ELEMENTS OF MUSICALLY CONVEYED EMOTION

ELEMENTS OF MUSICALLY CONVEYED EMOTION:

INSIGHTS FROM MUSICAL AND PERCEPTUAL ANALYSES OF HISTORIC PRELUDES

By CAMERON J. ANDERSON, B.Mus. (hons.)

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Master of Science

McMaster University

Descriptive Note

McMaster University MASTER OF SCIENCE (2021) Hamilton, Ontario

(Psychology, Neuroscience & Behaviour)

TITLE: Elements of musically conveyed emotion: Insights from musical and perceptual analyses of historic preludes

AUTHOR: Cameron J. Anderson, B.Mus. (hons.) (McMaster University)

SUPERVISOR: Dr. Michael Schutz

NUMBER OF PAGES: xiv, 84.

Lay Abstract

Music's capacity to express emotion has received considerable attention in psychological and musicological research. Whereas efforts from psychology clarify the musical cues for emotion through perceptual experiments, efforts from musicology track changes in compositional practice over time—finding changing relationships between music's cues for emotion in historically diverse compositions. To date, the implications of these changing musical relationships for emotion perception remain unclear. This thesis analyzes musical scores and listeners' emotion ratings to gain insight into music's structural changes throughout history and their implications for perceived emotion. By applying statistical techniques to (i) detect musical patterns in prelude sets by J.S. Bach and F. Chopin and (ii) clarify how cue relationships influence emotion perception, this thesis sheds light on the relationship between music's historic context and its emotional meaning.

Abstract

This thesis comprises two manuscripts prepared for scholarly journals. Chapter 2 comprises an article entitled “Exploring Historic Changes in Musical Communication: Deconstructing Emotional Cues in Preludes by Bach and Chopin.”, which examines emotion perception in historic prelude sets by J.S. Bach and F. Chopin. This work connects psychological research on perceived musical emotion to musicological research describing changes in music structure. Using a technique called commonality analysis to deconstruct cues’ individual and joint roles in predicting participants’ perceived emotions, the chapter clarifies how music’s conveyed emotion can differ in compositions from different eras. Chapter 3 comprises an article entitled “Parsing Musical Patterns in Prelude Sets: Bridging Qualitative and Quantitative Epistemologies in Historical Music Research”. This chapter bridges gaps between qualitative and quantitative research on music history through an analytical approach engaging with both fields. Specifically, cluster analyses of Bach and Chopin’s preludes reveal notable differences in the composers’ expressive toolkits, consistent with work from historical and empirical music research. Through a novel analytical framework, the chapter illustrates a method for detecting groups of pieces demarcated by salient musical differences, assessing cues’ importance within these groups, and determining the most influential cue values for each group. Together, these articles provide new insight into the subtle sonic relationships influencing musical meaning and emotion perception.

Acknowledgements

As a graduate of a music program, I had some doubts about my ability to adapt to the fast pace of psychological research. Throughout my undergraduate career I received guidance and support from my supervisor, Michael Schutz, who has provided excellent training, mentorship, and feedback on both my undergraduate and master's theses. Over nearly four years, his supervision has enabled me to grow as a researcher, and as a musician by providing me the opportunity to better understand my relationship to music through research. I am also indebted to Laurel Trainor, whose thoughtfully structured course on music cognition enabled me to develop essential writing skills before beginning my master's work and alerted me to fascinating research in the field of music cognition. Her feedback in committee meetings has made me to think beyond the scope of my own research project, helping me to refine my analyses and the structure of my ideas. I also thank Victor Kuperman, whose enthusiastic feedback and methodological suggestions during committee meetings informed some of the machine learning techniques I eagerly applied in the present thesis.

This research would not be possible without the work of many lab members. Thank you to past and present research assistants on the Emotional Piano project: Jamie Ling, William Zhang, Jackie Zhou, Fermin Retnavarathan, Julia Bissessar, Jessica LaMantia, Alessandra Lima, Vivian Li, Sarah Marshall, and Thomas Samson-Williams for their assistance with data collection and invaluable contributions to reading meetings; to Andres Elizondo Lopez for his kindness; to Max Delle Grazie and Benjamin Kelly for their prominent contributions to analyses and data management throughout the project, and for their infectiously positive attitudes; to Aimee Battcock for her inspirational past work on the project; and to my colleague Liam Foley, whose work ethic, camaraderie, and friendship have been a source of inspiration.

Thank you to the anonymous reviewers and editors of Chapter 2, whose insightful feedback has improved my writing skills and my ability to communicate information scientifically.

Finally, I thank my family. Their thoughtfulness and care provide the model for how I aim to approach research.

Table of Contents

List of Figures

Chapter 2

Chapter 3

List of Tables

Chapter 2

List of Abbreviations

α	Symbol for alpha/significance level
⋂	Union symbol
b	Musical flat symbol
♯	Musical sharp symbol
ALE	Accumulated local effect
APS	Attacks per second
AR	Attack rate
CA	Commonality Analysis
CC	Commonality Coefficient
dB	Decibels
M	Mean
MO	Mode
p	*p* value symbol
PH	Pitch height
PI	Percentile interval
RRA	Ridge Regression Analysis
r	Symbol for Pearson's correlation coefficient
r^2	Symbol for coefficient of determination
RF	Random forest
RMS	Root mean squared
SD	Standard deviation
WTC	*Well-Tempered Clavier*

Declaration of Academic Achievement

This thesis contains two articles intended for publication in scholarly journals. The first examines how historical considerations of compositions influence the emotions participants perceive while listening to them. The second performs quantitative musical analyses on the same musical sets to clarify expressive patterns in each composer's expressive toolkit. In both empirical chapters, I am primary author. Because these chapters cover similar topics, there is some overlap in background provided in the Introduction and Discussion sections. However, because they are intended for different journals with different audiences, differences exist in formatting, British vs. American conventions, and the extent of statistical reporting. I summarize my contributions to each chapter below.

Chapter 1: Introduction

Author: Cameron J. Anderson

Chapter 2: Exploring Historic Changes in Musical Communication: Deconstructing Emotional Cues in Preludes by Bach and Chopin.

Authors: Cameron J. Anderson & Michael Schutz

Publication: *Psychology of Music* (in press).

Comments: My contributions to this article include deploying perceptual experiments, collecting data from human participants, performing data analysis, and preparing the manuscript. My thesis supervisor is the second author of this manuscript. I am indebted to several research assistants for assisting with data collection, as well as my supervisory committee for their insightful feedback.

Chapter 3: Parsing Musical Patterns in Prelude Sets: Bridging Qualitative and Quantitative Epistemologies in Historical Music Research

Authors: Cameron J. Anderson & Michael Schutz

Publication: manuscript for submission.

Comments: My contributions to this article include researching and developing the analytical framework for detecting musical patterns, creating visualizations, and preparing the manuscript. My thesis supervisor is the second author of the manuscript. I am indebted to my supervisory committee members for their methodological suggestions which informed the analyses.

CHAPTER 1

GENERAL INTRODUCTION

Understanding how music communicates emotion to listeners requires disentangling relationships between its structural cues for emotion. Whereas perceptual analyses of participant data reveal the nuanced cue relationships eliciting perceived emotion, musical score analyses provide an understanding of salient musical patterns distinguishing pieces' expressive characteristics. Together, psychological and musical analyses provide a framework for not only understanding musical phenomena, but also how they shape perceptions of musical meaning.

The present research blends insights from psychological and musicological research through analyses focused on prelude sets by J.S. Bach and F. Chopin—two intensely-studied composers in the western musical canon. These sets, and the pieces contained within them, receive great attention from qualitative and quantitative areas of western music research, providing opportunity to engage with insights from disciplines with distinct epistemologies. Chapter 2 examines the perceptual implications of cue relationships for musical emotion, complementing past studies by demonstrating how compositions' historical context can provide greater understanding of differences in listeners' perceived emotions. Chapter 3 introduces novel musical analyses to the same prelude sets, providing detailed summaries of musical differences within and between sets. Through deep exploration of composers' musical decisions, this chapter demonstrates that quantitative analyses of composers' musical output can clarify insights from historical music research. Together, these articles represent my efforts to understand the complex and changing musical relationships underlying music's emotional meaning.

CHAPTER 2

Anderson, C.J. & Schutz, M. (in press). Exploring Historic Changes in Musical

Communication: Deconstructing Emotional Cues in Preludes by Bach and Chopin.

Exploring Historic Changes in Musical Communication:

Deconstructing Emotional Cues in Preludes by Bach and Chopin

Cameron J. Anderson[1], Michael Schutz[1,2]

[1] Department of Psychology, Neuroscience and Behaviour, McMaster University

[2] School of the Arts, McMaster University.

Author Note

Correspondence concerning this article should be addressed to Cameron Anderson.

Email: andersoc@mcmaster.ca

Abstract

A growing body of research analyzing musical scores suggests mode's relationship with other expressive cues has changed over time. However, to the best of our knowledge the perceptual implications of these changes have not been formally assessed. Here we explore how compositional choices of 17th and 19th century composers (J.S. Bach and F. Chopin, respectively) differentially affect emotional communication. This novel exploration builds on our team's previous techniques using commonality analysis to decompose intercorrelated cues in unaltered excerpts of influential compositions. In doing so, we offer an important naturalistic complement to traditional experimental work—often involving tightly controlled stimuli constructed to avoid the intercorrelations inherent to naturalistic music. Our data indicate intriguing changes in cues' effects between Bach and Chopin, consistent with score-based research suggesting mode's "meaning" changed across historical eras (Horn & Huron, 2015). For example, mode's unique effect accounts for the most variance in valence ratings of Chopin's preludes, whereas its shared use with attack rate plays a more prominent role in Bach's. We discuss the implications of these findings as part of our field's ongoing effort to understand the complexity of musical communication—addressing issues only visible when moving beyond stimuli created for scientific, rather than artistic, goals.

Keywords: emotion, perception, mode, musicology, psychology, commonality analysis

Introduction

Music's capacity to convey emotion has fascinated history's great thinkers, garnering attention from observers ranging from Plato (Stamou, 2002) to Darwin (Darwin, 1872). Although music treatises have related musical expression to its affective outcomes since the 1600s, empirical studies did not appear until the early twentieth century (Hevner, 1935, 1937). Mode—the cue describing sets of notes composers use to convey music's emotional character—has been of particular interest. The two most common modes in western music—major and minor—are generally recognized to convey positive and negative emotion states, respectively (Crowder, 1984; Gagnon & Peretz, 2003). Mode's associations with pitch and timing also influence emotion perception, with major music typically associated with higher, faster musical passages and minor with lower, slower ones (Huron, 2008; Rigg, 1964; Turner & Huron, 2008).

