No. 3, mov. 1	ABC
9	Beethoven, L. van	1799	String quartets, op. 18, No. 3, mov. 2	ABC
10	Beethoven, L. van	1799	String quartets, op. 18, No. 3, mov. 3	ABC
11	Beethoven, L. van	1799	String quartets, op. 18, No. 3, mov. 4	ABC
12	Beethoven, L. van	1799	String quartets, op. 18, No. 4, mov. 1	ABC
13	Beethoven, L. van	1799	String quartets, op. 18, No. 4, mov. 2	ABC
14	Beethoven, L. van	1799	String quartets, op. 18, No. 4, mov. 3	ABC
15	Beethoven, L. van	1799	String quartets, op. 18, No. 4, mov. 4	ABC
16	Beethoven, L. van	1799	String quartets, op. 18, No. 5, mov. 1	ABC
17	Beethoven, L. van	1799	String quartets, op. 18, No. 5, mov. 2	ABC
18	Beethoven, L. van	1799	String quartets, op. 18, No. 5, mov. 3	ABC
19	Beethoven, L. van	1799	String quartets, op. 18, No. 5, mov. 4	ABC
20	Beethoven, L. van	1800	String quartets, op. 18, No. 5, mov. 1	ABC
21	Beethoven, L. van	1800	String quartets, op. 18, No. 5, mov. 2	ABC
22	Beethoven, L. van	1800	String quartets, op. 18, No. 5, mov. 3	ABC
23	Beethoven, L. van	1800	String quartets, op. 18, No. 5, mov. 4	ABC
24	Beethoven, L. van	1806	String quartets, op. 59, No. 1, mov. 1	ABC
25	Beethoven, L. van	1806	String quartets, op. 59, No. 1, mov. 2	ABC
26	Beethoven, L. van	1806	String quartets, op. 59, No. 1, mov. 3	ABC
27	Beethoven, L. van	1806	String quartets, op. 59, No. 1, mov. 4	ABC
28	Beethoven, L. van	1806	String quartets, op. 59, No. 2, mov. 1	ABC
29	Beethoven, L. van	1806	String quartets, op. 59, No. 2, mov. 2	ABC
30	Beethoven, L. van	1806	String quartets, op. 59, No. 2, mov. 3	ABC
31	Beethoven, L. van	1806	String quartets, op. 59, No. 2, mov. 4	ABC
32	Beethoven, L. van	1806	String quartets, op. 59, No. 3, mov. 1	ABC
33	Beethoven, L. van	1806	String quartets, op. 59, No. 3, mov. 2	ABC
34	Beethoven, L. van	1806	String quartets, op. 59, No. 3, mov. 3	ABC
35	Beethoven, L. van	1806	String quartets, op. 59, No. 3, mov. 4	ABC
36	Beethoven, L. van	1809	String quartets, op. 74, mov. 1	ABC
37	Beethoven, L. van	1809	String quartets, op. 74, mov. 2	ABC
38	Beethoven, L. van	1809	String quartets, op. 74, mov. 3	ABC
39	Beethoven, L. van	1809	String quartets, op. 74, mov. 4	ABC
40	Beethoven, L. van	1810	String quartets, op. 95, mov. 1	ABC
41	Beethoven, L. van	1810	String quartets, op. 95, mov. 2	ABC
42	Beethoven, L. van	1810	String quartets, op. 95, mov. 3	ABC
43	Beethoven, L. van	1810	String quartets, op. 95, mov. 4	ABC
44	Beethoven, L. van	1823	String quartets, op. 127, mov. 1	ABC
45	Beethoven, L. van	1823	String quartets, op. 127, mov. 2	ABC
46	Beethoven, L. van	1823	String quartets, op. 127, mov. 3	ABC
47	Beethoven, L. van	1823	String quartets, op. 127, mov. 4	ABC
48	Beethoven, L. van	1825	String quartets, op. 130, mov. 1	ABC
49	Beethoven, L. van	1825	String quartets, op. 130, mov. 2	ABC
50	Beethoven, L. van	1825	String quartets, op. 130, mov. 3	ABC
51	Beethoven, L. van	1825	String quartets, op. 130, mov. 4	ABC
52	Beethoven, L. van	1825	String quartets, op. 130, mov. 5	ABC
53	Beethoven, L. van	1825	String quartets, op. 130, mov. 6	ABC

Continued on next page

Table B.2 – Corpus used in Part III ("Mesoanalysis"; cont.).

#	Name	Year	Title	Source
54	Beethoven, L. van	1825	String quartets, op. 132, mov. 1	ABC
55	Beethoven, L. van	1825	String quartets, op. 132, mov. 2	ABC
56	Beethoven, L. van	1825	String quartets, op. 132, mov. 3	ABC
57	Beethoven, L. van	1825	String quartets, op. 132, mov. 4	ABC
58	Beethoven, L. van	1825	String quartets, op. 132, mov. 5	ABC
59	Beethoven, L. van	1826	String quartets, op. 131, mov. 1	ABC
60	Beethoven, L. van	1826	String quartets, op. 131, mov. 2	ABC
61	Beethoven, L. van	1826	String quartets, op. 131, mov. 3	ABC
62	Beethoven, L. van	1826	String quartets, op. 131, mov. 4	ABC
63	Beethoven, L. van	1826	String quartets, op. 131, mov. 5	ABC
64	Beethoven, L. van	1826	String quartets, op. 131, mov. 6	ABC
65	Beethoven, L. van	1826	String quartets, op. 131, mov. 7	ABC
66	Beethoven, L. van	1826	String quartets, op. 135, mov. 1	ABC
67	Beethoven, L. van	1826	String quartets, op. 135, mov. 2	ABC
68	Beethoven, L. van	1826	String quartets, op. 135, mov. 3	ABC
69	Beethoven, L. van	1826	String quartets, op. 135, mov. 4	ABC
70	Chopin, F.	1827	Mazurkas, op. 68, No. 2	CCARH
71	Chopin, F.	1829	Mazurkas, op. 68, No. 1	CCARH
72	Chopin, F.	1829	Mazurkas, op. 68, No. 3	CCARH
73	Chopin, F.	1830	Mazurkas, op. 6, No. 1	CCARH
74	Chopin, F.	1830	Mazurkas, op. 6, No. 2	CCARH
75	Chopin, F.	1830	Mazurkas, op. 6, No. 3	CCARH
76	Chopin, F.	1830	Mazurkas, op. 6, No. 4	CCARH
77	Chopin, F.	1831	Mazurkas, op. 7, No. 1	CCARH
78	Chopin, F.	1831	Mazurkas, op. 7, No. 2	CCARH
79	Chopin, F.	1831	Mazurkas, op. 7, No. 3	CCARH
80	Chopin, F.	1831	Mazurkas, op. 7, No. 4	CCARH
81	Chopin, F.	1831	Mazurkas, op. 7, No. 5	CCARH
82	Chopin, F.	1832	Mazurkas, BI 71	CCARH
83	Chopin, F.	1832	Mazurkas, BI 73	CCARH
84	Chopin, F.	1833	Mazurkas, op. 17, No. 1	CCARH
85	Chopin, F.	1833	Mazurkas, op. 17, No. 2	CCARH
86	Chopin, F.	1833	Mazurkas, op. 17, No. 3	CCARH
87	Chopin, F.	1833	Mazurkas, op. 17, No. 4	CCARH
88	Chopin, F.	1833	Mazurkas, BI 82	CCARH
89	Chopin, F.	1834	Mazurkas, BI 85	CCARH
90	Chopin, F.	1835	Mazurkas, op. 24, No. 1	CCARH
91	Chopin, F.	1835	Mazurkas, op. 24, No. 2	CCARH
92	Chopin, F.	1835	Mazurkas, op. 24, No. 3	CCARH
93	Chopin, F.	1835	Mazurkas, op. 24, No. 4	CCARH
94	Chopin, F.	1835	Mazurkas, op. 67, No. 1	CCARH
95	Chopin, F.	1835	Mazurkas, op. 67, No. 3	CCARH
96	Chopin, F.	1836	Mazurkas, op. 30, No. 1	CCARH
97	Chopin, F.	1836	Mazurkas, op. 30, No. 2	CCARH
98	Chopin, F.	1836	Mazurkas, op. 30, No. 3	CCARH
99	Chopin, F.	1836	Mazurkas, op. 30, No. 4	CCARH
100	Chopin, F.	1838	Mazurkas, op. 33, No. 1	CCARH
101	Chopin, F.	1838	Mazurkas, op. 33, No. 2	CCARH
102	Chopin, F.	1838	Mazurkas, op. 33, No. 3	CCARH
103	Chopin, F.	1838	Mazurkas, op. 33, No. 4	CCARH
104	Chopin, F.	1838	Mazurkas, op. 41, No. 1	CCARH
105	Chopin, F.	1839	Mazurkas, op. 41, No. 2	CCARH
106	Chopin, F.	1839	Mazurkas, op. 41, No. 3	CCARH
107	Chopin, F.	1839	Mazurkas, op. 41, No. 4	CCARH

Continued on next page

Table B.2 – Corpus used in Part III ("Mesoanalysis"; cont.).

#	Name	Year	Title	Source
108	Chopin, F.	1840	Mazurkas, BI 134	CCARH
109	Chopin, F.	1840	Mazurkas, BI 140	CCARH
110	Chopin, F.	1842	Mazurkas, op. 50, No. 1	CCARH
111	Chopin, F.	1842	Mazurkas, op. 50, No. 2	CCARH
112	Chopin, F.	1842	Mazurkas, op. 50, No. 3	CCARH
113	Chopin, F.	1843	Mazurkas, op. 56, No. 1	CCARH
114	Chopin, F.	1843	Mazurkas, op. 56, No. 2	CCARH
115	Chopin, F.	1843	Mazurkas, op. 56, No. 3	CCARH
116	Chopin, F.	1845	Mazurkas, op. 59, No. 1	CCARH
117	Chopin, F.	1845	Mazurkas, op. 59, No. 2	CCARH
118	Chopin, F.	1845	Mazurkas, op. 59, No. 3	CCARH
119	Chopin, F.	1846	Mazurkas, op. 63, No. 1	CCARH
120	Chopin, F.	1846	Mazurkas, op. 63, No. 2	CCARH
121	Chopin, F.	1846	Mazurkas, op. 63, No. 3	CCARH
122	Chopin, F.	1846	Mazurkas, op. 67, No. 4	CCARH
123	Chopin, F.	1849	Mazurkas, op. 67, No. 2	CCARH
124	Chopin, F.	1849	Mazurkas, op. 68, No. 4	CCARH
125	Debussy, C.	1890	Suite bergamasque, No. 1	MS
126	Debussy, C.	1890	Suite bergamasque, No. 2	MS
127	Debussy, C.	1890	Suite bergamasque, No. 3	MS
128	Debussy, C.	1890	Suite bergamasque, No. 4	MS
129	Dvořák, A.	1870	Silhouettes, op. 8, No. 1	DCML
130	Dvořák, A.	1870	Silhouettes, op. 8, No. 2	DCML
131	Dvořák, A.	1870	Silhouettes, op. 8, No. 3	DCML
132	Dvořák, A.	1870	Silhouettes, op. 8, No. 4	DCML
133	Dvořák, A.	1870	Silhouettes, op. 8, No. 5	DCML
134	Dvořák, A.	1870	Silhouettes, op. 8, No. 6	DCML
135	Dvořák, A.	1870	Silhouettes, op. 8, No. 7	DCML
136	Dvořák, A.	1870	Silhouettes, op. 8, No. 8	DCML
137	Dvořák, A.	1870	Silhouettes, op. 8, No. 9	DCML
138	Dvořák, A.	1870	Silhouettes, op. 8, No. 10	DCML
139	Dvořák, A.	1870	Silhouettes, op. 8, No. 11	DCML
140	Dvořák, A.	1870	Silhouettes, op. 8, No. 12	DCML
141	Grieg, E.	1867	Lyrical Pieces, op. 12, No. 1, Arietta	DCML
142	Grieg, E.	1867	Lyrical Pieces, op. 12, No. 2, Vals	DCML
143	Grieg, E.	1867	Lyrical Pieces, op. 12, No. 3, Vektersang	DCML
144	Grieg, E.	1867	Lyrical Pieces, op. 12, No. 4, Alfedans	DCML
145	Grieg, E.	1867	Lyrical Pieces, op. 12, No. 5, Folkevise	DCML
146	Grieg, E.	1867	Lyrical Pieces, op. 12, No. 6, Norsk	DCML
147	Grieg, E.	1867	Lyrical Pieces, op. 12, No. 7, Albumblad	DCML
148	Grieg, E.	1867	Lyrical Pieces, op. 12, No. 8, Fedrelandssang	DCML
149	Grieg, E.	1883	Lyrical Pieces, op. 38, No. 1, Berceuse	DCML
150	Grieg, E.	1883	Lyrical Pieces, op. 38, No. 2, Folkevise	DCML
151	Grieg, E.	1883	Lyrical Pieces, op. 38, No. 3, Melodi	DCML
152	Grieg, E.	1883	Lyrical Pieces, op. 38, No. 4, Halling	DCML
153	Grieg, E.	1883	Lyrical Pieces, op. 38, No. 5, Springdans	DCML
154	Grieg, E.	1883	Lyrical Pieces, op. 38, No. 6, Elegi	DCML
155	Grieg, E.	1883	Lyrical Pieces, op. 38, No. 7, Vals	DCML
156	Grieg, E.	1883	Lyrical Pieces, op. 38, No. 8, Kanon	DCML
157	Grieg, E.	1886	Lyrical Pieces, op. 43, No. 1, Sommerfugl	DCML
158	Grieg, E.	1886	Lyrical Pieces, op. 43, No. 2, Ensom vandrer	DCML
159	Grieg, E.	1886	Lyrical Pieces, op. 43, No. 3, I hjemmet	DCML
160	Grieg, E.	1886	Lyrical Pieces, op. 43, No. 4, Liten fugl	DCML
161	Grieg, E.	1886	Lyrical Pieces, op. 43, No. 5, Erotikk	DCML

Continued on next page

Table B.2 – Corpus used in Part III ("Mesoanalysis"; cont.).

#	Name	Year	Title	Source
162	Grieg, E.	1886	Lyrical Pieces, op. 43, No. 6, Til våren	DCML
163	Grieg, E.	1888	Lyrical Pieces, op. 47, No. 1, Valse-Impromptu	DCML
164	Grieg, E.	1888	Lyrical Pieces, op. 47, No. 2, Albumblad	DCML
165	Grieg, E.	1888	Lyrical Pieces, op. 47, No. 3, Melodi	DCML
166	Grieg, E.	1888	Lyrical Pieces, op. 47, No. 4, Halling	DCML
167	Grieg, E.	1888	Lyrical Pieces, op. 47, No. 5, Melankoli	DCML
168	Grieg, E.	1888	Lyrical Pieces, op. 47, No. 6, Springdans	DCML
169	Grieg, E.	1888	Lyrical Pieces, op. 47, No. 7, Elegi	DCML
170	Grieg, E.	1891	Lyrical Pieces, op. 54, No. 1, Gjetergutt	DCML
171	Grieg, E.	1891	Lyrical Pieces, op. 54, No. 2, Gangar	DCML
172	Grieg, E.	1891	Lyrical Pieces, op. 54, No. 3, Trolltog	DCML
173	Grieg, E.	1891	Lyrical Pieces, op. 54, No. 4, Notturno	DCML
174	Grieg, E.	1891	Lyrical Pieces, op. 54, No. 5, Scherzo	DCML
175	Grieg, E.	1891	Lyrical Pieces, op. 54, No. 6, Klokkeklang	DCML
176	Grieg, E.	1893	Lyrical Pieces, op. 57, No. 1, Svundne dager	DCML
177	Grieg, E.	1893	Lyrical Pieces, op. 57, No. 2, Gade	DCML
178	Grieg, E.	1893	Lyrical Pieces, op. 57, No. 3, Illusjon	DCML
179	Grieg, E.	1893	Lyrical Pieces, op. 57, No. 4, Geheimniss	DCML
180	Grieg, E.	1893	Lyrical Pieces, op. 57, No. 5, Sie tanzt	DCML
181	Grieg, E.	1893	Lyrical Pieces, op. 57, No. 6, Heimweh	DCML
182	Grieg, E.	1895	Lyrical Pieces, op. 62, No. 1, Sylfide	DCML
183	Grieg, E.	1895	Lyrical Pieces, op. 62, No. 2, Takk	DCML
184	Grieg, E.	1895	Lyrical Pieces, op. 62, No. 3, Fransk Serenade	DCML
185	Grieg, E.	1895	Lyrical Pieces, op. 62, No. 4, Bekken	DCML
186	Grieg, E.	1895	Lyrical Pieces, op. 62, No. 5, Drømmesyn	DCML
187	Grieg, E.	1895	Lyrical Pieces, op. 62, No. 6, Hjemad	DCML
188	Grieg, E.	1896	Lyrical Pieces, op. 65, No. 1, Fra ungomsdagene	DCML
189	Grieg, E.	1896	Lyrical Pieces, op. 65, No. 2, Bondens sang	DCML
190	Grieg, E.	1896	Lyrical Pieces, op. 65, No. 3, Tungsinn	DCML
191	Grieg, E.	1896	Lyrical Pieces, op. 65, No. 4, Salong	DCML
192	Grieg, E.	1896	Lyrical Pieces, op. 65, No. 5, I balladetone	DCML
193	Grieg, E.	1896	Lyrical Pieces, op. 65, No. 6, Bryllupsdag på Troldhaugen	DCML
194	Grieg, E.	1899	Lyrical Pieces, op. 68, No. 1, Matrosenes oppsang	DCML
195	Grieg, E.	1899	Lyrical Pieces, op. 68, No. 2, Bestemors menuet	DCML
196	Grieg, E.	1899	Lyrical Pieces, op. 68, No. 3, For dine føtter	DCML
197	Grieg, E.	1899	Lyrical Pieces, op. 68, No. 4, Aften på højfjellet	DCML
198	Grieg, E.	1899	Lyrical Pieces, op. 68, No. 5, Bådnlåt	DCML
199	Grieg, E.	1899	Lyrical Pieces, op. 68, No. 6, Valse mélancolique	DCML
200	Grieg, E.	1901	Lyrical Pieces, op. 71, No. 1, Det var engang	DCML
201	Grieg, E.	1901	Lyrical Pieces, op. 71, No. 2, Sommeraften	DCML
202	Grieg, E.	1901	Lyrical Pieces, op. 71, No. 3, Småtroll	DCML
203	Grieg, E.	1901	Lyrical Pieces, op. 71, No. 4, Skogstillhet	DCML
204	Grieg, E.	1901	Lyrical Pieces, op. 71, No. 5, Halling	DCML
205	Grieg, E.	1901	Lyrical Pieces, op. 71, No. 6, Forbi	DCML
206	Grieg, E.	1901	Lyrical Pieces, op. 71, No. 7, Efterklang	DCML
207	Liszt, F.	1855	Anées de pèlerinage, S. 160, No. 1, Sposalizio	MS
208	Liszt, F.	1855	Anées de pèlerinage, S. 160, No. 2, Il Penseroso	MS
209	Liszt, F.	1855	Anées de pèlerinage, S. 160, No. 3, Canzonetta del Salvator Rosa	MS
210	Liszt, F.	1855	Anées de pèlerinage, S. 160, No. 4, Sonetto 47 del Petrarca	MS
211	Liszt, F.	1855	Anées de pèlerinage, S. 160, No. 5, Sonetto 104 del Petrarca	MS
212	Liszt, F.	1855	Anées de pèlerinage, S. 160, No. 6, Sonetto 123 del Petrarca	MS
213	Liszt, F.	1855	Anées de pèlerinage, S. 160, No. 7, Après une lecture de Dante / Fantasia quasi Sonata	MS
214	Liszt, F.	1855	Anées de pèlerinage, S. 160, No. 8, Gondoliera	MS

Continued on next page

Table B.2 – Corpus used in Part III ("Mesoanalysis"; cont.).

#	Name	Year	Title	Source
215	Liszt, F.	1855	Anées de pèlerinage, S. 160, No. 9, Canzone	MS
216	Medtner, N.	1905	Fairy Tales, op. 8, No. 1	DCML
217	Medtner, N.	1905	Fairy Tales, op. 8, No. 2	DCML
218	Medtner, N.	1905	Fairy Tales, op. 9, No. 1	DCML
219	Medtner, N.	1905	Fairy Tales, op. 9, No. 2	DCML
220	Medtner, N.	1905	Fairy Tales, op. 9, No. 3	DCML
221	Medtner, N.	1907	Fairy Tales, op. 14, No. 1	DCML
222	Medtner, N.	1907	Fairy Tales, op. 14, No. 2	DCML
223	Medtner, N.	1909	Fairy Tales, op. 20, No. 1	DCML
224	Medtner, N.	1909	Fairy Tales, op. 20, No. 2	DCML
225	Medtner, N.	1912	Fairy Tales, op. 26, No. 1	DCML
226	Medtner, N.	1912	Fairy Tales, op. 26, No. 2	DCML
227	Medtner, N.	1912	Fairy Tales, op. 26, No. 3	DCML
228	Medtner, N.	1912	Fairy Tales, op. 26, No. 4	DCML
229	Medtner, N.	1914	Fairy Tales, op. 31, No. 3	DCML
230	Medtner, N.	1915	Fairy Tales, op. deest	DCML
231	Medtner, N.	1917	Fairy Tales, op. 34, No. 1	DCML
232	Medtner, N.	1917	Fairy Tales, op. 34, No. 2	DCML
233	Medtner, N.	1917	Fairy Tales, op. 34, No. 3	DCML
234	Medtner, N.	1917	Fairy Tales, op. 34, No. 4	DCML
235	Medtner, N.	1917	Fairy Tales, op. 35, No. 1	DCML
236	Medtner, N.	1917	Fairy Tales, op. 35, No. 2	DCML
237	Medtner, N.	1917	Fairy Tales, op. 35, No. 3	DCML
238	Medtner, N.	1917	Fairy Tales, op. 35, No. 4	DCML
239	Medtner, N.	1924	Fairy Tales, op. 42, No. 1	DCML
240	Medtner, N.	1924	Fairy Tales, op. 42, No. 2	DCML
241	Medtner, N.	1924	Fairy Tales, op. 42, No. 3	DCML
242	Medtner, N.	1925	Fairy Tales, op. 48, No. 1	DCML
243	Medtner, N.	1925	Fairy Tales, op. 48, No. 2	DCML
244	Medtner, N.	1928	Fairy Tales, op. 51, No. 1	DCML
245	Medtner, N.	1928	Fairy Tales, op. 51, No. 2	DCML
246	Medtner, N.	1928	Fairy Tales, op. 51, No. 3	DCML
247	Medtner, N.	1928	Fairy Tales, op. 51, No. 4	DCML
248	Medtner, N.	1928	Fairy Tales, op. 51, No. 5	DCML
249	Medtner, N.	1928	Fairy Tales, op. 51, No. 6	DCML
250	Medtner, N.	1932	Fairy Tales, op. 54, No. 2	DCML
251	Medtner, N.	1932	Fairy Tales, op. 54, No. 4	DCML
252	Medtner, N.	1932	Fairy Tales, op. 54, No. 6	DCML
253	Medtner, N.	1932	Fairy Tales, op. 54, No. 8	DCML
254	Schubert, F.	1827	Winterreise, op. 89, No. 1, Gute Nacht	OSLC
255	Schubert, F.	1827	Winterreise, op. 89, No. 2, Die Wetterfahne	OSLC
256	Schubert, F.	1827	Winterreise, op. 89, No. 3, Gefror'ne Thränen	OSLC
257	Schubert, F.	1827	Winterreise, op. 89, No. 4, Erstarrung	OSLC
258	Schubert, F.	1827	Winterreise, op. 89, No. 5, Der Lindenbaum	OSLC
259	Schubert, F.	1827	Winterreise, op. 89, No. 6, Wasserfluth	OSLC
260	Schubert, F.	1827	Winterreise, op. 89, No. 7, Auf dem Flusse	OSLC
261	Schubert, F.	1827	Winterreise, op. 89, No. 8, Rückblick	OSLC
262	Schubert, F.	1827	Winterreise, op. 89, No. 9, Irrlicht	OSLC
263	Schubert, F.	1827	Winterreise, op. 89, No. 10, Rast	OSLC
264	Schubert, F.	1827	Winterreise, op. 89, No. 11, Frühlingstraum	OSLC
265	Schubert, F.	1827	Winterreise, op. 89, No. 12, Einsamkeit	OSLC
266	Schubert, F.	1827	Winterreise, op. 89, No. 13, Die Post	OSLC
267	Schubert, F.	1827	Winterreise, op. 89, No. 14, Der greise Kopf	OSLC
268	Schubert, F.	1827	Winterreise, op. 89, No. 15, Die Krähe	OSLC

Continued on next page

Table B.2 – Corpus used in Part III ("Mesoanalysis"; cont.).

#	Name	Year	Title	Source
269	Schubert, F.	1827	Winterreise, op. 89, No. 16, Letzte Hoffnung	OSLC
270	Schubert, F.	1827	Winterreise, op. 89, No. 17, Im Dorfe	OSLC
271	Schubert, F.	1827	Winterreise, op. 89, No. 18, Der stürmische Morgen	OSLC
272	Schubert, F.	1827	Winterreise, op. 89, No. 19, Täuschung	OSLC
273	Schubert, F.	1827	Winterreise, op. 89, No. 20, Der Wegweiser	OSLC
274	Schubert, F.	1827	Winterreise, op. 89, No. 21, Das Wirtshaus	OSLC
275	Schubert, F.	1827	Winterreise, op. 89, No. 22, Muth	OSLC
276	Schubert, F.	1827	Winterreise, op. 89, No. 23, Die Nebensonnen	OSLC
277	Schubert, F.	1827	Winterreise, op. 89, No. 24, Der Leidermann	OSLC
278	Tchaikovsky, P. I.	1886	Seasons, op. 37b, No. 1, January: At the fireside	MS
279	Tchaikovsky, P. I.	1886	Seasons, op. 37b, No. 2, February: Carnival	MS
280	Tchaikovsky, P. I.	1886	Seasons, op. 37b, No. 3, March: Song of the Lark	MS
281	Tchaikovsky, P. I.	1886	Seasons, op. 37b, No. 4, April: Snowdrop	DCML
282	Tchaikovsky, P. I.	1886	Seasons, op. 37b, No. 5, May: Starlit Nights	DCML
283	Tchaikovsky, P. I.	1886	Seasons, op. 37b, No. 6, June: Barcarole	MS
284	Tchaikovsky, P. I.	1886	Seasons, op. 37b, No. 7, July: Song of the Reaper	DCML
285	Tchaikovsky, P. I.	1886	Seasons, op. 37b, No. 8, August: Harvest	DCML
286	Tchaikovsky, P. I.	1886	Seasons, op. 37b, No. 9, September: The Hunt	DCML
287	Tchaikovsky, P. I.	1886	Seasons, op. 37b, No. 10, October: Autumn Song	MS
288	Tchaikovsky, P. I.	1886	Seasons, op. 37b, No. 11, November: Troika	DCML
289	Tchaikovsky, P. I.	1886	Seasons, op. 37b, No. 12, December: Christmas	DCML

Table B.3 – Corpus used in Part IV ("Macroanalysis").

#	Name	Year	Title	Source
0	Agricola, A.	1506	Missa Malheur me bat, Kyrie	ELVIS
1	Agricola, A.	1506	Missa Malheur me bat, Gloria	ELVIS
2	Agricola, A.	1506	Missa Malheur me bat, Credo	ELVIS
3	Agricola, A.	1506	Missa Malheur me bat, Sanctus	ELVIS
4	Agricola, A.	1506	Missa Malheur me bat, Agnus Dei	ELVIS
5	Alkan, C. V.	1833	Op. 12a, No. 1, Rondo Chromatique	MS
6	Alkan, C. V.	1833	Trois Improvisations dans le Style Brillante, Op. 12b, No. 3, Une Improvisations in B minor	MS
7	Alkan, C. V.	1837	Trois Andantes Romantiques, Op. 13, mov. 2	MS
8	Alkan, C. V.	1837	Trois Andantes Romantiques, Op. 13, mov. 3	MS
9	Alkan, C. V.	1837	Trois morceaux dans le genre pathétique, Op. 15, No. 1, Aime-moi	MS
10	Alkan, C. V.	1837	Trois morceaux dans le genre pathétique, Op. 15, No. 2, Le Vent	MS
11	Alkan, C. V.	1837	Trois morceaux dans le genre pathétique, Op. 15, No. 3, Morte	MS
12	Alkan, C. V.	1837	Scherzi di bravoure , Op. 16, No. 3, Etude de Bravour	MS
13	Alkan, C. V.	1838	Trois Grandes Études, Op. 76, No. 1, Fantaisie	MS
14	Alkan, C. V.	1838	Trois Grandes Études, Op. 76, No. 2, Introduction, Variations et Finale	MS
15	Alkan, C. V.	1838	Trois Grandes Études, Op. 76, No. 3, Mouvement semblable et perpetuel	MS
16	Alkan, C. V.	1840	Un Morceau Caractéristique, Op. 74, No. 1	MS
17	Alkan, C. V.	1840	Un Morceau Caractéristique, Op. 74, No. 10	MS
18	Alkan, C. V.	1840	Un Morceau Caractéristique, Op. 74, No. 11	MS
19	Alkan, C. V.	1840	Un Morceau Caractéristique, Op. 74, No. 12	MS
20	Alkan, C. V.	1840	Un Morceau Caractéristique, Op. 74, No. 2	MS
21	Alkan, C. V.	1840	Un Morceau Caractéristique, Op. 74, No. 3	MS
22	Alkan, C. V.	1840	Un Morceau Caractéristique, Op. 74, No. 4	MS
23	Alkan, C. V.	1840	Un Morceau Caractéristique, Op. 74, No. 5	MS
24	Alkan, C. V.	1840	Un Morceau Caractéristique, Op. 74, No. 6	MS
25	Alkan, C. V.	1840	Un Morceau Caractéristique, Op. 74, No. 7	MS
26	Alkan, C. V.	1840	Un Morceau Caractéristique, Op. 74, No. 8	MS
27	Alkan, C. V.	1840	Un Morceau Caractéristique, Op. 74, No. 9	MS
28	Alkan, C. V.	1840	Étude sans Opus, Op.	MS
29	Alkan, C. V.	1840	Fugue, Op., Jean qui pleure	MS
30	Alkan, C. V.	1840	Fugue, Op., Jean qui rit	MS
31	Alkan, C. V.	1844	Étude de Concert, Op. 17, Le Preux	MS
32	Alkan, C. V.	1844	Op. 23, Sartarelle	MS
33	Alkan, C. V.	1844	Gigue et Air de Ballet, Op. 24, No. 1, Gigue	MS
34	Alkan, C. V.	1844	Gigue et Air de Ballet, Op. 24, No. 2.0	MS
35	Alkan, C. V.	1844	Op. 25, Alleluia	MS
36	Alkan, C. V.	1844	Op. 26, Marche Funèbre	MS
37	Alkan, C. V.	1844	Op. 27, Marche Triomphale	MS
38	Alkan, C. V.	1844	Op. 27, Le chemin de fer	MS
39	Alkan, C. V.	1845	Impromptus, Op. 32, No. 1, mov. 2, L'Amitié	MS
40	Alkan, C. V.	1846	Op. 29, Bourrée d'Auvergne	MS
41	Alkan, C. V.	1846	Préludes, Op. 31, No. 1	MS
42	Alkan, C. V.	1846	Préludes, Op. 31, No. 10	MS
43	Alkan, C. V.	1846	Préludes, Op. 31, No. 11	MS
44	Alkan, C. V.	1846	Préludes, Op. 31, No. 12	MS
45	Alkan, C. V.	1846	Préludes, Op. 31, No. 13	MS
46	Alkan, C. V.	1846	Préludes, Op. 31, No. 14	MS
47	Alkan, C. V.	1846	Préludes, Op. 31, No. 15	MS
48	Alkan, C. V.	1846	Préludes, Op. 31, No. 16	MS
49	Alkan, C. V.	1846	Préludes, Op. 31, No. 17	MS
50	Alkan, C. V.	1846	Préludes, Op. 31, No. 18	MS

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
51	Alkan, C. V.	1846	Préludes, Op. 31, No. 19	MS
52	Alkan, C. V.	1846	Préludes, Op. 31, No. 2	MS
53	Alkan, C. V.	1846	Préludes, Op. 31, No. 20	MS
54	Alkan, C. V.	1846	Préludes, Op. 31, No. 21	MS
55	Alkan, C. V.	1846	Préludes, Op. 31, No. 22	MS
56	Alkan, C. V.	1846	Préludes, Op. 31, No. 23	MS
57	Alkan, C. V.	1846	Préludes, Op. 31, No. 24	MS
58	Alkan, C. V.	1846	Préludes, Op. 31, No. 25	MS
59	Alkan, C. V.	1846	Préludes, Op. 31, No. 3	MS
60	Alkan, C. V.	1846	Préludes, Op. 31, No. 4	MS
61	Alkan, C. V.	1846	Préludes, Op. 31, No. 5	MS
62	Alkan, C. V.	1846	Préludes, Op. 31, No. 6	MS
63	Alkan, C. V.	1846	Préludes, Op. 31, No. 7	MS
64	Alkan, C. V.	1846	Préludes, Op. 31, No. 8	MS
65	Alkan, C. V.	1846	Préludes, Op. 31, No. 9	MS
66	Alkan, C. V.	1847	Impromptus, Op. 32, No. 1, mov. 1, Vaghezza	MS
67	Alkan, C. V.	1847	Impromptus, Op. 32, No. 1, mov. 3, Fantasietta alla Moresca	MS
68	Alkan, C. V.	1847	Grande Sonate "Les Quatre Âges", Op. 33, mov. 1	MS
69	Alkan, C. V.	1847	Grande Sonate "Les Quatre Âges", Op. 33, mov. 2	MS
70	Alkan, C. V.	1847	Grande Sonate "Les Quatre Âges", Op. 33, mov. 3	MS
71	Alkan, C. V.	1847	Grande Sonate "Les Quatre Âges", Op. 33, mov. 4	MS
72	Alkan, C. V.	1847	Op. 34, Scherzo Focoso	MS
73	Alkan, C. V.	1847	Etudes, Op. 35, No. 1	MS
74	Alkan, C. V.	1847	Etudes, Op. 35, No. 10, Chant d'Amour - chant de Mort	MS
75	Alkan, C. V.	1847	Etudes, Op. 35, No. 11	MS
76	Alkan, C. V.	1847	Etudes, Op. 35, No. 12, Technique des Octaves	MS
77	Alkan, C. V.	1847	Etudes, Op. 35, No. 2	MS
78	Alkan, C. V.	1847	Etudes, Op. 35, No. 3	MS
79	Alkan, C. V.	1847	Etudes, Op. 35, No. 5	MS
80	Alkan, C. V.	1847	Etudes, Op. 35, No. 7, L'Incendie au village voisin	MS
81	Alkan, C. V.	1847	Etudes, Op. 35, No. 9, Contrapunctus	MS
82	Alkan, C. V.	1848	Impromptus, Op. 32, No. 1, mov. 4, La Foi	MS
83	Alkan, C. V.	1849	Deuxième Recueil d'Impromptus, Op. 32, No. 2, mov. 1	MS
84	Alkan, C. V.	1849	Deuxième Recueil d'Impromptus, Op. 32, No. 2, mov. 2	MS
85	Alkan, C. V.	1849	Deuxième Recueil d'Impromptus, Op. 32, No. 2, mov. 3	MS
86	Alkan, C. V.	1849	Deuxième Recueil d'Impromptus, Op. 32, No. 2, mov. 4	MS
87	Alkan, C. V.	1856	Op. 46, Menuetto alla Tedesca	MS
88	Alkan, C. V.	1857	Marches, Op. 37, No. 1, Quasi da Cavalleria	MS
89	Alkan, C. V.	1857	Marches, Op. 37, No. 2, Quasi da Cavalleria	MS
90	Alkan, C. V.	1857	Marches, Op. 37, No. 3, Quasi da Cavalleria	MS
91	Alkan, C. V.	1857	Op. 38b, No. 2, Chant de Guerre	MS
92	Alkan, C. V.	1857	Douze Études dans les tons Mineurs, Op. 39, No. 1, Comme le vent	MS
93	Alkan, C. V.	1857	Douze Études dans les tons Mineurs, Op. 39, No. 10, Concerto for Solo Piano mov. 3	MS
94	Alkan, C. V.	1857	Douze Études dans les tons Mineurs, Op. 39, No. 11, Ouverture in B minor	MS
95	Alkan, C. V.	1857	Douze Études dans les tons Mineurs, Op. 39, No. 12, Le Festin d'Ésope	MS
96	Alkan, C. V.	1857	Douze Études dans les tons Mineurs, Op. 39, No. 2, En rhythme mol-losique	MS
97	Alkan, C. V.	1857	Douze Études dans les tons Mineurs, Op. 39, No. 3, Scherzo Diabolico	MS
98	Alkan, C. V.	1857	Douze Études dans les tons Mineurs, Op. 39, No. 4, Symphony for Solo Piano mov. 1	MS
99	Alkan, C. V.	1857	Douze Études dans les tons Mineurs, Op. 39, No. 5, Symphony for Solo Piano mov. 2	MS