Although musical cultures throughout the world employ numerous modes and scales, listeners unacculturated to western music can still identify its conveyed emotions at above-chance levels (Balkwill & Thompson, 1999; Laukka et al., 2013). Listeners with minimal exposure to western music accurately decode happy/sad connotations in major and minor pieces, suggesting mode's emotional associations are to some degree universal (Fritz et al., 2009). Despite this expressive salience, mode's imbalanced use complicates efforts to explore its effects (Larue et al., 2015). Helmholtz (1877) first noticed this cultural preference for major pieces, arguing the major mode's popularity stems from its simpler acoustic structure (see Parncutt, 2014 for summary). This preference for major pieces appears in both classical and rock music (Horn & Huron, 2015; Temperley & de

Clercq, 2013), posing challenges for researchers seeking corpora with equal numbers of major and minor pieces.

A second challenge to understanding mode's role in musical emotion is that it intercorrelates with pitch, timing, and many other cues in musical performances. To account for this multicollinearity, researchers often use controlled approaches, such as assessing the effect of adjusting excerpts to parallel major/minor keys (Dalla Bella et al., 2001; Hevner, 1935). The obvious challenges of this approach for the complex, polyphonic music heard in concert halls leads researchers to frequently choose single-line melodies as experimental stimuli. These controlled investigations are useful for understanding how specific cues can affect emotion perception, however their relation to real-world listening is tenuous. Analysis of musical scores can offer greater understanding of nuanced cue relationships in more complex musical stimuli.

Historic Differences in Musical Emotion

Addressing mode's historic emotional context requires clarification of differing perspectives. Musicologists typically demarcate western common practice music into three or four distinct epochs between the 1600s and early 1900s (Post & Huron, 2009). In contrast, music cognition studies often use "classical music" as a catch-all term for Western tonal music from any of these epochs. This generalization is useful when distinguishing between western and non-western music but makes it easy to overlook historic changes in western music. For example, mode's changing relationship with loudness and timing in nominally "classical" music suggests generalization may not hold for all eras. Cluster analyses of cues encoded in musical scores illustrate differences between music from the early Classical (~1750) and late Romantic (~1850) eras, with mode most clearly distinguishing dissimilar Classical music, and musically-expressive cues such as loudness and timing distinguishing Romantic pieces (Horn & Huron, 2015). Additionally, examination of scores reveals changes in timing (Daniele & Patel, 2013; Hansen et al., 2016; Post & Huron, 2009), scale degree use (Perttu, 2007), and dynamics (Ladinig & Huron, 2010) between Baroque (~1600–1750) and Romantic (~1800–1910) era composition. Unfortunately, pinpointing the specific effects of any one cue is challenging, given the way composers tend to use them in intercorrelated ways (Schutz, 2017).

Intercorrelations in Musical Cue Use

Although cue intercorrelations are a natural consequence of music's complexity (and likely part of its appeal), they pose significant barriers to experimental approaches aimed at understanding cues' individual contributions. To deal with this issue, psychologists have primarily adopted two approaches: (1) creating highly controlled musical stimuli to avoid multicollinearity, and (2) using diagnostic techniques to assess multicollinearity. After summarizing these approaches, we discuss a third that has proven helpful in other fields dealing with problems concerning collinearity.

Accounting for Intercorrelations

Exhaustively combining factorized cue levels to assess effects is a popular method for preventing unwanted intercorrelations. This approach overcomes challenges with discerning cue effects from intercorrelated music by focusing on discretized cue levels in controlled stimuli (Juslin & Lindström, 2010). These methods have empowered researchers to develop and validate melodic stimuli expressing particular emotion states (Paquette et al., 2013; Vieillard et al., 2008), enabling meaningful interpretations of cues' effects by clarifying their relative importance for perceived emotion (Eerola et al., 2013; Juslin & Lindström, 2010). These controlled approaches reveal cues contribute additively to expressing specific emotion states. However, they implicitly treat music's intercorrelated structure as "hopelessly confounded" (Juslin & Lindström, 2010, p. 337) instead of an important perceptual feature. Consequently, it remains unclear how these findings generalize to real-world listening.

As an alternative to removing multicollinearity, other statistical techniques aim at managing some of its most problematic consequences. This gives researchers freedom to use naturalistic stimuli while minimizing the risk of inaccurate conclusions. One study applied ridge regression analysis (RRA) to improve estimated cue effects in intercorrelated stimuli to assess their emotional significance (Costa et al., 2004), replicating well-known associations between valence and mode. Diagnostic techniques such as the variance inflation factor can optimize selections of naturalistic musical stimuli, and have identified important emotional effects from pitch (Chordia & Rae, 2008) and timing information (Luck et al., 2008) in North Indian raag and piano improvisations.

One limitation to the abovementioned techniques is that they cannot assess precisely where, and to what extent, collinearities occur. Because music's complex structure contributes to its affective meaning, understanding how cues elicit diverse emotional effects provides insight into emotional listening experiences. This is particularly important for understanding how multicollinearity differs between composers from distant historical periods.

Exploring an Alternative to Dimension Reduction

To explore the nuanced cue relationships underlying emotional expression, we employ commonality analysis (CA) to decompose emotional cue effects into unique and combined contributions. Although used since the 1960s (Nimon & Oswald, 2013), to the best of our knowledge CA had not been applied to music prior to our team's recent explorations (Battcock & Schutz, in press, 2021, 2019). Nonetheless its ability to clarify the relative importance of interrelated cues has proven powerful in disciplines ranging from education (Mood, 1969; Werner et al., 2019) and clinical psychology (Gustavson et al., 2018; Marchetti et al., 2016) to evolutionary biology (Cuevas et al., 2021). Its utility in clarifying how social, cognitive and affective factors influence behaviours surrounding medical examinations (Seibold & Roper, 1979) illustrates this power. When applied to determine barriers preventing at-risk individuals from seeking timely treatment, regression analyses revealed all three factors contributed significantly. However, after accounting for unique and joint effects, CA revealed one set of factors to be most important—an insight crucial to obtaining the greatest benefit in public messaging campaigns. As other studies provide extensive historical and theoretical detail of CA (Ray-Mukherjee et al., 2014; Seibold & McPhee, 1979), we will now focus on its musical applications.

In the present study we perform CA to separate cues' unique and combined emotional effects on perceived emotion. This provides insight into (1) effects unique to each cue; (2) effects attributable to the 'combination' of two cues; and (3) effects jointly attributable to all three cues. Fig. 1 depicts a Venn diagram conceptually representing the relationship between shared and joint cues.

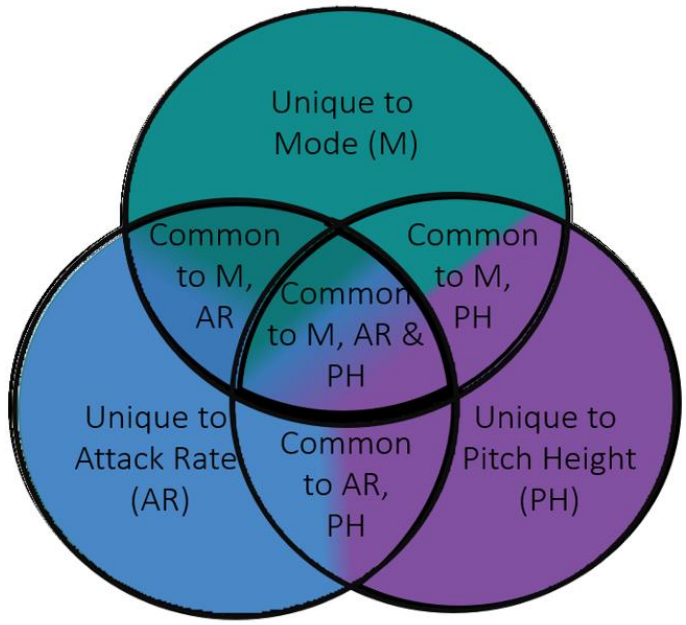


Figure 1. Venn diagram depicting the various subsets of independent variables producing coefficients in a commonality analysis performed on predictor variables mode (green), attack rate (blue) and pitch height (purple). The overlapping portions of (1) mode and attack rate, (2) mode and pitch height, and (3) attack rate and pitch height, signify their joint effect after controlling for their unique effects. The area where all three circles overlap indicates their combined contribution after controlling for their individual contributions and lower-order joint effects (e.g., mode and attack rate, pitch height and mode, etc.).

To build on score-based analyses of how cue use shifted over music history, here we compare participants' emotion ratings of music from Frederic Chopin's *Préludes* with those from a previous study of J.S Bach's *The Well-Tempered Clavier* (WTC) by Battcock and Schutz (2019). We see three benefits to using these corpora as the basis for our exploration. First, both sets contain 24 preludes written in each major and minor key, allowing for balanced assessment of how each composer uses mode to convey emotion. Second, each set was composed for a keyboard instrument, avoiding problematic differences in musical texture and timbre that are difficult to codify in scores. Third, Bach and Chopin composed during periods of clear stylistic contrast, making their music well suited for comparisons requiring music of different historical eras.

Previous findings from empirical musicology indicate notable changes in musical timing during the Romantic era, including the higher prevalence of fast minor, and slow major, music (Horn & Huron, 2015; Kelly et al., under review; Post & Huron, 2009). Here we complement and extend those findings in score-based analyses of cues, exploring differences in perceptual effects of cues (and cue combinations) in music from different eras. Our techniques are novel within music research, however our goal of relating musical structure to perceived emotion follows a long tradition within music psychology. Although this study is primarily exploratory, we hypothesized shifting patterns of cue weights with respect to mode between composers. This holds potential to complement and extend current knowledge of musically-expressed emotion—a topic fascinating several generations of musicians, psychologists, and neuroscientists alike.

Methods

Cue Extraction and Excerpt Preparation

To codify pitch and timing in these excerpts, a research assistant with advanced training in music analysis analyzed the first eight measures from the urtext edition of Frederic Chopin's *Préludes*, as well as the 24 preludes from Bach's *Well-Tempered Clavier*. This followed the method outlined in Poon and Schutz (2015) with the following four exceptions. First, here we included pick-up measures (previously omitted) as excerpts used in the experiment needed them for continuity. Second, we factored the album details into our timing calculations by summing note attacks (from the score) and dividing by the duration of the particular excerpt used in the experiment. Third, as this analysis aims to map cue values onto perceptual ratings of each excerpt, we used only a single value for pitch and timing (rather than the measure-by-measure values used by Poon and Schutz). Finally, in addition to the Kalmus edition of Bach's preludes (Bach, 1883), we analyzed the urtext edition of Chopin's (Chopin, 2007), which we have since learned is considered more authoritative.

For the Chopin experiment, we prepared eight measure excerpts of each piece from recordings by Vladimir Ashkenazy (Ashkenazy, 1993), appending a two-second fade to each excerpt using Amadeus Lite (HairerSoft, 2019) and exporting the audio at a sampling rate of 44.1kHz. The Chopin recordings took place between June 24th and 25th, 1993 at St. Charles Hall, Switzerland, providing some consistency in recording conditions (Ashkenazy, 1993 [liner notes]). The Bach recordings took place in April, 1972 in

Villingen, Germany (Gulda, 1973 [liner notes]). Additional technical details on methods can be found in this article's supplemental material.

Participants

To assess perceived emotion in Chopin's *Preludes*, we recruited 35 non-musicians (defined as having less than one year of musical training) through our institution's psychology participant pool. Participants (28 female, *M* = 18.39 years, *SD* = 1.47 years; 7 male, *M* = 17.86 years, *SD* = 0.38 years) reported normal hearing and corrected-to-normal vision. The experiment complied with our institution's research ethics board's ethics policy. The recruitment procedure of the comparison study (Battcock & Schutz, 2019) followed a similar procedure, comprising ratings of Bach's WTC from 30 participants (18 females; *M* = 19.1 years, *SD* = 3.0; 12 males, *M* = 19.7, *SD* = 2.9).

Procedure

Participants completed a consent form along with a brief demographic survey, indicating the number of hours spent each week listening to music or practicing an instrument. During the experiment, participants sat in a noise-attenuating sound booth where an experimenter provided instructions. We played each of the 24 prepared excerpts in a randomized order and participants rated the music's conveyed emotion along scales indicating (a) valence and (b) arousal of each excerpt (adapted from Russell, 1980). They listened to each excerpt through Sennheiser HDA-200 noise-cancelling headphones, registering ratings using experimental software designed in PsychoPy (Peirce et al., 2019).

We defined valence and arousal prior to the experiment, reminding participants to rate the emotion conveyed by the music as opposed to how they felt while listening. Participants completed practice trials rating four randomly selected pieces (two major, two minor). They rated valence on a scale from 1 (negative) to 7 (positive) and arousal on a scale from 1 (low) to 100 (high). Prior to the full experiment, comprising 24 trials, participants had the opportunity to ask the experimenter additional questions. After completing the experiment, the experimenter debriefed participants on the study.

Results

Replicating Cue Analyses

To assess consistency between the score-based cue quantification reported by Poon and Schutz (2015) and our approaches here (using both scores and information from audio files), we performed intraclass correlations. This revealed a high level of agreement. Additionally, we ran a series of assessments using only a single value for score-based calculations of pitch and timing (i.e., collapsing across the measure-by-measure values used in Poon & Schutz, 2015). We found strong agreement in all comparisons except one, which we attribute to statistical differences in power when collapsing across measures. Finally, we conducted Pearson correlations between mode, attack rate, and pitch height to gain insight into cue relationships in each set. This revealed differences between composers, including a stronger relationship between mode and attack rate for Bach (technical details are provided in the supplemental materials).

Perceptual Analysis

We visualized participant ratings of Chopin's *Préludes* and the preludes in Battcock and Schutz's (2019) analysis of Bach's WTC using Russell's circumplex model. See Fig. 2 for the mean valence and arousal ratings for each piece. A median split on averaged ratings suggested differences between the composers regarding mode's effect for one dimension. For valence, major pieces account for 11 of the 12 highest ratings for Bach, as well as 10 of the 12 highest for Chopin. However, the comparison for arousal ratings differed sharply, with major key pieces accounting for 8 of the 12 highest ratings for Bach, but only 4 of the 12 highest for Chopin.

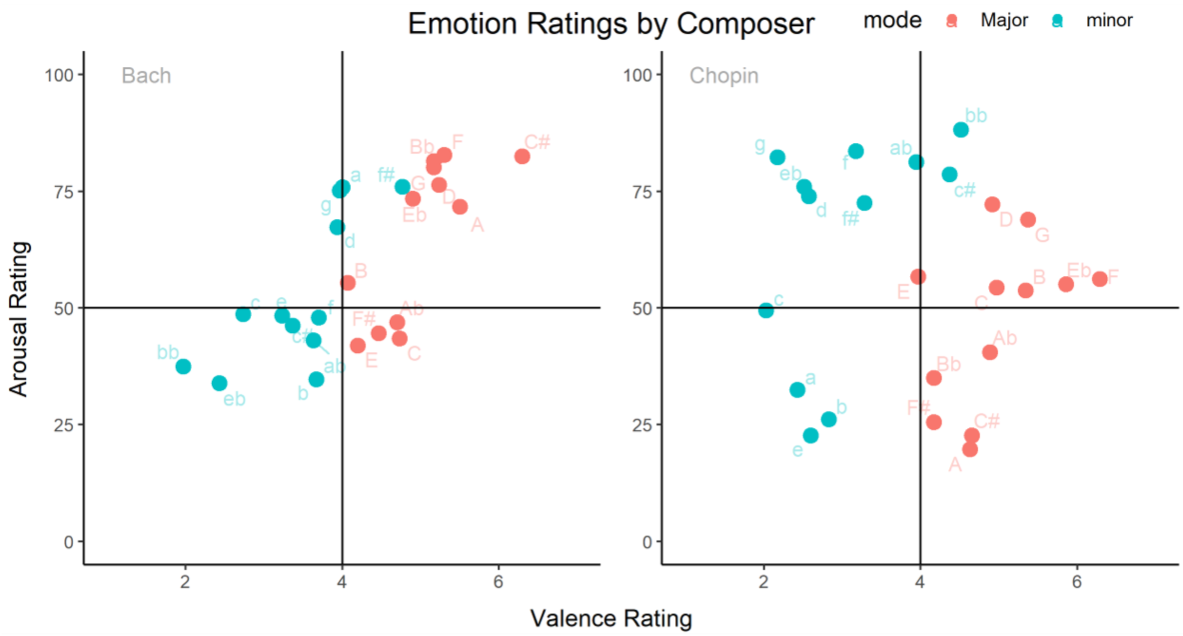


Figure 2. Visualization of mean valence and arousal ratings for each excerpt visualized using Russell's circumplex model for ratings of Bach's WTC (left) and Chopin's 24 Preludes (right). Major and minor pieces are denoted with red and blue text, respectively. Solid black lines indicate the middle of the rating scales for the valence and arousal dimensions.

Examining Cue Contributions to Emotion

We performed commonality analyses using the *yhat* package in R (Nimon et al., 2013), using nonparametric bootstrapping (averaging arousal and valence ratings within each new sample) to approximate normality through 10,000 simulated replications with replacement.[1] After bootstrapping, we estimated the 95th percentile interval (PI) for each commonality partition. Tables 1 and 2 list the commonality coefficients for each

---

[1] A minimum of 1,000 simulated replications is recommended (Berrar, 2019).

composer as percentages. Included are the commonality coefficients, 95th PI for each simulated commonality (column 4;
n = 10,000), and differences between commonalities (columns 5–10; for each, *n* = 10,000). This offers a range of simulated effect sizes along with PI estimates of statistical significance. For example, in Table 1(a) the PI indicates the difference in attack rate and pitch height's effect ranges from 4.63% to 13.51% for valence. As this does not contain 0%, we interpret this as indicating attack rate's effect on valence is significantly greater.

To afford comparison with Chopin's *Préludes*, we analyzed the 24 preludes from Bach's WTC (book 1)—a subset of the full analysis of preludes and fugues in Battcock and Schutz (2019), modeling the original participant ratings using the newly encoded cue information from the present study. We visualize these data first by showing a two-dimensional scatter plot illustrating both valence (x axis) and arousal (y axis) cue weights—derived from the 95% PIs of 10,000 sampled bootstrap replications (Fig. 3). In addition to only analyzing preludes, this method differs from Battcock and Schutz's (2021, 2019) analyses in two other ways: (1) our analyses comprise 10,000 rather than 1,000 replications; and (2) we do not perform averaging on participant ratings before bootstrapping.

Comparison with Bach

For Bach's set, attack rate uniquely plays a strong predictive role, accounting for 9.98% of variance explained in valence ratings and 29.15% of variance explained in arousal ratings. Mode accounts uniquely for 6.30% of the variance in valence ratings and only 0.79% in arousal ratings. Together, mode and attack rate jointly account for 18.4%

of variance in valence ratings and 5.27% in arousal ratings. Pitch height plays a smaller role, both uniquely (1.2% valence; 0.9% arousal), and through its joint explanation with mode (2.10% valence; -0.33% arousal) and attack rate (-0.49% valence; 5.11% arousal). Jointly, all three cues play a minimal role (-1.74% valence; 0.85% arousal). The explained variance from the cumulative effect of all commonalities is 35.70% [28.89%, 42.63%] for valence and 41.72% [32.55%, 50.08%] for arousal ratings.

For Chopin's set, attack rate accounts for 5.81% of the variance in valence and 40.5% for arousal. Mode's unique effect contributes strongly to valence (23.2%); but only accounts for 1.18% of the variance in arousal. In contrast to Bach, mode and attack rate's joint effect here contributes less prominently to valence (-4.09%) but similarly to arousal (6.26%). Pitch height contributes modestly (2.97% valence; 4.0% arousal), with its joint contributions with mode (12.31% valence; 2.88% arousal), and attack rate (3.07% valence; -3.89% arousal) explaining more variance in valence than arousal. Jointly, all three cues contribute minimally (valence: -0.46%; arousal: -1.29%). All commonalities cumulatively explain 42.84% [35.65%, 50.42%] of the variance in valence ratings and 49.60% [42.90%, 56.08%] in arousal ratings.