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
100	Alkan, C. V.	1857	Douze Études dans les tons Mineurs, Op. 39, No. 6, Symphony for Solo Piano mov. 3	MS
101	Alkan, C. V.	1857	Douze Études dans les tons Mineurs, Op. 39, No. 7, Symphony for Solo Piano mov. 4	MS
102	Alkan, C. V.	1857	Douze Études dans les tons Mineurs, Op. 39, No. 8, Concerto for Solo Piano mov. 1	MS
103	Alkan, C. V.	1857	Douze Études dans les tons Mineurs, Op. 39, No. 9, Concerto for Solo Piano mov. 2	MS
104	Alkan, C. V.	1859	Op. 50b, Le Tambour Bat aux Champs	MS
105	Alkan, C. V.	1859	Une petite Pièce pour Piano, Op. 60, No. 2, Ma Chère Servitude	MS
106	Alkan, C. V.	1859	Op. 60b, Le Grillon	MS
107	Alkan, C. V.	1861	Esquisses, Op. 63, No. 1	MS
108	Alkan, C. V.	1861	Esquisses, Op. 63, No. 10	MS
109	Alkan, C. V.	1861	Esquisses, Op. 63, No. 11	MS
110	Alkan, C. V.	1861	Esquisses, Op. 63, No. 12	MS
111	Alkan, C. V.	1861	Esquisses, Op. 63, No. 13	MS
112	Alkan, C. V.	1861	Esquisses, Op. 63, No. 14	MS
113	Alkan, C. V.	1861	Esquisses, Op. 63, No. 15	MS
114	Alkan, C. V.	1861	Esquisses, Op. 63, No. 16	MS
115	Alkan, C. V.	1861	Esquisses, Op. 63, No. 17	MS
116	Alkan, C. V.	1861	Esquisses, Op. 63, No. 18	MS
117	Alkan, C. V.	1861	Esquisses, Op. 63, No. 19	MS
118	Alkan, C. V.	1861	Esquisses, Op. 63, No. 2	MS
119	Alkan, C. V.	1861	Esquisses, Op. 63, No. 20	MS
120	Alkan, C. V.	1861	Esquisses, Op. 63, No. 21	MS
121	Alkan, C. V.	1861	Esquisses, Op. 63, No. 22	MS
122	Alkan, C. V.	1861	Esquisses, Op. 63, No. 23	MS
123	Alkan, C. V.	1861	Esquisses, Op. 63, No. 24	MS
124	Alkan, C. V.	1861	Esquisses, Op. 63, No. 25	MS
125	Alkan, C. V.	1861	Esquisses, Op. 63, No. 26	MS
126	Alkan, C. V.	1861	Esquisses, Op. 63, No. 27	MS
127	Alkan, C. V.	1861	Esquisses, Op. 63, No. 28	MS
128	Alkan, C. V.	1861	Esquisses, Op. 63, No. 29	MS
129	Alkan, C. V.	1861	Esquisses, Op. 63, No. 3	MS
130	Alkan, C. V.	1861	Esquisses, Op. 63, No. 30	MS
131	Alkan, C. V.	1861	Esquisses, Op. 63, No. 31	MS
132	Alkan, C. V.	1861	Esquisses, Op. 63, No. 32	MS
133	Alkan, C. V.	1861	Esquisses, Op. 63, No. 33	MS
134	Alkan, C. V.	1861	Esquisses, Op. 63, No. 34	MS
135	Alkan, C. V.	1861	Esquisses, Op. 63, No. 35	MS
136	Alkan, C. V.	1861	Esquisses, Op. 63, No. 36	MS
137	Alkan, C. V.	1861	Esquisses, Op. 63, No. 37	MS
138	Alkan, C. V.	1861	Esquisses, Op. 63, No. 38	MS
139	Alkan, C. V.	1861	Esquisses, Op. 63, No. 39	MS
140	Alkan, C. V.	1861	Esquisses, Op. 63, No. 4	MS
141	Alkan, C. V.	1861	Esquisses, Op. 63, No. 40	MS
142	Alkan, C. V.	1861	Esquisses, Op. 63, No. 41	MS
143	Alkan, C. V.	1861	Esquisses, Op. 63, No. 42	MS
144	Alkan, C. V.	1861	Esquisses, Op. 63, No. 43	MS
145	Alkan, C. V.	1861	Esquisses, Op. 63, No. 44	MS
146	Alkan, C. V.	1861	Esquisses, Op. 63, No. 45	MS
147	Alkan, C. V.	1861	Esquisses, Op. 63, No. 46	MS
148	Alkan, C. V.	1861	Esquisses, Op. 63, No. 47	MS
149	Alkan, C. V.	1861	Esquisses, Op. 63, No. 48	MS

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
150	Alkan, C. V.	1861	Esquisses, Op. 63, No. 49	MS
151	Alkan, C. V.	1861	Esquisses, Op. 63, No. 5	MS
152	Alkan, C. V.	1861	Esquisses, Op. 63, No. 6	MS
153	Alkan, C. V.	1861	Esquisses, Op. 63, No. 7	MS
154	Alkan, C. V.	1861	Esquisses, Op. 63, No. 8	MS
155	Alkan, C. V.	1861	Esquisses, Op. 63, No. 9	MS
156	Alkan, C. V.	1868	Quatrième Recueil de Chants, Op. 67, No. 6, Barcarolle	MS
157	Bach, J. S.	1721	Brandenburgische Konzerte, BWV 1050, mov. 1, Allegro	ELVIS
158	Bach, J. S.	1721	Brandenburgische Konzerte, BWV 1050, mov. 2, Affettuoso	ELVIS
159	Bach, J. S.	1721	Brandenburgische Konzerte, BWV 1050, mov. 3, Allegro	ELVIS
160	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 846, No. 1	MS
161	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 846, No. 2	MS
162	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 847, No. 1	MS
163	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 847, No. 2	MS
164	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 848, No. 1	MS
165	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 848, No. 2	MS
166	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 849, No. 1	MS
167	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 849, No. 2	MS
168	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 850, No. 1	MS
169	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 850, No. 2	MS
170	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 851, No. 1	MS
171	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 851, No. 2	MS
172	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 852, No. 1	MS
173	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 852, No. 2	MS
174	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 853, No. 1	MS
175	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 853, No. 2	MS
176	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 854, No. 1	MS
177	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 854, No. 2	MS
178	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 855, No. 1	MS
179	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 855, No. 2	MS
180	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 856, No. 1	MS
181	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 856, No. 2	MS
182	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 857, No. 1	MS
183	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 857, No. 2	MS
184	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 858, No. 1	MS
185	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 858, No. 2	MS
186	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 859, No. 1	MS
187	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 859, No. 2	MS
188	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 860, No. 1	MS
189	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 860, No. 2	MS
190	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 861, No. 1	MS
191	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 861, No. 2	MS
192	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 862, No. 1	MS
193	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 862, No. 2	MS
194	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 863, No. 1	MS
195	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 863, No. 2	MS
196	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 864, No. 1	MS
197	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 864, No. 2	MS
198	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 865, No. 1	MS
199	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 865, No. 2	MS
200	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 866, No. 1	MS
201	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 866, No. 2	MS
202	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 867, No. 1	MS
203	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 867, No. 2	MS

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
204	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 868, No. 1	MS
205	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 868, No. 2	MS
206	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 869, No. 1	MS
207	Bach, J. S.	1722	Wohltemperiertes Klavier I, BWV 869, No. 2	MS
208	Bach, J. S.	1723	Inventions and Sinfonias, BWV 772	MS
209	Bach, J. S.	1723	Inventions and Sinfonias, BWV 772a	MS
210	Bach, J. S.	1723	Inventions and Sinfonias, BWV 773	MS
211	Bach, J. S.	1723	Inventions and Sinfonias, BWV 774	MS
212	Bach, J. S.	1723	Inventions and Sinfonias, BWV 775	MS
213	Bach, J. S.	1723	Inventions and Sinfonias, BWV 776	MS
214	Bach, J. S.	1723	Inventions and Sinfonias, BWV 777	MS
215	Bach, J. S.	1723	Inventions and Sinfonias, BWV 778	MS
216	Bach, J. S.	1723	Inventions and Sinfonias, BWV 779	MS
217	Bach, J. S.	1723	Inventions and Sinfonias, BWV 780	MS
218	Bach, J. S.	1723	Inventions and Sinfonias, BWV 781	MS
219	Bach, J. S.	1723	Inventions and Sinfonias, BWV 782	MS
220	Bach, J. S.	1723	Inventions and Sinfonias, BWV 783	MS
221	Bach, J. S.	1723	Inventions and Sinfonias, BWV 784	MS
222	Bach, J. S.	1723	Inventions and Sinfonias, BWV 785	MS
223	Bach, J. S.	1723	Inventions and Sinfonias, BWV 786	MS
224	Bach, J. S.	1723	Inventions and Sinfonias, BWV 787	MS
225	Bach, J. S.	1723	Inventions and Sinfonias, BWV 788	MS
226	Bach, J. S.	1723	Inventions and Sinfonias, BWV 789	MS
227	Bach, J. S.	1723	Inventions and Sinfonias, BWV 790	MS
228	Bach, J. S.	1723	Inventions and Sinfonias, BWV 791	MS
229	Bach, J. S.	1723	Inventions and Sinfonias, BWV 792	MS
230	Bach, J. S.	1723	Inventions and Sinfonias, BWV 793	MS
231	Bach, J. S.	1723	Inventions and Sinfonias, BWV 794	MS
232	Bach, J. S.	1723	Inventions and Sinfonias, BWV 795	MS
233	Bach, J. S.	1723	Inventions and Sinfonias, BWV 796	MS
234	Bach, J. S.	1723	Inventions and Sinfonias, BWV 797	MS
235	Bach, J. S.	1723	Inventions and Sinfonias, BWV 798	MS
236	Bach, J. S.	1723	Inventions and Sinfonias, BWV 799	MS
237	Bach, J. S.	1723	Inventions and Sinfonias, BWV 800	MS
238	Bach, J. S.	1723	Inventions and Sinfonias, BWV 801	MS
239	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 870, No. 1	MS
240	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 870, No. 2	MS
241	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 871, No. 1	MS
242	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 871, No. 2	MS
243	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 872, No. 1	MS
244	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 872, No. 2	MS
245	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 873, No. 1	MS
246	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 873, No. 2	MS
247	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 874, No. 1	MS
248	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 874, No. 2	MS
249	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 875, No. 1	MS
250	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 875, No. 2	MS
251	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 876, No. 1	MS
252	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 876, No. 2	MS
253	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 877, No. 1	MS
254	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 877, No. 2	MS
255	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 878, No. 1	MS
256	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 878, No. 2	MS
257	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 879, No. 1	MS

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
258	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 879, No. 2	MS
259	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 880, No. 1	MS
260	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 880, No. 2	MS
261	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 881, No. 1	MS
262	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 881, No. 2	MS
263	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 882, No. 1	MS
264	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 882, No. 2	MS
265	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 883, No. 1	MS
266	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 883, No. 2	MS
267	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 884, No. 1	MS
268	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 884, No. 2	MS
269	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 885, No. 1	MS
270	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 885, No. 2	MS
271	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 886, No. 1	MS
272	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 886, No. 2	MS
273	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 887, No. 1	MS
274	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 887, No. 2	MS
275	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 888, No. 1	MS
276	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 888, No. 2	MS
277	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 889, No. 1	MS
278	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 889, No. 2	MS
279	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 890, No. 1	MS
280	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 890, No. 2	MS
281	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 891, No. 1	MS
282	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 891, No. 2	MS
283	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 892, No. 1	MS
284	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 892, No. 2	MS
285	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 893, No. 1	MS
286	Bach, J. S.	1740	Wohltemperiertes Klavier II, BWV 893, No. 2	MS
287	Balakirev, M.	1869	Islamey	MS
288	Bartók, B.	1915	Sonatina, Sz. 55, mov. 1	MS
289	Bartók, B.	1915	Sonatina, Sz. 55, mov. 2, Dance	MS
290	Bartók, B.	1915	Romanian Folk Dances, Sz. 56, mov. 1	MS
291	Bartók, B.	1915	Romanian Folk Dances, Sz. 56, mov. 2	MS
292	Bartók, B.	1915	Romanian Folk Dances, Sz. 56, mov. 3	MS
293	Bartók, B.	1915	Romanian Folk Dances, Sz. 56, mov. 4	MS
294	Bartók, B.	1915	Romanian Folk Dances, Sz. 56, mov. 5	MS
295	Bartók, B.	1915	Romanian Folk Dances, Sz. 56, mov. 6	MS
296	Beethoven, L. van	1795	Piano Sonatas, Op. 2, No. 1, mov. 1, Piano Sonata No. 1	MS
297	Beethoven, L. van	1795	Piano Sonatas, Op. 2, No. 1, mov. 2, Piano Sonata No. 1	MS
298	Beethoven, L. van	1795	Piano Sonatas, Op. 2, No. 1, mov. 3, Piano Sonata No. 1	MS
299	Beethoven, L. van	1795	Piano Sonatas, Op. 2, No. 1, mov. 4, Piano Sonata No. 1	MS
300	Beethoven, L. van	1795	Piano Sonatas, Op. 2, No. 2, mov. 1, Piano Sonata No. 2	MS
301	Beethoven, L. van	1795	Piano Sonatas, Op. 2, No. 2, mov. 2, Piano Sonata No. 2	MS
302	Beethoven, L. van	1795	Piano Sonatas, Op. 2, No. 2, mov. 3, Piano Sonata No. 2	MS
303	Beethoven, L. van	1795	Piano Sonatas, Op. 2, No. 2, mov. 4, Piano Sonata No. 2	MS
304	Beethoven, L. van	1795	Piano Sonatas, Op. 2, No. 3, mov. 1, Piano Sonata No. 3	MS
305	Beethoven, L. van	1795	Piano Sonatas, Op. 2, No. 3, mov. 2, Piano Sonata No. 3	MS
306	Beethoven, L. van	1795	Piano Sonatas, Op. 2, No. 3, mov. 3, Piano Sonata No. 3	MS
307	Beethoven, L. van	1795	Piano Sonatas, Op. 2, No. 3, mov. 4, Piano Sonata No. 3	MS
308	Beethoven, L. van	1797	Piano Sonatas, Op. 7, mov. 1, Piano Sonata No. 4	MS
309	Beethoven, L. van	1797	Piano Sonatas, Op. 7, mov. 2, Piano Sonata No. 4	MS
310	Beethoven, L. van	1797	Piano Sonatas, Op. 7, mov. 3, Piano Sonata No. 4	MS
311	Beethoven, L. van	1797	Piano Sonatas, Op. 7, mov. 4, Piano Sonata No. 4	MS

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
312	Beethoven, L. van	1798	Piano Sonatas, Op. 10, No. 1, mov. 1, Piano Sonata No. 5	MS
313	Beethoven, L. van	1798	Piano Sonatas, Op. 10, No. 1, mov. 2, Piano Sonata No. 5	MS
314	Beethoven, L. van	1798	Piano Sonatas, Op. 10, No. 1, mov. 3, Piano Sonata No. 5	MS
315	Beethoven, L. van	1798	Piano Sonatas, Op. 10, No. 2, mov. 1, Piano Sonata No. 6	MS
316	Beethoven, L. van	1798	Piano Sonatas, Op. 10, No. 2, mov. 2, Piano Sonata No. 6	MS
317	Beethoven, L. van	1798	Piano Sonatas, Op. 10, No. 2, mov. 3, Piano Sonata No. 6	MS
318	Beethoven, L. van	1798	Piano Sonatas, Op. 10, No. 3, mov. 1, Piano Sonata No. 7	MS
319	Beethoven, L. van	1798	Piano Sonatas, Op. 10, No. 3, mov. 2, Piano Sonata No. 7	MS
320	Beethoven, L. van	1798	Piano Sonatas, Op. 10, No. 3, mov. 3, Piano Sonata No. 7	MS
321	Beethoven, L. van	1798	Piano Sonatas, Op. 10, No. 3, mov. 4, Piano Sonata No. 7	MS
322	Beethoven, L. van	1798	Piano Sonatas, Op. 13, mov. 1, Piano Sonata No. 8	MS
323	Beethoven, L. van	1798	Piano Sonatas, Op. 13, mov. 2, Piano Sonata No. 8	MS
324	Beethoven, L. van	1798	Piano Sonatas, Op. 13, mov. 3, Piano Sonata No. 8	MS
325	Beethoven, L. van	1799	Piano Sonatas, Op. 14, No. 1, mov. 1, Piano Sonata No. 9	MS
326	Beethoven, L. van	1799	Piano Sonatas, Op. 14, No. 1, mov. 2, Piano Sonata No. 9	MS
327	Beethoven, L. van	1799	Piano Sonatas, Op. 14, No. 1, mov. 3, Piano Sonata No. 9	MS
328	Beethoven, L. van	1799	Piano Sonatas, Op. 14, No. 2, mov. 1, Piano Sonata No. 10	MS
329	Beethoven, L. van	1799	Piano Sonatas, Op. 14, No. 2, mov. 2, Piano Sonata No. 10	MS
330	Beethoven, L. van	1799	Piano Sonatas, Op. 14, No. 2, mov. 3, Piano Sonata No. 10	MS
331	Beethoven, L. van	1800	Piano Sonatas, Op. 22, mov. 1, Piano Sonata No. 11	MS
332	Beethoven, L. van	1800	Piano Sonatas, Op. 22, mov. 2, Piano Sonata No. 11	MS
333	Beethoven, L. van	1800	Piano Sonatas, Op. 22, mov. 3, Piano Sonata No. 11	MS
334	Beethoven, L. van	1800	Piano Sonatas, Op. 22, mov. 4, Piano Sonata No. 11	MS
335	Beethoven, L. van	1801	Piano Sonatas, Op. 26, mov. 1, Piano Sonata No. 12	MS
336	Beethoven, L. van	1801	Piano Sonatas, Op. 26, mov. 2, Piano Sonata No. 12	MS
337	Beethoven, L. van	1801	Piano Sonatas, Op. 26, mov. 3, Piano Sonata No. 12	MS
338	Beethoven, L. van	1801	Piano Sonatas, Op. 26, mov. 4, Piano Sonata No. 12	MS
339	Beethoven, L. van	1801	Piano Sonatas, Op. 27, No. 1, mov. 1, Piano Sonata No. 13	MS
340	Beethoven, L. van	1801	Piano Sonatas, Op. 27, No. 1, mov. 2, Piano Sonata No. 13	MS
341	Beethoven, L. van	1801	Piano Sonatas, Op. 27, No. 1, mov. 3, Piano Sonata No. 13	MS
342	Beethoven, L. van	1801	Piano Sonatas, Op. 27, No. 1, mov. 4, Piano Sonata No. 13	MS
343	Beethoven, L. van	1801	Piano Sonatas, Op. 27, No. 2, mov. 1, Piano Sonata No. 14	MS
344	Beethoven, L. van	1801	Piano Sonatas, Op. 27, No. 2, mov. 2, Piano Sonata No. 14	MS
345	Beethoven, L. van	1801	Piano Sonatas, Op. 27, No. 2, mov. 3, Piano Sonata No. 14	MS
346	Beethoven, L. van	1801	Piano Sonatas, Op. 28, mov. 1, Piano Sonata No. 15	MS
347	Beethoven, L. van	1801	Piano Sonatas, Op. 28, mov. 2, Piano Sonata No. 15	MS
348	Beethoven, L. van	1801	Piano Sonatas, Op. 28, mov. 3, Piano Sonata No. 15	MS
349	Beethoven, L. van	1801	Piano Sonatas, Op. 28, mov. 4, Piano Sonata No. 15	MS
350	Beethoven, L. van	1802	Piano Sonatas, Op. 31, No. 1, mov. 1, Piano Sonata No. 16	MS
351	Beethoven, L. van	1802	Piano Sonatas, Op. 31, No. 1, mov. 2, Piano Sonata No. 16	MS
352	Beethoven, L. van	1802	Piano Sonatas, Op. 31, No. 1, mov. 3, Piano Sonata No. 16	MS
353	Beethoven, L. van	1802	Piano Sonatas, Op. 31, No. 2, mov. 1, Piano Sonata No. 17	MS
354	Beethoven, L. van	1802	Piano Sonatas, Op. 31, No. 2, mov. 2, Piano Sonata No. 17	MS
355	Beethoven, L. van	1802	Piano Sonatas, Op. 31, No. 2, mov. 3, Piano Sonata No. 17	MS
356	Beethoven, L. van	1802	Piano Sonatas, Op. 31, No. 3, mov. 1, Piano Sonata No. 18	MS
357	Beethoven, L. van	1802	Piano Sonatas, Op. 31, No. 3, mov. 2, Piano Sonata No. 18	MS
358	Beethoven, L. van	1802	Piano Sonatas, Op. 31, No. 3, mov. 3, Piano Sonata No. 18	MS
359	Beethoven, L. van	1802	Piano Sonatas, Op. 31, No. 3, mov. 4, Piano Sonata No. 18	MS
360	Beethoven, L. van	1802	Sechs Lieder, Op. 48, No. 1, Bitten	OSLC
361	Beethoven, L. van	1802	Sechs Lieder, Op. 48, No. 2, Die Liebe des Nächsten	OSLC
362	Beethoven, L. van	1802	Sechs Lieder, Op. 48, No. 3, Vom Tode	OSLC
363	Beethoven, L. van	1802	Sechs Lieder, Op. 48, No. 4, Die Ehre Gottes aus der Natur	OSLC
364	Beethoven, L. van	1802	Sechs Lieder, Op. 48, No. 5, Gottes Macht und Vorsehung	OSLC
365	Beethoven, L. van	1802	Sechs Lieder, Op. 48, No. 6, Busslied	OSLC

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
366	Beethoven, L. van	1803	Piano Sonatas, Op. 53, mov. 1, Piano Sonata No. 21	MS
367	Beethoven, L. van	1803	Piano Sonatas, Op. 53, mov. 2, Piano Sonata No. 21	MS
368	Beethoven, L. van	1803	Piano Sonatas, Op. 53, mov. 3, Piano Sonata No. 21	MS
369	Beethoven, L. van	1804	Piano Sonatas, Op. 54, mov. 1, Piano Sonata No. 22	MS
370	Beethoven, L. van	1804	Piano Sonatas, Op. 54, mov. 2, Piano Sonata No. 22	MS
371	Beethoven, L. van	1805	Piano Sonatas, Op. 49, No. 1, mov. 1, Piano Sonata No. 19	MS
372	Beethoven, L. van	1805	Piano Sonatas, Op. 49, No. 1, mov. 2, Piano Sonata No. 19	MS
373	Beethoven, L. van	1805	Piano Sonatas, Op. 49, No. 2, mov. 1, Piano Sonata No. 20	MS
374	Beethoven, L. van	1805	Piano Sonatas, Op. 49, No. 2, mov. 2, Piano Sonata No. 20	MS
375	Beethoven, L. van	1805	Piano Sonatas, Op. 57, mov. 1, Piano Sonata No. 23	MS
376	Beethoven, L. van	1805	Piano Sonatas, Op. 57, mov. 2, Piano Sonata No. 23	MS
377	Beethoven, L. van	1805	Piano Sonatas, Op. 57, mov. 3, Piano Sonata No. 23	MS
378	Beethoven, L. van	1809	Piano Sonatas, Op. 78, mov. 1, Piano Sonata No. 24	MS
379	Beethoven, L. van	1809	Piano Sonatas, Op. 78, mov. 2, Piano Sonata No. 24	MS
380	Beethoven, L. van	1809	Piano Sonatas, Op. 79, mov. 1, Piano Sonata No. 25	MS
381	Beethoven, L. van	1809	Piano Sonatas, Op. 79, mov. 2, Piano Sonata No. 25	MS
382	Beethoven, L. van	1809	Piano Sonatas, Op. 79, mov. 3, Piano Sonata No. 25	MS
383	Beethoven, L. van	1810	Piano Sonatas, Op. 81a, mov. 1, Piano Sonata No. 26	MS
384	Beethoven, L. van	1810	Piano Sonatas, Op. 81a, mov. 2, Piano Sonata No. 26	MS
385	Beethoven, L. van	1810	Piano Sonatas, Op. 81a, mov. 3, Piano Sonata No. 26	MS
386	Beethoven, L. van	1814	Piano Sonatas, Op. 90, mov. 1, Piano Sonata No. 27	MS
387	Beethoven, L. van	1814	Piano Sonatas, Op. 90, mov. 2, Piano Sonata No. 27	MS
388	Beethoven, L. van	1816	Piano Sonatas, Op. 101, mov. 1, Piano Sonata No. 28	MS
389	Beethoven, L. van	1816	Piano Sonatas, Op. 101, mov. 2, Piano Sonata No. 28	MS
390	Beethoven, L. van	1816	Piano Sonatas, Op. 101, mov. 3, Piano Sonata No. 28	MS
391	Beethoven, L. van	1816	Piano Sonatas, Op. 101, mov. 4, Piano Sonata No. 28	MS
392	Beethoven, L. van	1818	Piano Sonatas, Op. 106, mov. 1, Piano Sonata No. 29	MS
393	Beethoven, L. van	1818	Piano Sonatas, Op. 106, mov. 2, Piano Sonata No. 29	MS
394	Beethoven, L. van	1818	Piano Sonatas, Op. 106, mov. 3, Piano Sonata No. 29	MS
395	Beethoven, L. van	1818	Piano Sonatas, Op. 106, mov. 4, Piano Sonata No. 29	MS
396	Beethoven, L. van	1820	Piano Sonatas, Op. 109, mov. 1, Piano Sonata No. 30	MS
397	Beethoven, L. van	1820	Piano Sonatas, Op. 109, mov. 2, Piano Sonata No. 30	MS
398	Beethoven, L. van	1820	Piano Sonatas, Op. 109, mov. 3, Piano Sonata No. 30	MS
399	Beethoven, L. van	1821	Piano Sonatas, Op. 110, mov. 1, Piano Sonata No. 31	MS
400	Beethoven, L. van	1821	Piano Sonatas, Op. 110, mov. 2, Piano Sonata No. 31	MS
401	Beethoven, L. van	1821	Piano Sonatas, Op. 110, mov. 3, Piano Sonata No. 31	MS
402	Beethoven, L. van	1822	Piano Sonatas, Op. 111, mov. 1, Piano Sonata No. 32	MS
403	Beethoven, L. van	1822	Piano Sonatas, Op. 111, mov. 2, Piano Sonata No. 32	MS
404	Binchois, G.	1436	Adieu, adieu	ELVIS
405	Binchois, G.	1436	Adieu, m'amour	ELVIS
406	Binchois, G.	1436	Amoureux suy	ELVIS
407	Binchois, G.	1436	Amour et qu'as tu	ELVIS
408	Binchois, G.	1436	Amours et souvenir	ELVIS
409	Binchois, G.	1436	Ay, douloureux	ELVIS
410	Binchois, G.	1436	En regardant	ELVIS
411	Binchois, G.	1436	Je me recommande	ELVIS
412	Binchois, G.	1440	Adieu, jusques	ELVIS
413	Binchois, G.	1440	Adieu ma doulce	ELVIS
414	Binchois, G.	1440	Adieu mes tres belles	ELVIS
415	Binchois, G.	1440	Jamais tant	ELVIS
416	Binchois, G.	1445	C'est assez	ELVIS
417	Binchois, G.	1445	De plus en plus	ELVIS
418	Binchois, G.	1445	Esclave puist yl	ELVIS
419	Binchois, G.	1450	Bien puist	ELVIS

Continued on next page

Table B.3 – Corpus used in Part IV (“Macroanalysis”; cont.).

#	Name	Year	Title	Source
420	Binchois, G.	1450	En sera il mieulx	ELVIS
421	Binchois, G.	1460	Adieu mon amoureuse joye	ELVIS
422	Binchois, G.	1475	Comme femme desconfortée	ELVIS
423	Brahms, J.	1868	Vier Gesänge, Op. 46, No. 1, Die Kränze	OSLC
424	Brahms, J.	1868	Vier Gesänge, Op. 46, No. 2, Magyarisch	OSLC
425	Brahms, J.	1868	Vier Gesänge, Op. 46, No. 3, Die Schale der Vergessenheit	OSLC
426	Brahms, J.	1868	Vier Gesänge, Op. 46, No. 4, An die Nachtigall	OSLC
427	Brahms, J.	1878	8 Klavierstücke, Op. 76, No. 1, Capriccio	DCML
428	Brahms, J.	1878	8 Klavierstücke, Op. 76, No. 2, Capriccio	DCML
429	Brahms, J.	1878	8 Klavierstücke, Op. 76, No. 3, Intermezzo	DCML
430	Brahms, J.	1878	8 Klavierstücke, Op. 76, No. 4, Intermezzo	DCML
431	Brahms, J.	1878	8 Klavierstücke, Op. 76, No. 5, Capriccio	DCML
432	Brahms, J.	1878	8 Klavierstücke, Op. 76, No. 7, Intermezzo	DCML
433	Brahms, J.	1878	8 Klavierstücke, Op. 76, No. 8, Capriccio	DCML
434	Brahms, J.	1892	7 Fantasien, Op. 116, No. 1, Capriccio	DCML
435	Brahms, J.	1892	7 Fantasien, Op. 116, No. 2, Intermezzo	DCML
436	Brahms, J.	1892	7 Fantasien, Op. 116, No. 3, Capriccio	DCML
437	Brahms, J.	1892	7 Fantasien, Op. 116, No. 4, Intermezzo	DCML
438	Brahms, J.	1892	7 Fantasien, Op. 116, No. 5, Intermezzo	DCML
439	Brahms, J.	1892	7 Fantasien, Op. 116, No. 6, Intermezzo	DCML
440	Brahms, J.	1892	7 Fantasien, Op. 116, No. 7, Capriccio	MS
441	Brahms, J.	1892	Drei Intermezzi, Op. 117, No. 1	MS
442	Brahms, J.	1892	Drei Intermezzi, Op. 117, No. 2	MS
443	Brahms, J.	1892	Drei Intermezzi, Op. 117, No. 3	MS
444	Brahms, J.	1893	Sechs Klavierstücke, Op. 118, No. 1, Intermezzo	DCML
445	Brahms, J.	1893	Sechs Klavierstücke, Op. 118, No. 2, Intermezzo	MS
446	Brahms, J.	1893	Sechs Klavierstücke, Op. 118, No. 3, Ballade	DCML
447	Brahms, J.	1893	Sechs Klavierstücke, Op. 118, No. 4, Intermezzo	DCML
448	Brahms, J.	1893	Sechs Klavierstücke, Op. 118, No. 5, Romanze	DCML
449	Brahms, J.	1893	Sechs Klavierstücke, Op. 118, No. 6, Intermezzo	DCML
450	Busnoys, A.	1475	A qui vens tu tes coquilles	ELVIS
451	Busnoys, A.	1475	Une dame j’ay fait veu	ELVIS
452	Busnoys, A.	1475	A vous sans autre me viens rendre	ELVIS
453	Busnoys, A.	1475	Au povre par necessité	ELVIS
454	Busnoys, A.	1475	Bel Acueil le sergent d’Amours	ELVIS
455	Busnoys, A.	1475	En soustenant vostre querelle	ELVIS
456	Busnoys, A.	1475	Enferme suys je en la tour	ELVIS
457	Busnoys, A.	1475	Ja que lui ne s’i attende	ELVIS
458	Busnoys, A.	1475	Je ne puis vivre ainsi tousjours	ELVIS
459	Busnoys, A.	1475	Joye me fuit	ELVIS
460	Busnoys, A.	1475	Le corps s’en va	ELVIS
461	Busnoys, A.	1475	O Fortune, trop tu es dure	ELVIS
462	Busnoys, A.	1475	Pour entretenir mes amours	ELVIS
463	Busnoys, A.	1475	Quant ce viendra au droit destraindre	ELVIS
464	Busnoys, A.	1475	Ung plus que tous	ELVIS
465	Busnoys, A.	1480	Une dame j’ay fait veu	ELVIS
466	Busnoys, A.	1480	Joye me fuit	ELVIS
467	Busnoys, A.	1480	Mon seul et celé souvenir	ELVIS
468	Busnoys, A.	1480	O Fortune, trop tu es dure	ELVIS
469	Busnoys, A.	1480	Quant ce viendra au droit destraindre	ELVIS
470	Buxtehude, D.	1694	Trio Sonata in F major, Op. 1, No. 1, Adagio	ELVIS
471	Buxtehude, D.	1694	Trio Sonata in F major, Op. 1, No. 1, Allegro	ELVIS
472	Buxtehude, D.	1694	Trio Sonata in F major, Op. 1, No. 1, Andante	ELVIS
473	Buxtehude, D.	1694	Trio Sonata in F major, Op. 1, No. 1, Grave	ELVIS

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
474	Buxtehude, D.	1694	Trio Sonata in F major, Op. 1, No. 1, Presto	ELVIS
475	Buxtehude, D.	1694	Trio Sonata in F major, Op. 1, No. 1, Vivace	ELVIS
476	Buxtehude, D.	1694	Trio Sonata in G major, Op. 1, No. 2, Adagio	ELVIS
477	Buxtehude, D.	1694	Trio Sonata in G major, Op. 1, No. 2, Allegro	ELVIS
478	Buxtehude, D.	1694	Trio Sonata in G major, Op. 1, No. 2, Grave	ELVIS
479	Buxtehude, D.	1694	Trio Sonata in G major, Op. 1, No. 2, Lento	ELVIS
480	Buxtehude, D.	1694	Trio Sonata in G major, Op. 1, No. 2, Vivace	ELVIS
481	Buxtehude, D.	1694	Trio Sonata in A minor, Op. 1, No. 3, Adagio	ELVIS
482	Buxtehude, D.	1694	Trio Sonata in A minor, Op. 1, No. 3, Allegro	ELVIS
483	Buxtehude, D.	1694	Trio Sonata in A minor, Op. 1, No. 3, Largo	ELVIS
484	Buxtehude, D.	1694	Trio Sonata in A minor, Op. 1, No. 3, Lento	ELVIS
485	Buxtehude, D.	1694	Trio Sonata in A minor, Op. 1, No. 3, Vivace	ELVIS
486	Buxtehude, D.	1694	Trio Sonata in Bb major, Op. 1, No. 4, Allegro	ELVIS
487	Buxtehude, D.	1694	Trio Sonata in Bb major, Op. 1, No. 4, Lento	ELVIS
488	Buxtehude, D.	1694	Trio Sonata in Bb major, Op. 1, No. 4, Vivace	ELVIS
489	Byrd, W.	1589	Ne irascaris domine	ELVIS
490	Byrd, W.	1605	Ave verum	ELVIS
491	Chaminade, C.	1883	Op. 22, Orientale	MS
492	Chaminade, C.	1884	Op. 29, Serenade	MS
493	Chaminade, C.	1887	Op. 39, Toccata	MS
494	Chaminade, C.	1890	Op. 54, Lolita – caprice espagnol	MS
495	Chaminade, C.	1892	Op. 60, Les sylvains	MS
496	Chaminade, C.	1892	Op. 61, Arabesque	MS
497	Chaminade, C.	1897	Op. 50, La Lisonjera	MS
498	Chaminade, C.	1898	Op. 89, Thème varié	MS
499	Chaminade, C.	1898	Op. 94, Danse créole	MS
500	Chaminade, C.	1899	Trois danses anciennes, Op. 95, No. 1	MS
501	Chaminade, C.	1899	Trois danses anciennes, Op. 95, No. 2	MS
502	Chaminade, C.	1899	Trois danses anciennes, Op. 95, No. 3	MS
503	Chaminade, C.	1901	Op. 105, Divertissement	MS
504	Chaminade, C.	1906	Album des enfants – Première serie, Op. 123, No. 1	MS
505	Chaminade, C.	1906	Album des enfants – Première serie, Op. 123, No. 10	MS
506	Chaminade, C.	1906	Album des enfants – Première serie, Op. 123, No. 11	MS
507	Chaminade, C.	1906	Album des enfants – Première serie, Op. 123, No. 12	MS
508	Chaminade, C.	1906	Album des enfants – Première serie, Op. 123, No. 2	MS
509	Chaminade, C.	1906	Album des enfants – Première serie, Op. 123, No. 3	MS
510	Chaminade, C.	1906	Album des enfants – Première serie, Op. 123, No. 4	MS
511	Chaminade, C.	1906	Album des enfants – Première serie, Op. 123, No. 5	MS
512	Chaminade, C.	1906	Album des enfants – Première serie, Op. 123, No. 6	MS
513	Chaminade, C.	1906	Album des enfants – Première serie, Op. 123, No. 7	MS
514	Chaminade, C.	1906	Album des enfants – Première serie, Op. 123, No. 8	MS
515	Chaminade, C.	1906	Album des enfants – Première serie, Op. 123, No. 9	MS
516	Chaminade, C.	1907	Album des enfants – Deuxième serie, Op. 126, No. 1	MS
517	Chaminade, C.	1907	Album des enfants – Deuxième serie, Op. 126, No. 10	MS
518	Chaminade, C.	1907	Album des enfants – Deuxième serie, Op. 126, No. 11	MS
519	Chaminade, C.	1907	Album des enfants – Deuxième serie, Op. 126, No. 12	MS
520	Chaminade, C.	1907	Album des enfants – Deuxième serie, Op. 126, No. 2	MS
521	Chaminade, C.	1907	Album des enfants – Deuxième serie, Op. 126, No. 3	MS
522	Chaminade, C.	1907	Album des enfants – Deuxième serie, Op. 126, No. 4	MS
523	Chaminade, C.	1907	Album des enfants – Deuxième serie, Op. 126, No. 5	MS
524	Chaminade, C.	1907	Album des enfants – Deuxième serie, Op. 126, No. 6	MS
525	Chaminade, C.	1907	Album des enfants – Deuxième serie, Op. 126, No. 7	MS
526	Chaminade, C.	1907	Album des enfants – Deuxième serie, Op. 126, No. 8	MS
527	Chaminade, C.	1907	Album des enfants – Deuxième serie, Op. 126, No. 9	MS