Dimension	Commonality	Variance %	95th Boot LL and UL CCs %	#–2 [95% Boot LL and UL CCs %]	#–3 [95% Boot LL and UL CCs %]	#–4 [95% Boot LL and UL CCs %]	#–5 [95% Boot LL and UL CCs %]	#–6 [95% Boot LL and UL CCs %]	#–7 [95% Boot LL and UL CCs %]
(a) Valence									
(1) U1	Attack Rate (AR)	9.98	**5.77, 14.83**	-4.18, 11.38	**4.63, 13.51**	**-13.11, -2.73**	**6.25, 15.37**	**3.13, 13.20**	**7.86, 16.13**
(2) U2	Mode (MO)	6.30	**2.91, 10.46**	-	**1.32, 9.46**	**-15.46, -7.98**	**2.59, 11.43**	**1.37, 7.76**	**3.93, 13.05**
(3) U3	Pitch Height (PH)	1.20	**0.42, 2.28**		-	**-20.82, -13.35**	-0.86, 4.41	**-1.71, 0.06**	**1.22, 4.92**
(4) C1	AR∩MO	18.35	**14.78, 22.02**			-	**15.04, 22.89**	**12.95, 19.54**	**16.11, 24.35**
(5) C2	AR∩PH	-0.49	-2.15, 1.36				-	-4.98, 0.07	**0.08, 2.17**
(6) C3	MO∩PH	2.10	**1.23, 3.10**					-	**1.94, 6.05**
(7) C-All	AR∩MO∩PH	-1.74	**-2.95, -.69**						-
(b) Arousal									
(1) U1	Attack Rate (AR)	29.15	**23.12, 34.49**	**22.50, 33.50**	**22.29, 33.51**	**17.68, 30.22**	**18.63, 29.40**	**23.41, 34.93**	**22.45, 33.48**
(2) U2	Mode (MO)	0.79	0.14, 1.85	-	-1.02, 0.91	**-8.48, 0.13**	**-7.77, -0.88**	**0.24, 2.46**	-0.61, 0.87
(3) U3	Pitch Height (PH)	0.90	**0.29, 1.75**		-	**-8.03, -0.46**	**-8.01, -0.32**	**0.42, 2.33**	-0.52, 0.82
(4) C1	AR∩MO	5.27	**1.66, 8.69**			-	-3.78, 3.97	**2.23, 8.83**	**0.71, 7.88**
(5) C2	AR∩PH	5.11	**1.96, 8.37**				-	**2.42, 8.58**	**1.08, 7.49**
(6) C3	MO∩PH	-0.33	-0.66, -0.09					-	**-1.70, -0.80**
(7) C-All	AR∩MO∩PH	0.85	**0.65, 1.06**						-

Table 1. Commonality analysis of Battcock & Schutz (2019) participant ratings of pieces in WTC, showing bootstrapped differences between each commonality coefficient (CC). Bolded text indicates a significant difference at the $\alpha = 0.05$ level. Coefficients reported as percentages. Table formatting adapted from Marchetti et al. (2016).

Dimension	Commonality	Variance %	95% Boot LL and UL CCs %	#–2 [95% Boot LL and UL CCs %]	#–3 [95% Boot LL and UL CCs %]	#–4 [95% Boot LL and UL CCs %]	#–5 [95% Boot LL and UL CCs %]	#–6 [95% Boot LL and UL CCs %]	#–7 [95% Boot LL and UL CCs %]
(a) Valence									
(1) U1	Attack Rate (AR)	5.81	**1.85, 10.81**	**-25.21, -8.92**	-2.48, 8.67	**3.67, 16.84**	-0.31, 6.92	-12.37, 0.28	**2.91, 10.77**
(2) U2	Mode (MO)	23.24	**17.06, 29.71**	-	**13.35, 27.35**	**20.43, 34.48**	**13.55, 26.99**	**5.80, 16.41**	**17.48, 30.25**
(3) U3	Pitch Height (PH)	2.97	**1.64, 4.74**		-	**5.38, 8.67**	-1.77, 2.19	**-11.45, -7.23**	**1.51, 5.76**
(4) C1	AR⋂MO	-4.09	**-6.08, -1.81**			-	**-10.03, -3.88**	**-18.98, -13.86**	**-6.18, -0.75**
(5) C2	AR⋂PH	3.07	**1.99, 4.24**				-	**-12.43, -6.18**	**2.67, 4.57**
(6) C3	MO⋂PH	12.31	**9.77, 14.98**					-	**9.77, 15.92**
(7) C-All	AR⋂MO⋂PH	-0.46	-1.11, 0.16						-
(b) Arousal									
(1) U1	Attack Rate (AR)	40.46	**33.49, 47.28**	**31.86, 46.56**	**29.02, 43.77**	**27.26, 41.26**	**37.49, 51.05**	**29.97, 45.07**	**34.69, 48.64**
(2) U2	Mode (MO)	1.18	**0.29, 2.39**	-	**-4.56, -1.18**	**-5.89, -4.20**	**3.46, 6.76**	**-2.37, -1.13**	**0.99, 4.21**
(3) U3	Pitch Height (PH)	4.00	**2.70, 5.53**		-	-4.35, 0.03	**5.38, 10.60**	-0.11, 2.55	**4.05, 6.70**
(4) C1	AR⋂MO	6.26	**4.61, 7.85**			-	**8.10, 12.10**	**2.32, 4.45**	**5.31, 9.72**
(5) C2	AR⋂PH	-3.89	**-5.08, -2.68**				-	**-9.00, -4.65**	**-4.13, -1.07**
(6) C3	MO⋂PH	2.88	**1.65, 4.26**					-	**2.46, 5.93**
(7) C-All	AR⋂MO⋂PH	-1.29	**-1.89, -0.66**						-

Table 2. Commonality analysis of results from current study, examining ratings of pieces in *Préludes* by Frederic Chopin, including bootstrapped differences between each commonality coefficient (CC). Bolded text indicates a significant difference at the α = 0.05 level. Coefficients reported as percentages.

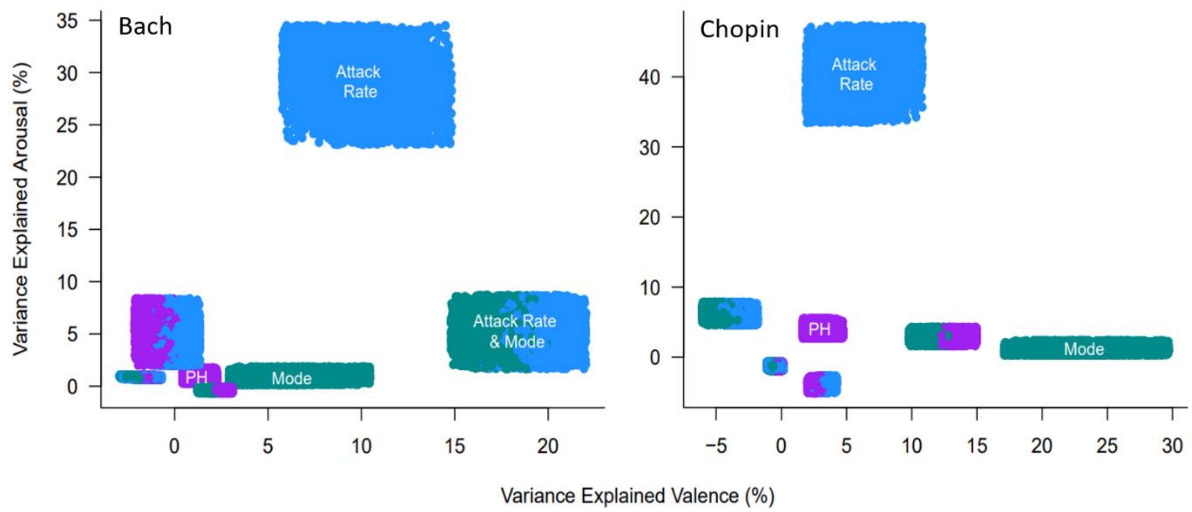


Figure 3. Scatterplot indicating cue's unique and joint contributions to valence (x axis) and arousal ratings (y axis) using 95th percentile intervals from 10,000 bootstrap simulations. Cues: attack rate (blue), mode (green), and pitch height (PH; purple). Mixed colours indicate joint contributions. Annotations appear above the three primary cues as well as the commonality between attack rate and mode.

Comparing Bach and Chopin's cue use

Comparing differences in how each commonality predicted participant ratings revealed all unique and combined effects except for attack rate's unique effect accounted for significant differences in valence ratings between the two composers. Similarly, all effects except mode's unique effect and joint effect with attack rate accounted for differences between composers in arousal ratings. Table 3 lists the commonalities and 95th PI estimates after subtracting the contributions of Bach from those of Chopin in accounting for emotion ratings. Whereas for valence the composers differed for most commonalities, the total variance explained by all cues and their combinations was not significantly different between composers (7.15% [-2.77%, 17.45%]). Similarly, the total variance explained for arousal ratings did not differ meaningfully between composers (7.88% [-2.75%, 19.10%]).

Dimension	Attack Rate (AR) %	Mode (MO) %	Pitch Height (PH) %	AR ∩ MO %	AR ∩ PH %	MO ∩ PH %	AR ∩ MO ∩ PH %	r^2 %
Valence	-4.17 [-10.41, 2.36]	**16.94 [9.50, 24.41]**	**1.77 [0.07, 3.72]**	**-22.44 [-26.69, -18.18]**	**3.56 [1.42, 5.56]**	**10.21 [7.46, 12.98]**	**1.28 [0.04, 2.61]**	7.15 [-2.77, 17.45]
Arousal	**11.31 [2.62, 20.35]**	0.39 [-.99, 1.80]	**3.10 [1.55, 4.74]**	1.00 [-2.77, 4.90]	**-9.00 [-12.44, -5.61]**	**3.21 [1.95, 4.61]**	**-2.14 [-2.78, -1.48]**	7.88 [-2.75, 19.10]

Table 3. Mean differences in explained variance after subtracting the distribution of bootstrapped commonality coefficients of Bach's WTC ratings from Chopin's 24 Preludes. Bold text indicates significance at the α = 0.05 level.