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
528	Chopin, F.	1827	Mazurkas, Op. 68, No. 2	CCARH
529	Chopin, F.	1829	Mazurkas, Op. 68, No. 1	CCARH
530	Chopin, F.	1829	Mazurkas, Op. 68, No. 3	CCARH
531	Chopin, F.	1830	Mazurkas, Op. 6, No. 1	CCARH
532	Chopin, F.	1830	Mazurkas, Op. 6, No. 2	CCARH
533	Chopin, F.	1830	Mazurkas, Op. 6, No. 3	CCARH
534	Chopin, F.	1830	Mazurkas, Op. 6, No. 4	CCARH
535	Chopin, F.	1830	Mazurkas, Op. 7, No. 1	CCARH
536	Chopin, F.	1830	Mazurkas, Op. 7, No. 2	CCARH
537	Chopin, F.	1830	Mazurkas, Op. 7, No. 3	CCARH
538	Chopin, F.	1830	Mazurkas, Op. 7, No. 4	CCARH
539	Chopin, F.	1830	Mazurkas, Op. 7, No. 5	CCARH
540	Chopin, F.	1831	Ballads, Op. 23	MS
541	Chopin, F.	1832	Mazurkas, Op. 17, No. 1	CCARH
542	Chopin, F.	1832	Mazurkas, Op. 17, No. 2	CCARH
543	Chopin, F.	1832	Mazurkas, Op. 17, No. 3	CCARH
544	Chopin, F.	1832	Mazurkas, Op. 17, No. 4	CCARH
545	Chopin, F.	1833	Nocturnes, Op. 15, No. 1, 3 Nocturnes	MS
546	Chopin, F.	1833	Nocturnes, Op. 15, No. 2, 3 Nocturnes	MS
547	Chopin, F.	1833	Nocturnes, Op. 15, No. 3, 3 Nocturnes	MS
548	Chopin, F.	1833	Nocturnes, Op. 9, No. 1, 3 Nocturnes	MS
549	Chopin, F.	1833	Nocturnes, Op. 9, No. 2, 3 Nocturnes	MS
550	Chopin, F.	1833	Nocturnes, Op. 9, No. 3, 3 Nocturnes	MS
551	Chopin, F.	1834	Mazurkas, Op. 24, No. 1	CCARH
552	Chopin, F.	1834	Mazurkas, Op. 24, No. 2	CCARH
553	Chopin, F.	1834	Mazurkas, Op. 24, No. 3	CCARH
554	Chopin, F.	1834	Mazurkas, Op. 24, No. 4	CCARH
555	Chopin, F.	1835	Polonaises, Op. 26, No. 1	MS
556	Chopin, F.	1835	Mazurkas, Op. 67, No. 1	CCARH
557	Chopin, F.	1835	Mazurkas, Op. 67, No. 3	CCARH
558	Chopin, F.	1836	Douze Études, Op. 25, No. 7	MS
559	Chopin, F.	1836	Nocturnes, Op. 27, No. 1, 2 Nocturnes	MS
560	Chopin, F.	1836	Nocturnes, Op. 27, No. 2, 2 Nocturnes	MS
561	Chopin, F.	1836	Mazurkas, Op. 30, No. 1	CCARH
562	Chopin, F.	1836	Mazurkas, Op. 30, No. 2	CCARH
563	Chopin, F.	1836	Mazurkas, Op. 30, No. 3	CCARH
564	Chopin, F.	1836	Mazurkas, Op. 30, No. 4	CCARH
565	Chopin, F.	1837	Op. 31, Scherzo No. 2	MS
566	Chopin, F.	1837	Nocturnes, Op. 32, No. 1, 2 Nocturnes	MS
567	Chopin, F.	1837	Nocturnes, Op. 32, No. 2, 2 Nocturnes	MS
568	Chopin, F.	1837	Mazurkas, Op. 33, No. 1	CCARH
569	Chopin, F.	1837	Mazurkas, Op. 33, No. 2	CCARH
570	Chopin, F.	1837	Mazurkas, Op. 33, No. 3	CCARH
571	Chopin, F.	1837	Mazurkas, Op. 33, No. 4	CCARH
572	Chopin, F.	1838	Mazurkas, Op. 41, No. 1	CCARH
573	Chopin, F.	1839	Préludes, Op. 28, No. 1, 24 Préludes	MS
574	Chopin, F.	1839	Préludes, Op. 28, No. 10, 24 Préludes	MS
575	Chopin, F.	1839	Préludes, Op. 28, No. 11, 24 Préludes	MS
576	Chopin, F.	1839	Préludes, Op. 28, No. 12, 24 Préludes	MS
577	Chopin, F.	1839	Préludes, Op. 28, No. 13, 24 Préludes	MS
578	Chopin, F.	1839	Préludes, Op. 28, No. 14, 24 Préludes	MS
579	Chopin, F.	1839	Préludes, Op. 28, No. 15, 24 Préludes	MS
580	Chopin, F.	1839	Préludes, Op. 28, No. 16, 24 Préludes	MS
581	Chopin, F.	1839	Préludes, Op. 28, No. 17, 24 Préludes	MS

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
582	Chopin, F.	1839	Préludes, Op. 28, No. 18, 24 Préludes	MS
583	Chopin, F.	1839	Préludes, Op. 28, No. 19, 24 Préludes	MS
584	Chopin, F.	1839	Préludes, Op. 28, No. 2, 24 Préludes	MS
585	Chopin, F.	1839	Préludes, Op. 28, No. 20, 24 Préludes	MS
586	Chopin, F.	1839	Préludes, Op. 28, No. 21, 24 Préludes	MS
587	Chopin, F.	1839	Préludes, Op. 28, No. 22, 24 Préludes	MS
588	Chopin, F.	1839	Préludes, Op. 28, No. 23, 24 Préludes	MS
589	Chopin, F.	1839	Préludes, Op. 28, No. 24, 24 Préludes	MS
590	Chopin, F.	1839	Préludes, Op. 28, No. 3, 24 Préludes	MS
591	Chopin, F.	1839	Préludes, Op. 28, No. 4, 24 Préludes	MS
592	Chopin, F.	1839	Préludes, Op. 28, No. 5, 24 Préludes	MS
593	Chopin, F.	1839	Préludes, Op. 28, No. 6, 24 Préludes	MS
594	Chopin, F.	1839	Préludes, Op. 28, No. 7, 24 Préludes	MS
595	Chopin, F.	1839	Préludes, Op. 28, No. 8, 24 Préludes	MS
596	Chopin, F.	1839	Préludes, Op. 28, No. 9, 24 Préludes	MS
597	Chopin, F.	1839	Ballads, Op. 38	MS
598	Chopin, F.	1839	Mazurkas, Op. 41, No. 2	CCARH
599	Chopin, F.	1839	Mazurkas, Op. 41, No. 3	CCARH
600	Chopin, F.	1839	Mazurkas, Op. 41, No. 4	CCARH
601	Chopin, F.	1840	Nocturnes, Op. 37, No. 1, 2 Nocturnes	DCML
602	Chopin, F.	1840	Nocturnes, Op. 37, No. 2, 2 Nocturnes	MS
603	Chopin, F.	1841	Ballads, Op. 47	MS
604	Chopin, F.	1841	Nocturnes, Op. 48, No. 1, 2 Nocturnes	MS
605	Chopin, F.	1841	Nocturnes, Op. 48, No. 2, 2 Nocturnes	DCML
606	Chopin, F.	1841	Mazurkas, Op. 50, No. 1	CCARH
607	Chopin, F.	1841	Mazurkas, Op. 50, No. 2	CCARH
608	Chopin, F.	1841	Mazurkas, Op. 50, No. 3	CCARH
609	Chopin, F.	1842	Ballads, Op. 52, No. 4	MS
610	Chopin, F.	1842	Op. 53, Polonaise	MS
611	Chopin, F.	1842	Op. 54, Scherzo No. 4	MS
612	Chopin, F.	1843	Mazurkas, Op. 56, No. 1	CCARH
613	Chopin, F.	1843	Mazurkas, Op. 56, No. 2	CCARH
614	Chopin, F.	1843	Mazurkas, Op. 56, No. 3	CCARH
615	Chopin, F.	1844	Nocturnes, Op. 55, No. 1, 2 Nocturnes	MS
616	Chopin, F.	1844	Nocturnes, Op. 55, No. 2, 2 Nocturnes	DCML
617	Chopin, F.	1844	Op. 58, mov. 1, Piano Sonata No. 3	MS
618	Chopin, F.	1844	Op. 58, mov. 2, Piano Sonata No. 3	MS
619	Chopin, F.	1844	Op. 58, mov. 3, Piano Sonata No. 3	MS
620	Chopin, F.	1844	Op. 58, mov. 4, Piano Sonata No. 3	MS
621	Chopin, F.	1845	Mazurkas, Op. 59, No. 1	CCARH
622	Chopin, F.	1845	Mazurkas, Op. 59, No. 2	CCARH
623	Chopin, F.	1845	Mazurkas, Op. 59, No. 3	CCARH
624	Chopin, F.	1846	Op. 61, Polonaise-Fantaisie	MS
625	Chopin, F.	1846	Nocturnes, Op. 62, No. 1, 2 Nocturnes	MS
626	Chopin, F.	1846	Nocturnes, Op. 62, No. 2, 2 Nocturnes	MS
627	Chopin, F.	1846	Mazurkas, Op. 63, No. 1	CCARH
628	Chopin, F.	1846	Mazurkas, Op. 63, No. 2	CCARH
629	Chopin, F.	1846	Mazurkas, Op. 63, No. 3	CCARH
630	Chopin, F.	1846	Mazurkas, Op. 67, No. 4	CCARH
631	Chopin, F.	1847	Trois Valses, Op. 64, No. 1	ELVIS
632	Chopin, F.	1847	Trois Valses, Op. 64, No. 2	ELVIS
633	Chopin, F.	1847	Trois Valses, Op. 64, No. 3	ELVIS
634	Chopin, F.	1849	Mazurkas, Op. 67, No. 2	CCARH
635	Chopin, F.	1855	Mazurkas, Op. 68, No. 4	CCARH

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
636	Chopin, F.	1855	Nocturnes, Op. 72, No. 1, Posthum	MS
637	Clementi, M.	1797	6 Sonatinas, Op. 36, No. 1	MS
638	Clementi, M.	1797	6 Sonatinas, Op. 36, No. 2	MS
639	Clementi, M.	1797	6 Sonatinas, Op. 36, No. 3	MS
640	Clementi, M.	1797	6 Sonatinas, Op. 36, No. 4	MS
641	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 1, mov. 1, Grave	CCARH
642	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 1, mov. 2, Allegro	CCARH
643	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 1, mov. 3, Adagio	CCARH
644	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 1, mov. 4, Allegro comodo	CCARH
645	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 10, mov. 1	CCARH
646	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 10, mov. 2	CCARH
647	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 10, mov. 3	CCARH
648	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 10, mov. 4	CCARH
649	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 10, mov. 5	CCARH
650	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 11, mov. 1	CCARH
651	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 11, mov. 2	CCARH
652	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 11, mov. 3	CCARH
653	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 11, mov. 4	CCARH
654	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 12, mov. 1	CCARH
655	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 12, mov. 2	CCARH
656	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 12, mov. 3	CCARH
657	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 12, mov. 4	CCARH
658	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 2, mov. 1, Grave	CCARH
659	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 2, mov. 2, Vivace	CCARH
660	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 2, mov. 3, Adagio	CCARH
661	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 2, mov. 4, Allegro	CCARH
662	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 3, mov. 1, Grave	CCARH
663	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 3, mov. 2, Allegro	CCARH
664	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 3, mov. 3, Adagio	CCARH
665	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 4, mov. 1	CCARH
666	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 4, mov. 2	CCARH
667	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 4, mov. 3	CCARH
668	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 4, mov. 4	CCARH
669	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 5, mov. 1, Grave	CCARH
670	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 5, mov. 2, Allegro	CCARH
671	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 5, mov. 3, Adagio	CCARH
672	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 5, mov. 4, Allegro	CCARH
673	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 6, mov. 1, Grave	CCARH
674	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 6, mov. 2, Largo	CCARH
675	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 6, mov. 3, Adagio	CCARH
676	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 6, mov. 4, Allegro	CCARH
677	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 7, mov. 1, Allegro	CCARH
678	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 7, mov. 2, Grave	CCARH
679	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 7, mov. 3, Allegro	CCARH
680	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 8, mov. 1, Grave	CCARH
681	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 8, mov. 2, Allegro	CCARH
682	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 8, mov. 3, Largo	CCARH
683	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 8, mov. 4, Vivace	CCARH
684	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 9, mov. 1	CCARH
685	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 9, mov. 2	CCARH
686	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 9, mov. 3	CCARH
687	Corelli, A.	1681	12 Trio Sonatas, Op. 1, No. 9, mov. 4	CCARH
688	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 1, mov. 1	CCARH
689	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 1, mov. 2	CCARH

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
690	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 1, mov. 3	CCARH
691	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 1, mov. 4	CCARH
692	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 10, mov. 1	CCARH
693	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 10, mov. 2	CCARH
694	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 10, mov. 3	CCARH
695	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 10, mov. 4	CCARH
696	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 11, mov. 1	CCARH
697	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 11, mov. 2	CCARH
698	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 11, mov. 3	CCARH
699	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 11, mov. 4	CCARH
700	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 12, mov. 1	CCARH
701	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 12, mov. 1	CCARH
702	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 12, mov. 2	CCARH
703	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 12, mov. 2	CCARH
704	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 12, mov. 3	CCARH
705	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 12, mov. 3	CCARH
706	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 12, mov. 4	CCARH
707	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 2, mov. 1	CCARH
708	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 2, mov. 2	CCARH
709	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 2, mov. 3	CCARH
710	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 2, mov. 4	CCARH
711	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 3, mov. 1	CCARH
712	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 3, mov. 2	CCARH
713	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 3, mov. 3	CCARH
714	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 3, mov. 4	CCARH
715	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 4, mov. 1	CCARH
716	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 4, mov. 1	CCARH
717	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 4, mov. 2	CCARH
718	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 4, mov. 2	CCARH
719	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 4, mov. 3	CCARH
720	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 4, mov. 3	CCARH
721	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 4, mov. 4	CCARH
722	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 4, mov. 4	CCARH
723	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 5, mov. 1	CCARH
724	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 5, mov. 2	CCARH
725	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 5, mov. 3	CCARH
726	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 5, mov. 4	CCARH
727	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 6, mov. 1	CCARH
728	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 6, mov. 2	CCARH
729	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 6, mov. 3	CCARH
730	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 6, mov. 4	CCARH
731	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 7, mov. 1	CCARH
732	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 7, mov. 2	CCARH
733	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 7, mov. 3	CCARH
734	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 7, mov. 4	CCARH
735	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 8, mov. 1	CCARH
736	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 8, mov. 2	CCARH
737	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 8, mov. 3	CCARH
738	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 8, mov. 4	CCARH
739	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 9, mov. 1	CCARH
740	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 9, mov. 2	CCARH
741	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 9, mov. 3	CCARH
742	Corelli, A.	1689	12 Trio Sonatas, Op. 3, No. 9, mov. 4	CCARH
743	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 1, mov. 1	CCARH

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
744	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 1, mov. 3	CCARH
745	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 10, mov. 2	CCARH
746	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 10, mov. 3	CCARH
747	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 11, mov. 1	CCARH
748	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 11, mov. 2	CCARH
749	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 11, mov. 3	CCARH
750	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 12, mov. 1	CCARH
751	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 12, mov. 2	CCARH
752	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 12, mov. 3	CCARH
753	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 2, mov. 1	CCARH
754	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 2, mov. 3	CCARH
755	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 3, mov. 1	CCARH
756	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 3, mov. 2	CCARH
757	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 3, mov. 3	CCARH
758	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 3, mov. 4	CCARH
759	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 4, mov. 1	CCARH
760	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 4, mov. 2	CCARH
761	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 4, mov. 3	CCARH
762	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 5, mov. 1	CCARH
763	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 5, mov. 2	CCARH
764	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 5, mov. 3	CCARH
765	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 5, mov. 4	CCARH
766	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 6, mov. 2	CCARH
767	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 6, mov. 3	CCARH
768	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 7, mov. 1	CCARH
769	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 7, mov. 3	CCARH
770	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 7, mov. 4	CCARH
771	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 8, mov. 1	CCARH
772	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 8, mov. 2	CCARH
773	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 8, mov. 3	CCARH
774	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 9, mov. 1	CCARH
775	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 9, mov. 2	CCARH
776	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 9, mov. 3	CCARH
777	Corelli, A.	1694	12 Trio Sonatas, Op. 4, No. 9, mov. 4	CCARH
778	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 1, mov. 2	CCARH
779	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 1, mov. 3	CCARH
780	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 1, mov. 4	CCARH
781	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 10, mov. 1	CCARH
782	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 10, mov. 2	CCARH
783	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 10, mov. 3	CCARH
784	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 10, mov. 4	CCARH
785	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 10, mov. 5	CCARH
786	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 11, mov. 1	CCARH
787	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 11, mov. 2	CCARH
788	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 11, mov. 3	CCARH
789	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 11, mov. 4	CCARH
790	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 11, mov. 5	CCARH
791	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 2, mov. 1	CCARH
792	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 2, mov. 2	CCARH
793	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 2, mov. 3	CCARH
794	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 2, mov. 4	CCARH
795	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 2, mov. 5	CCARH
796	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 3, mov. 1	CCARH
797	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 3, mov. 2	CCARH

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
798	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 3, mov. 3	CCARH
799	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 3, mov. 4	CCARH
800	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 3, mov. 5	CCARH
801	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 4, mov. 1	CCARH
802	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 4, mov. 2	CCARH
803	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 4, mov. 3	CCARH
804	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 4, mov. 4	CCARH
805	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 4, mov. 5	CCARH
806	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 5, mov. 1	CCARH
807	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 5, mov. 3	CCARH
808	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 5, mov. 4	CCARH
809	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 5, mov. 5	CCARH
810	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 6, mov. 1	CCARH
811	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 6, mov. 2	CCARH
812	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 6, mov. 3	CCARH
813	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 6, mov. 4	CCARH
814	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 6, mov. 5	CCARH
815	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 7, mov. 1	CCARH
816	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 7, mov. 2	CCARH
817	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 7, mov. 3	CCARH
818	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 7, mov. 4	CCARH
819	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 8, mov. 1	CCARH
820	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 8, mov. 2	CCARH
821	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 8, mov. 3	CCARH
822	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 8, mov. 4	CCARH
823	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 9, mov. 1	CCARH
824	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 9, mov. 2	CCARH
825	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 9, mov. 3	CCARH
826	Corelli, A.	1700	12 Violin Sonatas, Op. 5, No. 9, mov. 4	CCARH
827	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 1, mov. 1	CCARH
828	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 1, mov. 2	CCARH
829	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 1, mov. 3	CCARH
830	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 1, mov. 4	CCARH
831	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 1, mov. 5	CCARH
832	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 1, mov. 6	CCARH
833	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 10, mov. 1	CCARH
834	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 10, mov. 2	CCARH
835	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 10, mov. 3	CCARH
836	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 10, mov. 4	CCARH
837	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 10, mov. 5	CCARH
838	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 10, mov. 6	CCARH
839	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 11, mov. 1	CCARH
840	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 11, mov. 2	CCARH
841	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 11, mov. 3	CCARH
842	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 11, mov. 4	CCARH
843	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 11, mov. 5	CCARH
844	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 12, mov. 1	CCARH
845	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 12, mov. 2	CCARH
846	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 12, mov. 3	CCARH
847	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 12, mov. 4	CCARH
848	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 12, mov. 5	CCARH
849	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 2, mov. 2	CCARH
850	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 2, mov. 3	CCARH
851	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 2, mov. 4	CCARH

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
852	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 3, mov. 1	CCARH
853	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 3, mov. 2	CCARH
854	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 3, mov. 3	CCARH
855	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 3, mov. 4	CCARH
856	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 4, mov. 1	CCARH
857	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 4, mov. 2	CCARH
858	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 4, mov. 3	CCARH
859	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 4, mov. 4	CCARH
860	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 4, mov. 5	CCARH
861	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 5, mov. 1	CCARH
862	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 5, mov. 2	CCARH
863	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 5, mov. 4	CCARH
864	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 5, mov. 5	CCARH
865	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 6, mov. 1	CCARH
866	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 6, mov. 2	CCARH
867	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 6, mov. 3	CCARH
868	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 6, mov. 4	CCARH
869	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 6, mov. 5	CCARH
870	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 7, mov. 1	CCARH
871	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 7, mov. 2	CCARH
872	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 7, mov. 3	CCARH
873	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 7, mov. 4	CCARH
874	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 7, mov. 5	CCARH
875	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 7, mov. 6	CCARH
876	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 8, mov. 1	CCARH
877	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 8, mov. 2	CCARH
878	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 8, mov. 3	CCARH
879	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 8, mov. 4	CCARH
880	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 8, mov. 5	CCARH
881	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 8, mov. 6	CCARH
882	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 8, mov. 7	CCARH
883	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 9, mov. 1	CCARH
884	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 9, mov. 2	CCARH
885	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 9, mov. 3	CCARH
886	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 9, mov. 4	CCARH
887	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 9, mov. 5	CCARH
888	Corelli, A.	1714	12 concerti grossi, Op. 6, No. 9, mov. 6	CCARH
889	Cornelius, P.	1849	Sechs Lieder, Op. 1, No. 1, Untreu	OSLC
890	Cornelius, P.	1849	Sechs Lieder, Op. 1, No. 2, Veilchen	OSLC
891	Cornelius, P.	1849	Sechs Lieder, Op. 1, No. 3, Wiegenlied	OSLC
892	Cornelius, P.	1849	Sechs Lieder, Op. 1, No. 4, Schmetterling	OSLC
893	Cornelius, P.	1849	Sechs Lieder, Op. 1, No. 5, Nachts	OSLC
894	Cornelius, P.	1849	Sechs Lieder, Op. 1, No. 6, Denkst du an mich?	OSLC
895	Couperin, F.	1713	Premier livre de Clavecin, Ordre II, No. 16, La Florentine	MS
896	Couperin, F.	1716	Ordre 6ième du Clavecin, No. 5, Rondeau: Les Barricades Mystérieuses	MS
897	Couperin, F.	1722	Troisième livre de pièces de Clavecin, No. 18, mov. 4, Le petit rien	MS
898	Debussy, C.	1890	L. 68, Rêverie	MS
899	Debussy, C.	1891	Deux Arabesques, L. 66, mov. 1	MS
900	Debussy, C.	1891	Deux Arabesques, L. 66, mov. 2	MS
901	Debussy, C.	1905	Suite Bergamasque, L. 75, mov. 1, Prélude	MS
902	Debussy, C.	1905	Suite Bergamasque, L. 75, mov. 2, Menuet	MS
903	Debussy, C.	1905	Suite Bergamasque, L. 75, mov. 3, Clair de Lune	MS
904	Debussy, C.	1905	Suite Bergamasque, L. 75, mov. 4, Passepied	MS

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
905	Debussy, C.	1908	Children's corner, L. 113, No. 1, Doctor Gradus ad Parnassum	MS
906	Debussy, C.	1908	Children's corner, L. 113, No. 2, Jimbo's Lullaby	MS
907	Debussy, C.	1908	Children's corner, L. 113, No. 5, The Little Shepherd	MS
908	Debussy, C.	1908	Children's corner, L. 113, No. 6, Golliwogg's Cakewalk	MS
909	Debussy, C.	1909	Préludes I, The girl with the flaxen hair	MS
910	Desprez, J.	1486	Missa Ave Maris Stella, Agnus II	ELVIS
911	Desprez, J.	1486	Missa Ave Maris Stella, Benedictus	ELVIS
912	Desprez, J.	1486	Missa Ave Maris Stella, Qui venit	ELVIS
913	Desprez, J.	1486	Credo de tous biens playne, Et in spiritum	ELVIS
914	Desprez, J.	1486	Missa Gaudeamus, Agnus II	ELVIS
915	Desprez, J.	1486	Missa Gaudeamus, Benedictus	ELVIS
916	Desprez, J.	1486	Missa Gaudeamus, In nomine	ELVIS
917	Desprez, J.	1486	Missa Hercules Dux Ferrarie, Benedictus	ELVIS
918	Desprez, J.	1486	Missa Hercules Dux Ferrarie, Pleni	ELVIS
919	Desprez, J.	1486	Missa Hercules Dux Ferrarie, In nomine	ELVIS
920	Desprez, J.	1486	Missa La Sol Fa Re, Agnus II	ELVIS
921	Desprez, J.	1486	Missa Malheur me bat, Benedictus	ELVIS
922	Desprez, J.	1486	Missa Pange Lingua, Agnus II	ELVIS
923	Desprez, J.	1486	Missa Pange Lingua, Benedictus	ELVIS
924	Desprez, J.	1486	Missa Pange Lingua, Pleni	ELVIS
925	Desprez, J.	1486	Missa Sine Nomine, Agnus II	ELVIS
926	Desprez, J.	1486	Missa Sine Nomine, Benedictus	ELVIS
927	Desprez, J.	1486	Missa Sine Nomine, In nomine	ELVIS
928	Desprez, J.	1486	Missa Sine Nomine, Pleni	ELVIS
929	Desprez, J.	1486	Missa Sine Nomine, Qui venit	ELVIS
930	Desprez, J.	1486	Missa de Beata Virgine, Agnus II	ELVIS
931	Desprez, J.	1486	Missa l'Homme Armé Sexti Toni, Benedictus	ELVIS
932	Desprez, J.	1486	Missa l'Homme Armé Sexti Toni, Gloria Tua	ELVIS
933	Desprez, J.	1486	Missa l'Homme Armé Sexti Toni, In nomine	ELVIS
934	Desprez, J.	1486	Missa l'Homme Armé Sexti Toni, Pleni	ELVIS
935	Desprez, J.	1486	Missa l'Homme Armé Sexti Toni, Qui venit	ELVIS
936	Desprez, J.	1486	Missa Malheur me bat, In nomine	ELVIS
937	Dufay, G.	1472	Du tout m'estoie abandonnee	ELVIS
938	Dufay, G.	1474	Ad cenam agni providi I	ELVIS
939	Dufay, G.	1474	Ad cenam agni providi II	ELVIS
940	Dufay, G.	1474	Anima mea liquefacta est	ELVIS
941	Dufay, G.	1474	Ave maris stella	ELVIS
942	Dufay, G.	1474	Ave regina coelorum (3)	ELVIS
943	Dufay, G.	1474	Ave regina coelorum (4)	ELVIS
944	Dufay, G.	1474	Belle, que vous ay je meffait	ELVIS
945	Dufay, G.	1474	Bon jour, bon mois, bon an et bonne estraine	ELVIS
946	Dufay, G.	1474	Belle, vueillés moy vengier	ELVIS
947	Dufay, G.	1474	Conditor alme siderum	ELVIS
948	Dufay, G.	1474	Dieu gard la dame sans reprise	ELVIS
949	Dufay, G.	1474	Dona gentile, bella come l'oro	ELVIS
950	Dufay, G.	1474	Du tout m'estoie abandonné	ELVIS
951	Dufay, G.	1474	Ecclesiae militantis	ELVIS
952	Dufay, G.	1474	Gloria	ELVIS
953	Dufay, G.	1474	Gloria ad modum tube	ELVIS
954	Dufay, G.	1474	Hic iocundus sumit mundus	ELVIS
955	Dufay, G.	1474	Je nay doubte fors que des envieux	ELVIS
956	Dufay, G.	1474	Je ne suy plus teil que soloye	ELVIS
957	Dufay, G.	1474	Magnanimae gentis	ELVIS
958	Dufay, G.	1474	Missa Ava Maria coelorum, Gloria	ELVIS

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
959	Dufay, G.	1474	Missa Ava Maria coelorum, Kyrie	ELVIS
960	Dufay, G.	1474	Missa l'homme armé, Agnus Dei I	ELVIS
961	Dufay, G.	1474	Missa l'homme armé, Agnus Dei II	ELVIS
962	Dufay, G.	1474	Missa l'homme armé, Benedictus	ELVIS
963	Dufay, G.	1474	Missa l'homme armé, Christe	ELVIS
964	Dufay, G.	1474	Missa l'homme armé, Creado	ELVIS
965	Dufay, G.	1474	Missa l'homme armé, Et incarnatus est	ELVIS
966	Dufay, G.	1474	Missa l'homme armé, Gloria	ELVIS
967	Dufay, G.	1474	Missa l'homme armé, Kyrie	ELVIS
968	Dufay, G.	1474	Missa l'homme armé, Osanna II	ELVIS
969	Dufay, G.	1474	Missa l'homme armé, Osanna II	ELVIS
970	Dufay, G.	1474	Missa l'homme armé, Pleni sunt	ELVIS
971	Dufay, G.	1474	Missa l'homme armé, Sanctus	ELVIS
972	Dufay, G.	1474	Mon bien, m'amour	ELVIS
973	Dufay, G.	1474	Ne je dors, ne je veille	ELVIS
974	Dufay, G.	1474	Ou lit des pleurs	ELVIS
975	Dufay, G.	1474	Par le regard de vos beaux yeux	ELVIS
976	Dufay, G.	1474	Proles de celo prodiit	ELVIS
977	Dufay, G.	1474	Puisque celle qui me tient en prison	ELVIS
978	Dufay, G.	1474	Quel fronte signorille in paradiso	ELVIS
979	Dufay, G.	1474	Resistera	ELVIS
980	Dufay, G.	1474	Resvelons nous, resvelons, amoureux	ELVIS
981	Dufay, G.	1474	Salve flos Tusce	ELVIS
982	Dufay, G.	1474	Si queris miracula	ELVIS
983	Dufay, G.	1474	Vergene bella	ELVIS
984	Dufay, G.	1474	Vostre bruit et vostre grant fame	ELVIS
985	Dvorak, A.	1870	Silhouettes, Op. 8, No. 1, Volume 1	DCML
986	Dvorak, A.	1870	Silhouettes, Op. 8, No. 2, Volume 1	DCML
987	Dvorak, A.	1870	Silhouettes, Op. 8, No. 3, Volume 1	DCML
988	Dvorak, A.	1870	Silhouettes, Op. 8, No. 4, Volume 1	DCML
989	Dvorak, A.	1870	Silhouettes, Op. 8, No. 5, Volume 1	DCML
990	Dvorak, A.	1870	Silhouettes, Op. 8, No. 6, Volume 1	DCML
991	Dvorak, A.	1875	Silhouettes, Op. 8, No. 10, Volume 2	DCML
992	Dvorak, A.	1875	Silhouettes, Op. 8, No. 11, Volume 2	DCML
993	Dvorak, A.	1875	Silhouettes, Op. 8, No. 12, Volume 2	DCML
994	Dvorak, A.	1875	Silhouettes, Op. 8, No. 7, Volume 2	DCML
995	Dvorak, A.	1875	Silhouettes, Op. 8, No. 8, Volume 2	DCML
996	Dvorak, A.	1875	Silhouettes, Op. 8, No. 9, Volume 2	DCML
997	Elgar, E.	1888	Op. 12, Salut d'Amour	MS
998	Fauré, G.	1870	Op. 4, No. 1, La chanson du pêcheur	MS
999	Fauré, G.	1871	Op. 3, No. 2, Sérénade Toscane	MS
1000	Fauré, G.	1873	Op. 10, No. 1, Puisqu'ici bas tout âme	MS
1001	Fauré, G.	1876	Op. 2, No. 2, Les Matelots	MS
1002	Fauré, G.	1878	Op. 18, No. 3, Automne	MS
1003	Fauré, G.	1879	Op. 23, No. 1, Les berceaux	MS
1004	Fauré, G.	1880	Op. 6, No. 2, Tristesse	MS
1005	Fauré, G.	1887	Op. 8, No. 1, Au bord de l'eau	MS
1006	Fauré, G.	1914	Op. 106, Le jardin clos	MS
1007	Finzi, G.	1942	Let us Garlands bring, Op. 18, No. 1, Come away, come away, death	MS
1008	Finzi, G.	1942	Let us Garlands bring, Op. 18, No. 2, Who is Silvia?	MS
1009	Finzi, G.	1942	Let us Garlands bring, Op. 18, No. 3, Fear no more the heat o' the sun	MS
1010	Finzi, G.	1942	Let us Garlands bring, Op. 18, No. 4, O Mistress Mine	MS
1011	Finzi, G.	1942	Let us Garlands bring, Op. 18, No. 5, It was a lover and his lass	MS
1012	Franz, R.	1851	Sechs Gesänge, Op. 14, No. 1, Widmung	OSLC

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
1013	Franz, R.	1851	Sechs Gesänge, Op. 14, No. 2, Lenz	OSLC
1014	Franz, R.	1851	Sechs Gesänge, Op. 14, No. 3, Waldfahrt	OSLC
1015	Franz, R.	1851	Sechs Gesänge, Op. 14, No. 4, Volkslied	OSLC
1016	Franz, R.	1851	Sechs Gesänge, Op. 14, No. 5, Liebesfrühling	OSLC
1017	Franz, R.	1851	Sechs Gesänge, Op. 14, No. 6, Frage nicht	OSLC
1018	Frescobaldi, G.	1635	Fiori Musicali, No. 1, Toccata avanti la Messa della Domenica	MS
1019	Frescobaldi, G.	1635	Fiori Musicali, No. 10, Bergamasca	MS
1020	Frescobaldi, G.	1635	Fiori Musicali, No. 11, Capriccio sopra la Girolmeta	MS
1021	Frescobaldi, G.	1635	Fiori Musicali, No. 2, Canzon dopo l'Epistola	MS
1022	Frescobaldi, G.	1635	Fiori Musicali, No. 3, Toccata cromaticha per l'Elevatione	MS
1023	Frescobaldi, G.	1635	Fiori Musicali, No. 4, Recercar cromaticho post il Credo	MS
1024	Frescobaldi, G.	1635	Fiori Musicali, No. 5, Altro recercar	MS
1025	Frescobaldi, G.	1635	Fiori Musicali, No. 6, Toccata per l'Elevatione	MS
1026	Frescobaldi, G.	1635	Fiori Musicali, No. 7, Canzon quarti toni dopo il post Comune	MS
1027	Frescobaldi, G.	1635	Fiori Musicali, No. 8, Recercar con obligo di cantare la quinta parte senza toccarla	MS
1028	Frescobaldi, G.	1635	Fiori Musicali, La Messa della Madonna, No. 9, Toccata per l'Elevatione (1635)	MS
1029	Gesualdo, C.	1603	Sacrae cantiones I, No. 1, Ave regina coelorum	CPDL
1030	Gesualdo, C.	1603	Sacrae cantiones I, No. 10, Peccantum me	CPDL
1031	Gesualdo, C.	1603	Sacrae cantiones I, No. 11, O vos omnes	CPDL
1032	Gesualdo, C.	1603	Sacrae cantiones I, No. 12, Exaudi Deus	CPDL
1033	Gesualdo, C.	1603	Sacrae cantiones I, No. 13, Precibus et meritis	CPDL
1034	Gesualdo, C.	1603	Sacrae cantiones I, No. 14, O crux benedicta	CPDL
1035	Gesualdo, C.	1603	Sacrae cantiones I, No. 15, Tribularer si nescirem	CPDL
1036	Gesualdo, C.	1603	Sacrae cantiones I, No. 16, Deus refugium et virtus	CPDL
1037	Gesualdo, C.	1603	Sacrae cantiones I, No. 17, Tribulationem et dolorem	CPDL
1038	Gesualdo, C.	1603	Sacrae cantiones I, No. 18, Illumina faciem tuam	CPDL
1039	Gesualdo, C.	1603	Sacrae cantiones I, No. 19, Maria, mater gratiae	CPDL
1040	Gesualdo, C.	1603	Sacrae cantiones I, No. 3, Ave, dulcissima Maria	CPDL
1041	Gesualdo, C.	1603	Sacrae cantiones I, No. 4, Reminiscere	CPDL
1042	Gesualdo, C.	1603	Sacrae cantiones I, No. 5, Dignare me	CPDL
1043	Gesualdo, C.	1603	Sacrae cantiones I, No. 6, Sancti Spiritus Domine	CPDL
1044	Gesualdo, C.	1603	Sacrae cantiones I, No. 7, Domine ne despicias	CPDL
1045	Gesualdo, C.	1603	Sacrae cantiones I, No. 9, Laboravi in gemitu meo	CPDL
1046	Gesualdo, C.	1611	Responses of Tenebrae Responsorium, No. 1, Sicut ovis ad occisionem	CPDL
1047	Gesualdo, C.	1611	Responses of Tenebrae Responsorium, No. 3, Plange quasi virgo	CPDL
1048	Gesualdo, C.	1611	Responses of Tenebrae Responsorium, No. 4, Recessit pastor noster	CPDL
1049	Gesualdo, C.	1611	Responses of Tenebrae Responsorium, No. 5, O vos omnes	CPDL
1050	Gesualdo, C.	1611	Responses of Tenebrae Responsorium, No. 6, Ecce quomodo moritur justus	CPDL
1051	Gesualdo, C.	1611	Responses of Tenebrae Responsorium, No. 7, Astiterunt reges terrae	CPDL
1052	Gesualdo, C.	1611	Responses of Tenebrae Responsorium, No. 8, Aestimatus sum	CPDL
1053	Gesualdo, C.	1611	Responses of Tenebrae Responsorium, No. 9, Sepulto Domino	CPDL
1054	Gesualdo, C.	1611	Il sesti libri di madrigali, Beltà pio che t'assenti	CPDL
1055	Gesualdo, C.	1613	Moro lasso	MS
1056	Granados, E.	1894	Valses poéticos, mov. 1, Melodico	MS
1057	Granados, E.	1894	Valses poéticos, mov. 2, Tempo de Vals noble	MS
1058	Granados, E.	1894	Valses poéticos, mov. 3, Tempo de Vals lento	MS
1059	Granados, E.	1894	Valses poéticos, Introducción	MS
1060	Granados, E.	1900	Valses poéticos, mov. 4, Allegro humoristico	MS
1061	Granados, E.	1900	Valses poéticos, mov. 5, Allegretto elegante	MS
1062	Granados, E.	1900	Valses poéticos, mov. 6, Quasi ad libitum sentimental	MS

Continued on next page

Table B.3 – Corpus used in Part IV (“Macroanalysis”; cont.).