Interpreting Negative Commonality Values

Negative coefficients in commonality analysis are thought to indicate either (1) the presence of a suppressor variable removing irrelevant variance from another independent variable, or (2) a null effect equivalent to zero (Seibold & McPhee, 1979). A suppressor variable improves a predictor's estimates by removing some of its irrelevant variance. It strongly correlates with the predictor while yielding a coefficient close to zero with the dependent variable (Ray-Mukherjee et al., 2014). We conducted several assessments to determine the best interpretation of negative commonalities for these data.

To investigate negative commonalities in Chopin related to pitch, we assessed its Pearson correlations with attack rate and the pieces' average arousal ratings. For arousal ratings of Chopin's pieces, pitch appeared in multiple commonalities yielding negative values, suggesting it may have suppressed some of the variance associated with attack rate. Pitch exhibited a nonsignificant positive correlation with attack rate, $r\ (22) = .20$, $p = 0.70$ and a weak negative correlation with arousal ratings, $r\ (22) = -\ .17$, $p = 0.70$. In contrast, attack rate and arousal ratings exhibited a strong positive correlation, $r\ (22) = .84$, $p < 0.01$.[2] The positive Pearson coefficient and lack of significance with arousal suggests pitch's negative value reflects a negligible effect instead of suppression. For Bach, pitch height also appeared in negative commonalities involving mode and attack rate. It exhibited a strong negative correlation with attack rate, $r\ (22) = -.50$, $p = 0.04$, but significant correlations with neither mode, r (22) = .04, $p = 0.98$, nor valence, $r\ (22) = -.15$, p = 0.98. We interpret pitch's weak associations with mode and valence to signify pitch explains minimal variance in valence ratings.

Assessing Cues' Overall Effects

To assess the cumulative effect of each cue we performed a second bootstrap simulation, randomly sampling the unique and common effects of each cue 10,000 times, excluding commonalities with negative mean values. This allows comparing cue contributions accounting for unique and joint variance. Fig. 4 reorganizes information from Fig. 3 to convey each cue's full role. Whereas attack rate (28.32% [22.81%, 34.36%]) and mode (26.74% [21.65%, 32.34%]) contribute roughly equally to valence ratings in Bach, pitch height plays a much smaller role (3.29% [2.06%, 4.73%]). For Chopin, mode (35.56% [28.76%, 42.63%]) contributes more than both attack rate (8.87%

[2] Holm-corrected *p* values reported throughout.

[4.79%, 14.04%]) and pitch height (18.36% [15.26%, 21.62%]). The combined cue contributions are more similar for arousal, with attack rate (and joint contributions) accounting for the most variance for both composers (Bach: 40.32% [32.80%. 47.33%]; Chopin: 46.73% [39.61%, 53.84%]). Similarly, for both composers, mode (Bach: 6.91% [3.23%. 10.47%]; Chopin: 10.33% [8.08%, 12.69%]) and pitch height (Bach: 6.86% [3.26%. 10.16%]; Chopin: 6.87% [5.05%, 8.88%]) explained small but significant proportions of variance in arousal.

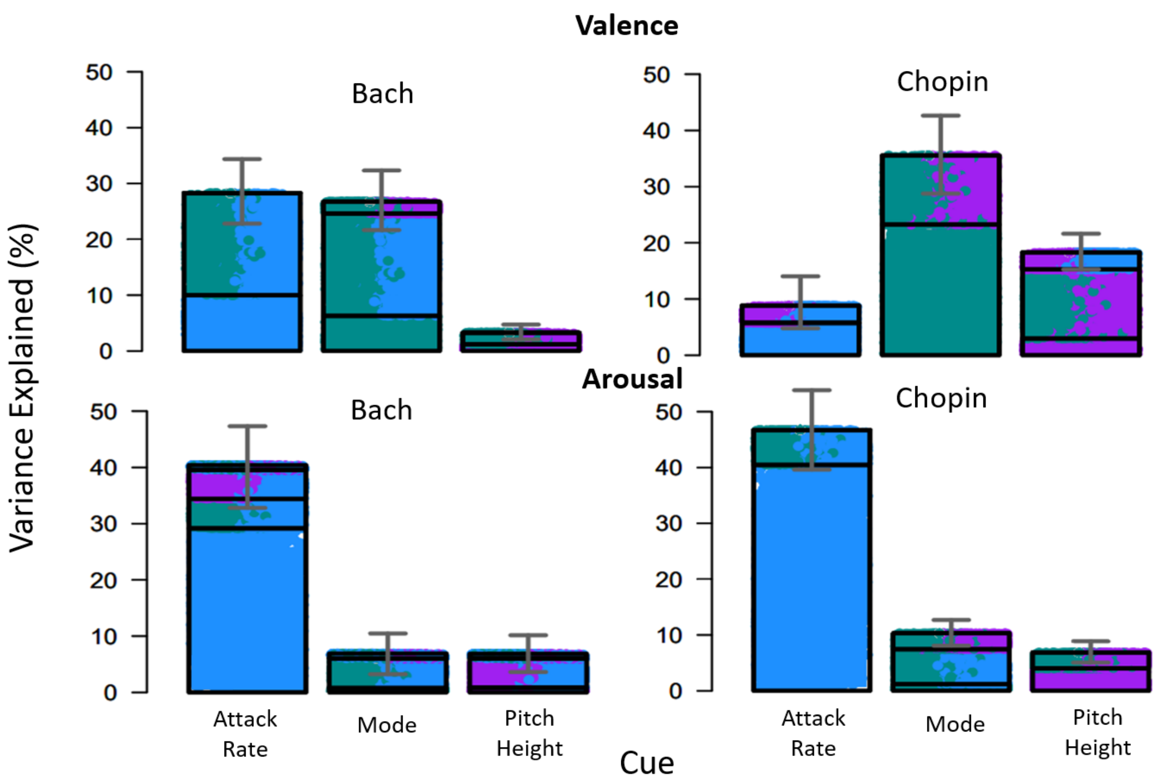


Figure 4. Bar charts indicating cue's unique (solid colour) and joint (mixed colours) contributions to valence (top) and arousal (bottom) ratings from 10,000 bootstrap simulations. Commonalities yielding a mean value of ≤ 0% are excluded. Simulated 95% percentile intervals from 10,000 summed bootstrapped resamples represent the total effect

of each cue's unique and joint contributions (grey bars). Cues: attack rate (blue), mode (green), and pitch height (purple).

Discussion

Our data provide new insight into the unique and joint effects of pitch, timing, and mode in musical sets by Bach (1722) and Chopin (1836). We believe the most important outcomes of these extensive analyses are (a) the visualizations of each composer's "emotional palate" shown in Fig. 2, and (b) insight into each cue's unique and joint contributions to musical communication shown in Figs 3 and 4. Together these findings shed new light on how emotion is communicated in unaltered musical passages by two renowned composers. Crucially, the use of natural music also affords exploring changes over musical history. This complements and extends a literature often based on simplified stimuli to avoid problems with cue collinearity. Consequently, this approach allows for new insight regarding cue collinearity—showing that it not only plays a crucial role in conveying musical emotion, but that Bach and Chopin may have used it in different ways.

Circumplex Analyses

Visualizing participant ratings on the circumplex (see Fig. 2) illustrates each set's emotional palate. Mode clearly distinguished pieces' positive/negative connotations, with minor pieces rated lower in valence than major pieces. Although both composers' major and minor pieces split relatively evenly along the valence dimension, Chopin's exhibit more variability in arousal, with twice as many minor pieces receiving high arousal ratings. This difference in the use of the minor mode is compelling in light of the increased prevalence of fast minor pieces during the Romantic era (Horn & Huron, 2015;

Post & Huron, 2009). Consequently our findings provide a useful perceptual counterpart to musicological work tracking mode's changing use and function (Pedneault-Deslauriers, 2017), contextualizing explorations of mode's aesthetic and expressive meaning by music theorists (Hatten, 2004; Parncutt, 2014).

Although intriguing, differences in the circumplex visualizations must be interpreted carefully. Despite each set containing 12 pieces in each mode (major/minor), we cannot assume equivalence across sets—Bach's Prelude in B^b is not "equivalent" to Chopin's Prelude in B^b. Therefore, understanding differences between cues' specific mappings to emotional responses is essential for understanding composer-related differences.

Investigating Cues' Unique and Joint Contributions

Commonality analysis enables disentangling how mode, timing and pitch influence valence and arousal ratings. For valence, mode's joint effect with attack rate explains most of the variance in ratings of Bach's pieces, whereas its unique effect more strongly influences ratings of Chopin's pieces (along with its correlated use with pitch height). This stronger importance of collinearity for ratings of Bach's music also affects arousal ratings—attack rate's joint effects with pitch and timing explain more variance for Bach than for Chopin.

Recomposing the CA variance through bootstrapping enabled deriving estimates of each cue's total effect (inclusive of intercorrelations). For Bach, mode and attack rate similarly affected valence (each explaining over 20% of the variance in participants' ratings). Fig. 4 reveals this similarity stems from their joint effect influencing ratings more than either cue's unique effect. In contrast, mode (35.6%) contributed more prominently than attack rate (8.9%) for Chopin, suggesting greater independence in conveying valence (although its joint contribution with pitch also affected ratings). For arousal, attack rate's cumulative effect (Bach: 40.3%, Chopin: 46.7%) explained more variance than either mode (Bach: 6.9%, Chopin: 10.3%) or pitch height (Bach: 6.9%, Chopin: 6.9%) for both composers. Comparing unique and joint contributions reveals the unique effects of mode and pitch explain less than 1% of the variance in arousal for Bach; similarly, mode uniquely explains just over 1% of the variance in arousal ratings of Chopin's pieces—suggesting its contribution to arousal largely stem from its relationship with attack rate. Consequently, deconstructing and reconstituting cues clarifies the importance of collinearity, revealing how renowned composers weave cues to convey complex emotional messages.

General Discussion

Consistent with past research, we observe associations between timing and arousal (Carpentier & Potter, 2007; Husain, Thompson, & Schellenberg, 2002) and between mode and valence (Dalla Bella et al., 2001; Gagnon & Peretz, 2003; Kastner & Crowder, 1990). However, unravelling the contributions of correlated cues offers novel insight. Most cue combinations affecting valence ratings differed between Bach and Chopin (except attack rate). Similarly, most cue combinations affected arousal ratings differently between composers (except for mode and its shared variance with attack rate). In the following sections we summarize the importance of exploring music's complex structure, suggesting steps for refining explorations of music's changing emotional implications.