#	Name	Year	Title	Source
1063	Granados, E.	1900	Valses poéticos, mov. 7, Vivo	MS
1064	Granados, E.	1900	Valses poéticos, Coda	MS
1065	Granados, E.	1909	Goyescas, Op. 11, No. 1, Los requiebros	MS
1066	Granados, E.	1909	Goyescas, Op. 11, No. 2, Coloquio en la reja	MS
1067	Granados, E.	1909	Goyescas, Op. 11, No. 3, El fandango de candil	MS
1068	Granados, E.	1909	Goyescas, Op. 11, No. 4, Quejas, o La Maya y el ruiseñor	MS
1069	Granados, E.	1909	Goyescas, Op. 11, No. 5, El Amor y la muerte (balada)	MS
1070	Grieg, E.	1865	Piano Sonata, Op. 7, No. 1	MS
1071	Grieg, E.	1865	Piano Sonata, Op. 7, No. 2	MS
1072	Grieg, E.	1865	Piano Sonata, Op. 7, No. 3	MS
1073	Grieg, E.	1865	Piano Sonata, Op. 7, No. 4	MS
1074	Grieg, E.	1866	Lyrical Pieces, Op. 12, No. 1, Arietta	DCML
1075	Grieg, E.	1866	Lyrical Pieces, Op. 12, No. 2	DCML
1076	Grieg, E.	1866	Lyrical Pieces, Op. 12, No. 2, Waltz	MS
1077	Grieg, E.	1866	Lyrical Pieces, Op. 12, No. 3	DCML
1078	Grieg, E.	1866	Lyrical Pieces, Op. 12, No. 4, Alfedans	DCML
1079	Grieg, E.	1866	Lyrical Pieces, Op. 12, No. 6	DCML
1080	Grieg, E.	1866	Lyrical Pieces, Op. 12, No. 7	DCML
1081	Grieg, E.	1866	Lyrical Pieces, Op. 12, No. 8	DCML
1082	Grieg, E.	1886	Lyrical Pieces, Op. 43, No. 1, Sommerfugl	DCML
1083	Grieg, E.	1886	Lyrical Pieces, Op. 47, No. 2, Albumblad	DCML
1084	Grieg, E.	1889	Lyrical Pieces, Op. 54, No. 3, Trolltog	DCML
1085	Grieg, E.	1896	Lyrical Pieces, Op. 65, No. 1, Fra ungdomsdagene	DCML
1086	Grieg, E.	1896	Lyrical Pieces, Op. 65, No. 6, Bryllupsdag	DCML
1087	Grieg, E.	1898	Lyrical Pieces, Op. 68, No. 1, Matrosenes oppsang	DCML
1088	Haydn, J.	1793	String Quartet no. 57, Op. 74, No. 1, Allegro	ELVIS
1089	Haydn, J.	1793	String Quartet no. 57, Op. 74, No. 1, Andante grazioso	ELVIS
1090	Haydn, J.	1793	String Quartet no. 57, Op. 74, No. 1, Minuetto - Allegro - Trio	ELVIS
1091	Haydn, J.	1793	String Quartet no. 57, Op. 74, No. 1, Vivace	ELVIS
1092	Haydn, J.	1793	String Quartet no. 58, Op. 74, No. 2, Allegro spritoso	ELVIS
1093	Haydn, J.	1793	String Quartet no. 58, Op. 74, No. 2, Andante grazioso	ELVIS
1094	Haydn, J.	1793	String Quartet no. 58, Op. 74, No. 2, Menuet	ELVIS
1095	Haydn, J.	1793	String Quartet no. 58, Op. 74, No. 2, Finale: Vivace	ELVIS
1096	Hensel, F.	1846	Sechs Lieder, Op. 1, No. 1, Schwanenlied	OSLC
1097	Hensel, F.	1846	Sechs Lieder, Op. 1, No. 2, Wanderlied	OSLC
1098	Hensel, F.	1846	Sechs Lieder, Op. 1, No. 3, Warum sind denn die Rosen so blass?	OSLC
1099	Hensel, F.	1846	Sechs Lieder, Op. 1, No. 4, Mayenlied	OSLC
1100	Hensel, F.	1846	Sechs Lieder, Op. 1, No. 5, Morgenständchen	OSLC
1101	Hensel, F.	1846	Sechs Lieder, Op. 1, No. 6, Gondellied	OSLC
1102	Hensel, F.	1850	Fünf Lieder, Op. 10, No. 1, Nach Süden	OSLC
1103	Hensel, F.	1850	Fünf Lieder, Op. 10, No. 2, Vorwurf	OSLC
1104	Hensel, F.	1850	Fünf Lieder, Op. 10, No. 3, Abendbild	OSLC
1105	Hensel, F.	1850	Fünf Lieder, Op. 10, No. 4, Im Herbste	OSLC
1106	Hensel, F.	1850	Fünf Lieder, Op. 10, No. 5, Bergeslust	OSLC
1107	Hensel, F.	1850	Sechs Lieder, Op. 9, No. 1, Die Ersehnte	OSLC
1108	Hensel, F.	1850	Sechs Lieder, Op. 9, No. 2, Ferne	OSLC
1109	Hensel, F.	1850	Sechs Lieder, Op. 9, No. 3, Der Rosenkranz	OSLC
1110	Hensel, F.	1850	Sechs Lieder, Op. 9, No. 4, Die frühen Gräber	OSLC
1111	Hensel, F.	1850	Sechs Lieder, Op. 9, No. 5, Der Maiabend	OSLC
1112	Hensel, F.	1850	Sechs Lieder, Op. 9, No. 6, Die Mainacht	OSLC
1113	Henselt, A. von	1838	Études Caracteristiques, Op. 2, No. 1.0, Orage, tu ne saurais m’abattre	MS
1114	Henselt, A. von	1838	Études Caracteristiques, Op. 2, No. 2.0, Pensez un peu à moi, qui pense toujours à vous	MS
1115	Henselt, A. von	1838	Études Caracteristiques, Op. 2, No. 3.0, Exauce mes vœux!	MS

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
1116	Hummel, J. N.	1815	Preludes, Op. 67, No. 1, Quasi Improvisazione	ELVIS
1117	Hummel, J. N.	1815	Preludes, Op. 67, No. 10, Energico	ELVIS
1118	Hummel, J. N.	1815	Preludes, Op. 67, No. 11, Allegro	ELVIS
1119	Hummel, J. N.	1815	Preludes, Op. 67, No. 12, Allegro moderato	ELVIS
1120	Hummel, J. N.	1815	Preludes, Op. 67, No. 13, Allegro con fuoco	ELVIS
1121	Hummel, J. N.	1815	Preludes, Op. 67, No. 14, Alla cappella	ELVIS
1122	Hummel, J. N.	1815	Preludes, Op. 67, No. 15, Allegro moderato	ELVIS
1123	Hummel, J. N.	1815	Preludes, Op. 67, No. 16, Sostenuto	ELVIS
1124	Hummel, J. N.	1815	Preludes, Op. 67, No. 17, Allegro moderato	ELVIS
1125	Hummel, J. N.	1815	Preludes, Op. 67, No. 18, Con molto fuoco	ELVIS
1126	Hummel, J. N.	1815	Preludes, Op. 67, No. 19, Allegro comodo	ELVIS
1127	Hummel, J. N.	1815	Preludes, Op. 67, No. 2, Allegro moderato	ELVIS
1128	Hummel, J. N.	1815	Preludes, Op. 67, No. 20, Andante con moto	ELVIS
1129	Hummel, J. N.	1815	Preludes, Op. 67, No. 21, Allegro con brio	ELVIS
1130	Hummel, J. N.	1815	Preludes, Op. 67, No. 22, Allegro vivace	ELVIS
1131	Hummel, J. N.	1815	Preludes, Op. 67, No. 23, Allegro con fuoco	ELVIS
1132	Hummel, J. N.	1815	Preludes, Op. 67, No. 24, Allegro spritoso	ELVIS
1133	Hummel, J. N.	1815	Preludes, Op. 67, No. 3, Allegro	ELVIS
1134	Hummel, J. N.	1815	Preludes, Op. 67, No. 4, Molto allegro	ELVIS
1135	Hummel, J. N.	1815	Preludes, Op. 67, No. 5, Allegro con fuoco	ELVIS
1136	Hummel, J. N.	1815	Preludes, Op. 67, No. 6, Allegro non troppo	ELVIS
1137	Hummel, J. N.	1815	Preludes, Op. 67, No. 7, Moderato	ELVIS
1138	Hummel, J. N.	1815	Preludes, Op. 67, No. 8, Allegro molto animato	ELVIS
1139	Hummel, J. N.	1815	Preludes, Op. 67, No. 9, Allegro	ELVIS
1140	Händel, G. F.	1733	Chaconne in G major, HWV 435	MS
1141	Händel, G. F.	1737	Funeral Anthem for Queen Caroline, HWV 264, mov. 10, Their bodies are buried	KPM
1142	Händel, G. F.	1737	Funeral Anthem for Queen Caroline, HWV 264, mov. 11, The people will tell	KPM
1143	Händel, G. F.	1737	Funeral Anthem for Queen Caroline, HWV 264, mov. 13, The merci-full goodness	KPM
1144	Händel, G. F.	1737	Funeral Anthem for Queen Caroline, HWV 264, mov. 2, The ways of Zion do mourn	KPM
1145	Händel, G. F.	1737	Funeral Anthem for Queen Caroline, HWV 264, mov. 4, She put on righteousness	KPM
1146	Händel, G. F.	1737	Funeral Anthem for Queen Caroline, HWV 264, mov. 9, The righteous shall be	KPM
1147	Isaac, H.	1517	A la battaglia	ELVIS
1148	Isaac, H.	1517	Alma Redemptoris Mater	ELVIS
1149	Isaac, H.	1517	Helas que pourras devenir	ELVIS
1150	Isaac, H.	1517	Innsbruck ich muss dich lassen	ELVIS
1151	Joplin, S.	1896	Ragtimes, Combination March	CCARH
1152	Joplin, S.	1896	Ragtimes, The Crush Collision March	CCARH
1153	Joplin, S.	1896	Ragtimes, Harmony Club Waltz	CCARH
1154	Joplin, S.	1899	Ragtimes, Maple Leaf Rag	CCARH
1155	Joplin, S.	1899	Ragtimes, Original Rags	CCARH
1156	Joplin, S.	1899	Ragtimes, Palm Leaf Rag	CCARH
1157	Joplin, S.	1900	Ragtimes, Swipesy	CCARH
1158	Joplin, S.	1901	Ragtimes, Augustan Club Waltz	CCARH
1159	Joplin, S.	1901	Ragtimes, The Easy Winners	CCARH
1160	Joplin, S.	1901	Ragtimes, Peacherine Rag	CCARH
1161	Joplin, S.	1901	Ragtimes, Sunflower Slow Drag	CCARH
1162	Joplin, S.	1902	Ragtimes, A Breeze from Alabama	CCARH
1163	Joplin, S.	1902	Ragtimes, Cleopha	CCARH

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
1164	Joplin, S.	1902	Ragtimes, Elite Syncopations	CCARH
1165	Joplin, S.	1902	Ragtimes, The Entertainer	CCARH
1166	Joplin, S.	1902	Ragtimes, March Majestic	CCARH
1167	Joplin, S.	1903	Ragtimes, Something Doing	CCARH
1168	Joplin, S.	1903	Ragtimes, Weeping Willow	CCARH
1169	Joplin, S.	1904	Ragtimes, The Cascades	CCARH
1170	Joplin, S.	1904	Ragtimes, The Chrysanthemum	CCARH
1171	Joplin, S.	1905	Ragtimes, Bethena	CCARH
1172	Joplin, S.	1905	Ragtimes, Binks's Waltz	CCARH
1173	Joplin, S.	1905	Ragtimes, Eugenia	CCARH
1174	Joplin, S.	1905	Ragtimes, Leola	CCARH
1175	Joplin, S.	1905	Ragtimes, The Rosebud March	CCARH
1176	Joplin, S.	1906	Ragtimes, Antoinette	CCARH
1177	Joplin, S.	1907	Ragtimes, Gladiolus Rag	CCARH
1178	Joplin, S.	1907	Ragtimes, Lily Queen	CCARH
1179	Joplin, S.	1907	Ragtimes, The Nonpareil	CCARH
1180	Joplin, S.	1907	Ragtimes, Rose Leaf Rag	CCARH
1181	Joplin, S.	1907	Ragtimes, Searchlight Rag	CCARH
1182	Joplin, S.	1908	Ragtimes, Fig Leaf Rag	CCARH
1183	Joplin, S.	1908	Ragtimes, Pine Apple Rag	CCARH
1184	Joplin, S.	1908	Ragtimes, Sugar Cane	CCARH
1185	Joplin, S.	1909	Ragtimes, Country Club	CCARH
1186	Joplin, S.	1909	Ragtimes, Paragon Rag	CCARH
1187	Joplin, S.	1909	Ragtimes, Pleasant Moments	CCARH
1188	Joplin, S.	1909	Ragtimes, Wall Street Rag	CCARH
1189	Joplin, S.	1910	Ragtimes, Stoptime Rag	CCARH
1190	Joplin, S.	1912	Ragtimes, Scott Joplin's New Rag	CCARH
1191	Joplin, S.	1914	Ragtimes, Magnetic Rag	CCARH
1192	Joplin, S.	1917	Ragtimes, Reflection Rag	CCARH
1193	Koželuh, L.	1784	Piano Sonata, Op. 10, No. 1, mov. 1, Allegro molto	DB
1194	Koželuh, L.	1784	Piano Sonata, Op. 13, No. 2, mov. 1, Allegro	DB
1195	Koželuh, L.	1784	Piano Sonata, Op. 13, No. 2, mov. 2, Andante	DB
1196	Koželuh, L.	1784	Piano Sonata, Op. 13, No. 2, mov. 3, Rondeau	DB
1197	Koželuh, L.	1784	Piano Sonata, Op. 13, No. 3, mov. 1, Allegro molto	DB
1198	Koželuh, L.	1784	Piano Sonata, Op. 13, No. 3, mov. 2, Cantabile	DB
1199	Koželuh, L.	1784	Piano Sonata, Op. 13, No. 3, mov. 3, Presto	DB
1200	Koželuh, L.	1784	Piano Sonata, Op. 15, No. 1, mov. 1, Largo - Allegro molto - Largo	DB
1201	Koželuh, L.	1784	Piano Sonata, Op. 15, No. 1, mov. 2, Rondeau - Allegro	DB
1202	Koželuh, L.	1784	Piano Sonata, Op. 8, No. 1, mov. 1, Allegro molto	DB
1203	Koželuh, L.	1784	Piano Sonata, Op. 8, No. 1, mov. 2, Rondo - Andante	DB
1204	Koželuh, L.	1784	Piano Sonata, Op. 8, No. 2, mov. 1, Allegro maestoso	DB
1205	Koželuh, L.	1784	Piano Sonata, Op. 8, No. 2, mov. 2, Rondo - Andante	DB
1206	Koželuh, L.	1784	Piano Sonata, Op. 8, No. 2, mov. 3b, Aria con variatione	DB
1207	Koželuh, L.	1785	Piano Sonata, Op. 15, No. 2, mov. 1, Allegro	DB
1208	Koželuh, L.	1785	Piano Sonata, Op. 15, No. 2, mov. 2, Poco adagio	DB
1209	Koželuh, L.	1785	Piano Sonata, Op. 15, No. 2, mov. 3, Presto	DB
1210	Koželuh, L.	1785	Piano Sonata, Op. 17, No. 1, mov. 1, Largo	DB
1211	Koželuh, L.	1785	Piano Sonata, Op. 17, No. 1, mov. 2, Allegro agitato	DB
1212	Koželuh, L.	1785	Piano Sonata, Op. 17, No. 1, mov. 3, Finale - Allegretto	DB
1213	Koželuh, L.	1785	Piano Sonata, Op. 17, No. 2, mov. 1, Allegro molto	DB
1214	Koželuh, L.	1785	Piano Sonata, Op. 17, No. 2, mov. 2, Adagio	DB
1215	Koželuh, L.	1785	Piano Sonata, Op. 17, No. 2, mov. 3, Allegro molto	DB
1216	Koželuh, L.	1785	Piano Sonata, Op. 17, No. 3, mov. 1, Allegro	DB
1217	Koželuh, L.	1785	Piano Sonata, Op. 17, No. 3, mov. 2, Rondeau - Allegretto	DB

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
1218	Koželuh, L.	1786	Piano Sonata, Op. 20, No. 1, mov. 1, Allegro	DB
1219	Koželuh, L.	1786	Piano Sonata, Op. 20, No. 1, mov. 2, Adagio	DB
1220	Koželuh, L.	1786	Piano Sonata, Op. 20, No. 1, mov. 3, Rondeau - Allegretto	DB
1221	Koželuh, L.	1786	Piano Sonata, Op. 20, No. 2, mov. 1, Allegro	DB
1222	Koželuh, L.	1786	Piano Sonata, Op. 20, No. 2, mov. 2, Adagio	DB
1223	Koželuh, L.	1786	Piano Sonata, Op. 20, No. 2, mov. 3, Rondeau - Allegretto	DB
1224	Koželuh, L.	1786	Piano Sonata, Op. 20, No. 3, mov. 1, Moderato	DB
1225	Koželuh, L.	1786	Piano Sonata, Op. 20, No. 3, mov. 2, Poco adagio	DB
1226	Koželuh, L.	1786	Piano Sonata, Op. 20, No. 3, mov. 3, Rondeau - Allegretto	DB
1227	Koželuh, L.	1788	Piano Sonata, Op. 26, No. 1, mov. 1, Allegro	DB
1228	Koželuh, L.	1788	Piano Sonata, Op. 26, No. 1, mov. 2, Adagio	DB
1229	Koželuh, L.	1788	Piano Sonata, Op. 26, No. 1, mov. 3, Rondeau - Allegretto	DB
1230	Koželuh, L.	1788	Piano Sonata, Op. 26, No. 2, mov. 1, Allegro	DB
1231	Koželuh, L.	1788	Piano Sonata, Op. 26, No. 2, mov. 2, Andante con variazione	DB
1232	Koželuh, L.	1788	Piano Sonata, Op. 26, No. 3, mov. 1, Allegro	DB
1233	Koželuh, L.	1788	Piano Sonata, Op. 26, No. 3, mov. 2, Larghetto alla Siziliana	DB
1234	Koželuh, L.	1788	Piano Sonata, Op. 26, No. 3, mov. 3, Rondeau Allegro con Fuoco	DB
1235	Koželuh, L.	1789	Piano Sonata, Op. 30, No. 1, mov. 1, Allegro	DB
1236	Koželuh, L.	1789	Piano Sonata, Op. 30, No. 1, mov. 2, Poco adagio	DB
1237	Koželuh, L.	1789	Piano Sonata, Op. 30, No. 1, mov. 3, Rondeau - Allegretto	DB
1238	Koželuh, L.	1789	Piano Sonata, Op. 30, No. 2, mov. 1, Allegro	DB
1239	Koželuh, L.	1789	Piano Sonata, Op. 30, No. 2, mov. 2, Andante	DB
1240	Koželuh, L.	1789	Piano Sonata, Op. 30, No. 2, mov. 3, Rondeau - Allegretto	DB
1241	Koželuh, L.	1789	Piano Sonata, Op. 30, No. 3, mov. 1, Largo - Allegro - Largo	DB
1242	Koželuh, L.	1789	Piano Sonata, Op. 30, No. 3, mov. 2, Rondeau - Allegretto	DB
1243	Koželuh, L.	1791	Piano Sonata, Op. 35, No. 1, mov. 1, Allegro	DB
1244	Koželuh, L.	1791	Piano Sonata, Op. 35, No. 1, mov. 2, Adagio	DB
1245	Koželuh, L.	1791	Piano Sonata, Op. 35, No. 1, mov. 3, Rondo - Allegretto	DB
1246	Koželuh, L.	1791	Piano Sonata, Op. 35, No. 2, mov. 1, Allegro	DB
1247	Koželuh, L.	1791	Piano Sonata, Op. 35, No. 2, mov. 2, Adagio	DB
1248	Koželuh, L.	1791	Piano Sonata, Op. 35, No. 2, mov. 3, Rondo - Allegro	DB
1249	Koželuh, L.	1791	Piano Sonata, Op. 35, No. 3, mov. 1, Largo - Allegro agitato	DB
1250	Koželuh, L.	1791	Piano Sonata, Op. 35, No. 3, mov. 2, Allegretto	DB
1251	Koželuh, L.	1793	Piano Sonata, Op. 38, No. 3, mov. 1, Largo - Allegro agitato - Allegretto	DB
1252	Koželuh, L.	1793	Piano Sonata, Op. 38, No. 3, mov. 2, Allegretto	DB
1253	Koželuh, L.	1803	Piano Sonata, Op. 51, No. 1, mov. 1, Allegro	DB
1254	Koželuh, L.	1803	Piano Sonata, Op. 51, No. 1, mov. 2, Adagio	DB
1255	Koželuh, L.	1803	Piano Sonata, Op. 51, No. 1, mov. 3, Rondo - Vivace	DB
1256	Koželuh, L.	1803	Piano Sonata, Op. 51, No. 2, mov. 1, Largo - Allegro molto	DB
1257	Koželuh, L.	1803	Piano Sonata, Op. 51, No. 2, mov. 2, Rondeau - Allegretto	DB
1258	Koželuh, L.	1803	Piano Sonata, Op. 51, No. 3, mov. 1, Largo - Allegro molto e agitato	DB
1259	Koželuh, L.	1803	Piano Sonata, Op. 51, No. 3, mov. 2, Rondeau - Allegretto	DB
1260	Lang, J.	1860	Sechs Lieder, Op. 25, No. 1, Frühlings-Glaube	OSLC
1261	Lang, J.	1860	Sechs Lieder, Op. 25, No. 2, Winterseufzer	OSLC
1262	Lang, J.	1860	Sechs Lieder, Op. 25, No. 3, Barcarole	OSLC
1263	Lang, J.	1860	Sechs Lieder, Op. 25, No. 4, Lied (Immer sich rein kindlich erfreu'n)	OSLC
1264	Lang, J.	1860	Sechs Lieder, Op. 25, No. 5, Die Wolken	OSLC
1265	Lang, J.	1860	Sechs Lieder, Op. 25, No. 6, Das Paradies	OSLC
1266	Lasso, O. di	1577	Beatus homo	ELVIS
1267	Lasso, O. di	1577	Beatus vir	ELVIS
1268	Lasso, O. di	1577	Expectatio justorum justitia	ELVIS
1269	Lasso, O. di	1577	Fulgebunt justi	ELVIS
1270	Lasso, O. di	1577	Justi tulerunt spolia	ELVIS
1271	Lasso, O. di	1577	Justus cor suum	ELVIS

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
1272	Lasso, O. di	1577	Oculos non vidit	ELVIS
1273	Lasso, O. di	1577	Qui sequitur me	ELVIS
1274	Lasso, O. di	1577	Qui vult venire post me	ELVIS
1275	Lasso, O. di	1577	Sancti mei	ELVIS
1276	Lasso, O. di	1577	Servi bone et fidelis	ELVIS
1277	Lasso, O. di	1577	Sicut rosa	ELVIS
1278	Lasso, O. di	1581	Matona mia cara	ELVIS
1279	Lasso, O. di	1582	Salve regina	ELVIS
1280	Lasso, O. di	1583	Missa „Je suis desherite", Kyrie	ELVIS
1281	Liszt, F.	1841	S. 394, Réminiscences de Norma	MS
1282	Liszt, F.	1847	Harmonies Poètiques et Réligieuses, S. 173, No. 3, Bénédiction de Dieu dans la Solitude	MS
1283	Liszt, F.	1847	Hungarian Rhapsodies, S. 244, No. 6, Hungarian Rhapsodie No. 6	MS
1284	Liszt, F.	1848	Liebesträume, S. 541, No. 3	MS
1285	Liszt, F.	1849	Trois Etudes de Concert, S. 144, No. 1, Il Lamento	MS
1286	Liszt, F.	1849	Trois Etudes de Concert, S. 144, No. 2, La Leggierezza	MS
1287	Liszt, F.	1849	Trois Etudes de Concert, S. 144, No. 3, Un Sospiro	MS
1288	Liszt, F.	1849	Années de pèlerinage II, S. 161, No. 7, Après une Lecture du Dante	MS
1289	Liszt, F.	1849	Consolations, S. 172, No. 3, Lento placido	MS
1290	Liszt, F.	1849	Harmonies Poètiques et Réligieuses, S. 173, No. 7, Funérailles	MS
1291	Liszt, F.	1851	12 Transcendental Etudes, S. 139, No. 1, Preludio	MS
1292	Liszt, F.	1851	12 Transcendental Etudes, S. 139, No. 10, Allegro Agitato Molto	MS
1293	Liszt, F.	1851	12 Transcendental Etudes, S. 139, No. 11, Harmonies du Soir	MS
1294	Liszt, F.	1851	12 Transcendental Etudes, S. 139, No. 12, Chasse-Neige	MS
1295	Liszt, F.	1851	12 Transcendental Etudes, S. 139, No. 2, Molto Vivace	MS
1296	Liszt, F.	1851	12 Transcendental Etudes, S. 139, No. 3, Paysage	MS
1297	Liszt, F.	1851	12 Transcendental Etudes, S. 139, No. 4, Mazeppa	MS
1298	Liszt, F.	1851	12 Transcendental Etudes, S. 139, No. 5, Feux Follets	MS
1299	Liszt, F.	1851	12 Transcendental Etudes, S. 139, No. 6, Vision	MS
1300	Liszt, F.	1851	12 Transcendental Etudes, S. 139, No. 7, Eroica	MS
1301	Liszt, F.	1851	12 Transcendental Etudes, S. 139, No. 8, Wilde Jagd	MS
1302	Liszt, F.	1851	12 Transcendental Etudes, S. 139, No. 9, Ricordanza	MS
1303	Liszt, F.	1853	Ballades, S. 171, Ballade No. 2	MS
1304	Liszt, F.	1853	Sonata, S. 178, B minor Sonata	MS
1305	Liszt, F.	1853	Hungarian Rhapsodies, S. 244, No. 15, Hungarian Rhapsodie No. 15	MS
1306	Liszt, F.	1855	Années de pèlerinage I, S. 160, No. 1, Chapelle de Guillaume Tell	MS
1307	Liszt, F.	1855	Années de pèlerinage I, S. 160, No. 2, Au lac de Wallenstadt	MS
1308	Liszt, F.	1855	Années de pèlerinage I, S. 160, No. 3, Pastorale	MS
1309	Liszt, F.	1855	Années de pèlerinage I, S. 160, No. 4, Au bord d'une source	MS
1310	Liszt, F.	1855	Années de pèlerinage I, S. 160, No. 5, Orage	MS
1311	Liszt, F.	1855	Années de pèlerinage I, S. 160, No. 6, Vallée d'Obermann	MS
1312	Liszt, F.	1855	Années de pèlerinage I, S. 160, No. 7, Eglogue	MS
1313	Liszt, F.	1855	Années de pèlerinage I, S. 160, No. 8, Le mal du pays	MS
1314	Liszt, F.	1855	Années de pèlerinage I, S. 160, No. 9, Les cloches de Genève	MS
1315	Liszt, F.	1856	S. 514, Mephisto Waltz No. 1	MS
1316	Liszt, F.	1858	Années de pèlerinage II, S. 161, No. 1, Sposalizio	MS
1317	Liszt, F.	1861	Venezia e Napoli, S. 162, No. 3, Tarantella	ELVIS
1318	Liszt, F.	1863	Zwei Konzertetüden, S. 145, No. 1, Waldesrauschen	MS
1319	Liszt, F.	1863	Zwei Konzertetüden, S. 145, No. 2, Gnomenreigen	MS
1320	Liszt, F.	1872	S. 191, Impromptu in F♯ major	MS
1321	Liszt, F.	1881	Late Pieces, S. 198, Wiegenlied	MS
1322	Liszt, F.	1881	Late Pieces, S. 199, Nuages gris	MS
1323	Liszt, F.	1882	Late Pieces, S. 200, No. 1, La lugubre gondola	DCML
1324	Liszt, F.	1882	Late Pieces, S. 200, No. 2, La lugubre gondola	DCML

Continued on next page

Table B.3 – Corpus used in Part IV (“Macroanalysis”; cont.).

#	Name	Year	Title	Source
1325	Liszt, F.	1883	Late Pieces, S. 201, R. W. Venezia	MS
1326	Liszt, F.	1883	Late Pieces, S. 202, Am Grabe Richard Wagners	DCML
1327	Liszt, F.	1883	Late Pieces, S. 203, Schlaflos! Frage und Antwort	DCML
1328	Liszt, F.	1885	Late Pieces, S. 206, Trauervorspiel und Trauermarsch	DCML
1329	Liszt, F.	1885	Late Pieces, S. 207, En rêve	DCML
1330	Liszt, F.	1885	Bagatelle sans Tonalité, S. 216a	MS
1331	Lyapunov, S. M.	1893	Op. 8, Nocturne	MS
1332	Lyapunov, S. M.	1896	7 Preludes, Op. 6, No. 1	MS
1333	Lyapunov, S. M.	1896	7 Preludes, Op. 6, No. 3	MS
1334	Lyapunov, S. M.	1896	7 Preludes, Op. 6, No. 6	MS
1335	Lyapunov, S. M.	1900	12 Études d’exécution transcendante, Op. 11, No. 1, Berceuse	MS
1336	Lyapunov, S. M.	1906	Op. 26, Chant d’automne	MS
1337	Lyapunov, S. M.	1906	Op. 27, Sonata	MS
1338	Machaut, G. de	1370	Gloria	ELVIS
1339	Machaut, G. de	1370	Messe de Nostre Dame - Kyrie	ELVIS
1340	Machaut, G. de	1377	Dame, ce vous n’aperceu	ELVIS
1341	Machaut, G. de	1377	Comment puet on mieus	ELVIS
1342	Machaut, G. de	1377	De toutes flours	ELVIS
1343	Machaut, G. de	1377	Cinc, un, treze	ELVIS
1344	Mahler, G.	1897	Lieder eines fahrenden Gesellen, No. 1, Wenn mein Schatz heut Hochzeit macht	OSLC
1345	Mahler, G.	1897	Lieder eines fahrenden Gesellen, No. 2, Ging heut morgen übers Feld	OSLC
1346	Mahler, G.	1897	Lieder eines fahrenden Gesellen, No. 3, Ich hab’ ein glühend Messer	OSLC
1347	Mahler, G.	1897	Lieder eines fahrenden Gesellen, No. 4, Die zwei blauen Augen von meinem Schatz	OSLC
1348	Mahler, G.	1904	Kindertotenlieder, No. 1, Nun will die Sonn’ so hell aufgeh’n	OSLC
1349	Mahler, G.	1904	Kindertotenlieder, No. 2, Nun seh’ ich wohl, warum so dunkle Flammen	OSLC
1350	Mahler, G.	1904	Kindertotenlieder, No. 3, Wenn dein Mütterlein	OSLC
1351	Mahler, G.	1904	Kindertotenlieder, No. 4, Oft denk’ ich sie sind nur ausgegangen	OSLC
1352	Mendelssohn, F.	1835	WoO 3, Scherz à Capriccio	MS
1353	Mendelssohn, F.	1837	String Quartet, Op. 44, No. 2, Scherzo: Allegro di molto	MS
1354	Mendelssohn, F.	1838	String Quartet, Op. 44, No. 1, mov. 1, Molto allegro vivace	MS
1355	Mendelssohn, F.	1838	String Quartet, Op. 44, No. 1, mov. 4, Presto con brio	MS
1356	Mendelssohn, F.	1847	String Quartet, Op. 80, Allegro vivace assai	MS
1357	Mendelssohn, F.	1847	String Quartet, Op. 80, Allegro assai	MS
1358	Mendelssohn, F.	1847	String Quartet, Op. 80, Adagio	MS
1359	Mendelssohn, F.	1847	String Quartet, Op. 80, Finale: Allegro molto	MS
1360	Monteverdi, C.	1607	Felle amaro / Cruda Amarilli	ELVIS
1361	Monteverdi, C.	1607	Maria quid ploras / Dorinda	ELVIS
1362	Monteverdi, C.	1607	Ma tu più che mai	ELVIS
1363	Monteverdi, C.	1607	Pulchrae sunt / Ferrir quel pretto	ELVIS
1364	Monteverdi, C.	1607	Qui pependit in cruce / Ecce Silvio	ELVIS
1365	Monteverdi, C.	1607	Sancta Maria / Deh bella e cara	ELVIS
1366	Monteverdi, C.	1607	Stabat Virgo Maria / Era l’anima mea	ELVIS
1367	Monteverdi, C.	1607	Te Jesu Christe / Ecco piegando	ELVIS
1368	Monteverdi, C.	1607	Ure me / Troppo ben può	ELVIS
1369	Monteverdi, C.	1607	Viues in corde / Ahi come un vago sol	ELVIS
1370	Monteverdi, C.	1641	SV 272, Laudate Dominum omnes gentes	ELVIS
1371	Monteverdi, C.	1641	Gloria tua / T’amo mia vita	ELVIS
1372	Mozart, W. A.	1774	Sonaten, KV 279, No. 1, mov. 1	CCARH
1373	Mozart, W. A.	1774	Sonaten, KV 279, No. 1, mov. 2	CCARH
1374	Mozart, W. A.	1774	Sonaten, KV 279, No. 1, mov. 3	CCARH
1375	Mozart, W. A.	1774	Sonaten, KV 280, No. 2, mov. 1	CCARH

Continued on next page

Table B.3 – Corpus used in Part IV (“Macroanalysis”; cont.).

#	Name	Year	Title	Source
1376	Mozart, W. A.	1774	Sonaten, KV 280, No. 2, mov. 2	CCARH
1377	Mozart, W. A.	1774	Sonaten, KV 280, No. 2, mov. 3	CCARH
1378	Mozart, W. A.	1774	Sonaten, KV 281, No. 3, mov. 1	CCARH
1379	Mozart, W. A.	1774	Sonaten, KV 281, No. 3, mov. 2	CCARH
1380	Mozart, W. A.	1774	Sonaten, KV 281, No. 3, mov. 3	CCARH
1381	Mozart, W. A.	1774	Sonaten, KV 282, No. 4, mov. 1	CCARH
1382	Mozart, W. A.	1774	Sonaten, KV 282, No. 4, mov. 2	CCARH
1383	Mozart, W. A.	1774	Sonaten, KV 282, No. 4, mov. 3	CCARH
1384	Mozart, W. A.	1774	Sonaten, KV 283, No. 5, mov. 1	CCARH
1385	Mozart, W. A.	1774	Sonaten, KV 283, No. 5, mov. 2	CCARH
1386	Mozart, W. A.	1774	Sonaten, KV 283, No. 5, mov. 3	CCARH
1387	Mozart, W. A.	1774	Sonaten, KV 284, No. 6, mov. 1	CCARH
1388	Mozart, W. A.	1774	Sonaten, KV 284, No. 6, mov. 2	CCARH
1389	Mozart, W. A.	1774	Sonaten, KV 284, No. 6, mov. 3, Thema	CCARH
1390	Mozart, W. A.	1774	Sonaten, KV 284, No. 6, mov. 3, Var. 1	CCARH
1391	Mozart, W. A.	1774	Sonaten, KV 284, No. 6, mov. 3, Var. 2	CCARH
1392	Mozart, W. A.	1774	Sonaten, KV 284, No. 6, mov. 3, Var. 3	CCARH
1393	Mozart, W. A.	1774	Sonaten, KV 284, No. 6, mov. 3, Var. 4	CCARH
1394	Mozart, W. A.	1774	Sonaten, KV 284, No. 6, mov. 3, Var. 5	CCARH
1395	Mozart, W. A.	1774	Sonaten, KV 284, No. 6, mov. 3, Var. 6	CCARH
1396	Mozart, W. A.	1774	Sonaten, KV 284, No. 6, mov. 3, Var. 7	CCARH
1397	Mozart, W. A.	1774	Sonaten, KV 284, No. 6, mov. 3, Var. 8	CCARH
1398	Mozart, W. A.	1774	Sonaten, KV 284, No. 6, mov. 3, Var. 9	CCARH
1399	Mozart, W. A.	1774	Sonaten, KV 284, No. 6, mov. 3, Var. 10	CCARH
1400	Mozart, W. A.	1774	Sonaten, KV 284, No. 6, mov. 3, Var. 11	CCARH
1401	Mozart, W. A.	1774	Sonaten, KV 284, No. 6, mov. 3, Var. 12	CCARH
1402	Mozart, W. A.	1777	Sonaten, KV 309, No. 7, mov. 1	CCARH
1403	Mozart, W. A.	1777	Sonaten, KV 309, No. 7, mov. 2	CCARH
1404	Mozart, W. A.	1777	Sonaten, KV 309, No. 7, mov. 3	CCARH
1405	Mozart, W. A.	1777	Sonaten, KV 310, No. 9, mov. 1	CCARH
1406	Mozart, W. A.	1777	Sonaten, KV 310, No. 9, mov. 2	CCARH
1407	Mozart, W. A.	1777	Sonaten, KV 310, No. 9, mov. 3	CCARH
1408	Mozart, W. A.	1777	Sonaten, KV 311, No. 8, mov. 1	CCARH
1409	Mozart, W. A.	1777	Sonaten, KV 311, No. 8, mov. 2	CCARH
1410	Mozart, W. A.	1777	Sonaten, KV 311, No. 8, mov. 3	CCARH
1411	Mozart, W. A.	1783	Sonaten, KV 330, No. 10, mov. 1	CCARH
1412	Mozart, W. A.	1783	Sonaten, KV 330, No. 10, mov. 2	CCARH
1413	Mozart, W. A.	1783	Sonaten, KV 330, No. 10, mov. 3	CCARH
1414	Mozart, W. A.	1783	Sonaten, KV 331, No. 11, mov. 1, Thema	CCARH
1415	Mozart, W. A.	1783	Sonaten, KV 331, No. 11, mov. 1, Var. 1	CCARH
1416	Mozart, W. A.	1783	Sonaten, KV 331, No. 11, mov. 1, Var. 2	CCARH
1417	Mozart, W. A.	1783	Sonaten, KV 331, No. 11, mov. 1, Var. 3	CCARH
1418	Mozart, W. A.	1783	Sonaten, KV 331, No. 11, mov. 1, Var. 4	CCARH
1419	Mozart, W. A.	1783	Sonaten, KV 331, No. 11, mov. 1, Var. 5	CCARH
1420	Mozart, W. A.	1783	Sonaten, KV 331, No. 11, mov. 1, Var. 6	CCARH
1421	Mozart, W. A.	1783	Sonaten, KV 331, No. 11, mov. 2	CCARH
1422	Mozart, W. A.	1783	Sonaten, KV 331, No. 11, mov. 3	CCARH
1423	Mozart, W. A.	1783	Sonaten, KV 332, No. 12, mov. 1	CCARH
1424	Mozart, W. A.	1783	Sonaten, KV 332, No. 12, mov. 2	CCARH
1425	Mozart, W. A.	1783	Sonaten, KV 332, No. 12, mov. 3	CCARH
1426	Mozart, W. A.	1783	Sonaten, KV 333, No. 13, mov. 1	CCARH
1427	Mozart, W. A.	1783	Sonaten, KV 333, No. 13, mov. 2	CCARH
1428	Mozart, W. A.	1783	Sonaten, KV 333, No. 13, mov. 3	CCARH
1429	Mozart, W. A.	1785	Sonaten, KV 457, No. 14, mov. 1	CCARH

Continued on next page

Table B.3 – Corpus used in Part IV (“Macroanalysis”; cont.).

#	Name	Year	Title	Source
1430	Mozart, W. A.	1785	Sonaten, KV 457, No. 14, mov. 2	CCARH
1431	Mozart, W. A.	1785	Sonaten, KV 457, No. 14, mov. 3	CCARH
1432	Mozart, W. A.	1788	Sonaten, KV 533, No. 15, mov. 1	CCARH
1433	Mozart, W. A.	1788	Sonaten, KV 533, No. 15, mov. 2	CCARH
1434	Mozart, W. A.	1788	Sonaten, KV 533, No. 15, mov. 3	CCARH
1435	Mozart, W. A.	1788	Sonaten, KV 570, No. 16, mov. 1	CCARH
1436	Mozart, W. A.	1788	Sonaten, KV 570, No. 16, mov. 2	CCARH
1437	Mozart, W. A.	1788	Sonaten, KV 570, No. 16, mov. 3	CCARH
1438	Mozart, W. A.	1789	Sonaten, KV 576, No. 17, mov. 1	CCARH
1439	Mozart, W. A.	1789	Sonaten, KV 576, No. 17, mov. 2	CCARH
1440	Mozart, W. A.	1789	Sonaten, KV 576, No. 17, mov. 3	CCARH
1441	Mozart, W. A.	1791	Requiem (Franz Beyer), KV 626, Introitus	KPM
1442	Mozart, W. A.	1791	Requiem (Franz Beyer), KV 626, Kyrie	KPM
1443	Mozart, W. A.	1791	Requiem (Franz Beyer), KV 626, Dies irae	KPM
1444	Mozart, W. A.	1791	Requiem (Franz Beyer), KV 626, Rex tremendae	KPM
1445	Mozart, W. A.	1791	Requiem (Franz Beyer), KV 626, Confutatis	KPM
1446	Mozart, W. A.	1791	Requiem (Franz Beyer), KV 626, Lacrimosa	KPM
1447	Mozart, W. A.	1791	Requiem (Franz Beyer), KV 626, Domine Jesu	KPM
1448	Mozart, W. A.	1791	Requiem (Franz Beyer), KV 626, Hostias	KPM
1449	Mozart, W. A.	1791	Requiem (Franz Beyer), KV 626, Sanctus	KPM
1450	Mozart, W. A.	1791	Requiem (Franz Beyer), KV 626, Benedictus	KPM
1451	Mozart, W. A.	1791	Requiem (Franz Beyer), KV 626, Agnus Dei	KPM
1452	Mozart, W. A.	1791	Requiem (Franz Beyer), KV 626, Communio	KPM
1453	Mussorgsky, M.	1874	Pictures at an Exhibition, Promenade I	MS
1454	Mussorgsky, M.	1874	Pictures at an Exhibition, Gnomus	MS
1455	Mussorgsky, M.	1874	Pictures at an Exhibition, Promenade II	MS
1456	Mussorgsky, M.	1874	Pictures at an Exhibition, Il vecchio castello	MS
1457	Mussorgsky, M.	1874	Pictures at an Exhibition, Promenade III	MS
1458	Mussorgsky, M.	1874	Pictures at an Exhibition, Tuileries	MS
1459	Mussorgsky, M.	1874	Pictures at an Exhibition, Bydlo	MS
1460	Mussorgsky, M.	1874	Pictures at an Exhibition, Promenade IV	MS
1461	Mussorgsky, M.	1874	Pictures at an Exhibition, Ballet of Unhatched Chicks	MS
1462	Mussorgsky, M.	1874	Pictures at an Exhibition, Samuel Goldenberg und Schmuyle	MS
1463	Mussorgsky, M.	1874	Pictures at an Exhibition, Promenade V	MS
1464	Mussorgsky, M.	1874	Pictures at an Exhibition, Limoges, le Marché	MS
1465	Mussorgsky, M.	1874	Pictures at an Exhibition, Catacumbae (Sepulchrum Romanum)	MS
1466	Mussorgsky, M.	1874	Pictures at an Exhibition, Cum Mortuis in Lingua Morta	MS
1467	Mussorgsky, M.	1874	Pictures at an Exhibition, The Hut on Hen’s Legs	MS
1468	Mussorgsky, M.	1874	Pictures at an Exhibition, The Great Gate of Kiev	MS
1469	Ockeghem, J.	1475	L’autre d’antan	ELVIS
1470	Ockeghem, J.	1475	Ma bouche rit	ELVIS
1471	Ockeghem, J.	1475	S’elle m’amera	ELVIS
1472	Ockeghem, J.	1497	Alius discantus super O rosa bella	ELVIS
1473	Ockeghem, J.	1497	Alma Redemptoris Mater	ELVIS
1474	Ockeghem, J.	1497	Fors seulement contre	ELVIS
1475	Ockeghem, J.	1497	Il ne m’en chault plus	ELVIS
1476	Ockeghem, J.	1497	Je n’ay deuil	ELVIS
1477	Ockeghem, J.	1497	L’autre d’antan	ELVIS
1478	Ockeghem, J.	1497	Ma bouche rit	ELVIS
1479	Ockeghem, J.	1497	Ma bouche rit	ELVIS
1480	Ockeghem, J.	1497	Malheur me bat	ELVIS
1481	Ockeghem, J.	1497	Ma maistresse	ELVIS
1482	Ockeghem, J.	1497	Missa Au travail suis, Credo	ELVIS
1483	Ockeghem, J.	1497	Missa Au travail suis, Gloria	ELVIS

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
1484	Ockeghem, J.	1497	Missa Au travail suis, Kyrie	ELVIS
1485	Ockeghem, J.	1497	Missa Au travail suis, Sanctus	ELVIS
1486	Ockeghem, J.	1497	Missa Cuiusvis toni, Agnus Dei	ELVIS
1487	Ockeghem, J.	1497	Missa Cuiusvis toni, Credo	ELVIS
1488	Ockeghem, J.	1497	Missa Cuiusvis toni, Gloria	ELVIS
1489	Ockeghem, J.	1497	Missa Cuiusvis toni, Kyrie	ELVIS
1490	Ockeghem, J.	1497	Missa Cuiusvis toni, Sanctus	ELVIS
1491	Ockeghem, J.	1497	Missa Ecce Ancilla Domini, Agnus Dei	ELVIS
1492	Ockeghem, J.	1497	Missa Ecce Ancilla Domini, Credo	ELVIS
1493	Ockeghem, J.	1497	Missa Ecce Ancilla Domini, Gloria	ELVIS
1494	Ockeghem, J.	1497	Missa Ecce Ancilla Domini, Kyrie	ELVIS
1495	Ockeghem, J.	1497	Missa Fors Seulement, Credo	ELVIS
1496	Ockeghem, J.	1497	Missa Fors Seulement, Gloria	ELVIS
1497	Ockeghem, J.	1497	Missa Fors Seulement, Kyrie	ELVIS
1498	Ockeghem, J.	1497	Missa Ma maistresse, Gloria	ELVIS
1499	Ockeghem, J.	1497	Missa Ma maistresse, Kyrie	ELVIS
1500	Ockeghem, J.	1497	Missa prolationum, Agnus Dei	ELVIS
1501	Ockeghem, J.	1497	Missa prolationum, Credo	ELVIS
1502	Ockeghem, J.	1497	Missa prolationum, Gloria	ELVIS
1503	Ockeghem, J.	1497	Missa prolationum, Kyrie	ELVIS
1504	Ockeghem, J.	1497	Missa prolationum, Sanctus	ELVIS
1505	Ockeghem, J.	1497	Missa Quinti toni, Agnus Dei	ELVIS
1506	Ockeghem, J.	1497	Missa Quinti toni, Credo	ELVIS
1507	Ockeghem, J.	1497	Missa Quinti toni, Gloria	ELVIS
1508	Ockeghem, J.	1497	Missa Quinti toni, Kyrie	ELVIS
1509	Ockeghem, J.	1497	Missa Quinti toni, Sanctus	ELVIS
1510	Ockeghem, J.	1497	Missa Sine Nomine, Credo	ELVIS
1511	Ockeghem, J.	1497	Missa Sine Nomine, Gloria	ELVIS
1512	Ockeghem, J.	1497	Missa Sine Nomine, Kyrie	ELVIS
1513	Ockeghem, J.	1497	Mort tu as navré	ELVIS
1514	Ockeghem, J.	1497	Mort tu as navré	ELVIS
1515	Ockeghem, J.	1497	Prenez sur moi	ELVIS
1516	Ockeghem, J.	1497	Presque transi	ELVIS
1517	Ockeghem, J.	1497	Quant de vous seul	ELVIS
1518	Ockeghem, J.	1497	Qu'es mi vida preguntays	ELVIS
1519	Ockeghem, J.	1497	Requiem, Graduale	ELVIS
1520	Ockeghem, J.	1497	Requiem, Introitus	ELVIS
1521	Ockeghem, J.	1497	Requiem, Kyrie	ELVIS
1522	Ockeghem, J.	1497	Requiem, Offertorium	ELVIS
1523	Ockeghem, J.	1497	Requiem, Tract	ELVIS
1524	Ockeghem, J.	1497	S'elle m'amera	ELVIS
1525	Ockeghem, J.	1497	Se vostre cuer eslongne	ELVIS
1526	Ockeghem, J.	1497	Tant fuz gentement resjouy	ELVIS
1527	Ockeghem, J.	1497	Ung aultre l'a	ELVIS
1528	Ockeghem, J.	1797	Les desléaux ont la saison	ELVIS
1529	Palestrina, P.	1555	Missa Papae Marcelli, Kyrie	CPDL
1530	Palestrina, P.	1555	Missa Papae Marcelli, Gloria	CPDL
1531	Palestrina, P.	1555	Missa Papae Marcelli, Credo	CPDL
1532	Palestrina, P.	1555	Missa Papae Marcelli, Sanctus	CPDL
1533	Palestrina, P.	1555	Missa Papae Marcelli, Benedictus	CPDL
1534	Palestrina, P.	1555	Missa Papae Marcelli, Agnus Dei I	CPDL
1535	Palestrina, P.	1555	Missa Papae Marcelli, Agnus Dei II	CPDL
1536	Quilter, R.	1921	Five Shakespeare songs, Op. 23, No. 1, Fear no more the heat of the sun	MS

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
1537	Quilter, R.	1921	Five Shakespeare songs, Op. 23, No. 2, Under the greenwood tree	MS
1538	Quilter, R.	1921	Five Shakespeare songs, Op. 23, No. 3, It was a lover and his lass	MS
1539	Quilter, R.	1921	Five Shakespeare songs, Op. 23, No. 4, Take, o take those lips away	MS
1540	Quilter, R.	1921	Five Shakespeare songs, Op. 23, No. 5, Hey, ho, the wind and the rain	MS
1541	Rachmaninoff, S.	1887	4 pieces, No. 4, Gavotte	MS
1542	Rachmaninoff, S.	1892	Morceaux de fantaisie, Op. 3, No. 1, Elegie	MS
1543	Rachmaninoff, S.	1903	10 Préludes, Op. 23, No. 1	MS
1544	Rachmaninoff, S.	1903	10 Préludes, Op. 23, No. 10	MS
1545	Rachmaninoff, S.	1903	10 Préludes, Op. 23, No. 2	MS
1546	Rachmaninoff, S.	1903	10 Préludes, Op. 23, No. 3	MS
1547	Rachmaninoff, S.	1903	10 Préludes, Op. 23, No. 4	MS
1548	Rachmaninoff, S.	1903	10 Préludes, Op. 23, No. 5	MS
1549	Rachmaninoff, S.	1903	10 Préludes, Op. 23, No. 6	MS
1550	Rachmaninoff, S.	1903	10 Préludes, Op. 23, No. 7	MS
1551	Rachmaninoff, S.	1903	10 Préludes, Op. 23, No. 8	MS
1552	Rachmaninoff, S.	1903	10 Préludes, Op. 23, No. 9	MS
1553	Rameau, J. P.	1727	Suite, RCT 5, Gavotte et six doubles (Sixième double)	MS
1554	Rameau, J. P.	1727	Nouvelles suites de pièces de clavecin, RCT 6, No. 5, Menuet	MS
1555	Rameau, J. P.	1735	Les Indes Galantes, Rondeau	MS
1556	Ravel, M.	1903	Sonatine, M. 40, No. 1, Modéré	MS
1557	Ravel, M.	1903	Sonatine, M. 40, No. 2, Mouvement de Menuet	MS
1558	Ravel, M.	1903	Sonatine, M. 40, No. 3, Animé	MS
1559	Ravel, M.	1904	Mirroirs, M. 43, No. 3	MS
1560	Ravel, M.	1904	Mirroirs, M., No. 1, Noctuelles	MS
1561	Ravel, M.	1909	Gaspar de la nuit, M. 55, No. 1, Ondine	MS
1562	Ravel, M.	1909	Gaspar de la nuit, M. 55, No. 2, Le Gibet	MS
1563	Ravel, M.	1909	Gaspar de la nuit, M. 55, No. 3, Scarbo	MS
1564	Ravel, M.	1912	M. 63, No. 1, À la manière de Borodin	MS
1565	Ravel, M.	1912	M. 63, No. 2, À la manière de Chabrier	MS
1566	Ravel, M.	1913	M. 65, Prelude	MS
1567	Reger, M.	1915	Fünf neue Kinderlieder, Op. 142, No. 1, Wiegenlied	MS
1568	Reichardt, L.	1802	Zwölf Gesänge, Op. 3, No. 1, Frühlingsblumen	OSLC
1569	Reichardt, L.	1802	Zwölf Gesänge, Op. 3, No. 10, Der Mond	OSLC
1570	Reichardt, L.	1802	Zwölf Gesänge, Op. 3, No. 11, An Maria (aus Novalis geistlichen Liedern)	OSLC
1571	Reichardt, L.	1802	Zwölf Gesänge, Op. 3, No. 12, Duettino	OSLC
1572	Reichardt, L.	1802	Zwölf Gesänge, Op. 3, No. 2, Der traurige Wanderer	OSLC
1573	Reichardt, L.	1802	Zwölf Gesänge, Op. 3, No. 3, Die Blume der Blumen	OSLC
1574	Reichardt, L.	1802	Zwölf Gesänge, Op. 3, No. 4, Wachtelwacht	OSLC
1575	Reichardt, L.	1802	Zwölf Gesänge, Op. 3, No. 5, Betteley der Vögel	OSLC
1576	Reichardt, L.	1802	Zwölf Gesänge, Op. 3, No. 6, Kriegslied des Mays	OSLC
1577	Reichardt, L.	1802	Zwölf Gesänge, Op. 3, No. 7, Die Wiese	OSLC
1578	Reichardt, L.	1802	Zwölf Gesänge, Op. 3, No. 8, Kaeuzlein	OSLC
1579	Reichardt, L.	1802	Zwölf Gesänge, Op. 3, No. 9, Hier liegt ein Spielmann begraben	OSLC
1580	Reichardt, L.	1802	Sechs Lieder von Novalis, Op. 4, No. 1, Sehnsucht nach dem Vaterlande	OSLC
1581	Reichardt, L.	1802	Sechs Lieder von Novalis, Op. 4, No. 2, Frühlingslied	OSLC
1582	Reichardt, L.	1802	Sechs Lieder von Novalis, Op. 4, No. 3a, Geistliches Lied	OSLC
1583	Reichardt, L.	1802	Sechs Lieder von Novalis, Op. 4, No. 3b, Geistliches Lied	OSLC
1584	Reichardt, L.	1802	Sechs Lieder von Novalis, Op. 4, No. 4, Bergmannslied	OSLC
1585	Reichardt, L.	1802	Sechs Lieder von Novalis, Op. 4, No. 5, Noch ein Bergmannslied	OSLC
1586	Reichardt, L.	1802	Sechs Lieder von Novalis, Op. 4, No. 6, Er besucht den Klostergarten	OSLC
1587	Reichardt, L.	1802	Zwölf Deutsche und Italiänische Romantische Gesänge, Op., No. 1, Frühlingslied	OSLC

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
1588	Reichardt, L.	1802	Zwölf Gesänge, Op., No. 1, Erinnrung am Bach	OSLC
1589	Reichardt, L.	1802	Zwölf Deutsche und Italiänische Romantische Gesänge, Op., No. 10, Ida (aus Ariels Offenbarungen)	OSLC
1590	Reichardt, L.	1802	Zwölf Gesänge, Op., No. 10, Die Veilchen	OSLC
1591	Reichardt, L.	1802	Zwölf Deutsche und Italiänische Romantische Gesänge, Op., No. 11, Aus Tiecks Genoveva	OSLC
1592	Reichardt, L.	1802	Zwölf Gesänge, Op., No. 11, Daphne am Bach	OSLC
1593	Reichardt, L.	1802	Zwölf Deutsche und Italiänische Romantische Gesänge, Op., No. 12, Heymdal	OSLC
1594	Reichardt, L.	1802	Zwölf Gesänge, Op., No. 12, Aus Novalis Hymnen der Nacht	OSLC
1595	Reichardt, L.	1802	Zwölf Deutsche und Italiänische Romantische Gesänge, Op., No. 2, Wenn ich ihn nur habe	OSLC
1596	Reichardt, L.	1802	Zwölf Gesänge, Op., No. 2, Der Sänger geht	OSLC
1597	Reichardt, L.	1802	Zwölf Deutsche und Italiänische Romantische Gesänge, Op., No. 3, Durch die bunten Rosenhecken	OSLC
1598	Reichardt, L.	1802	Zwölf Gesänge, Op., No. 3, Nach Sevilla	OSLC
1599	Reichardt, L.	1802	Zwölf Deutsche und Italiänische Romantische Gesänge, Op., No. 4, Wohle dem Mann	OSLC
1600	Reichardt, L.	1802	Zwölf Gesänge, Op., No. 4, Vaters Klage	OSLC
1601	Reichardt, L.	1802	Zwölf Deutsche und Italiänische Romantische Gesänge, Op., No. 5, Poesia di Metastasio I	OSLC
1602	Reichardt, L.	1802	Zwölf Gesänge, Op., No. 5, Für die Laute componirt	OSLC
1603	Reichardt, L.	1802	Zwölf Deutsche und Italiänische Romantische Gesänge, Op., No. 6, Notturno	OSLC
1604	Reichardt, L.	1802	Zwölf Gesänge, Op., No. 6, Unruhiger Schlaf	OSLC
1605	Reichardt, L.	1802	Zwölf Deutsche und Italiänische Romantische Gesänge, Op., No. 7, Poesia di Metastasio II	OSLC
1606	Reichardt, L.	1802	Zwölf Gesänge, Op., No. 7, Volkslied	OSLC
1607	Reichardt, L.	1802	Zwölf Deutsche und Italiänische Romantische Gesänge, Op., No. 8, Poesie von Tieck	OSLC
1608	Reichardt, L.	1802	Zwölf Gesänge, Op., No. 8, Ein recht Gemüth	OSLC
1609	Reichardt, L.	1802	Zwölf Deutsche und Italiänische Romantische Gesänge, Op., No. 9, Aus Ariels Offenbarungen	OSLC
1610	Reichardt, L.	1802	Zwölf Gesänge, Op., No. 9, Der Spinnerin Nachtlied	OSLC
1611	Rue, P. de l.	1485	Missa Incessament, Benedictus	ELVIS
1612	Rue, P. de l.	1485	Missa Incessament, In nomine	ELVIS
1613	Rue, P. de l.	1485	Missa Incessament, Pleni	ELVIS
1614	Rue, P. de l.	1485	Missa Alamana, Benedictus	ELVIS
1615	Rue, P. de l.	1485	Missa Assumpta est Maria, Gloria Tua	ELVIS
1616	Rue, P. de l.	1485	Missa Assumpta est Maria, Pleni	ELVIS
1617	Rue, P. de l.	1485	Ave Sanctissima Maria, Cruxifixus	ELVIS
1618	Rue, P. de l.	1485	Ave Sanctissima Maria, Et resurrexit	ELVIS
1619	Rue, P. de l.	1485	Ave Sanctissima Maria, Gloria - Qui tollis	ELVIS
1620	Rue, P. de l.	1485	Ave Sanctissima Maria, Pleni	ELVIS
1621	Rue, P. de l.	1485	Conceptio tua, Benedictus	ELVIS
1622	Rue, P. de l.	1485	Conceptio tua, In nomine	ELVIS
1623	Rue, P. de l.	1485	Missa de Feria, Agnus II	ELVIS
1624	Rue, P. de l.	1485	Missa de Feria, Pleni	ELVIS
1625	Rue, P. de l.	1485	Missa de Sancta Anna, Benedictus	ELVIS
1626	Rue, P. de l.	1485	Missa de Sancta Anna, In nomine	ELVIS
1627	Rue, P. de l.	1485	Missa de Sancto Antonio, In nomine	ELVIS
1628	Rue, P. de l.	1485	Missa de Sancto Job, Benedictus	ELVIS
1629	Rue, P. de l.	1485	Missa de Sancto Job, Pleni	ELVIS
1630	Rue, P. de l.	1485	Missa de septem doloribus, In nomine	ELVIS

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
1631	Rue, P. de l.	1485	Missa de septem doloribus, Pleni	ELVIS
1632	Rue, P. de l.	1485	Missa de virginibus, Benedictus	ELVIS
1633	Rue, P. de l.	1485	Missa de virginibus, In nomine	ELVIS
1634	Rue, P. de l.	1485	Missa de virginibus, Pleni	ELVIS
1635	Rue, P. de l.	1485	Missa Inviolata, Pleni	ELVIS
1636	Rue, P. de l.	1485	Missa Ista est Speziosa, Pleni	ELVIS
1637	Rue, P. de l.	1485	Missa l'homme armé, Pleni	ELVIS
1638	Rue, P. de l.	1485	Missa Nunca fue pena mayor, Benedictus	ELVIS
1639	Rue, P. de l.	1485	Missa O Gloriosa Domina, Agnus II	ELVIS
1640	Rue, P. de l.	1485	Missa O Gloriosa Domina, Benedictus	ELVIS
1641	Rue, P. de l.	1485	Missa O Gloriosa Domina, In nomine	ELVIS
1642	Rue, P. de l.	1485	Missa O salutaris hostia, Benedictus	ELVIS
1643	Rue, P. de l.	1485	Missa O salutaris hostia, Pleni	ELVIS
1644	Rue, P. de l.	1485	Missa Pascale, Benedictus	ELVIS
1645	Rue, P. de l.	1485	Missa Pascale, In nomine	ELVIS
1646	Rue, P. de l.	1485	Missa Pro fidelibus defunctis, Sicut cervus	ELVIS
1647	Rue, P. de l.	1485	Missa Pro fidelibus defunctis, Sitivit anima mea	ELVIS
1648	Rue, P. de l.	1485	Missa Sancta Dei Genetrix, Benedictus	ELVIS
1649	Rue, P. de l.	1485	Missa Sine Nomine I, Benedictus	ELVIS
1650	Rue, P. de l.	1485	Missa Sine Nomine I, In nomine	ELVIS
1651	Rue, P. de l.	1485	Missa Sub tuum presidium, Benedictus	ELVIS
1652	Rue, P. de l.	1485	Missa Tandernaken, Agnus II	ELVIS
1653	Rue, P. de l.	1485	Missa Tandernaken, Benedictus	ELVIS
1654	Rue, P. de l.	1485	Missa tous les regretz, Agnus II	ELVIS
1655	Rue, P. de l.	1485	Missa de septem doloribus, Pleni	ELVIS
1656	Satie, E.	1887	Sarabandes, No. 1	MS
1657	Satie, E.	1887	Sarabandes, No. 2	MS
1658	Satie, E.	1887	Sarabandes, No. 3	MS
1659	Satie, E.	1888	Gymnopédies, No. 1	MS
1660	Satie, E.	1888	Gymnopédies, No. 2	MS
1661	Satie, E.	1888	Gymnopédies, No. 3	MS
1662	Satie, E.	1889	Ogives, No. 1	MS
1663	Satie, E.	1889	Ogives, No. 2	MS
1664	Satie, E.	1889	Ogives, No. 3	MS
1665	Satie, E.	1889	Ogives, No. 4	MS
1666	Satie, E.	1889	Gnossiennes, No. 5	MS
1667	Satie, E.	1891	Gnossiennes, No. 4	MS
1668	Satie, E.	1891	Gnossiennes, No. 7	MS
1669	Satie, E.	1893	Gnossiennes, No. 1	MS
1670	Satie, E.	1893	Gnossiennes, No. 2	MS
1671	Satie, E.	1893	Gnossiennes, No. 3	MS
1672	Satie, E.	1897	Gnossiennes, No. 6	MS
1673	Satie, E.	1923	Ludions, mov. 1, Air du rat	MS
1674	Satie, E.	1923	Ludions, mov. 2, Spleen	MS
1675	Satie, E.	1923	Ludions, mov. 3, La grenouille américaine	MS
1676	Satie, E.	1923	Ludions, mov. 4, Air du poète	MS
1677	Satie, E.	1923	Ludions, mov. 5, Chanson du chat	MS
1678	Scarlatti, D.	1721	Sonata, K 116	MS
1679	Scarlatti, D.	1721	Sonata, K 138	MS
1680	Scarlatti, D.	1721	Sonata, K 159	MS
1681	Scarlatti, D.	1721	Sonata, K 162	MS
1682	Scarlatti, D.	1721	Sonata, K 19	MS
1683	Scarlatti, D.	1721	Sonata, K 208	MS
1684	Scarlatti, D.	1721	Sonata, K 223	MS

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
1685	Scarlatti, D.	1721	Sonata, K 23	MS
1686	Scarlatti, D.	1721	Sonata, K 27	MS
1687	Scarlatti, D.	1721	Sonata, K 431	MS
1688	Scarlatti, D.	1721	Sonata, K 455	MS
1689	Scarlatti, D.	1721	Sonata, K 526	MS
1690	Scarlatti, D.	1721	Sonata, K 531	MS
1691	Scarlatti, D.	1721	Sonata, K 63	MS
1692	Scarlatti, D.	1721	Sonata, L 63	MS
1693	Scarlatti, D.	1721	Sonata, K 64	MS
1694	Scarlatti, D.	1721	Sonata, K 64	MS
1695	Scarlatti, D.	1721	Sonata, K 87	MS
1696	Scheidt, S.	1620	Zion spricht	KPM
1697	Schein, J. H.	1623	Israelsbrünnlein, mov. 10, Da Jakob vollendet hatte	KPM
1698	Schein, J. H.	1623	Israelsbrünnlein, mov. 21, Was betrübst du dich, meine Seele	KPM
1699	Schein, J. H.	1623	Israelsbrünnlein, mov. 3, Die mit Tränen säen	KPM
1700	Schein, J. H.	1623	Israelsbrünnlein, mov. 6, Wende dich, Herr, und sei mir gnädig	KPM
1701	Schoenberg, A.	1911	Sechs kleine Klavierstücke, Op. 19, No. 2	MS
1702	Schoenberg, A.	1911	Sechs kleine Klavierstücke, Op. 19, No. 6	MS
1703	Schubert, F.	1822	Wanderer-Fantasie, D. 760	MS
1704	Schubert, F.	1823	D. 778 60, No. 1, Greisengesang	OSLC
1705	Schubert, F.	1823	D. 801 60, No. 2, Dithyrambe	OSLC
1706	Schubert, F.	1823	Die schöne Müllerin, D. 795, No. 1, Das Wandern	OSLC
1707	Schubert, F.	1823	Die schöne Müllerin, D. 795, No. 10, Tränenregen	OSLC
1708	Schubert, F.	1823	Die schöne Müllerin, D. 795, No. 11, Mein!	OSLC
1709	Schubert, F.	1823	Die schöne Müllerin, D. 795, No. 12, Pause	OSLC
1710	Schubert, F.	1823	Die schöne Müllerin, D. 795, No. 13, Mit dem grünen Lautenbande	OSLC
1711	Schubert, F.	1823	Die schöne Müllerin, D. 795, No. 14, Der Jäger	OSLC
1712	Schubert, F.	1823	Die schöne Müllerin, D. 795, No. 15, Eifersucht und Stolz	OSLC
1713	Schubert, F.	1823	Die schöne Müllerin, D. 795, No. 16, Die liebe Farbe	OSLC
1714	Schubert, F.	1823	Die schöne Müllerin, D. 795, No. 17, Die böse Farbe	OSLC
1715	Schubert, F.	1823	Die schöne Müllerin, D. 795, No. 18, Trockne Blumen	OSLC
1716	Schubert, F.	1823	Die schöne Müllerin, D. 795, No. 19, Der Müller und der Bach	OSLC
1717	Schubert, F.	1823	Die schöne Müllerin, D. 795, No. 2, Wohin?	OSLC
1718	Schubert, F.	1823	Die schöne Müllerin, D. 795, No. 20, Des Baches Wiegenlied	OSLC
1719	Schubert, F.	1823	Die schöne Müllerin, D. 795, No. 3, Halt!	OSLC
1720	Schubert, F.	1823	Die schöne Müllerin, D. 795, No. 4, Danksagung an den Bach	OSLC
1721	Schubert, F.	1823	Die schöne Müllerin, D. 795, No. 5, Am Feierabend	OSLC
1722	Schubert, F.	1823	Die schöne Müllerin, D. 795, No. 6, Der Neugierige	OSLC
1723	Schubert, F.	1823	Die schöne Müllerin, D. 795, No. 7, Ungeduld	OSLC
1724	Schubert, F.	1823	Die schöne Müllerin, D. 795, No. 8, Morgengruß	OSLC
1725	Schubert, F.	1823	Die schöne Müllerin, D. 795, No. 9, Des Müllers Blumen	OSLC
1726	Schubert, F.	1826	Vier Gesänge aus 'Wilhelm Meister', D. 877 62, No. 1, Mignon und der Harfner	OSLC
1727	Schubert, F.	1826	Vier Gesänge aus 'Wilhelm Meister', D. 877 62, No. 2, Lied der Mignon (Heiss mich nicht reden)	OSLC
1728	Schubert, F.	1826	Vier Gesänge aus 'Wilhelm Meister', D. 877 62, No. 3, Lied der Mignon (So lasst mich scheinen)	OSLC
1729	Schubert, F.	1826	Vier Gesänge aus 'Wilhelm Meister', D. 877 62, No. 4, Lied der Mignon (Nur wer die Sehnsucht kennt)	OSLC
1730	Schubert, F.	1827	Impromptus, Op. 142, No. 1	DCML
1731	Schubert, F.	1827	Impromptus, Op. 142, No. 2	DCML
1732	Schubert, F.	1827	Impromptus, Op. 142, No. 3	DCML
1733	Schubert, F.	1827	Impromptus, Op. 142, No. 4	DCML
1734	Schubert, F.	1827	Winterreise, D. 911 89, No. 1, Gute Nacht	OSLC

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
1735	Schubert, F.	1827	Winterreise, D. 911 89, No. 10, Rast	OSLC
1736	Schubert, F.	1827	Winterreise, D. 911 89, No. 11, Frühlingstraum	OSLC
1737	Schubert, F.	1827	Winterreise, D. 911 89, No. 12, Einsamkeit	OSLC
1738	Schubert, F.	1827	Winterreise, D. 911 89, No. 13, Die Post	OSLC
1739	Schubert, F.	1827	Winterreise, D. 911 89, No. 14, Der greise Kopf	OSLC
1740	Schubert, F.	1827	Winterreise, D. 911 89, No. 15, Die Krähe	OSLC
1741	Schubert, F.	1827	Winterreise, D. 911 89, No. 16, Letzte Hoffnung	OSLC
1742	Schubert, F.	1827	Winterreise, D. 911 89, No. 17, Im Dorfe	OSLC
1743	Schubert, F.	1827	Winterreise, D. 911 89, No. 18, Der stürmische Morgen	OSLC
1744	Schubert, F.	1827	Winterreise, D. 911 89, No. 19, Täuschung	OSLC
1745	Schubert, F.	1827	Winterreise, D. 911 89, No. 2, Die Wetterfahne	OSLC
1746	Schubert, F.	1827	Winterreise, D. 911 89, No. 20, Der Wegweiser	OSLC
1747	Schubert, F.	1827	Winterreise, D. 911 89, No. 21, Das Wirtshaus	OSLC
1748	Schubert, F.	1827	Winterreise, D. 911 89, No. 22, Muth	OSLC
1749	Schubert, F.	1827	Winterreise, D. 911 89, No. 23, Die Nebensonnen	OSLC
1750	Schubert, F.	1827	Winterreise, D. 911 89, No. 24, Der Leiermann	OSLC
1751	Schubert, F.	1827	Winterreise, D. 911 89, No. 3, Gefror'ne Tränen	OSLC
1752	Schubert, F.	1827	Winterreise, D. 911 89, No. 4, Erstarrung	OSLC
1753	Schubert, F.	1827	Winterreise, D. 911 89, No. 5, Der Lindenbaum	OSLC
1754	Schubert, F.	1827	Winterreise, D. 911 89, No. 6, Wasserfluth	OSLC
1755	Schubert, F.	1827	Winterreise, D. 911 89, No. 7, Auf dem Flusse	OSLC
1756	Schubert, F.	1827	Winterreise, D. 911 89, No. 8, Rückblick	OSLC
1757	Schubert, F.	1827	Winterreise, D. 911 89, No. 9, Irrlicht	OSLC
1758	Schubert, F.	1827	Impromptus, Op. 90, No. 1	DCML
1759	Schubert, F.	1827	Impromptus, Op. 90, No. 2	DCML
1760	Schubert, F.	1827	Impromptus, Op. 90, No. 3	DCML
1761	Schubert, F.	1827	Impromptus, Op. 90, No. 4	DCML
1762	Schubert, F.	1828	Klavierstücke, D. 946, No. 1	MS
1763	Schubert, F.	1828	Schwanengesang, D. 957, No. 1, Liebesbotschaft	OSLC
1764	Schubert, F.	1828	Schwanengesang, D. 957, No. 10, Das Fischermädchen	OSLC
1765	Schubert, F.	1828	Schwanengesang, D. 957, No. 11, Die Stadt	OSLC
1766	Schubert, F.	1828	Schwanengesang, D. 957, No. 12, Am Meer	OSLC
1767	Schubert, F.	1828	Schwanengesang, D. 957, No. 13, Der Doppelgänger	OSLC
1768	Schubert, F.	1828	Schwanengesang, D. 957, No. 14, Die Taubenpost	OSLC
1769	Schubert, F.	1828	Klavierstücke, D. 946, No. 2	MS
1770	Schubert, F.	1828	Schwanengesang, D. 957, No. 2, Kriegers Ahnung	OSLC
1771	Schubert, F.	1828	Schwanengesang, D. 957, No. 3, Frühlingssehnsucht	OSLC
1772	Schubert, F.	1828	Schwanengesang, D. 957, No. 4, Ständchen	OSLC
1773	Schubert, F.	1828	Schwanengesang, D. 957, No. 5, Aufenthalt	OSLC
1774	Schubert, F.	1828	Schwanengesang, D. 957, No. 6, In der Ferne	OSLC
1775	Schubert, F.	1828	Schwanengesang, D. 957, No. 7, Abschied	OSLC
1776	Schubert, F.	1828	Schwanengesang, D. 957, No. 8, Der Atlas	OSLC
1777	Schubert, F.	1828	Schwanengesang, D. 957, No. 9, Ihr Bild	OSLC
1778	Schubert, F.	1828	Sonata in Bb, D. 960	MS
1779	Schubert, F.	1895	D. 756 59, No. 1, Du liebst mich nicht	OSLC
1780	Schubert, F.	1895	D. 775 59, No. 2, Dass sie hier gewesen	OSLC
1781	Schubert, F.	1895	D. 776 59, No. 3, Du bist die Ruh	OSLC
1782	Schubert, F.	1895	D. 775 59, No. 4, Lachen und Weinen	OSLC
1783	Schumann, C.	1844	Sechs Lieder, Op. 13, No. 1, Ich stand in dunklein Träumen	OSLC
1784	Schumann, C.	1844	Sechs Lieder, Op. 13, No. 2, Sie liebten sich beide	OSLC
1785	Schumann, C.	1844	Sechs Lieder, Op. 13, No. 3, Liebeszauber	OSLC
1786	Schumann, C.	1844	Sechs Lieder, Op. 13, No. 4, Der Mond kommt still gegangen	OSLC
1787	Schumann, C.	1844	Sechs Lieder, Op. 13, No. 5, Ich hab' in Deinem Auge	OSLC
1788	Schumann, C.	1844	Sechs Lieder, Op. 13, No. 6, Die stille Lotosblume	OSLC