Limitations and Future Directions

Our study attempts to address a long-standing fundamental challenge with using naturalistic music stimuli by combining rigorous musical analyses characteristic of empirical musicology studies with statistical analyses capable of disentangling nuanced relationships between analyzed cues and perceptual responses. Ironically, these complex methods enable clear insights into the emotional messages encoded in Bach and Chopin's preludes. Although we believe this offers exciting new possibilities for broader musical inquiry, recognizing its limitations is crucial to both interpreting our findings and guiding future research.

First, our study focused on 24-piece prelude sets from two composers of historical renown. Although their prominence makes them valuable, using only two sets precludes clearly disambiguating between composers and the eras they lived in, along with how representative these sets are of the composers' complete oeuvre. Although our results are consistent with previous findings tracking changes in musical structure during the Romantic era, further research must explore whether this extension into perceptual experimentation holds for a broader range of pieces, composers, and performer interpretations. Second, emotion judgments of a convenience sample of university students may not reflect those of a demographic varied in age, education or emotion processing abilities (see Henrich et al., 2010; Kret & Ploeger, 2015). To better understand how diverse participants perceive music's conveyed emotion, we plan to conduct online studies recruiting from a larger pool of participants using extensive inclusion criteria. Third, as both sets fit squarely within the western musical canon, future research should explore how these findings compare across cultures. Investigating diverse music in emotion research will facilitate developing informed hypotheses and accurate conclusions about music's emotional associations (see Ewell, 2020). Finally, we recognize the complexity in our combination of bootstrapping techniques with commonality analysis, acknowledging this reflects our team's efforts to consolidate a diverse toolset exploring the historic changes in musical communication. We hope this study serves as a starting point for further inquiry, and recognize dialogue with statisticians, psychologists, and musicologists will drive further musical insights. Despite these limitations, this study provides a valuable step toward sharpening our understanding of the diverse emotions elicited by music's historic shifts.

The Case for Exploring Intercorrelations

We believe disentangling music's intercorrelated structure is helpful for understanding its emotional impact. Whereas removing multicollinearity in stimuli created for experimental purposes lends insight into which cues can influence emotion perception, deconstructing multicollinearity reveals how composers *actually use* these cues in real musical compositions. This approach enables exploring questions beyond those available with stimuli constructed for psychological experiments—such as historical changes in musical cue use. Additionally, it offers the possibility of formally assessing how developments in instruments' technology influence their emotional affordances. Studies exploring the relationship between instruments' design and their emotional palate highlight the importance of these inquiries (de Souza, 2017; Huron et al., 2014; Schutz et al., 2008). Several aspects of our data complement more traditional experimental approaches—such as the strong contribution of attack rate to arousal ratings and mode's influence on valence for both composers. However, they also offer novel insight unavailable with more controlled stimuli—namely that mode's effects stem from its covariation with timing in Bach's set. Curiously, pitch's unique effect explained little variance for either emotion dimension for either composer, contrasting previous work finding strong relations to valence (Ilie & Thompson, 2006) and arousal (Jaquet et al., 2014). Instead, its joint contributions provided its strongest effects, suggesting pitch's expressivity in musical practice stems from its relation to other cues.

Our findings provide a snapshot of how two renowned composers elicit emotional responses using different cue combinations. These differences have important implications for the perception of music's meaning—and suggest it might ultimately prove beneficial to embrace music's structural complexity as a rich source of information—rather than a problematic feature to be avoided. We hope these approaches applied to a variety of musically expressive styles and music from different historical periods will shed new light on the complexity of musical emotion—unveiling how composers' musical choices continue to captivate audiences separated by centuries.

References

Ashkenazy, V. (1993). Chopin, 24 Preludes, Op.28. *[CD-ROM]*, *DECCA*.

Bach, J. S. (1883). *The Well-Tempered Clavier Book I ed. Hans Bischoff*. Kalmus & Co. (Music Publisher).

Balkwill, L.-L., & Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception: An Interdisciplinary Journal*, *17*(1), 43–64.

Battcock, A., & Schutz, M. (in press). *Individualized interpretation: Exploring structural and interpretive effects on evaluations of emotional content in Bach's Well Tempered Clavier.*

Battcock, A., & Schutz, M. (2019). Acoustically Expressing Affect. *Music Perception*, *37*(1), 66–91.

Battcock, A., & Schutz, M. (2021). Emotion and expertise : How listeners with formal music training use cues to perceive emotion. *Psychological Research*.

Berrar, D. (2019). Introduction to the Non-Parametric Bootstrap. *Encyclopedia of Bioinformatics and Computational Biology*, 766–773. https://doi.org/10.1016/B978-0-12-809633-8

Carpentier, Francesca R Dillman, Potter, R. F. (2007). Effects of Music on Physiological Arousal: Explorations into Tempo and Genre. *Effects of Music on Physiological Arousal: Explorations into Tempo and Genre*, *10*(3), 339–363.

https://doi.org/10.1080/15213260701533045

Chopin, F. (2007). *Préludes ed. Norbert Müllemann*. G Henle Verlag.

Chordia, P., & Rae, A. (2008). Understanding emotion in raag: An empirical study of listener responses. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 4969 LNCS*, 110–124. https://doi.org/10.1007/978-3-540-85035-9_7

Costa, M., Fine, P., Enrico, P., & Bitti, P. E. R. (2004). Interval distributions, mode, and tonal strength of melodies as predictors of perceived emotion. *Music Perception: An Interdisciplinary Journal*, *22*(1), 1–14.

Crowder, R. G. (1984). Perception of the major/minor distinction: I. Historical and theoretical foundations. *Psychomusicology: A Journal of Research in Music Cognition*, *4*(1–2), 3–12. https://doi.org/10.1037/h0094207

Cuevas, A., Ravinet, M., Sætre, G., & Eroukhmanoff, F. (2021). Intraspecific genomic variation and local adaptation in a young hybrid species. *Molecular Ecology*, *30*(3), 791–809. https://doi.org/10.1111/mec.15760

Dalla Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition*, *80*(3), B1–B10.

Daniele, J. R., & Patel, A. D. (2013). An Empirical Study of Historical Patterns in Musical Rhythm: Analysis of German & Italian Classical Music Using the nPVI Equation. *Music Perception: An Interdisciplinary Journal*, *31*(1), 10–18.

Darwin, C. (1872). The Expression of the Emotions in Man and Animals. In *Animals*. John Murray.

de Souza, J. (2017). *Music at Hand: Instruments, Bodies, And Cognition*. Oxford University Press.

Eerola, T., Friberg, A., & Bresin, R. (2013). Emotional expression in music: Contribution, linearity, and additivity of primary musical cues. *Frontiers in Psychology*, *4*(July), 1–12. https://doi.org/10.3389/fpsyg.2013.00487

Ewell, P. A. (2020). Music Theory and the White Racial Frame. *Music Theory Online*, *26*(2). https://doi.org/10.30535/mto.26.2.4

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., Friederici, A. D., & Koelsch, S. (2009). Universal recognition of three basic emotions in music. *Current Biology*, *19*(7), 573–576.

Gagnon, L., & Peretz, I. (2003). Mode and tempo relative contributions to" happy-sad" judgements in equitone melodies. *Cognition and Emotion*, *17*(1), 25–40.

Gulda, F. (1973). Bach: The Well-Tempered Clavier, Book I - Prelude #24 in B minor, BWV 869 [Recorded by F. Gulda]. In *Bach: The Well-Tempered Clavier, Bk I*. [CD]. MPS-Tonstudio, Villingen, Germany: Decca (1973).

Gustavson, D. E., Pont, A. du, Whisman, M. A., & Miyake, A. (2018). Evidence for transdiagnostic repetitive negative thinking and its association with rumination, worry, and depression and anxiety symptoms: A commonality analysis. *Collabra:*

Psychology, *4*(1), 1–18. https://doi.org/10.1525/collabra.128

HairerSoft. (2019). *Amadeus Lite*.

Hansen, N. C., Sadakata, M., & Pearce, M. T. (2016). Nonlinear changes in the rhythm of european art music: Quantitative support for historical musicology. In *Music Perception* (Vol. 33, Issue 4, pp. 414–431). University of California Press. https://doi.org/10.1525/MP.2016.33.4.414

Hatten, R. S. (2004). *Interpreting musical gestures, topics, and tropes: Mozart, Beethoven, and Schubert*. Indiana University Press.

Helmholtz, H. Von. (1877). *On the sensations of tone (English translation AJ Ellis, 1885, 1954)*.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. In *Nature* (Vol. 466, Issue 7302, p. 29). Nature Publishing Group. https://doi.org/10.1038/466029a

Hevner, K. (1935). The affective character of the major and minor modes in music. *The American Journal of Psychology*, *47*, 103–118. http://www.jstor.org/stable/1416710

Hevner, K. (1937). The affective value of pitch and tempo in music. *The American Journal of Psychology*, *49*(4), 621–630.

Horn, K., & Huron, D. (2015). On the changing use of the major and minor modes 1750–1900. *Music Theory Online*, *21*(1), 1–11. http://www.mtosmt.org/issues/mto.15.21.1/mto.15.21.1.horn_huron.html

Huron, D. (2008). A comparison of average pitch height and interval size in major- and minor-key themes : Evidence consistent with affect-related pitch prosody. *Empirical Musicology Review*, *3*(2), 59–63.

Huron, D., Anderson, N., & Shanahan, D. (2014). "You Can't Play a Sad Song on the Banjo:" Acoustic Factors in the Judgment of Instrument Capacity to Convey Sadness. *Empirical Musicology Review*, *9*(1), 29. https://doi.org/10.18061/emr.v9i1.4085

Husain, G., Thompson, W. F., & Schellenberg, E. G. (2002). Effects of Musical Tempo and Mode on Arousal, Mood, and Spatial Abilities. *Music Perception*, *20*(2), 151–171. https://doi.org/10.1525/mp.2002.20.2.151

Ilie, G., & Thompson, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception: An Interdisciplinary Journal*, *23*(4), 319–330.

Jaquet, L., Danuser, B., & Gomez, P. (2014). Music and felt emotions: How systematic pitch level variations affect the experience of pleasantness and arousal. *Psychology of Music*, *42*(1), 51–70. https://doi.org/10.1177/0305735612456583

Juslin, P. N., & Lindström, E. (2010). Musical expression of emotions: Modelling listeners' judgements of composed and performed features. *Music Analysis*, *29*(1–3), 334–364. https://doi.org/10.1111/j.1468-2249.2011.00323.x

Kastner, M. P., & Crowder, R. G. (1990). Perception of the major/minor distinction: IV.

Emotional connotations in young children. *Music Perception: An Interdisciplinary Journal*, *8*(2), 189–202.

Kelly, B. O., Anderson, C. J., & Schutz, M. (under review). *Perceptual Categorization of Music between Eras*.

Kret, M. E., & Ploeger, A. (2015). Emotion processing deficits: A liability spectrum providing insight into comorbidity of mental disorders. In *Neuroscience and Biobehavioral Reviews* (Vol. 52, pp. 153–171). Elsevier Ltd. https://doi.org/10.1016/j.neubiorev.2015.02.011

Ladinig, O., & Huron, D. (2010). Dynamic levels in Classical and Romantic keyboard music: Effect of musical mode. *Empirical Musicology Review*, *5*(2), 51–56.

Larue, J., Wolf, E., Bonds, M. E., Walsh, S., & Wilson, C. (2015). Symphony. In *Grove Music Online*. http://www.oxfordmusiconline.com.

Laukka, P., Eerola, T., Thingujam, N. S., Yamasaki, T., & Beller, G. (2013). Universal and culture-specific factors in the recognition and performance of musical affect expressions. *Emotion*, *13*(3), 434–449. https://doi.org/10.1037/a0031388

Luck, G., Toiviainen, P., Erkkilä, J., Lartillot, O., Riikkilä, K., Mäkelä, A., Pyhäluoto, K., Raine, H., Varkila, L., & Värri, J. (2008). Modelling the relationships between emotional responses to, and musical content of, music therapy improvisations. *Psychology of Music*, *36*(1), 25–45. https://doi.org/10.1177/0305735607079714

Marchetti, I., Loeys, T., Alloy, L. B., & Koster, E. H. W. (2016). Unveiling the structure of cognitive vulnerability for depression: Specificity and overlap. *PLoS ONE*, *11*(12), 1–16. https://doi.org/10.1371/journal.pone.0168612

Mood, A. M. (1969). Macro-Analysis of the American Educational System. *Operations Research*, *17*(5), 770–784. https://doi.org/10.1287/opre.17.5.770

Nimon, K. F., & Oswald, F. L. (2013). Understanding the Results of Multiple Linear Regression: Beyond Standardized Regression Coefficients. *Organizational Research Methods*, *16*(4), 650–674. https://doi.org/10.1177/1094428113493929

Nimon, K., Oswald, F., Yhat, J. R. https://CRAN. R. org/package=, & 2013, U. (2020). *yhat: Interpreting Regression Effects. R package version 2.0-0*. https://cran.r-project.org/package=yhat

Paquette, S., Peretz, I., & Belin, P. (2013). The "Musical Emotional Bursts": A validated set of musical affect bursts to investigate auditory affective processing. *Frontiers in Psychology*, *4*, 1–7. https://doi.org/10.3389/fpsyg.2013.00509

Parncutt, R. (2014). The emotional connotations of major versus minor tonality: One or more origins? *Musicae Scientiae*, *18*(3), 324–353. https://doi.org/10.1177/1029864914542842

Pedneault-Deslauriers, J. (2017). The French Path: Early Major-Minor Theory from Jean Rousseau to Saint-Lambert. *Music Theory Online*, *23*(1).

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Perttu, D. (2007). A Quantitative Study of Chromaticism: Changes Observed in Historical Eras and Individual Composers. In *Empirical Musicology Review* (Vol. 2, Issue 2).

Poon, M., & Schutz, M. (2015). Cueing musical emotions: An empirical analysis of 24-piece sets by Bach and Chopin documents parallels with emotional speech. *Frontiers in Psychology*, *6*(November), 1–13. https://doi.org/10.3389/fpsyg.2015.01419

Post, O., & Huron, D. (2009). Western Classical music in the minor mode is slower (Except in the Romantic period). *Empirical Musicology Review*, *4*(1), 2–10.

Ray-Mukherjee, J., Nimon, K., Mukherjee, S., Morris, D. W., Slotow, R., & Hamer, M. (2014). Using commonality analysis in multiple regressions: A tool to decompose regression effects in the face of multicollinearity. *Methods in Ecology and Evolution*, *5*(4), 320–328. https://doi.org/10.1111/2041-210X.12166

Rigg, M. G. (1964). The Mood Effects of Music: A Comparison of Data from Four Investigators. *The Journal of Psychology*, *58*(2), 427–438. https://doi.org/10.1080/00223980.1964.9916765

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178.

Schutz, M., Huron, D., Keeton, K., & Loewer, G. (2008). The happy xylophone: Acoustics affordances restrict an emotional palate. *Empirical Musicology Review*, *3*(3), 126–135.

Seibold, D. R., & McPhee, R. D. (1979). Commonality Analysis: a Method for Decomposing Explained Variance in Multiple Regression Analyses. *Human Communication Research, 5*(4), 355–365. https://doi.org/10.1111/j.1468-2958.1979.tb00649.x

Seibold, D. R., & Roper, R. E. (1979). Psychosocial Determinants of Health Care Intentions: Test of the Triandis and Fishbein Models. *Annals of the International Communication Association, 3*(1), 625–643. https://doi.org/10.1080/23808985.1979.1192378

Stamou, L. (2002). Plato and Aristotle on Music and Music Education: Lessons From Ancient Greece. *International Journal of Music Education*, *os-39*(1), 3–16. https://doi.org/10.1177/025576140203900102

Temperley, D., & de Clercq, T. (2013). Statistical Analysis of Harmony and Melody in Rock Music. *Journal of New Music Research*, *42*(3), 187–204. https://doi.org/10.1080/09298215.2013.839525

Turner, B., & Huron, D. (2008). A comparison of dynamics in major-and minor-key works. *Empirical Musicology Review*, *3*(2), 64–68.

Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., & Bouchard, B. (2008). Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cognition and Emotion*, *22*(4), 720–752. https://doi.org/10.1080/02699930701503567

Werner, K. M., Milyavskaya, M., Klimo, R., & Levine, S. L. (2019). Examining the unique and combined effects of grit, trait self-control, and conscientiousness in predicting motivation for academic goals: A commonality analysis. *Journal of Research in Personality*, *81*, 168–175. https://doi.org/10.1016/j.jrp.2019.06.003

Supplementary Material for

Exploring Historic Changes in Musical Communication:

Deconstructing Emotional Cues in Preludes by Bach and Chopin

Agreement Measures

Our score encodings differed slightly from those of Poon and Schutz (2015) as we used a more authoritative urtext score edition for Chopin. To assess the effect of this differences, as well as consistency between analyses, we employed a two-way mixed effects models from the irr package in R (Gamer et al., 2019).[3] As expected, this revealed high inter-rater agreement for extractions of Bach (attack rate: 0.99, 95% confidence interval [0.99, 1.0]; pitch height: 0.97 [0.96, 0.98]) and Chopin (attack rate: 0.98 [0.98,0.99]; pitch height: 0.99 [0.98,0.99]).

Having determined the timing information to be consistent with Poon and Schutz's score analysis, we next compared the score-based timing information of our score analysis to the recorded performance timing participants heard in the experiments. We found moderately high inter-class consistency between these approaches for Bach (attack rate: 0.88, [0.75, 0.95]) and Chopin (attack rate: 0.94, [0.86, 0.97]). See Table 1 for summary statistics of attack rate, pitch height, and excerpt duration between each composer's major and minor pieces.

[3] We excluded anacruses from these comparisons as they are unique to the present study's analysis.

Supplementary Table 1. Summary of attack rate, pitch height, and excerpt durations for major and minor pieces including the number of pieces analyzed in each mode for both composers (*n*), as well as their mean, standard deviation (*SD*) and median values.

Composer	Mode	Attack Rate		Pitch Height		Duration (seconds)	
		Mean (*SD*)	**Median**	**Mean (*SD*)**	**Median**	**Mean (*SD*)**	**Median**
Bach	Major (*n* = 12)	6.78 (2.61)	7.15	43.08 (3.43)	43.86	23.43 (6.46)	24.86
	Minor (*n* = 12)	4.09 (2.19)	3.86	43.34 (2.71)	43.39	35.96 (13.21)	35.15
Chopin	Major (*n* = 12)	4.87 (2.70)	4.66	41.99 (3.63)	42.21	21.69 (12.11)	17.20
	Minor (*n* = 12)	5.81 (3.13)	6.65	38.72 (5.11)	39.13	24.03 (12.11)	18.01

Comparing Cue Analyses

In addition to encoding differences, our quantification of attack rate differed from Poon and Schutz's study by using performance-based timing information. Whereas their theoretical analysis used rhythms from scores, we used a slightly different method for quantifying timing to align with participants' listening experience. This involved (a) including anacruses participants heard in the experiment, and (b) calculating attack rate by dividing the number of note attacks by the duration of the corresponding recording.

This resulted in a single attack rate measure per piece (rather than per measure). Similarly, with respect to pitch, we averaged pitch height across measures for each excerpt, whereas that study used raw values.