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
1789	Schumann, C.	1853	Sechs Lieder, Op. 23, No. 1, Was weinst du, Blümlein?	OSLC
1790	Schumann, C.	1853	Sechs Lieder, Op. 23, No. 2, An einem lichten Morgen	OSLC
1791	Schumann, C.	1853	Sechs Lieder, Op. 23, No. 3, Geheimes Flüstern hier und dort	OSLC
1792	Schumann, C.	1853	Sechs Lieder, Op. 23, No. 4, Auf einem grünen Hügel	OSLC
1793	Schumann, C.	1853	Sechs Lieder, Op. 23, No. 5, Das ist der Tag, der klingen mag	OSLC
1794	Schumann, C.	1853	Sechs Lieder, Op. 23, No. 6, O Lust, o Lust	OSLC
1795	Schumann, R.	1840	Liederkreis, Op. 39, No. 1, In der Fremde I	OSLC
1796	Schumann, R.	1840	Liederkreis, Op. 39, No. 10, Zwielicht	OSLC
1797	Schumann, R.	1840	Liederkreis, Op. 39, No. 11, Im Walde	OSLC
1798	Schumann, R.	1840	Liederkreis, Op. 39, No. 12, Frühlingsnacht	OSLC
1799	Schumann, R.	1840	Liederkreis, Op. 39, No. 2, Intermezzo	OSLC
1800	Schumann, R.	1840	Liederkreis, Op. 39, No. 3, Waldesgespräch	OSLC
1801	Schumann, R.	1840	Liederkreis, Op. 39, No. 4, Die Stille	OSLC
1802	Schumann, R.	1840	Liederkreis, Op. 39, No. 5, Mondnacht	OSLC
1803	Schumann, R.	1840	Liederkreis, Op. 39, No. 6, Schöne Fremde	OSLC
1804	Schumann, R.	1840	Liederkreis, Op. 39, No. 7, Auf einer Burg	OSLC
1805	Schumann, R.	1840	Liederkreis, Op. 39, No. 8, In der Fremde II	OSLC
1806	Schumann, R.	1840	Liederkreis, Op. 39, No. 9, Wehmuth	OSLC
1807	Schumann, R.	1840	Frauenliebe und Leben, Op. 42, No. 1, Seit ich ihn gesehen	OSLC
1808	Schumann, R.	1840	Frauenliebe und Leben, Op. 42, No. 2, Er, der Herrlichste von Allen	OSLC
1809	Schumann, R.	1840	Frauenliebe und Leben, Op. 42, No. 3, Ich kann's nicht fassen, nicht glauben	OSLC
1810	Schumann, R.	1840	Frauenliebe und Leben, Op. 42, No. 4, Du Ring an meinem Finger	OSLC
1811	Schumann, R.	1840	Frauenliebe und Leben, Op. 42, No. 5, Helft mir, ihr Schwestern	OSLC
1812	Schumann, R.	1840	Frauenliebe und Leben, Op. 42, No. 6, Süßer Freund, du blickest	OSLC
1813	Schumann, R.	1840	Frauenliebe und Leben, Op. 42, No. 7, An meinem Herzen, an meiner Brust	OSLC
1814	Schumann, R.	1840	Frauenliebe und Leben, Op. 42, No. 8, Nun hast du mir den ersten Schmerz getan	OSLC
1815	Schumann, R.	1840	Dichterliebe, Op. 48, No. 1, Im wunderschönen Monat Mai	OSLC
1816	Schumann, R.	1840	Dichterliebe, Op. 48, No. 10, Hör' ich das Liedchen klingen	OSLC
1817	Schumann, R.	1840	Dichterliebe, Op. 48, No. 11, Ein Jüngling liebt ein Mädchen	OSLC
1818	Schumann, R.	1840	Dichterliebe, Op. 48, No. 12, Am leuchtenden Sommermorgen	OSLC
1819	Schumann, R.	1840	Dichterliebe, Op. 48, No. 13, Ich hab' im Traum geweinet	OSLC
1820	Schumann, R.	1840	Dichterliebe, Op. 48, No. 14, Allnächtlich im Träume	OSLC
1821	Schumann, R.	1840	Dichterliebe, Op. 48, No. 15, Aus alten Märchen	OSLC
1822	Schumann, R.	1840	Dichterliebe, Op. 48, No. 16, Die alten, bösen Lieder	OSLC
1823	Schumann, R.	1840	Dichterliebe, Op. 48, No. 2, Aus meinen Thränen sprießen	OSLC
1824	Schumann, R.	1840	Dichterliebe, Op. 48, No. 3, Die Rose, die Lilie, die Taube	OSLC
1825	Schumann, R.	1840	Dichterliebe, Op. 48, No. 4, Wenn ich in deine Augen seh'	OSLC
1826	Schumann, R.	1840	Dichterliebe, Op. 48, No. 5, Ich will meine Seele tauchen	OSLC
1827	Schumann, R.	1840	Dichterliebe, Op. 48, No. 6, Im Rhein, im heil'gen Strome	OSLC
1828	Schumann, R.	1840	Dichterliebe, Op. 48, No. 7, Ich grolle nicht	OSLC
1829	Schumann, R.	1840	Dichterliebe, Op. 48, No. 8, Und wüssten's die Blumen	OSLC
1830	Schumann, R.	1840	Dichterliebe, Op. 48, No. 9, Das ist ein Flöten und Geigen	OSLC
1831	Schumann, R.	1842	String Quartet, Op. 41, No. 1, mov. 1, Introduzione - Andante espressivo	ELVIS
1832	Schumann, R.	1842	String Quartet, Op. 41, No. 1, mov. 2, Allegro	ELVIS
1833	Schumann, R.	1842	String Quartet, Op. 41, No. 1, mov. 3, Scherzo - Presto	ELVIS
1834	Schumann, R.	1842	String Quartet, Op. 41, No. 1, mov. 4, Adagio	ELVIS
1835	Schumann, R.	1842	String Quartet, Op. 41, No. 1, mov. 5, Presto	ELVIS
1836	Schütz, H.	1648	SWV 387, Herzlich lieb hab ich dich	KPM
1837	Schütz, H.	1648	SWV 388, Das ist je gewißlich wahr	KPM
1838	Scriabin, A.	1894	12 Etudes, Op. 8, No. 1	MS

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
1839	Scriabin, A.	1894	12 Etudes, Op. 8, No. 10	MS
1840	Scriabin, A.	1894	12 Etudes, Op. 8, No. 11	MS
1841	Scriabin, A.	1894	12 Etudes, Op. 8, No. 12	MS
1842	Scriabin, A.	1894	12 Etudes, Op. 8, No. 2	MS
1843	Scriabin, A.	1894	12 Etudes, Op. 8, No. 3	MS
1844	Scriabin, A.	1894	12 Etudes, Op. 8, No. 4	MS
1845	Scriabin, A.	1894	12 Etudes, Op. 8, No. 5	MS
1846	Scriabin, A.	1894	12 Etudes, Op. 8, No. 6	MS
1847	Scriabin, A.	1894	12 Etudes, Op. 8, No. 7	MS
1848	Scriabin, A.	1894	12 Etudes, Op. 8, No. 8	MS
1849	Scriabin, A.	1894	12 Etudes, Op. 8, No. 9	MS
1850	Scriabin, A.	1894	Op. 9, Nocturne for the left hand	MS
1851	Scriabin, A.	1895	Préludes, Op. 13, No. 1, 6 Preludes	DCML
1852	Scriabin, A.	1895	Préludes, Op. 13, No. 2, 6 Preludes	DCML
1853	Scriabin, A.	1895	Préludes, Op. 13, No. 3, 6 Preludes	DCML
1854	Scriabin, A.	1895	Préludes, Op. 13, No. 4, 6 Preludes	DCML
1855	Scriabin, A.	1895	Préludes, Op. 13, No. 5, 6 Preludes	DCML
1856	Scriabin, A.	1895	Préludes, Op. 13, No. 6, 6 Preludes	DCML
1857	Scriabin, A.	1895	Préludes, Op. 16, No. 1, 5 Preludes	DCML
1858	Scriabin, A.	1895	Préludes, Op. 16, No. 2, 5 Preludes	DCML
1859	Scriabin, A.	1895	Préludes, Op. 16, No. 3, 5 Preludes	DCML
1860	Scriabin, A.	1895	Préludes, Op. 16, No. 4, 5 Preludes	DCML
1861	Scriabin, A.	1895	Préludes, Op. 16, No. 5, 5 Preludes	DCML
1862	Scriabin, A.	1896	Préludes, Op. 15, No. 1, 5 Preludes	DCML
1863	Scriabin, A.	1896	Préludes, Op. 15, No. 2, 5 Preludes	DCML
1864	Scriabin, A.	1896	Préludes, Op. 15, No. 3, 5 Preludes	DCML
1865	Scriabin, A.	1896	Préludes, Op. 15, No. 4, 5 Preludes	DCML
1866	Scriabin, A.	1896	Préludes, Op. 15, No. 5, 5 Preludes	DCML
1867	Scriabin, A.	1896	Préludes, Op. 17, No. 1, 7 Preludes	DCML
1868	Scriabin, A.	1896	Préludes, Op. 17, No. 2, 7 Preludes	DCML
1869	Scriabin, A.	1896	Préludes, Op. 17, No. 3, 7 Preludes	DCML
1870	Scriabin, A.	1896	Préludes, Op. 17, No. 4, 7 Preludes	DCML
1871	Scriabin, A.	1896	Préludes, Op. 17, No. 5, 7 Preludes	DCML
1872	Scriabin, A.	1896	Préludes, Op. 17, No. 6, 7 Preludes	DCML
1873	Scriabin, A.	1896	Préludes, Op. 17, No. 7, 7 Preludes	DCML
1874	Scriabin, A.	1897	Préludes, Op. 22, No. 1, 4 Preludes	DCML
1875	Scriabin, A.	1897	Préludes, Op. 22, No. 2, 4 Preludes	DCML
1876	Scriabin, A.	1897	Préludes, Op. 22, No. 3, 4 Preludes	DCML
1877	Scriabin, A.	1897	Préludes, Op. 22, No. 4, 4 Preludes	DCML
1878	Scriabin, A.	1898	Préludes, Op. 11, No. 1, 24 Preludes	DCML
1879	Scriabin, A.	1898	Préludes, Op. 11, No. 10, 24 Preludes	DCML
1880	Scriabin, A.	1898	Préludes, Op. 11, No. 11, 24 Preludes	DCML
1881	Scriabin, A.	1898	Préludes, Op. 11, No. 12, 24 Preludes	DCML
1882	Scriabin, A.	1898	Préludes, Op. 11, No. 13, 24 Preludes	DCML
1883	Scriabin, A.	1898	Préludes, Op. 11, No. 14, 24 Preludes	DCML
1884	Scriabin, A.	1898	Préludes, Op. 11, No. 15, 24 Preludes	DCML
1885	Scriabin, A.	1898	Préludes, Op. 11, No. 17, 24 Preludes	DCML
1886	Scriabin, A.	1898	Préludes, Op. 11, No. 18, 24 Preludes	DCML
1887	Scriabin, A.	1898	Préludes, Op. 11, No. 2, 24 Preludes	DCML
1888	Scriabin, A.	1898	Préludes, Op. 11, No. 21, 24 Preludes	DCML
1889	Scriabin, A.	1898	Préludes, Op. 11, No. 22, 24 Preludes	DCML
1890	Scriabin, A.	1898	Préludes, Op. 11, No. 23, 24 Preludes	DCML
1891	Scriabin, A.	1898	Préludes, Op. 11, No. 3, 24 Preludes	DCML
1892	Scriabin, A.	1898	Préludes, Op. 11, No. 4, 24 Preludes	DCML

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
1893	Scriabin, A.	1898	Préludes, Op. 11, No. 5, 24 Preludes	DCML
1894	Scriabin, A.	1898	Préludes, Op. 11, No. 6, 24 Preludes	DCML
1895	Scriabin, A.	1898	Préludes, Op. 11, No. 7, 24 Preludes	DCML
1896	Scriabin, A.	1898	Préludes, Op. 11, No. 8, 24 Preludes	DCML
1897	Scriabin, A.	1898	Préludes, Op. 11, No. 9, 24 Preludes	DCML
1898	Scriabin, A.	1901	Préludes, Op. 27, No. 1, 2 Preludes	DCML
1899	Scriabin, A.	1901	Préludes, Op. 27, No. 2, 2 Preludes	DCML
1900	Scriabin, A.	1903	Préludes, Op. 31, No. 1, 4 Preludes	DCML
1901	Scriabin, A.	1903	Préludes, Op. 31, No. 2, 4 Preludes	DCML
1902	Scriabin, A.	1903	Préludes, Op. 31, No. 3, 4 Preludes	DCML
1903	Scriabin, A.	1903	Préludes, Op. 31, No. 4, 4 Preludes	DCML
1904	Scriabin, A.	1903	Préludes, Op. 33, No. 1, 4 Preludes	DCML
1905	Scriabin, A.	1903	Préludes, Op. 33, No. 2, 4 Preludes	DCML
1906	Scriabin, A.	1903	Préludes, Op. 33, No. 3, 4 Preludes	DCML
1907	Scriabin, A.	1903	Préludes, Op. 33, No. 4, 4 Preludes	DCML
1908	Scriabin, A.	1903	Préludes, Op. 35, No. 1, 3 Preludes	DCML
1909	Scriabin, A.	1903	Préludes, Op. 35, No. 2, 3 Preludes	DCML
1910	Scriabin, A.	1903	Préludes, Op. 35, No. 3, 3 Preludes	DCML
1911	Scriabin, A.	1903	Préludes, Op. 37, No. 1, 4 Preludes	DCML
1912	Scriabin, A.	1903	Préludes, Op. 37, No. 2, 4 Preludes	DCML
1913	Scriabin, A.	1903	Préludes, Op. 37, No. 3, 4 Preludes	DCML
1914	Scriabin, A.	1903	Préludes, Op. 37, No. 4, 4 Preludes	DCML
1915	Scriabin, A.	1903	Préludes, Op. 39, No. 1, 4 Preludes	DCML
1916	Scriabin, A.	1903	Préludes, Op. 39, No. 2, 4 Preludes	DCML
1917	Scriabin, A.	1903	Préludes, Op. 39, No. 3, 4 Preludes	DCML
1918	Scriabin, A.	1903	Préludes, Op. 39, No. 4, 4 Preludes	DCML
1919	Scriabin, A.	1905	Préludes, Op. 48, No. 1, 4 Preludes	DCML
1920	Scriabin, A.	1905	Préludes, Op. 48, No. 2, 4 Preludes	DCML
1921	Scriabin, A.	1905	Préludes, Op. 48, No. 3, 4 Preludes	DCML
1922	Scriabin, A.	1905	Préludes, Op. 48, No. 4, 4 Preludes	DCML
1923	Scriabin, A.	1913	Préludes, Op. 67, No. 1, 2 Preludes	DCML
1924	Scriabin, A.	1913	Préludes, Op. 67, No. 2, 2 Preludes	DCML
1925	Scriabin, A.	1914	Préludes, Op. 74, No. 1, 5 Preludes	DCML
1926	Sibelius, J.	1900	Op. 26, Finlandia	MS
1927	Sibelius, J.	1914	Op. 75, Granen	MS
1928	Strauss, R.	1894	Lieder, Op. 27, No. 1, Ruhe, meine Seele	MS
1929	Strauss, R.	1894	Lieder, Op. 27, No. 2, Cäcilie	MS
1930	Strauss, R.	1894	Lieder, Op. 27, No. 3, Heimliche Aufforderung	MS
1931	Strauss, R.	1894	Lieder, Op. 27, No. 4, Morgen	MS
1932	Tchaikovsky, P.	1876	The Seasons, Op. 37a, No. 1, January: By the Hearth	MS
1933	Tchaikovsky, P.	1876	The Seasons, Op. 37a, No. 10, October: Autumn Song	MS
1934	Tchaikovsky, P.	1876	The Seasons, Op. 37a, No. 11, November: On the Troika	DCML
1935	Tchaikovsky, P.	1876	The Seasons, Op. 37a, No. 12, December: Christmas	DCML
1936	Tchaikovsky, P.	1876	The Seasons, Op. 37a, No. 2, February: The Carnival	MS
1937	Tchaikovsky, P.	1876	The Seasons, Op. 37a, No. 3, March: Song of the Lark	MS
1938	Tchaikovsky, P.	1876	The Seasons, Op. 37a, No. 4, April: Snowdrop	DCML
1939	Tchaikovsky, P.	1876	The Seasons, Op. 37a, No. 5, May: White Nights	DCML
1940	Tchaikovsky, P.	1876	The Seasons, Op. 37a, No. 6, June: Barcarolle	MS
1941	Tchaikovsky, P.	1876	The Seasons, Op. 37a, No. 7, July: Reaper's Song	DCML
1942	Tchaikovsky, P.	1876	The Seasons, Op. 37a, No. 8, August: The Harvest	DCML
1943	Tchaikovsky, P.	1876	The Seasons, Op. 37a, No. 9, September: The Hunt	DCML
1944	Vaughan Williams, R.	1904	Songs of Travel, mov. 1, The Vagabond	MS
1945	Vaughan Williams, R.	1904	Songs of Travel, mov. 2, Let Beauty Awake	MS
1946	Vaughan Williams, R.	1904	Songs of Travel, mov. 3, The Roadside Fire	MS

Continued on next page

Table B.3 – Corpus used in Part IV (“Macroanalysis”; cont.).

#	Name	Year	Title	Source
1947	Vaughan Williams, R.	1904	Songs of Travel, mov. 4, Youth and Love	MS
1948	Vaughan Williams, R.	1904	Songs of Travel, mov. 5, In Dreams	MS
1949	Vaughan Williams, R.	1904	Songs of Travel, mov. 6, The Infinite Shining Heavens	MS
1950	Vaughan Williams, R.	1904	Songs of Travel, mov. 7, Whither Must I Wander?	MS
1951	Vaughan Williams, R.	1925	Four Poems by Fredegond Shove, mov. 1, Motion and Stillness	MS
1952	Vaughan Williams, R.	1925	Four Poems by Fredegond Shove, mov. 2, Four Nights	MS
1953	Vaughan Williams, R.	1925	Four Poems by Fredegond Shove, mov. 3, The New Ghost	MS
1954	Vaughan Williams, R.	1925	Four Poems by Fredegond Shove, mov. 4, The Water Mill	MS
1955	Victoria, T.	1572	O magnum mysterium	ELVIS
1956	Victoria, T.	1585	Caligaverunt oculi mei	ELVIS
1957	Victoria, T.	1585	Eram quasi agnus	ELVIS
1958	Victoria, T.	1585	Vau. Et egressus est	ELVIS
1959	Victoria, T.	1585	Lamed. Matribus suis dixerunt	ELVIS
1960	Victoria, T.	1585	Aleph. Quomodo obscuratum	ELVIS
1961	Victoria, T.	1585	Animam meam dilectam	ELVIS
1962	Victoria, T.	1585	Astiterunt reges	ELVIS
1963	Victoria, T.	1585	Amicus meus	ELVIS
1964	Victoria, T.	1585	Tamquam ad latronem	ELVIS
1965	Victoria, T.	1585	Judas mercator pessimus	ELVIS
1966	Victoria, T.	1585	Aestimatus sunt	ELVIS
1967	Victoria, T.	1585	Unus ex discipulis	ELVIS
1968	Victoria, T.	1585	Una hora	ELVIS
1969	Victoria, T.	1585	Jod. Manum suam	ELVIS
1970	Victoria, T.	1585	Jesum tradidit impius	ELVIS
1971	Victoria, T.	1585	O vos omnes	ELVIS
1972	Victoria, T.	1585	Ecce quomodo moritur	ELVIS
1973	Victoria, T.	1585	Heth. Cogitavit dominus	ELVIS
1974	Victoria, T.	1585	Incipit lamentatio Jeremiae	ELVIS
1975	Victoria, T.	1585	Recessit pastor noster	ELVIS
1976	Victoria, T.	1585	Aleph. Ego vir videns	ELVIS
1977	Victoria, T.	1585	Incipit oratio Jeremiae	ELVIS
1978	Victoria, T.	1585	Seniores populi	ELVIS
1979	Victoria, T.	1585	Sepulto Domino	ELVIS
1980	Victoria, T.	1585	Heth. Misericordiae Domini	ELVIS
1981	Victoria, T.	1585	Tenebrae factae sunt	ELVIS
1982	Victoria, T.	1592	Missa quarti toni, Gloria	ELVIS
1983	Vierne, L.	1914	12 Préludes, Op. 36, No. 12, Seul	MS
1984	Vitry, P.	1361	Virtutibus laudabilis	ELVIS
1985	Vitry, P.	1361	Gratissima virginis species	ELVIS
1986	Vitry, P.	1361	Lugentium siccentur	ELVIS
1987	Vitry, P.	1361	Rex quem metrorum	ELVIS
1988	Wagner, R.	1839	WWV 57, Mignonne	MS
1989	Webern, A.	1936	Variationen für Klavier, Op. 27, No. 1	MS
1990	Webern, A.	1936	Variationen für Klavier, Op. 27, No. 2	MS
1991	Webern, A.	1936	Variationen für Klavier, Op. 27, No. 3	MS
1992	Wolf, H.	1889	Eichendorff-Lieder, No. 1, Der Freund	OSLC
1993	Wolf, H.	1889	Eichendorff-Lieder, No. 10, Der Glücksritter	OSLC
1994	Wolf, H.	1889	Eichendorff-Lieder, No. 11, Lieber Alles	OSLC
1995	Wolf, H.	1889	Eichendorff-Lieder, No. 12, Heimweh	OSLC
1996	Wolf, H.	1889	Eichendorff-Lieder, No. 13, Der Scholar	OSLC
1997	Wolf, H.	1889	Eichendorff-Lieder, No. 14, Der verzweifelte Liebhaber	OSLC
1998	Wolf, H.	1889	Eichendorff-Lieder, No. 15, Unfall	OSLC
1999	Wolf, H.	1889	Eichendorff-Lieder, No. 16, Liebesglück	OSLC
2000	Wolf, H.	1889	Eichendorff-Lieder, No. 17, Seemanns Abschied	OSLC

Continued on next page

Table B.3 – Corpus used in Part IV ("Macroanalysis"; cont.).

#	Name	Year	Title	Source
2001	Wolf, H.	1889	Eichendorff-Lieder, No. 18, Erwartung	OSLC
2002	Wolf, H.	1889	Eichendorff-Lieder, No. 19, Die Nacht	OSLC
2003	Wolf, H.	1889	Eichendorff-Lieder, No. 2, Der Musikant	OSLC
2004	Wolf, H.	1889	Eichendorff-Lieder, No. 20, Waldmädchen	OSLC
2005	Wolf, H.	1889	Eichendorff-Lieder, No. 3, Verschwiegene Liebe	OSLC
2006	Wolf, H.	1889	Eichendorff-Lieder, No. 4, Das Ständchen	OSLC
2007	Wolf, H.	1889	Eichendorff-Lieder, No. 5, Der Soldat I	OSLC
2008	Wolf, H.	1889	Eichendorff-Lieder, No. 6, Der Soldat II	OSLC
2009	Wolf, H.	1889	Eichendorff-Lieder, No. 7, Die Zigeunerin	OSLC
2010	Wolf, H.	1889	Eichendorff-Lieder, No. 8, Nachtzauber	OSLC
2011	Wolf, H.	1889	Eichendorff-Lieder, No. 9, Der Schreckenberger	OSLC

Table B.4 – Tonal pitch-class distribution for $K = 2$ topics.

	ϕ_1	ϕ_2
F♭♭	0.000	0.000
C♭♭	0.000	0.000
G♭♭	0.000	0.000
D♭♭	0.000	0.000
A♭♭	0.000	0.000
E♭♭	0.000	0.001
B♭♭	0.000	0.002
F♭	0.000	0.004
C♭	0.000	0.010
G♭	0.000	0.021
D♭	0.000	0.038
A♭	0.000	0.061
E♭	0.000	0.087
B♭	0.000	0.115
F	0.003	0.141
C	0.018	0.151
G	0.054	0.135
D	0.096	0.105
A	0.125	0.070
E	0.149	0.039
B	0.146	0.015
F♯	0.122	0.004
C♯	0.100	0.001
G♯	0.073	0.000
D♯	0.050	0.000
A♯	0.031	0.000
E♯	0.017	0.000
B♯	0.009	0.000
F♯♯	0.005	0.000
C♯♯	0.003	0.000
G♯♯	0.001	0.000
D♯♯	0.000	0.000
A♯♯	0.000	0.000
E♯♯	0.000	0.000
B♯♯	0.000	0.000

Table B.5 – Tonal pitch-class distribution for $K = 3$ topics.

	ϕ_1	ϕ_2	ϕ_3
F♭♭	0.000	0.000	0.000
C♭♭	0.000	0.000	0.000
G♭♭	0.000	0.000	0.000
D♭♭	0.000	0.000	0.000
A♭♭	0.000	0.000	0.000
E♭♭	0.000	0.000	0.001
B♭♭	0.000	0.000	0.003
F♭	0.000	0.000	0.009
C♭	0.000	0.000	0.021
G♭	0.000	0.000	0.045
D♭	0.000	0.000	0.082
A♭	0.000	0.000	0.130
E♭	0.000	0.002	0.184
B♭	0.000	0.038	0.181
F	0.000	0.099	0.135
C	0.000	0.144	0.106
G	0.008	0.177	0.067
D	0.042	0.184	0.026
A	0.084	0.154	0.006
E	0.129	0.111	0.002
B	0.154	0.063	0.000
F♯	0.161	0.020	0.000
C♯	0.141	0.007	0.000
G♯	0.108	0.002	0.000
D♯	0.075	0.000	0.000
A♯	0.046	0.000	0.000
E♯	0.025	0.000	0.000
B♯	0.013	0.000	0.000
F♯♯	0.007	0.000	0.000
C♯♯	0.004	0.000	0.000
G♯♯	0.001	0.000	0.000
D♯♯	0.000	0.000	0.000
A♯♯	0.000	0.000	0.000
E♯♯	0.000	0.000	0.000
B♯♯	0.000	0.000	0.000

Table B.6 – Tonal pitch-class distribution for $K = 5$ topics.

	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5
F♭♭	0.000	0.000	0.000	0.000	0.000
C♭♭	0.000	0.000	0.000	0.000	0.000
G♭♭	0.000	0.000	0.000	0.000	0.000
D♭♭	0.000	0.000	0.000	0.000	0.000
A♭♭	0.001	0.000	0.000	0.000	0.000
E♭♭	0.002	0.000	0.000	0.000	0.000
B♭♭	0.005	0.000	0.000	0.000	0.000
F♭	0.014	0.000	0.000	0.000	0.000
C♭	0.032	0.000	0.000	0.000	0.000
G♭	0.070	0.000	0.000	0.000	0.000
D♭	0.126	0.000	0.000	0.000	0.000
A♭	0.197	0.000	0.000	0.000	0.002
E♭	0.193	0.000	0.000	0.000	0.077
B♭	0.128	0.000	0.000	0.000	0.203
F	0.106	0.102	0.000	0.000	0.179
C	0.085	0.206	0.000	0.002	0.137
G	0.032	0.167	0.000	0.094	0.170
D	0.004	0.109	0.005	0.213	0.155
A	0.001	0.145	0.038	0.201	0.068
E	0.001	0.175	0.108	0.129	0.005
B	0.002	0.091	0.145	0.122	0.000
F♯	0.001	0.001	0.146	0.140	0.002
C♯	0.000	0.000	0.147	0.084	0.000
G♯	0.000	0.004	0.149	0.015	0.000
D♯	0.000	0.000	0.113	0.000	0.000
A♯	0.000	0.000	0.070	0.000	0.000
E♯	0.000	0.000	0.039	0.000	0.000
B♯	0.000	0.000	0.020	0.000	0.000
F♯♯	0.000	0.000	0.011	0.000	0.000
C♯♯	0.000	0.000	0.006	0.000	0.000
G♯♯	0.000	0.000	0.002	0.000	0.000
D♯♯	0.000	0.000	0.000	0.000	0.000
A♯♯	0.000	0.000	0.000	0.000	0.000
E♯♯	0.000	0.000	0.000	0.000	0.000
B♯♯	0.000	0.000	0.000	0.000	0.000

Table B.7 – Tonal pitch-class distribution for $K = 7$ topics.

	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7
F♭♭	0.000	0.000	0.000	0.000	0.000	0.000	0.000
C♭♭	0.000	0.000	0.000	0.000	0.000	0.000	0.000
G♭♭	0.000	0.000	0.000	0.000	0.000	0.000	0.000
D♭♭	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A♭♭	0.000	0.000	0.001	0.000	0.000	0.000	0.000
E♭♭	0.000	0.000	0.003	0.000	0.000	0.000	0.000
B♭♭	0.000	0.000	0.007	0.000	0.000	0.000	0.000
F♭	0.000	0.000	0.020	0.000	0.000	0.000	0.000
C♭	0.000	0.000	0.044	0.000	0.000	0.000	0.000
G♭	0.000	0.000	0.098	0.000	0.000	0.000	0.000
D♭	0.000	0.000	0.176	0.000	0.000	0.001	0.000
A♭	0.000	0.000	0.195	0.000	0.000	0.071	0.000
E♭	0.000	0.000	0.141	0.000	0.000	0.218	0.000
B♭	0.000	0.000	0.121	0.000	0.102	0.224	0.000
F	0.000	0.000	0.122	0.080	0.209	0.119	0.000
C	0.007	0.000	0.061	0.279	0.125	0.133	0.000
G	0.186	0.000	0.004	0.217	0.109	0.154	0.000
D	0.322	0.000	0.002	0.058	0.192	0.075	0.047
A	0.147	0.001	0.006	0.069	0.195	0.003	0.201
E	0.018	0.037	0.001	0.175	0.065	0.000	0.273
B	0.113	0.116	0.000	0.120	0.000	0.001	0.162
F♯	0.168	0.173	0.000	0.000	0.000	0.001	0.082
C♯	0.036	0.156	0.000	0.000	0.003	0.000	0.126
G♯	0.000	0.144	0.000	0.000	0.000	0.000	0.091
D♯	0.000	0.149	0.000	0.001	0.000	0.000	0.019
A♯	0.003	0.104	0.000	0.000	0.000	0.000	0.000
E♯	0.000	0.059	0.000	0.000	0.000	0.000	0.000
B♯	0.000	0.031	0.000	0.000	0.000	0.000	0.000
F♯♯	0.000	0.017	0.000	0.000	0.000	0.000	0.000
C♯♯	0.000	0.009	0.000	0.000	0.000	0.000	0.000
G♯♯	0.000	0.003	0.000	0.000	0.000	0.000	0.000
D♯♯	0.000	0.001	0.000	0.000	0.000	0.000	0.000
A♯♯	0.000	0.000	0.000	0.000	0.000	0.000	0.000
E♯♯	0.000	0.000	0.000	0.000	0.000	0.000	0.000
B♯♯	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table B.8 – Tonal pitch-class distribution for $K = 10$ topics.

	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8	ϕ_9	ϕ_{10}
F♭♭	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
C♭♭	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
G♭♭	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
D♭♭	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A♭♭	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
E♭♭	0.002	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000
B♭♭	0.004	0.000	0.000	0.000	0.014	0.000	0.000	0.000	0.000	0.000
F♭	0.019	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000
C♭	0.044	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000
G♭	0.085	0.000	0.000	0.000	0.062	0.000	0.000	0.000	0.000	0.000
D♭	0.148	0.000	0.000	0.000	0.144	0.000	0.000	0.000	0.000	0.000
A♭	0.212	0.000	0.000	0.001	0.081	0.000	0.058	0.000	0.000	0.000
E♭	0.193	0.000	0.000	0.000	0.020	0.000	0.224	0.000	0.000	0.000
B♭	0.124	0.000	0.192	0.000	0.051	0.000	0.228	0.000	0.000	0.000
F	0.096	0.136	0.314	0.034	0.087	0.000	0.096	0.000	0.000	0.000
C	0.062	0.203	0.209	0.257	0.044	0.000	0.084	0.000	0.001	0.000
G	0.006	0.078	0.123	0.373	0.009	0.106	0.145	0.000	0.096	0.000
D	0.001	0.090	0.062	0.129	0.030	0.368	0.132	0.000	0.182	0.006
A	0.000	0.183	0.073	0.001	0.101	0.361	0.027	0.007	0.086	0.116
E	0.001	0.231	0.028	0.051	0.099	0.124	0.000	0.042	0.040	0.248
B	0.000	0.075	0.000	0.134	0.038	0.004	0.002	0.075	0.155	0.236
F♯	0.000	0.000	0.000	0.021	0.046	0.001	0.004	0.124	0.308	0.127
C♯	0.000	0.000	0.000	0.000	0.077	0.031	0.000	0.168	0.130	0.099
G♯	0.000	0.005	0.000	0.000	0.057	0.005	0.000	0.167	0.000	0.106
D♯	0.000	0.000	0.000	0.000	0.017	0.000	0.000	0.150	0.000	0.051
A♯	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.119	0.002	0.012
E♯	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.073	0.000	0.000
B♯	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.038	0.000	0.000
F♯♯	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.021	0.000	0.000
C♯♯	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.000
G♯♯	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.000
D♯♯	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000
A♯♯	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
E♯♯	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
B♯♯	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table B.9 – Tonal pitch-class distribution for $K = 12$ topics.

	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8	ϕ_9	ϕ_{10}	ϕ_{11}	ϕ_{12}
F♭♭	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
C♭♭	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
G♭♭	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
D♭♭	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A♭♭	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
E♭♭	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000
B♭♭	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000
F♭	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.031	0.000	0.000	0.000	0.000
C♭	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.069	0.000	0.000	0.000	0.000
G♭	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.151	0.001	0.000	0.000	0.000
D♭	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.207	0.000	0.000	0.053	0.000
A♭	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.154	0.004	0.000	0.214	0.000
E♭	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.146	0.275	0.000	0.161	0.000
B♭	0.043	0.000	0.000	0.000	0.000	0.000	0.232	0.134	0.323	0.000	0.068	0.000
F	0.195	0.000	0.003	0.133	0.000	0.000	0.336	0.075	0.044	0.000	0.166	0.021
C	0.079	0.000	0.119	0.278	0.000	0.000	0.179	0.012	0.026	0.001	0.224	0.135
G	0.046	0.070	0.326	0.085	0.000	0.000	0.067	0.001	0.152	0.005	0.108	0.319
D	0.214	0.220	0.169	0.038	0.001	0.049	0.075	0.002	0.155	0.000	0.002	0.317
A	0.314	0.169	0.039	0.143	0.001	0.236	0.088	0.001	0.009	0.027	0.000	0.160
E	0.101	0.069	0.083	0.249	0.015	0.291	0.021	0.001	0.000	0.185	0.002	0.040
B	0.002	0.100	0.224	0.072	0.053	0.101	0.000	0.000	0.004	0.328	0.001	0.006
F♯	0.000	0.237	0.038	0.000	0.133	0.008	0.000	0.000	0.007	0.211	0.000	0.002
C♯	0.005	0.125	0.000	0.000	0.188	0.154	0.002	0.000	0.000	0.028	0.000	0.000
G♯	0.000	0.000	0.000	0.001	0.176	0.153	0.000	0.000	0.000	0.056	0.000	0.000
D♯	0.000	0.000	0.000	0.000	0.146	0.005	0.000	0.000	0.000	0.119	0.000	0.000
A♯	0.000	0.006	0.000	0.000	0.116	0.000	0.000	0.000	0.000	0.041	0.000	0.000
E♯	0.000	0.005	0.000	0.000	0.082	0.000	0.000	0.000	0.000	0.000	0.000	0.000
B♯	0.000	0.000	0.000	0.000	0.046	0.001	0.000	0.000	0.000	0.000	0.000	0.000
F♯♯	0.000	0.000	0.000	0.000	0.025	0.000	0.000	0.000	0.000	0.000	0.000	0.000
C♯♯	0.000	0.000	0.000	0.000	0.013	0.000	0.000	0.000	0.000	0.000	0.000	0.000
G♯♯	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000
D♯♯	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A♯♯	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
E♯♯	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
B♯♯	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table B.10 – Tonal pitch-class distribution for $K = 24$ topics.

	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8	ϕ_9	ϕ_{10}	ϕ_{11}	ϕ_{12}	ϕ_{13}	ϕ_{14}	ϕ_{15}	ϕ_{16}	ϕ_{17}	ϕ_{18}	ϕ_{19}	ϕ_{20}	ϕ_{21}	ϕ_{22}	ϕ_{23}	ϕ_{24}
F♭♭	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
C♭♭	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
G♭♭	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
D♭♭	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A♭♭	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000
E♭♭	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000
B♭♭	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.013	0.000
F♭	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.035	0.000
C♭	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.073	0.000
G♭	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.166	0.000
D♭	0.000	0.000	0.094	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.212	0.000
A♭	0.000	0.000	0.309	0.000	0.000	0.000	0.046	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.144	0.000
E♭	0.000	0.000	0.211	0.000	0.000	0.000	0.413	0.001	0.000	0.001	0.000	0.000	0.000	0.000	0.176	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.123	0.000
B♭	0.001	0.000	0.072	0.003	0.000	0.000	0.125	0.225	0.000	0.001	0.000	0.000	0.061	0.000	0.381	0.000	0.297	0.000	0.012	0.000	0.000	0.000	0.141	0.000
F	0.214	0.000	0.096	0.198	0.000	0.011	0.004	0.051	0.000	0.000	0.000	0.000	0.439	0.043	0.185	0.011	0.223	0.157	0.035	0.000	0.000	0.000	0.077	0.000
C	0.455	0.000	0.154	0.144	0.002	0.286	0.102	0.043	0.000	0.036	0.007	0.000	0.260	0.065	0.054	0.178	0.116	0.161	0.028	0.000	0.009	0.000	0.008	0.000
G	0.246	0.000	0.061	0.197	0.160	0.342	0.186	0.157	0.000	0.212	0.310	0.001	0.005	0.055	0.074	0.028	0.256	0.016	0.201	0.000	0.100	0.000	0.000	0.001
D	0.012	0.108	0.001	0.209	0.398	0.111	0.072	0.278	0.000	0.326	0.308	0.026	0.031	0.097	0.113	0.001	0.082	0.118	0.474	0.000	0.005	0.000	0.000	0.090
A	0.002	0.300	0.000	0.179	0.233	0.041	0.015	0.220	0.001	0.084	0.030	0.032	0.144	0.270	0.012	0.102	0.006	0.344	0.223	0.000	0.093	0.062	0.001	0.166
E	0.054	0.256	0.000	0.069	0.095	0.111	0.000	0.022	0.003	0.000	0.017	0.113	0.060	0.277	0.000	0.452	0.020	0.190	0.025	0.042	0.177	0.201	0.000	0.064
B	0.016	0.065	0.001	0.001	0.095	0.098	0.013	0.000	0.007	0.016	0.259	0.283	0.001	0.190	0.000	0.208	0.000	0.014	0.001	0.159	0.319	0.173	0.000	0.014
F♯	0.000	0.069	0.000	0.000	0.016	0.000	0.009	0.002	0.097	0.285	0.068	0.377	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.190	0.198	0.031	0.000	0.168
C♯	0.000	0.130	0.000	0.000	0.000	0.000	0.001	0.001	0.168	0.039	0.000	0.127	0.000	0.001	0.000	0.000	0.000	0.000	0.001	0.118	0.009	0.092	0.000	0.399
G♯	0.000	0.072	0.000	0.000	0.000	0.000	0.000	0.000	0.218	0.000	0.000	0.006	0.000	0.003	0.000	0.019	0.000	0.000	0.000	0.049	0.003	0.294	0.000	0.077
D♯	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.123	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.189	0.078	0.138	0.000	0.000
A♯	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.088	0.000	0.000	0.027	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.199	0.009	0.000	0.000	0.000
E♯	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.131	0.000	0.000	0.008	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.031	0.000	0.000	0.000	0.019
B♯	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.091	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.008	0.000	0.000
F♯♯	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.046	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.008	0.000	0.001	0.000	0.000
C♯♯	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.016	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.000	0.000	0.000	0.000
G♯♯	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.007	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000
D♯♯	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A♯♯	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
E♯♯	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
B♯♯	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

C Musical Scores

Example C.1 – F. Liszt, *Sonetto 47 del Petrarca.*

Example C.1 – F. Liszt, *Sonetto 47 del Petrarca* (cont.).

Example C.1 – F. Liszt, *Sonetto 47 del Petrarca* (cont.).

*) Hier und in den nachfolgenden Takten bis *quasi cadenza* wechseln sich *recitando* und *quasi in tempo* ab, und zwar so, daß die in den unteren zwei Liniensystemen notierten Töne stets *recitando* und die in den oberen notierten immer *in tempo* zu spielen sind.

*) Here and in the following bars right up to the *quasi cadenza*, *recitando* and *quasi in tempo* alternate with one another in such a way that the notes written on the two lower staves are always *recitando* and those on the two top staves are always to be played *in tempo*.

Example C.1 – F. Liszt, *Sonetto 47 del Petrarca* (cont.).

Example C.1 – F. Liszt, *Sonetto 47 del Petrarca* (cont.).

Example C.1 – F. Liszt, *Sonetto 47 del Petrarca* (cont.).

Bibliography

Aarden, B. J. (2003). *Dynamic melodic expectancy* (Doctoral dissertation, Ohio State University, Columbus, OH). doi:10.1024/0301-1526.32.1.54

Abdallah, S., Gold, N., & Marsden, A. (2016). Analysing symbolic music with probabilistic grammars. In D. Meredith (Ed.), *Computational music analysis* (Chap. 7, pp. 157–189). Berlin: Springer.

Albrecht, J., & Huron, D. (2014). A statistical approach to tracing the historical development of major and minor pitch distributions, 1400–1750. *Music Perception: An Interdisciplinary Journal*, *31*(3), 223–243. doi:10.1525/mp.2014.31.3.223

Albrecht, J., & Shanahan, D. (2013). The use of large corpora to train a new type of key-finding algorithm: an improved treatment of the minor mode. *Music Perception: An Interdisciplinary Journal*, *31*(1), 59–67. doi:10.1525/mp.2013.31.1.59

Albrecht, J., & Shanahan, D. (2016). Can I have the keys?: Key validation using a MIDI database. In *Proceedings of the 14th International Conference for Music Perception and Cognition (ICMPC 2016)* (pp. 752–755). San Fransico, CA.

Aldwell, E., Schachter, C., & Cadwallader, A. (2010). *Harmony and voice leading* (4th ed.). Cengage Learning.

Arppe, A., Gilquin, G., Glynn, D., Hilpert, M., & Zeschel, A. (2010). Cognitive corpus linguistics: Five points of debate on current theory and methodology. *Corpora*, *5*(1), 1–27. doi:10.3366/cor.2010.0001

Atlas, R. (1990). Spelling in early 19th-century piano music: A guide to performance. *Journal of Musicological Research*, *10*(3–4), 199–237. doi:10.1080/01411899108574635

Baayen, R. H. (2001). *Word frequency distributions*. Berlin: Springer.

Backer, E., & Van Kranenburg, P. (2005). On musical stylometry: a pattern recognition approach. *Pattern Recognition Letters*, *26*(3), 299–309. doi:10.1016/j.patrec.2004.10.016

Bárdos, L. (1956). Natürliche Tonsysteme: Methode ihrer Messung. In B. Rajeczky & L. Vargyas (Eds.), *Studia memoriae Belae Bartók sacra* (pp. 209–248). Budapest, Hungary: Akadémiai Kiadó.

Bárdos, L. (1979). *Modális harmóniák*. Budapest: Zeneműkiadó.

Belkin, M., & Niyogi, P. (2003). Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Computation*, *15*(6), 1373–1396.

Bellmann, H. G. (2011). *Categorization of tonal music style: a quantitative investigation* (Doctoral dissertation, Griffith University, Mount Gravatt, Australia).

Beykikhoshk, A., Arandjelović, O., Phung, D., & Venkatesh, S. (2018). Discovering topic structures of a temporally evolving document corpus. *Knowledge and Information Systems*, *55*(3), 599–632. doi:10.1007/s10115-017-1095-4

Biamonte, N. (2008). Augmented-sixth chords vs. tritone substitutes. *Music Theory Online*, *14*(2).

Bigo, L., & Andreatta, M. (2016). Topological structures in computer-aided music analysis. In D. Meredith (Ed.), *Computational music analysis* (Chap. 3, pp. 57–80). Berlin: Springer.

Binder, J. M. (2016). Alien reading: Text mining, language standardization, and the humanities. In M. K. Gold & L. F. Klein (Eds.), *Debates in the Digital Humanities 2016* (Chap. 18). Minneapolis: University of Minnesota Press.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Berlin: Springer.

Blei, D. M. (2012a). Probabilistic topic models. *Communications of the ACM [Association for Computing Machinery]*, *55*(4), 77–84. doi:10.1145/2133806.2133826

Blei, D. M. (2012b). Topic modeling and Digital Humanities. *Journal of Digital Humanities*, *2*(1), 8–11.

Blei, D. M., & Lafferty, J. D. (2006). Dynamic topic models. In *Proceedings of the 23rd International Conference on Machine Learning* (pp. 113–120). Pittsburgh, PA. doi:10.1145/1143844.1143859

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*, 993–1022. doi:10.1162/jmlr.2003.3.4-5.993

Bonte, T. (2009). MuseScore: Open source music notation and composition software. Technical Report, Free and Open Source Software Developers' European Meeting. Retrieved from https://de.slideshare.net/thomasbonte/musescore-at-fosdem-2009

Bragg, J., Chew, E., & Shieber, S. (2011). Neo-Riemannian cycle detection with weighted finite-state transducers. In A. Klapuri & C. Leider (Eds.), *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011)* (pp. 399–404). Miami, FL.

Brinkman, A., & Huron, D. (2017). The leading sixth scale degree: A test of Day-O'Connell's theory. *Journal of New Music Research*, *47*(2), 166–175. doi:10.1080/09298215.2017.1407345

Brinkman, A., Shanahan, D., & Sapp, C. (2014). Musical stylometry, machine learning, and attribution studies: a semi-supervised approach to the works of Josquin. In *Proceedings of the 14th International Conference on Music Perception and Cognition (ICMPC 2014)*. San Fransisco, CA.

Broido, A. D., & Clauset, A. (2018). Scale-free networks are rare. *Nature Communications*, *10*(1), 1017. doi:10.1038/s41467-019-08746-5

Brower, C. (2008). Paradoxes of pitch space. *Music Analysis*, *27*(1), 51–106. doi:10.1111/j.1468-2249.2008.00268.x

Brown, M. (2005). *Explaining Tonality: Schenkerian Theory and Beyond*. Rochester: University of Rochester Press.

Broze, Y., & Shanahan, D. (2013). Diachronic changes in Jazz harmony: a cognitive perspective. *Music Perception: An Interdisciplinary Journal*, *31*(1), 32–45. doi:10.1525/mp.2013.31.1.32

Buchler, M. (2016). Are there any bad (or good) transformational analyses? *Intégral*, *30*, 41–51.

Budge, H. (1943). *A study of chord frequencies: Based on the music of representative composers of the eighteenth and nineteenth centuries* (Doctoral dissertation, Columbia University, New York, NY).

Burgoyne, J. A. (2011). *Stochastic processes & database-driven musicology* (Doctoral dissertation, McGill University, Montreal, Canada).

Burgoyne, J. A., Wild, J., & Fujinaga, I. (2011). An expert ground-truth set for audio chord recognition and music analysis. In A. Klapuri & C. Leider (Eds.), *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011)* (pp. 633–638). Miami, FL.

Burgoyne, J. A., Wild, J., & Fujinaga, I. (2013). Compositional data analysis of harmonic structures in Popular music. In J. Yust, J. Wild, & J. A. Burgoyne (Eds.), *Mathematics and computation in music* (pp. 52–63). doi:10.1007/978-3-642-39357-0_4

Butt, J. (2015). What is a 'musical work'? Reflections on the origins of the 'work concept' in Western art music. In A. Rahmatian (Ed.), *Concepts of music and copyright* (Chap. 1, pp. 1–22). doi:10.4337/9781783478194.00006

Calvo-Zaragoza, J., Hajič Jr., J., & Pacha, A. (2019). Understanding optical music recognition. *arXiv:1908.03608 [cs, eess]*. Retrieved from http://arxiv.org/abs/1908.03608

Cambouropoulos, E. (2003). Pitch Spelling: A Computational Model. *Music Perception: An Interdisciplinary Journal*, *20*(4), 411–429. doi:10.1525/mp.2003.20.4.411

Cambouropoulos, E., Kaliakatsos-Papakostas, M., & Tsougras, C. (2014). An idiom-independent representation of chords for computational music analysis and generation. In *Proceedings of the Joint 40th International Computer Music Conference (ICMC) and 11th Sound and Music Computing (SMC) Conference (ICMC-SMC 2014)* (pp. 1002–1009). Athens, Greece. Athens, Greece.

Caplin, W. E. (1998). *Classical form: a theory of formal functions for the instrumental music of Haydn, Mozart and Beethoven*. Oxford: Oxford University Press.

Chang, J., Boyd-Graber, J., Wang, C., Gerrish, S., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. In *Neural information processing systems*. Vancouver, Canada.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. doi:https://doi.org/10.1613/jair.953

Chew, E. (2014). *Mathematical and computational modeling of tonality*. doi:10.1007/978-1-4614-9475-1

Chew, E., & Chen, Y.-C. (2005). Real-time pitch spelling using the spiral array. *Computer Music Journal*, *29*(2), 61–76. doi:10.1162/0148926054094378

Choron, A.-É. (1810). Introduction: sommaire de l'histoire de la musique. In A.-É. Choron & J. M. Fayolle (Eds.), *Dictionnaire historique des musiciens* (pp. xi–xcii). Paris: Chimot.

Christensen, T. (Ed.). (2002). *The Cambridge history of Western music theory*. Cambridge: Cambridge University Press.

Christensen, T. (2019). *Stories of tonality in the age of François-Joseph Fétis*. Chicago: University of Chicago Press.

Clauset, A., Rohilla Shalizi, C., & Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review*, *51*(4), 661–703.

Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, *83*(403), 596–610. doi:10.2307/2289282

Cohn, R. (1996). Maximally Smooth Cycles, Hexatonic Systems, and the Analysis of Late-Romantic Triadic Progressions. *Music Analysis*, *15*(1), 9–40.

Cohn, R. (1997). Neo-Riemannian operations, parsimonious trichords, and their "Tonnetz" representations. *Journal of Music Theory*, *41*(1), 1–66. doi:10.2307/843761

Cohn, R. (1998). Introduction to Neo-Riemannian theory: A survey and a historical perspective. *Journal of Music Theory*, *42*(2), 167–180. doi:10.2307/843871

Cohn, R. (1999). As wonderful as star clusters: Instruments for gazing at tonality in schubert. *19th-Century Music*, *22*(3), 213–232. doi:10.2307/746799

Cohn, R. (2012). *Audacious euphony: Chromatic harmony and the triad's second nature*. Oxford: Oxford University Press.

Conard, N. J., Malina, M., & Münzel, S. C. (2009). New flutes document the earliest musical tradition in southwestern Germany. *Nature*, *460*, 737. doi:10.1038/nature08169

Cook, N. (2014). Close and distant listening. In *Beyond the score: Music as performance* (Chap. 5, pp. 135–175). Oxford: Oxford University Press.

Cortens, E. (2014). The expositions of Haydn's string quartets: A corpus analysis. *HAYDN: Online Journal of the Haydn Society of North America*, *4*(1), 1–42.

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed.). Hoboken, NJ: Wiley.

Dahlhaus, C. (1968). *Untersuchungen über die Entstehung der harmonischen Tonalität*. Kassel: Bärenreiter.

Dahlhaus, C. (1978). Über den Zerfall des musikalischen Werkbegriffs. In *Schönberg und andere. Gesammelte Aufsätze zur Neuen Musik* (pp. 279–290). Mainz: Schott.

Dahlhaus, C. (1989). *Nineteenth-century music* (J. Bradford Robinson, Trans.). Berkeley: University of California Press.

Damschroder, D. (2008). *Thinking about harmony: Historical perspectives on analysis*. Cambridge: Cambridge University Press.

de Clercq, T. (2016). Big data, big questions: A closer look at the Yale-Classical Archives Corpus, (c. 2005). *Empirical Musicology Review*, *11*(1), 59–67. doi:10.18061/emr.v11i1.5274

de Clercq, T., & Temperley, D. (2011). A corpus analysis of rock harmony. *Popular Music*, *30*(1), 47–70. doi:10.1017/S026114301000067X

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

de Haas, W. B., Robine, M., Hanna, P., Veltkamp, R. C., & Wiering, F. (2011). Exploring music contents. In S. Ystad, M. Aramaki, R. Kronland-Martinet, & K. Jensen (Eds.), *Proceedings of the 7th International Symposium on Computer Music Modeling and Retrieval (CMMR 2010)* (pp. 242–258). Revised papers. Málaga, Spain. Berlin: Springer.

de Jong, K., & Noll, T. (2011). Fundamental passacaglia: Harmonic functions and the modes of the musical Tetractys. *Journal of Mathematics and Music*, 98–114.

Dickensheets, J. (2012). The topical vocabulary of the nineteenth century. *Journal of Musicological Research*, *31*, 97–137. doi:10.1080/01411896.2012.682887

Diergarten, F., & Neuwirth, M. (2019). *Formenlehre: Ein Lese- und Arbeitsbuch zur Instrumentalmusik des 18. und 19. Jahrhunderts*. Laaber: Laaber.

Doll, C. (2017). *Hearing harmony: Toward a tonal theory for the Rock era*. University of Michigan Press.

Efron, B. (1979). Bootstrap methods: Another look at the Jackknife. *The Annals of Statistics*, *7*(1), 1–26. doi:10.1214/aos/1176344552

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the Bootstrap*. London: Chapman and Hall/CRC.

Endres, D., & Schindelin, J. (2003). A new metric for probability distributions. *IEEE Transactions on Information Theory*, *49*(7), 1858–1860. doi:10.1109/TIT.2003.813506

Erauw, W. (1998). Canon formation: Some more reflections on Lydia Goehr's Imaginary Museum of Musical Works. *Acta Musicologica*, *70*(2), 109–115. doi:10.2307/932705

Euler, L. (1739). *Tentamen novæ theoriæ musicæ ex certissimis harmoniæ principiis dilucide expositæ*. St. Petersburg: Ex Typographia Academiae Scientiarum.

Euler, L. (1774). De harmoniae veris principiis per speculum musicum repraesentatis. In *Novi Commentarii academiae scientiarum Petropolitanae* (Vol. 18, pp. 330–353). St. Petersburg: Ex Typographia Academiae Scientiarum.

Farbood, M. M., Marcus, G., & Poeppel, D. (2013). Temporal dynamics and the identification of musical key. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(4), 911–918. doi:10.1037/a0031087

Fétis, F.-J. (1844). *Traité complet de la théorie et de la pratique de l'harmonie* (2nd ed.). Paris: Maurice Schlesinger.

Finkensiep, C., Neuwirth, M., & Rohrmeier, M. (2018). Generalized skipgrams for pattern discovery in polyphonic streams. In E. Gómez, X. H. Hu, E. Humphrey, & E. Benetos (Eds.), *Proceedings of the 19th International Society for Music Information Retrieval Conference (ISMIR 2018)* (pp. 547–553). Paris, France.

Fiore, T. M., & Noll, T. (2011). Commuting groups and the topos of triads. In C. Agon, E. Amiot, M. Andreatta, G. Assayag, J. Bresson, & J. Manderau (Eds.), *Mathematics and Computation in Music* (Vol. 6726). Berlin: Springer.

Fiore, T. M., Noll, T., & Satyendra, R. (2013). Morphisms of generalized interval systems and PR-groups. *Journal of Mathematics and Music*, *7*(1), 3–27. doi:10.1080/17459737.2013.785724

Fitsioris, G., & Conklin, D. (2008). Parallel successions of perfect fifths in the Bach chorales. In *Proceedings of the fourth Conference on Interdisciplinary Musicology (CIM 2008)* (pp. 1–10). Thessaloniki, Greece.

Forte, A. (1991). Debussy and the Octatonic. *Music Analysis*, *10*(1/2), 125–169.

Gárdonyi, Z., & Nordhoff, H. (2002). *Harmonik*. Wolfenbüttel: Möseler Verlag.

Gauldin, R. (2004). *Harmonic practice in tonal music* (2nd ed.). New York, NY: Norton.

Gauvin, H. L. (2015). "The times they were a-changin"': A database-driven approach to the evolution of harmonic syntax in Popular music from the 1960s. *Empirical Musicology Review*, *10*(3), 215–238. doi:10.18061/emr.v10i3.4467

Gillies, M. (1983). Bartók's notation: Tonality and modality. *Tempo*, (145), 4–9.

Goehr, L. (1994). *The imaginary museum of musical works: An essay in the philosophy of music*. Oxford: Oxford University Press.

Goldstone, A., & Underwood, T. (2012). What can topic models of PMLA teach us about the history of literary scholarship? *Journal of Digital Humanities*, *2*(1).

Goldstone, A., & Underwood, T. (2014). The quiet transformations of literary studies: What thirteen thousand scholars could tell us. *New Literary History*, *45*(3), 359–384. doi:10.1353/nlh.2014.0025

Gollin, E. (2005). Neo-Riemannian Theory. *Zeitschrift der Gesellschaft für Musiktheorie*, *2*(2/3), 153–155.

Gollin, E. (2006). Some further notes on the history of the *Tonnetz*. *Theoria: Historical Aspects of Music Theory*, *13*, 99–111.

Gollin, E., & Rehding, A. (2011). *The Oxford handbook of Neo-Riemannian music theories*. Oxford: Oxford University Press.

Griffiths, T. (2002). *Gibbs sampling in the generative model of Latent Dirichlet Allocation*. Stanford University. doi:10.1.1.7.8022

Haas, B. (2004). *Die neue Tonalität von Schubert bis Webern: Hören und Analysieren nach Albert Simon*. Wilhelmshaven: Florian Noetzel.

Handschin, J. (1948). *Der Toncharakter: Eine Einführung in die Tonpsychologie*. Zürich: Atlantis Verlag.

Hanninen, D. A. (2012). *A theory of music analysis: On segmentation and associative organization*. Rochester, NY: University Rochester Press.

Hanslick, E. (1922). *Vom Musikalisch-Schönen. Ein Beitrag zur Revision der Ästhetik der Tonkunst*. Leipzig: Breitkopf und Härtel.

Harasim, D., Moss, F. C., Ramirez, M., & Rohrmeier, M. (submitted). Cognitive modeling reveals evolution of major and minor in Western music.

Harasim, D., Schmidt, S. E., & Rohrmeier, M. (2016). Bridging scale theory and geometrical approaches to harmony: The voice-leading duality between complementary chords. *Journal of Mathematics and Music*, *10*(3), 193–209. doi:10.1080/17459737.2016.1216186

Harris, Z. S. (1954). Distributional structure. *WORD*, *10*(3), 146–162. doi:10.1080/00437956.1954.11659520

Harrison, D. (1994). *Harmonic function in chromatic music: A renewed dualist theory and an account of its precedents*. Chicago: University of Chicago Press.

Harte, C. (2010). *Towards automatic extraction of harmony information from music signals* (Doctoral dissertation, Queen Mary University of London, London, UK).

Harte, C. A., Sandler, M., Abdallah, S., & Gómez, E. (2005). Symbolic representation of musical chords: A proposed syntax for text annotations. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)* (pp. 66–71). London, UK.

Hauptmann, M. (1853). *Die Natur der Harmonik und der Metrik*. Leipzig: Breitkopf und Härtel.

Hauptmann, M. (1888). *The nature of harmony and metre* (W. E. Heathcote, Trans.). London: Swan Sonnenschein & Co.

Heaps, H. S. (1978). *Information retrieval, computational and theoretical aspects*. New York, NY: Academic Press.

Hedges, T., & Rohrmeier, M. (2011). Exploring Rameau and beyond: A corpus study of root progression theories. In C. Agon, M. Andreatta, G. Assayag, E. Amiot, J. Bresson, & J. Mandereau (Eds.), *Mathematics and computation in music* (pp. 334–337). Paris, France. Berlin: Springer.

Herrmann, B., & Lauer, G. (2018). Korpusliteraturwissenschaft. Zur Konzeption und Praxis am Beispiel eines Korpus zur literarischen Moderne. *Osnabrücker Beiträge zur Sprachtheorie*, *92*, 127–156.

Hostinský, O. (1879). *Die Lehre von den musikalischen Klängen: ein Beitrag zur aesthetischen Begründung der Harmonielehre*. Prague: H. Dominicus.

Hu, D. J., & Saul, L. K. (2009a). A probabilistic topic model for music analysis. In Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, & A. Culotta (Eds.), *Proceedings of the 22nd Conference on Neural Information Processing Systems (NIPS 2009)* (pp. 1–4).

Hu, D. J., & Saul, L. K. (2009b). A probabilistic topic model for unsupervised learning of musical key-profiles. In J. S. Downie & R. C. Veltkamp (Eds.), *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009)* (pp. 441–446). doi:10.1.1.174.7644

Huang, P., Wilson, M., Mayfield-Jones, D., Coneva, V., & Frank, M. (2017). The evolution of Western tonality: A corpus analysis of 24,000 songs from 190 composers over six centuries. *SocArXiv*. doi:10.31235/osf.io/btshk

Humphrey, E., & Bello, J. P. (2012). Rethinking automatic chord recognition with convolutional neural networks. In M. A. Wani, T. Khoshgoftaar, X. Zhu, & N. Seliya (Eds.), *11th International Conference on Machine Learning and Applications (ICMLA 2012)* (Vol. 2, pp. 357–362). Boca Raton, FL. doi:10.1109/ICMLA.2012.220

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, *9*(3), 90–95. doi:10.1109/mcse.2007.55

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. Cambridge, MA: The MIT Press.

Huron, D. (2013). On the virtuous and the vexatious in an age of Big Data. *Music Perception: An Interdisciplinary Journal*, *31*(1), 4–9. doi:10.1525/MP.2013.31.1.4

Hyer, B. (2001). Tonality. In S. Sadie & J. Tyrrell (Eds.), *The new Grove dictionary of music and musicians* (2nd ed., pp. 583–594). London: Macmillan Publishers.

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks*, *13*(4-5), 411–430. doi:10.1016/S0893-6080(00)00026-5

Izenman, A. J. (2008). *Modern multivariate statistical techniques: Regression, classification, and manifold learning*. Berlin: Springer.

Janata, P., Birk, J. L., Van Horn, J., Leman, M., Tillmann, B., & Bharucha, J. J. (2002). The cortical topography of tonal structures underlying Western music. *Science*, *298*(5601), 2167–2170. doi:10.1126/science.1076262

Jockers, M. L. (2013). *Macroanalysis: Digital methods and literary history*. doi:10.1093/llc/fqt056

Jockers, M. L., & Mimno, D. (2013). Significant themes in 19th-century literature. *Poetics*, *41*(6), 750–769. doi:10.1016/j.poetic.2013.08.005

Johnson, T. (2017). Tonality as topic: Opening a world of analysis for early twentieth-century modernist music. *Music Theory Online*, *23*(4).

Jolliffe, I. T. (2002). *Principal Component Analysis* (2nd ed.). doi:10.1007/b98835

Jones, E., Oliphant, T., Peterson, P. et al. (2001). SciPy: Open source scientific tools for Python. [Online; accessed November 29, 2019]. Retrieved from http://www.scipy.org/

Jurafsky, D., & Martin, J. H. (2009). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (2nd ed.). London, UK: Pearson Education.

Katz, J. (2017). Harmonic syntax of the twelve-bar Blues form. *Music Perception: An Interdisciplinary Journal*, *35*(2), 165–192. doi:10.1525/mp.2017.35.2.165

Kessy, A., Lewin, A., & Strimmer, K. (2018). Optimal whitening and decorrelation. *The American Statistician*, *72*(4), 309–314. doi:10.1080/00031305.2016.1277159

Kinderman, W., & Krebs, H. (Eds.). (1996). *The second practice of nineteenth-century tonality*. Lincoln: University of Nebraska Press.

Klumpenhouwer, H. (1994). Some remarks on the use of Riemann transformations. *Music Theory Online*, *0*(9).

Knopoff, L., & Hutchinson, W. (1983). Entropy as a measure of style: the influence of sample length. *Journal of Music Theory*, *27*(1), 75–97. doi:10.2307/843561

Koelsch, S., Busch, T., Jentschke, S., & Rohrmeier, M. (2016). Under the hood of statistical learning: A statistical MMN reflects the magnitude of transitional probabilities in auditory sequences. *Scientific Reports*, *6*(19741). doi:10.1038/srep19741

Koelsch, S., Rohrmeier, M., Torrecuso, R., & Jentschke, S. (2013). Processing of hierarchical syntactic structure in music. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(38), 15443–8. doi:10.1073/pnas.1300272110

Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: Principles and techniques*. doi:10.1017/CBO9781107415324.004

Koops, H. V., de Haas, W. B., Burgoyne, J. A., Bransen, J., Kent-Muller, A., & Volk, A. (2019). Annotator subjectivity in harmony annotations of popular music. *Journal of New Music Research*, *48*(3). doi:10.1080/09298215.2019.1613436

Koplenig, A. (2017). Against statistical significance testing in corpus linguistics. *Corpus Linguistics and Linguistic Theory*, 1–26. doi:10.1515/cllt-2016-0036

Kopp, D. (2002). *Chromatic transformations in nineteenth-century music* (2nd ed.). Cambridge: Cambridge University Press.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New York, NY: Oxford University Press.

Krumhansl, C. L. (2004). The cognition of tonality—as we know it today. *Journal of New Music Research*, *33*(3), 253–268. doi:10.1080/0929821042000317831

Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, *89*(4), 334–368. doi:10.1037/0033-295X.89.4.334

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, *29*(1), 1–27. doi:10.1007/BF02289565

Laitz, S. G. (2012). *The complete musician: An integrated approach to tonal theory, analysis, and listening*. Oxford: Oxford University Press.

Landnes, K., Mehrabyan, L., Wiklund, V., Lieck, R., Moss, F. C., & Rohrmeier, M. (2019). A Model Comparison for Chord Prediction on the Annotated Beethoven Corpus. In I. Barbancho, L. J. Tardón, A. Peinado, & A. M. Barbancho (Eds.), *Proceedings of the 16th Sound and Music Computing Conference (SMC 2019)* (pp. 250–254). Málaga, Spain.

Lehman, F. M. (2014). Film music and Neo-Riemannian theory. In *Oxford handbooks online* (pp. 1–25). doi:10.1093/oxfordhb/9780199935321.013.002

Lehman, F. M. (2018). *Hollywood harmony: Musical wonder and the sound of cinema*. Oxford: Oxford University Press.

Lerdahl, F. (1988). Tonal Pitch Space. *Music Perception: An Interdisciplinary Journal*, *5*(3), 315–349. doi:10.2307/40285402

Lerdahl, F. (2001). *Tonal pitch space*. Oxford: Oxford University Press.

Lerdahl, F., & Jackendoff, R. S. (1983). *A generative theory of tonal music*. Cambridge, MA: The MIT Press.

Lester, J. (2002). Rameau and eighteenth-century harmonic theory. In T. Christensen (Ed.), *The Cambridge history of Western music theory* (Chap. 24, pp. 753–777). Cambridge, UK: Cambridge University Press.

Levandowsky, M., & Winter, D. (1971). Distance between sets. *Nature*, *234*, 34–35.

Levine, M. (2011). *The Jazz theory book*. Sebastopol, CA: O'Reilly Media.

Lewin, D. (1987). *Generalized musical intervals and transformations*. Oxford: Oxford University Press.

Lieck, R., Moss, F. C., & Rohrmeier, M. (in preparation). The tonal diffusion model.

Liszt, F. (1978). *Années de pèlerinage: Deuxième année, Italie* (E. Herttrich, Ed.). München: Henle Verlag.

Liu, X.-Y., Wu, J., & Zhou, Z.-H. (2009). Exploratory undersampling for class-imbalance learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *39*(2), 539–550. doi:10.1109/tsmcb.2008.2007853

London, J. (2013). Building a representative corpus of classical music. *Music Perception: An Interdisciplinary Journal*, *31*(1), 68–90. doi:10.1525/mp.2013.31.1.68

Longuet-Higgins, C. (1987a). The three dimensions of harmony. In *Mental processes: Studies in cognitive science* (Chap. 7, pp. 59–63). Cambridge, MA: The MIT Press.

Longuet-Higgins, C. (1987b). Two letters to a musical friend. In *Mental processes: Studies in cognitive science* (Chap. 8, pp. 64–81). Cambridge, MA: The MIT Press.

MacKay, D. (2003). *Information theory, inference, and learning algorithms*. Cambridge, UK: Cambridge University Press.

Madsen, S. T., & Widmer, G. (2007). Key-finding with interval profiles. In *Proceedings of the International Computer Music Conference (ICMC 07)* (pp. 27–31). Copenhagen, Denmark.

Manaris, B., Romero, J., Machado, P., Krehbiel, D., Hirzel, T., Pharr, W., & Davis, R. B. (2005). Zipf's law, music classification, and aesthetics. *Computer Music Journal*, *29*(1), 55–69. doi:10.1162/comj.2005.29.1.55

Mandelbrot, B. (1953). An informational theory of the statistical structure of language. In B. W. Jackson (Ed.), *Communication theory* (Chap. 36, pp. 486–502). London: Butterworths.

Manning, C., & Schütze, H. (2003). *Foundations of statistical natural language processing* (6th ed.). Cambridge, MA: The MIT Press.

Marco, G. A. (1958). A musical task in the "surprise" symphony. *Journal of the American Musicological Society*, *11*(1), 41–44. doi:10.2307/830138

Markus Neuwirth, & Bergé, P. (Eds.). (2015). *What Is a Cadence?: Theoretical and Analytical Perspectives on Cadences in the Classical Repertoire*. Leuven: Leuven University Press.

Masada, K., & Bunescu, R. (2019). Chord recognition in symbolic music: A segmental CRF model, segment-level features, and comparative evaluations on classical and popular music. *Transactions of the International Society for Music Information Retrieval*, *2*(1), 1. doi:10.5334/tismir.18

Mauch, M. (2010). *Automatic chord transcription from audio using computational models of musical context* (Doctoral dissertation, Queen Mary, University of London, London, UK).

Mauch, M., Dixon, S., Harte, C. A., Casey, M., & Fields, B. (2007). Discovering chord idioms through Beatles and Real Book songs. In S. Dixon, D. Bainbridge, & R. Typke (Eds.), *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007)* (pp. 255–258). Vienna, Austria. doi:10.1.1.125.6638

Mauch, M., MacCallum, R. M., Levy, M., & Leroi, A. M. (2015). The evolution of popular music: USA 1960–2010. *Royal Society Open Science*, *2*(5), 150081. doi:10.1098/rsos.150081

Mauch, M., Müllensiefen, D., Dixon, S., & Wiggins, G. A. (2008). Can statistical language models be used for the analysis of harmonic progressions. In *Proceedings of the 10th International Conference on Music Perception and Cognition (ICMPC 2008)*. Sapporo, Japan.

Mazzola, G. (1985). *Geometrie der Töne: Elemente der Mathematischen Musiktheorie*. Basel: Birkhäuser.

Mazzola, G. (2002). *The Topos of Music: Geometric Logic of Concepts, Theory, and Performance*. Basel: Birkhäuser.

McFee, B., Nieto, O., Farbood, M. M., & Bello, J. P. (2017). Evaluating hierarchical structure in music annotations. *Frontiers in Psychology*, *8*(1337). doi:10.3389/fpsyg.2017.01337

McKinney, W. (2010). Data structures for statistical computing in Python. In S. van der Walt & J. Millman (Eds.), *Proceedings of the 9th Python in Science Conference* (pp. 51–56).

Mehr, S. A., Singh, M., Knox, D., Ketter, D. M., Pickens-Jones, D., Atwood, S., … Glowacki, L. (2019). Universality and diversity in human song. *Science*, *366*(6468). doi:10.1126/science.aax0868

Meredith, D. (2006). The ps13 pitch spelling algorithm. *Journal of New Music Research*, *35*(2), 121–159. doi:10.1080/09298210600834961

Meredith, D. (Ed.). (2016). *Computational music analysis*. Berlin: Springer.

Messiaen, O. (1944). *Technique de mon langage musical*. Paris: A. Leduc.

Meyer, L. B. (1957). Meaning in music and information theory. *The Journal of Aesthetics and Art Criticism*, *15*(4), 412–424.

Meyer, L. B. (1989). *Style and music: Theory, history, and ideology*. Chicago: University of Chicago Press.

Milička, J. (2009). Type-token & hapax-token relation: A combinatorial model. *Glottotheory*, *2*(1), 99–110. doi:10.1515/glot-2009-0009

Milička, J. (2012). Rank-frequency relation & type-token relation: Two sides of the same coin. In *Methods and Applications of Quantitative Linguistics: Selected papers of the 8th International Conference on Quantitative Linguistics (QUALICO)* (pp. 163–171). Belgrade, Serbia.

Mirka, D. (2014a). Introduction. In D. Mirka (Ed.), *The Oxford handbook of Topic Theory* (pp. 1–60). Oxford: Oxford University Press.

Mirka, D. (Ed.). (2014b). *The Oxford handbook of Topic Theory*. Oxford: Oxford University Press.

Mohr, J. W. (2013). Introduction–topic models: What they are and why they matter. *Poetics*, *41*(6), 545–569. doi:10.1016/J.POETIC.2013.10.001

Morehen, J., & Bent, I. (1979). Computer applications in musicology. *The Musical Times*, *120*(1637), 563–566. doi:10.2307/963094

Moretti, F. (2000). The Slaughterhouse of Literature. *MLQ: Modern Language Quarterly*, *61*(1), 207–227. doi:10.1215/00267929-61-1-207

Moretti, F. (2005). *Graphs, maps, trees: Abstract models for literary history*. New York, NY: Verso.

Moretti, F. (2013). *Distant reading*. New York, NY: Verso.

Móricz, K. (1993). The ambivalent connection between theory and practice in the relationship of F. Liszt & F.-J. Fétis. *Studia Musicologica Academiae Scientiarum Hungaricae*, *35*(4), 399–420. doi:10.2307/902315

Moss, F. C. (2014). Tonality and functional equivalence: A multi-level model for the cognition of triadic progressions in 19th century music. In *Proceedings of the International Conference of Students of Systematic Musicology (SysMus 2014)* (pp. 1–8). London.

Moss, F. C., Loayza, T., & Rohrmeier, M. (2019a). pitchplots. doi:10.5281/zenodo.3265393

Moss, F. C., de Souza, W. F., & Rohrmeier, M. (submitted). Harmony and form in Brazilian Choro: A corpus-driven approach to musical style analysis.

Moss, F. C., Neuwirth, M., Harasim, D., & Rohrmeier, M. (2019b). Statistical characteristics of tonal harmony: A corpus study of Beethoven's string quartets. *PLOS ONE*, *14*(6), e0217242. doi:10.1371/journal.pone.0217242

Neuwirth, M. (2015). Is there a "musical task" in the first movement of Haydn's "Oxford" symphony? Voice-leading schemata and intrinsic formal functions. *Music Theory & Analysis*, *2*(11), 194–203. doi:10.11116/MTA.2.2.3

Neuwirth, M., Harasim, D., Moss, F. C., & Rohrmeier, M. (2018). The Annotated Beethoven Corpus (ABC): A dataset of harmonic analyses of all Beethoven string quartets. *Frontiers in Digital Humanities*, *5*(16). doi:10.3389/fdigh.2018.00016

Neuwirth, M., & Rohrmeier, M. (2016). Wie wissenschaftlich muss Musiktheorie sein? Chancen und Herausforderungen musikalischer Korpusforschung. *Zeitschrift der Gesellschaft für Musiktheorie*, *13*(2), 171–193. doi:10.31751/915

Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, *46*(5), 323–351.

Niemöller, W. K. (1991). Werkbegriff und Werkverzeichnis bei Liszt. *Wissenschaftliche Arbeiten aus dem Burgenland*, *087*, 37–46.

Nocedal, J., & Wright, S. J. (2006). *Numerical Optimization* (2nd ed.). Berlin: Springer.

Noll, T. (2009). Der Halbton im Zwielicht von Leitton- und Moduswechsel: Zum Stellenwert von Enharmonik in Schuberts B-Dur Sonate (D 960). In A. Moths (Ed.), *Musiktheorie an ihren Grenzen: Neue und Alte Musik: 3. Internationaler Kongress für Musiktheorie, 10.-12. Oktober 2003, Musik-Akademie der Stadt Basel.* (pp. 335–347). Peter Lang Verlag.

Osborn, B. (2017). Rock harmony reconsidered: Tonal, modal and contrapuntal voice-leading systems in Radiohead. *Music Analysis*, *36*(1), 59–93. doi:10.1111/musa.12085

Panteli, M., Benetos, E., & Dixon, S. (2018). A review of manual and computational approaches for the study of World music corpora. *Journal of New Music Research*, *47*(2), 176–189. doi:10.1080/09298215.2017.1418896

Perle, G. (1984). Scriabin's self-analyses. *Music Analysis*, *3*(2), 101–122.

Perry, S. (1995). Rummaging through the 'Catacombs': Clues in Musorgsky's pitch notations. *Music Analysis*, *14*(2/3), 221–255. doi:10.2307/854014

Perry, S. (1998). Musorgsky's 'Gnomus:' Composer's score as analytical text. *Context: Journal of Music Research*, (15/16), 5.

Pfleiderer, M., Frieler, K., Abeßer, J., Zaddach, W. G., & Burkhart, B. (Eds.). (2017). *Inside the Jazzomat: New perspectives for Jazz research.* Mainz: Schott.

Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin & Review*, *21*(5), 1112–1130. doi:10.3758/s13423-014-0585-6

Piotrowski, M. (2019). Accepting and modeling uncertainty. In A. Kuczera, T. Wübbena, & T. Kollatz (Eds.), *Die Modellierung des Zweifels – Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten* (Vol. 4, Sonderband der Zeitschrift für digitale Geisteswissenschaften). doi:10.17175/sb004_006

Piston, W. (1948). *Harmony*. New York: Norton.

Polth, M. (2001). Nicht System – nicht Resultat. Zur Bestimmung von harmonischer Tonalität. *Musik & Ästhetik*, *18*, 12–36.

Polth, M. (2016). Tonalität als geschichtliches System: ›Dogmatische Denkform‹ und historischer Nachweis. *Zeitschrift der Gesellschaft für Musiktheorie*, *13*(Sonderausgabe), 53–70. doi:10.31751/862

Polth, M. (2018). The individual tone and musical context in Albert Simon's *Tonfeldtheorie*. *Music Theory Online*, *24*(4). doi:10.30535/mto.24.4.15

Pugin, L. (2015). The challenge of data in digital musicology. *Frontiers in Digital Humanities*, *2*(4). doi:10.3389/fdigh.2015.00004

Purwins, H., Blankertz, B., & Obermayer, K. (2007). Toroidal models in tonal theory and pitch-class analysis. In W. B. Hewlett, E. Selfridge-Field, & E. J. Correia (Eds.), *Tonal theory for the digital age* (Chap. 5, Vol. 15, pp. 73–98). Stanford: Center for Computer Assisted Research in the Humanities.

Quinn, I., & White, C. W. (2017). Corpus-derived key profiles are not transpositionally equivalent. *Music Perception: An Interdisciplinary Journal*, *34*(5), 531–540. doi:10.1525/mp.2017.34.5.531

Rameau, J.-P. (1722). *Traité de l'harmonie réduite à ses principes naturels*. Paris: Imp. de J.-B.-C. Ballard.

Ratner, L. G. (1980). *Classic music: Expression, form, and style*. New York, NY: Schirmer.

Rebelo, A., Fujinaga, I., Paszkiewicz, F., Marcal, A. R. S., Guedes, C., & Cardoso, J. S. (2012). Optical music recognition: State-of-the-art and open issues. *International Journal of Multimedia Information Retrieval*, *1*(3), 173–190. doi:10.1007/s13735-012-0004-6

Renaud, O., & Victoria-Feser, M.-P. (2010). A robust coefficient of determination for regression. *Journal of Statistical Planning and Inference*, *140*(7), 1852–1862. doi:10.1016/J.JSPI.2010.01.008

Rhody, L. M. (2012). Topic modeling and figurative language. *Journal of Digital Humanities*, *2*(1).

Riemann, H. (1880). *Skizze einer neuen Methode der Harmonielehre*. Wiesbaden, Germany: Breitkopf und Härtel.

Riemann, H. (1896). *Dictionary of music* (J. S. Shedlock, Trans.). London: Augener.

Riemann, H. (1900). *Musik-Lexikon* (5th ed.). Leipzig: Max Hesse.

Riemann, H. (1916). *Ideen zu einer „Lehre von den Tonvorstellungen"*. Leipzig: C. F. Peters.

Rings, S. (2011). *Tonality and transformation*. New York: Oxford University Press.

Rohrmeier, M. (2007). Modelling dynamics of key induction in harmony progressions. In C. Spyridis, A. Georgaki, G. Kouroupetroglou, & C. Anagnostopoupou (Eds.), *Proceedings of the 4th Sound and Music Computing Conference (SMC 2007)* (pp. 82–89). Lefkada, Greece.

Rohrmeier, M. (2011). Towards a generative syntax of tonal harmony. *Journal of Mathematics and Music*, *5*(1), 35–53. doi:10.1080/17459737.2011.573676

Rohrmeier, M. (accepted). A generative syntax of Jazz harmony: Diatonic and extended tonality, phrase structure, and form. *Music Theory & Analysis*.

Rohrmeier, M., & Cross, I. (2008). Statistical properties of tonal harmony in Bach's chorales. In K. Miyazaki (Ed.), *Proceedings of the 10th International Conference on Music Perception and Cognition (ICMPC 2008)* (pp. 619–627). Sapporo, Japan.

Rohrmeier, M., & Graepel, T. (2012). Comparing feature-based models of harmony. In M. Aramaki, M. Barthet, R. Kronland-Martinet, & S. Ystad (Eds.), *Proceedings of the 9th International Symposium on Computer Music Modeling and Retrieval (CMMR 2012)* (pp. 357–370). Revised selected papers. London, UK.

Rohrmeier, M., & Neuwirth, M. (2015). Towards a syntax of the classical cadence. In M. Neuwirth & P. Bergé (Eds.), *What is a cadence?* (pp. 285–336). Leuven: Leuven University Press.

Rohrmeier, M., & Rebuschat, P. (2012). Implicit learning and acquisition of music. *Topics in Cognitive Science*, *4*(4), 525–53. doi:10.1111/j.1756-8765.2012.01223.x

Rom, U. B. (2011). *Tonartbezogenes Denken in Mozarts Werken unter besonderer Berücksichtigung des Instrumentalwerks* (Doctoral dissertation, Technische Universität Berlin, Berlin, Germany). doi:10.14279/depositonce-2915

Rosenbaum, P. R. (2010). *Design of observational studies.* Berlin: Springer.

Rousseau, J.-J. (1768). *Dictionnaire de musique.* Paris: Chez la veuve Duchesne.

Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, *290*(5500), 2323–2326. doi:10.1126/science.290.5500.2323

Salzer, F. (1952). *Structural hearing: Tonal coherence in music.* New York, NY: Dover.

Samson, J. (1977). *Music in transition: A study of tonal expansion and atonality, 1900–1920.* Oxford: Oxford University Press.

Schaffer, K. (2016). What is computational musicology? [Blog post]. Accessed: September 21, 2019. Retrieved from https://pushpullfork.com/computational-musicology

Schenker, H. (1906). *Harmonielehre.* Stuttgart: Cotta.

Schenker, H. (1935). *Der freie Satz.* Vienna, Austria: Universal Edition.

Schild, J. (2010). „… zum Raum wird hier die Zeit.“: Tonfelder in Wagners Parsifal. In B. Haas & B. Haas (Eds.), *Funktionale Analyse: Musik – Malerei – Antike Literatur* (pp. 313–373). Hildesheim: Georg Olms Verlag.

Schmidt, B. M. (2012). Words Alone: Dismantling Topic Models in the Humanities. *Journal of Digital Humanities*, *2*(1), 49–65.

Schoenberg, A. (1969). *Structural functions of harmony.* New York, NY: Norton.

Schoenberg, A. (1975). Opinion or insight (L. Black, Trans.). In L. Stein (Ed.), *Style and idea* (pp. 258–264). London, UK: Faber and Faber.

Schubert, P., & Cumming, J. (2015). Another lesson from Lassus: Using computers to analyse counterpoint. *Early Music*, *43*(4), 577–586. doi:10.1093/em/cav088

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. In S. van der Walt & J. Millman (Eds.), *Proceedings of the 9th Python in Science Conference* (pp. 57–61).

Sears, D. (2016). *The classical cadence as a closing schema: Learning, memory, & perception* (Doctoral dissertation, McGill University, Montreal, Canada).

Sechter, S. (1853). *Die Grundsätze der musikalischen Komposition*. Leipzig: Breitkopf und Härtel.

Serrà, J., Corral, L., Boguñá, M., Haro, M., & Arcos, J. L. (2012). Measuring the evolution of contemporary Western popular music. *Scientific Reports*, *2*(521), 1–6. doi:10.1038/srep00521

Serra, X. (2014). Creating research corpora for the computational study of music: The case of the CompMusic project. In *Proceedings of the 53rd AES International Conference on Semantic Audio* (pp. 1–9). London, UK.

Sethares, W. A. (2005). *Tuning, timbre, spectrum, scale* (2nd ed.). Berlin: Springer.

Shanahan, D. (2016). Editor's note. *Empirical Musicology Review [Special Issue in Corpus Studies: No. 1]*, *11*(1), 1. doi:10.18061/emr.v11i1.5305

Shanahan, D. (2017). Editor's note. *Empirical Musicology Review [Special Issue on Corpus Studies: No. 2]*, *11*(2), 108. doi:10.18061/emr.v11i2.5552

Shanahan, D., & Broze, Y. (2012). A diachronic analysis of harmonic schemata in Jazz. In E. Cambouropoulos, K. Tsougkras, P. Mavromatis, & K. Pastiadis (Eds.), *Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music (ICMPC-ESCOM 2012)* (pp. 909–917). Thessaloniki, Greece.

Shanahan, D., Burgoyne, J. A., & Quinn, I. (Eds.). (forthcoming). *The Oxford handbook of music and corpus studies*. Oxford: Oxford University Press.

Sheh, A., & Ellis, D. P. W. (2003). Chord segmentation and recognition using EM-trained Hidden Markov Models. In *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR 2003)*. Baltimore, MD.

Shmulevich, I., & Yli-Harja, O. (2000). Localized key finding: Algorithms and applications. *Music Perception: An Interdisciplinary Journal*, *17*(4), 531–544. doi:10.2307/40285832

Smith, J. B. L., Burgoyne, J. A., Fujinaga, I., De Roure, D., & Downie, J. S. (2011). Design and creation of a large-scale database of structural annotations. In A. Klapuri & C. Leider (Eds.), *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011)* (pp. 555–560). Miami, FL.

Snyder, J. L. (1990). Entropy as a measure of musical style: The influence of a priori assumptions. *Music Theory Spectrum*, *12*(1), 121–160. doi:10.2307/746148

Steyvers, M., & Griffiths, T. (2007). Probabilistic topic models. In T. Landauer, D. McNamara, S. Dennis, & W. Kintsch (Eds.), *Latent semantic analysis: A road to meaning*. Mahwah, NJ: Laurence Erlbaum.

Stoddard, J., Raphael, C., & Utgoff, P. E. (2004). Well-tempered spelling: A key-invariant pitch spelling algorithm. In *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR 2004)*. Barcelona, Spain.

Straus, J. N. (2005). *Introduction to post-tonal theory* (3rd ed.). New York: Pearson Prentice Hall.

Stumpf, M. P. H., & Porter, M. A. (2012). Critical truths about power laws. *Science*, *335*(6069), 665–666. doi:10.1126/science.1216142

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). Cambridge, MA: The MIT Press.

Tan, I., Lustig, E., & Temperley, D. (2019). Anticipatory syncopation in Rock: A corpus study. *Music Perception: An Interdisciplinary Journal*, *36*(4), 353–370. doi:10.1525/mp.2019.36.4.353

Tang, J., Meng, Z., Nguyen, X., Mei, Q., & Zhang, M. (2014). Understanding the limiting factors of topic modeling via Posterior Contraction Analysis. In *Proceedings of the 31st International Conference on Machine Learning*. Beijing, China.

Teh, Y. W., Jordan, M. I., Beal, M. J., & Blei, D. M. (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, *101*(476), 1566–1581. doi:10.1198/016214506000000302

Temperley, D. (1999). What's Key for Key? The Krumhansl-Schmuckler Key-Finding Algorithm Reconsidered. *Music Perception: An Interdisciplinary Journal*, *17*(1), 65–100. doi:10.2307/40285812

Temperley, D. (2000). The line of fifths. *Music Analysis*, *19*(3), 289–319. doi:10.2307/854457

Temperley, D. (2001). *The cognition of basic musical structures*. Cambridge, MA: The MIT Press.

Temperley, D. (2002). A bayesian approach to key-finding. In C. Anagnostopoulou, M. Ferrand, & A. Smaill (Eds.), *Music and artificial intelligence* (195–206´). Berlin: Springer.

Temperley, D. (2009). A statistical analysis of tonal harmony. Accessed: 2019, August 17. Retrieved from http://davidtemperley.com/kp-stats

Temperley, D. (2011). The cadential IV in Rock. *Music Theory Online*, *17*(1).

Temperley, D. (2018). *The musical language of Rock*. Oxford: Oxford University Press.

Temperley, D., & de Clercq, T. (2013). Statistical analysis of harmony and melody in Rock music. *Journal of New Music Research*, *43*(2), 187–204. doi:10.1080/09298215.2013.839525

Temperley, D., & Marvin, E. W. (2008). Pitch-class distribution and the identification of key. *Music Perception: An Interdisciplinary Journal*, *25*(3), 193–212. doi:10.1525/mp.2008.25.3.193

Temperley, D., & VanHandel, L. (2013). Introduction to the special issues on corpus methods. *Music Perception: An Interdisciplinary Journal*, *31*(1), 1–3. doi:10.1525/mp.2013.31.1.1

Tenenbaum, J. B., Silva, V. d., & Langford, J. C. (2000). A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science*, *290*(5500), 2319–2323. doi:10.1126/science.290.5500.2319

Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: A self-organizing approach. *Psychological review*, *107*(4), 885. doi:10.1037/0033-295x.107.4.885

Tymoczko, D. (2002). Stravinsky and the Octatonic: A reconsideration. *Music Theory Spectrum*, *24*(1), 68–102. doi:10.1525/mts.2002.24.1.68

Tymoczko, D. (2003). Function theories: A statistical approach. *Musurgia. Analyse et Pratique Musicales*, *10*(3–4), 35–64. Retrieved from https://dmitri.mycpanel.princeton.edu/files/publications/tonaltheories.pdf

Tymoczko, D. (2011). *A geometry of music: Harmony and counterpoint in the extended common practice*. Oxford: Oxford University Press.

Underwood, T. (2019). *Distant horizons: Digital evidence and literary change*. Chicago: University of Chicago Press.

Van Den Toorn, P. C. (2003). Stravinsky and the Octatonic: The sounds of Stravinsky. *Music Theory Spectrum*, *25*(1), 167–202. doi:10.1525/mts.2003.25.1.167

Van Der Maaten, L., & Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, *9*, 2579–2605.

van der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: A structure for efficient numerical computation. *Computing in Science & Engineering*, *13*(2), 22–30. doi:10.1109/mcse.2011.37

van Leijenhorst, D., & van der Weide, T. (2005). A formal derivation of Heaps' law. *Information Sciences*, *170*, 263–272. doi:10.1016/J.INS.2004.03.006

VanHandel, L., & Temperley, D. (2014). Introduction to the second special issue on corpus methods. *Music Perception: An Interdisciplinary Journal*, *31*(3), 191–191. doi:10.1525/mp.2014.31.3.191

Volk, A., & Honingh, A. (2012). Mathematical and computational approaches to music: Challenges in an interdisciplinary enterprise. *Journal of Mathematics and Music*, *6*(2), 73–81. doi:10.1080/17459737.2012.704154

von Oettingen, A. (1866). *Harmoniesystem in dualer Entwicklung*. Dorpat und Leipzig: W. Gläser.

Wai-Ling, C. (1993). Orthography in Scriabin's late works. *Music Analysis*, *12*(1), 47–69. doi:10.2307/854075

Wallach, H. M., Murray, I., Salakhutdinov, R., & Mimno, D. (2009). Evaluation methods for topic models. In *Proceedings of the 26th Annual International Conference on Machine Learning (ICML 2009)* (pp. 1105–1112). Montreal, Canada.

Waskom, M., Botvinnik, O., O'Kane, D., Hobson, P., Lukauskas, S., Gemperline, D. C., ... Qalieh, A. (2017). Seaborn (v0.8.1). doi:10.5281/zenodo.883859

Wason, R. W., & Marvin, E. W. (1992). Riemann's "Ideen zu Einer 'Lehre von den Tonvorstellungen"': An Annotated Translation. *Journal of Music Theory*, *36*(1), 69–79.

Weber, G. (1851). *The theory of musical composition, treated with a view to a naturally consecutive arrangement of topics* (J. F. Warner, Trans.). London, UK: Messrs. Robert Cocks and Co.

Weiß, C., Mauch, M., & Dixon, S. (2018). Investigating style evolution of Western classical music: A computational approach. *Musicae Scientiae*, *23*(4), 486–507. doi:10.1177/1029864918757595

Weitin, T. (2017). Scalable reading. *Zeitschrift für Literaturwissenschaft und Linguistik*, *47*(1), 1–6. doi:10.1007/s41244-017-0048-4

Weitzmann, C. F. (1860). *Harmoniesystem*. Leipzig: C. F. Kahnt.

White, C. W. (2014). Changing styles, changing corpora, changing tonal models. *Music Perception: An Interdisciplinary Journal*, *31*(3), 224–253. doi:10.1525/MP.2014.31.3.244

White, C. W., & Quinn, I. (2016). The Yale-Classical Archives Corpus. *Empirical Musicology Review*, *11*(1), 50–58. doi:10.18061/emr.v11i1.4958

Wintner, S. (2002). Formal language theory for natural language processing. (Vol. 1, pp. 71–76). Philadelphia, PA. doi:10.3115/1118108.1118118

Wundervald, B. D., & Zeviani, W. M. (2019). Machine learning and chord based feature engineering for genre prediction in popular Brazilian music. *arXiv e-prints*, arXiv:1902.03283.

Yang, C. (2013). Ontogeny and phylogeny of language. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(16), 6324–6327. doi:10.1073/pnas.1216803110

Youngblood, J. E. (1958). Style as information. *Journal of Music Theory*, *2*(1), 24–35. doi:10.2307/842928

Yust, J. (2019). Stylistic information in pitch-class distributions. *Journal of New Music Research*, *48*(3), 217–231. doi:10.1080/09298215.2019.1606833

Zanette, D. H. (2006). Zipf's law and the creation of musical context. *Musicae Scientiae*, *10*(1), 3–18. doi:10.1177/102986490601000101

Zbikowski, L. M. (2002). *Conceptualizing music: Cognitive structure, theory, and analysis*. Oxford: Oxford University Press.

Zhu, M., Zhang, X., & Wang, H. (2016). A LDA based model for topic evolution: Evidence from information science journals. In *Proceedings of the 2016 International Conference on Modeling, Simulation and Optimization Technologies and Applications (MSOTA 2016)* (pp. 49–54). doi:10.2991/msota-16.2016.12

Zipf, G. K. (1949). *Human behaviour and the principle of least effort*. Cambridge, MA: Addison-Wesley.

Zivic, P. H. R., Shifres, F., & Cecchi, G. A. (2013). Perceptual basis of evolving Western musical styles. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(24), 10034–10038. doi:10.1073/pnas.1222336110

Fabian C. Moss

Chemin de Renens 18
CH-1004 Lausanne
(+41) 78 700 8485
fabian.moss@epfl.ch
www.fabian-moss.de

Education

2017–2019 **École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland**, *Digital and Cognitive Musicology Lab (DCML)*, Doctoral Assistant.

Jan–Mar 2016 **Massachusetts Institute of Technology, (MIT), Cambridge, MA**, *Department of Linguistics and Philosophy*, Visiting Student, supervision: Martin Rohrmeier & David Pesetsky.

2015–2017 **Technische Univserität Dresden (TUD), Dresden, Germany**, *Dresden Music Cognition Lab (DMCL)*, Doctoral Assistant.

Jan–Apr 2012 **Escola Superior de Musica de Catalunya (ESMUC), Barcelona, Spain**, ERASMUS Exchange Student, supervision: Thomas Noll.

2011–2013 **Hochschule für Musik und Tanz Köln (HfMT), Cologne, Germany**, *Musicology*, Master of Arts.

2008–2013 **Hochschule für Musik und Tanz Köln (HfMT), Cologne, Germany**, *Music Education*, Staatexamen (State Examination).

2006–2016 **Universität zu Köln (UzK), Cologne, Germany**, *Mathematics and Educational Sciences*, Staatexamen (State Examination).

2002–2005 **Friedrich-Wilhelm-Gymnasium Köln (FWG), Cologne, Germany**, Abitur.

Theses

PhD

Title *Transitions of Tonality: A Model-Based Corpus Study* (2019)

Supervisors Martin Rohrmeier & Markus Neuwirth, DCML, EPFL

Master

Title *"Theorie der Tonfelder" nach Simon und "Neo-Riemannian Theory": Systematik, historische Bezüge und analytische Praxis im Vergleich* (2012)

Supervisor Hans Neuhoff, HfMT

Publications

submitted Harasim, D., **Moss, F. C.**, Ramirez, M., & Rohrmeier, M. Cognitive modeling reveals history of major and minor in Western classical music.

under review **Moss, F. C.**, de Souza, W. F., & Rohrmeier, M. Harmony and Form in Brazilian Choro: A Corpus-Driven Approach to Musical Style Analysis.

2019 Popescu, T., Neuser, M. P., Neuwirth, M., Bravo, F., Mende, W., Boneh, O., **Moss, F. C.**, & Rohrmeier, M. (2019). The pleasantness of sensory dissonance is mediated by musical style and expertise. *Scientific Reports*, 9(1), 1070. https://doi.org/10.1038/s41598-018-35873-8

Moss, F. C., Neuwirth, M., Harasim, D., & Rohrmeier, M. (2019). Statistical characteristics of tonal harmony: A corpus study of Beethoven's string quartets. *PLOS ONE*, 14(6), e0217242. https://doi.org/10.1371/journal.pone.0217242

Landnes, K., Mehrabyan, L., Wiklund, V., Lieck, R., **Moss, F. C.**, & Rohrmeier, M. (2019). A Model Comparison for Chord Prediction on the Annotated Beethoven Corpus. In I. Barbancho, L. J. Tardón, A. Peinado, & A. M. Barbancho (Eds.), *Proceedings of the 16th Sound and Music Computing Conference (SMC 2019)* (pp. 250–254). Málaga, Spain.

2018 Neuwirth, M., Harasim, D., **Moss, F. C.**, & Rohrmeier, M. (2018). The Annotated Beethoven Corpus (ABC): A Dataset of Harmonic Analyses of All Beethoven String Quartets. *Frontiers in Digital Humanities*, 5(July), 1–5. https://doi.org/10.3389/fdigh.2018.00016

Moss, F. C., de Souza, W. F., & Rohrmeier, M.. (2018). Choro Songbook Corpus (Version 1.0) [Data set]. *Zenodo*. https://doi.org/10.5281/zenodo.1442765.

2017 **Moss, F. C.** (2017). [Review of David Huron. Voice Leading: The Science behind a Musical Art]. Music Theory & Analysis, 4(1), 119–130. https://doi.org/10.11116/MTA.4.1.7

2014 **Moss, F. C.** (2014). Tonality and functional equivalence: A multi-level model for the cognition of triadic progressions in 19th century music. In *International Conference of Students of Systematic Musicology – Proceedings* (pp. 1–8). London, UK.

Talks, Conference Presentations, and Posters

2019 **Moss, F. C.** *Transitions of Tonality: Perspectives on the Historical Changes of Tonal Pitch Relations from Computational Musicology, Music Theory, and the Digital Humanities*. University of Cologne, November 29, 2019, Cologne, Germany.

Moss, F. C. *Tracing the History of Tonality with Note Distributions*. "Corpus Research as a Means of Unlocking Musical Grammar" International Research Workshop, July 1–4, 2019, Tel-Aviv, Israel.

Moss, F. C. *Inferring Tonality from Note Distributions – Why Models Matter (Poster)*. SEMPRE Graduate Conference 2019, Cambridge, UK.

Moss, F. C. *Analyzing Tonality with Note Distributions*. First Swiss Digital Humanities Student Exchange DHX2019, Basel, Switzerland.

2018 **Moss, F. C.**, Souza, W. F. & Rohrmeier, M. *Harmony and Form in Brazilian Choro: A Corpus Study*. 15th International Conference on Music Perception and Cognition & 10th triennial conference of the European Society for the Cognitive Sciences of Music, Graz, Austria.

Aitken, C., O'Donnell, T. & Rohrmeier, M. [Poster presented by **Moss, F. C.**]. *A Maximum Likelihood Model for the Harmonic Analysis of Symbolic Music*. 15th Sound and Music Computing Conference "Sonic Crossings". Limassol, Cyprus.

Moss, F. C. *Corpus Research in Digital Musicology* (Talk and Tutorial). Seminar "Willkommen in der Matrix: Digitale Anwendungen für die Musikanalyse in Theorie und Praxis", University of Basel, Basel, Switzerland.

Harasim, D., **Moss, F. C.** & Ramirez, M. *A Brief History of Tonality* (Poster). Applied Machine Learning Days, EPFL, Switzerland.

2017 **Moss, F. C.** *Formal Grammars and Ambiguity in Extended Tonality*. Workshop and Symposium on Schenkerian Analysis "Wege der Kreativität – Zwischen Erfindung und Rekonstruktion", Universität der Künste, Berlin, Germany.

Moss, F. C., Souza, W. F. & Rohrmeier, M. *Brazilian Choro: A New Data Set of Chord Transcriptions and Analyses of Harmonic and Formal Features*. 17. Jahreskongress der Gesellschaft für Musiktheorie (GMTH) & 27. Arbeitstagung der Gesellschaft für Popularmusikforschung (GfPM) "Populäre Musik und ihre Theorien: Begegnungen – Perspektivwechsel – Transfers", Graz, Austria.

Moss, F. C., Harasim, D., Neuwirth, M. & Rohrmeier, M. *Beethovens Streichquartette – ein XML-basierter Korpus harmonischer Analysen in einem neuen Annotationssystem*. Jahrestagung der Gesellschaft für Musikforschung, Kassel, Germany.

Moss, F. C., Rohrmeier, M. *Integrating Transformational and Hierarchical Models of Extended Tonality*. 9th European Music Analysis Conference (EuroMAC), Strasbourg, France.

Rom, U., Jeßulat, A., **Moss, F. C.** & Guter, I. *Ambiguity, Illusion & Timelessness in Late and Post-Tonal Harmony*. Panel discussion at the 9th European Music Analysis Conference (EuroMAC), Strasbourg, France.

Moss, F. C. *Musik und Sprache*. Talk for Student Association "Denkzettel", TUD, Dresden, Germany.

Moss, F. C., Rohrmeier, M. & Bravo, F. *Emotional Associations Evoked by Structural Properties of Musical Scales and Abstract Visual Shapes*. KOSMOS Dialogue "Music, Emotion, and Visual Imagery", Berlin, Germany.

Harasim, D., **Moss, F. C.**, Neuwirth, M. & Rohrmeier M. *Beethoven's String Quartets: Introducing an XML-Based Corpus of Harmonic Labels Using a New Annotation System*. Music Encoding Conference, Tours, France.

2016 **Moss, F. C.** *Extended Tonality: Theoretical Challenges and their Relation to the Neuroscientific Study of Musical Syntax*. Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany.

Moss, F. C., Rohrmeier, M. *Structural Ambiguities in Language and Music* (Poster). Helsinki Summer School for Cognitive Neuroscience 2016 (HSSCN 2016).

Moss, F. C., Rohrmeier, M. *A grammatical approach to tension-resolution patterns in extended tonal harmony*. Meeting of the Computational Cognitive Science Group, Massachusetts Institute of Technology, Department of Brain and Cognitive Sciences, Cambridge, USA.

Moss, F. C., Rohrmeier, M. *Towards a syntactic account for harmonic sequences in extended tonality*. Syntax Square Meeting, Massachusetts Institute of Technology, Department of Linguistics and Philosophy, Cambridge, USA.

Moss, F. C. *Syntax of Extended Tonality: Towards a Grammar of Generalized Harmonic Functions*. Music Theory Colloquium, Boston University, College of Fine Arts, School of Music, Boston, USA.

Moss, F. C. *Generalizing Harmonic Functions: A Grammatical Approach to Extended Tonality*. Yale University, Department of Music, New Haven, USA.

Moss, F. C. & Harasim, D. *Extended Tonality and Music Cognition*. Symposium "Towards a World Music Theory", University of Hamburg, Institute for Systematic Musicology, Hamburg, Germany.

Moss, F. C. *Music Cognition and Extended Tonality: Theoretical Challenges and Empirical Implications*. Research Colloquium, University of Cologne, Cologne, Germany.

2015 **Moss, F. C.** *On generative modelling of musical form*. Seminar "Mathematics and Music", TUD, Dresden, Germany.

Moss, F. C. *'The terror of sanctity.' Tonal cues for resolving dramatic ambiguities in Wagner's Parsifal*. Seminar "Understanding Musical Structures", TUD, Dresden Germany.

2014 **Moss, F. C.** *Tonality and functional equivalence: A multi-level model for the cognition of triadic progressions in 19th century music*. International conference of Students of Systematic Musicology, Goldsmiths University, London, UK.

Moss, F. C. *Language, music and the brain: a resource-sharing framework (Patel, 2012)*. Seminar "Cognitive Neuroscience of Music", Institut for Musicology, University of Cologne, Cologne, Germany.

Awards and Scholarships

2016–2107 **Konrad Adenauer Foundation**, *PhD Scholarship*.

Aug 2016 **TUD Graduate Academy**, *Travel Award*.

Jan–Mar 2016 **Deutscher Akademischer Austauschdienst (DAAD)**, *great!$_{ipid4all}$ (group2group exchange for academic talents)*.

Sep 2014 **Society for Education and Music Psychology (SEMPRE)**, *Travel Award*.

Jan–Apr 2012 **European Union (EU)**, *ERASMUS Scholarship*.

2008–2013 **Konrad Adenauer Foundation**, *Student Scholarship*.

Teaching and Supervision

2019 Supervision of an MSc student project for "Machine Learning" course, EPFL.

2018 Teaching Assistant, "Digital Musicology" (MSc; tutorials and excercises), EPFL.

Supervision of three MSc student projects for “Machine Learning” course, EPFL.

Supervision of four MSc student projects for “Digital Musicology” course, EPFL.

2017 Peer-mentoring visiting PhD student in music theory/composition, TUD.

2016–2017 “Reading Class Musicology”, (BA; with Christoph Wald), TUD.

2015–2016 “Introduction to Musicology”, (BA; with Christoph Wald), TUD.

2015 Joint supervision of interdisciplinary project of technical design undergraduate, TUD.

2013 “Academic Writing and Research Techniques” (MA), HfMT.

Administration

Organization

2019 Workshop “Schenkerian and Tonfeld Theory for Music Analysis”, DCML, EPFL.

2015 Co-organization of lecture series “Systematic Musicology: Perception and Cognition of Music”, DMCL, TUD.

2013 Co-organization of the international conference “Musical Meter in Comparative Perspective”, HfMT.

Reviewer Activity

since 2019 *Music and Science*.

since 2016 *International Conference of Students of Systematic Musicology* (SysMus).

Memberships

since 2019 UNIL-EPFL Centre for Digital Humanities (dhCenter).

EPFL Data Champions Community.

Gesellschaft für Musikforschung (GfM).

since 2018 Co-founder and vice-president of the Digital Humanities Student Association *dhelta* at EPFL.

since 2017 Gesellschaft für Musiktheorie (GMTH).

Relevant Courses

Workshops and Summer Schools

2019 Workshop “Research Data Management: introduction”, EPFL Library, October 10, 2019.

2018 Workshop “Voice-leading schemata in theory, corpus research, and practical composition (Compose your own Chopin!)”, EPFL, September 18–20, 2018

Symposium “Archiving Intangible Cultural Heritage & Performing Arts: A Symposium and Summer School for Living Traditions”, EPFL, August 6–7, 2018

2017 Workshop “Meaning in Music: Bridging Musicological, Linguistic, and Neuroscientific Perspectives”, EPFL, December 4–6, 2017.

Summer School “Exploring Edges: An International Colloquium between the Digital Humanities, Architecture, Artistic Research, and Critical Technical Practice”, EPFL, July 11–14, 2017.

2016 Summer School “Cognitive Neuroscience of Music”, University of Helsinki, August 11–17, 2016

University Courses

2017 “Applied Data Analysis” (Robert West), EPFL.

2016 “Introduction to Schenkerian Theory” (Oliver Schwab-Felisch), TUD; “Cognitive Science” (P. Sinha, J. Tenenbaum, E. Gibson), MIT; “Computational Modeling of Phonology and Morphology” (T. O’Donnell, A. Albright), MIT.

2015 “Generative Modeling” (T. O’Donnell), TUD; “Introduction to Quantitative Methods for the Social Sciences” (Bernhard Schipp), TUD.

2012-13 “Cognitive Neuroscience of Music”, “Cognitive Musicology: Theoretical Foundations”, “Cognitive Modeling” (Uwe Seifert), UzK.

Skills

Languages	Python, Latex, HTML/CSS; German (native), English (fluent), French, Spanish (basic).
Utilities	Git, Anaconda, Jupyter Notebook/Lab.

Extra-Curricular Activities

Voice

2014–2017	Classical a-capella octet *Vokalexkursion*.
2013–2015	Cologne Cathedral Chamber Choir.
2012–2014	Conductor of several children's choirs at Musikschule Leverkusen.
2011–2013	Cologne Conservatory Chamber Choir.
2008–2013	Pop a-capella group *gezwungenermaßen*.

Instruments

since 1993	Piano
since 1994	Guitar