To assess whether these differences affected the present study's findings, we compared pitch and timing across sets (correcting for multiple comparisons with Holm's method). Replicating that study's findings comparing differences between composers, we found differences for neither major, nor minor, preludes in attack rate (*T*-Tests) or pitch height (Mann-Whitney *U*-Tests). Similarly, for attack rate our analysis replicated significant differences between Bach's, but not Chopin's, major and minor preludes. However, for pitch height our analysis revealed significant differences between Bach's major and minor pieces, but not Chopin's, contrasting their finding of significant differences in pitch height for Chopin. This ultimately reflects a difference in statistical power resulting from varied encoding approaches—here we coded only a single pitch value per piece, rather than per measure as in the original analysis. Nevertheless, the resulting values are highly similar, differing by less than 0.1 semitones.[4]

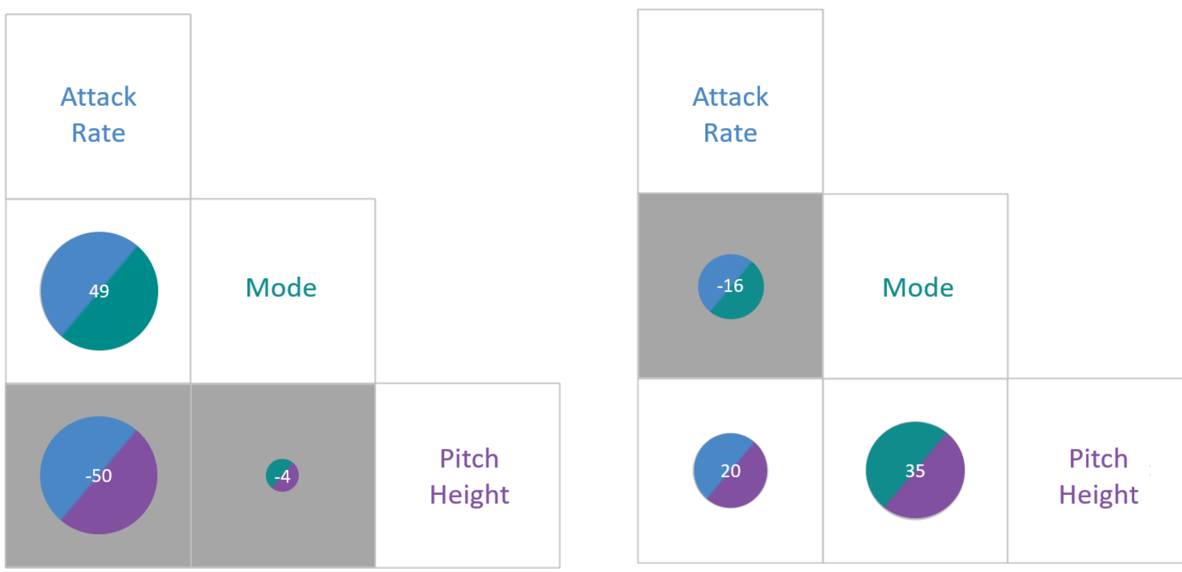
Cue Correlations

To assess differences in cue collinearity between sets we performed Pearson correlations for each composer's cue use. For Bach, attack rate moderately correlated positively with mode,

[4] Poon and Schutz (2015) reported a difference of 3.2 semitones whereas we find a difference of 3.27 semitones.

r (22) = 0.49, *p* = 0.04, and negatively correlated with pitch height, *r* (22) = -.50, *p* = 0.04, whereas mode and pitch height did not correlate significantly, *r* (22) = -.04, *p* = 0.85. For Chopin, attack rate neither significantly correlated with mode, *r* (22) = -.16, *p* = 0.70; nor pitch,

r (22) = .20, *p* = 0.70. Similarly, pitch and mode did not significantly correlate after adjusting for multiple comparisons, *r* (22) = .35, *p* = 0.29 (Supplement Figure 1).



Supplementary Figure 1. Correlation plot indicating Pearson correlations between attack rate (blue), mode (green) and pitch height (purple) for Bach (left) and Chopin (right). Colours correspond to compared cues two cues (e.g., the blue-green circle depicts the Pearson correlation between attack rate and mode). Circle size reflects correlation strength, with grey and white backgrounds indicating negative and positive correlations, respectively. *r* values reported as percentages.

References

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2019). *irr: Various Coefficients of Interrater Reliability and Agreement*. R package version 0.84.1. https://cran.r-project.org/package=irr

Poon, M., & Schutz, M. (2015). Cueing musical emotions: An empirical analysis of 24-piece sets by Bach and Chopin documents parallels with emotional speech. *Frontiers in Psychology, 6*(November), 1–13. https://doi.org/10.3389/fpsyg.2015.01419

CHAPTER 3

Anderson, C.J. & Schutz, M. (manuscript prepared for submission). Parsing Musical Patterns in Prelude Sets: Bridging Qualitative and Quantitative Epistemologies in Historical Music Research.

Parsing Expressive Patterns in Prelude Sets:

Bridging Qualitative and Quantitative Analyses of Bach and Chopin

Cameron J. Anderson[1], Michael Schutz[1,2]

[1] Department of Psychology, Neuroscience and Behaviour, McMaster University

[2] School of the Arts, McMaster University.

Author Note

Correspondence concerning this article should be addressed to Cameron Anderson.

Email: andersoc@mcmaster.ca

Abstract

Cluster analysis is a popular technique among empirical musicologists for its ability to find meaningful patterns in musical works' expressive cues. Although several studies use cluster analysis to track large-scale stylistic differences between periods in music history, few have applied this technique to explore the corpora of individual composers. The present analysis demonstrates how cluster analysis can provide insight into composers' expressive cue use, providing an epistemological link between insights from quantitative studies of large musical corpora and qualitative studies examining specific composers and scores. Specifically, we examine how pitch, timing, and dynamic information cluster into groups demarcated by expressive differences in recordings of J.S. Bach and F. Chopin's preludes. We also explore how accumulated local effects—powerful visualizations clarifying the complex decisions of machine learning algorithms—provide insight into the nuanced cue patterns distinguishing clusters. To the best of our knowledge, this is the first application of such techniques to music. Finally, we highlight our findings using an interactive data exploration tool, allowing readers to engage with the analyzed musical recordings through listening.

Keywords: musicology, corpus analysis, cluster analysis, accumulated local effects, preludes, Well-Tempered Clavier

[1.0] **Introduction**

On Musicology's Changing Epistemological Framework

[1.1] The pursuit to develop an objective understanding of musical expression has a long history dating back to Pythagoras (ca. 570–495 BCE) who purportedly first related consonant and dissonant sounds to the ratio relationship of plucked strings' wavelengths (Caleon and Ramanathan 2008; but see Rehding 2016 for a critical counterpart). Despite growing interest in chronicling innovations in instrument design and theories of musical phenomena over the centuries, the formal study of music did not emerge as a field independent of general knowledge until the mid-late 1800s (Duckles et al. 2001). Comprising both historic and systematic subdisciplines, music researchers from different backgrounds combined efforts applying qualitative and quantitative techniques to understand historic musical developments. Whereas qualitative analyses classified corpora, composers, and instruments into stylistically distinct epochs; quantitative techniques shed light on acoustic phenomena, along with scientific, aesthetic, and psychological accounts of music practice (Adler 1885; Duckles et al. 2001).

[1.2] In recent decades, an expanded analytical toolkit for exploring musical eras empowers researchers to conduct large-scale analyses using notational and audio features derived from specialized software and databases. Taking advantage of tools for automating analyses, psychologists and technical specialists working in music information retrieval can investigate musicological phenomena conventionally explored by music historians. These efforts harness advances in technology to derive classifications in a "bottom-up" manner, providing a data-driven counterpart to the "top-down" approaches of historians merging score analysis and deep historical knowledge. These quantitative avenues in historical music research offer a new lens for examining western music's stylistic evolution throughout history.

Evidence for Expressive Changes Across Music History

[1.3] Empirical research examining music's stylistic changes describes marked differences in the relationships between music's expressive cues across eras. Studies examining mode—the structural cue often associated with music's conveyed emotional tenor—find nuanced changes in its relationships with other cues, such as loudness and timing during the Romantic era (Horn and Huron 2015; Post and Huron 2009; Ladinig and Huron 2010). These findings provide a critical counterpart to music theory textbooks' treatment of the major and minor modes' expressive functions (see Hatten 2004 for a theoretical account of nuance in mode's expressive meaning).[5] Other research provides descriptive accounts of differences in how composers use specific pitches, intervals, and chords in their tonal vocabulary—capturing changes in tonality across music history (Albrecht and Shanahan 2013; Broze and Shanahan 2013; Weiß et al. 2019; Harasim et al. 2021). Studies examining rhythmic variability in music between 1600 and 1900 find notable differences between composers from different countries and historical eras (Daniele and Patel 2013; Hansen, Sadakata, and Pearce 2016; Patel and Daniele 2003; but see Condit-Schultz 2019 for criticism of the measure commonly used to evaluate rhythmic contrast). This growing interest in exploring musicological questions with empirical methods can clarify changes to music's structure and function, enabling music researchers to enhance and refine historical accounts with scientific evidence.

[5] In the introduction to a widely used textbook among music theory classrooms, Aldwell, Schachter, and Cadwallader (2002) describe how some musicians object to conventional wisdom claiming that major and minor music express "happiness" and "sadness", respectively. Empirical methods provide a way to assess assumed knowledge taught in music theory classrooms through an empirical lens.

A Method for Discovering Patterns in Musical Expression

[1.4] Although empirical explorations of musical structure employ diverse statistical tools, cluster analysis has long been favored for its flexible and statistically agnostic approach to detecting contrasting expressive groups in musical corpora. This powerful technique identifies distinct data patterns while reducing biases associated with more subjective methodologies. Early applications of the technique in the biological sciences enabled researchers to avoid subjective judgments in the development of taxonomic hierarchies (Sokal 1963; Sokal and Michener 1958).[6] Since then, modern computing technology has streamlined previously laborious calculations offering rapid analysis and classifications of large and diverse datasets. A result of this increased accessibility is a rapid expansion of cluster analysis studies in diverse disciplines including biology (Sharma, Ali, and Ghersi 2018; Ge et al. 2005), psychology (Shensa et al. 2018; Stenlund, Lyrén, and Eklöf 2018), and sociology (Stylidis 2018; Kabók, Radišić, and Kuzmanović 2017).

[1.5] Music researchers have employed cluster analysis to classify similar sounds (Howard and Silverman 1976); identify learning styles (Zhukov 2007) and classes of injuries among musicians (Cruder et al. 2021); and explore the prevalence of musical styles across cultures (Savage and Brown 2014). More recent adoption of the technique to map expressive features and structural elements across history has proven particularly valuable for gaining insight into western musical practice (Albrecht and Huron 2014; Albrecht and Shanahan 2013; White 2014; Weiß et al. 2019).

[6] Descriptions of cluster analysis date as far back as 1911 to the botanist Czekanowski (Blashfield and Aldenderfer 1988).

Exploring Expressive Shifts with Cluster Analysis

[1.6] Cluster analysis offers a data-driven method to critically assess theories about music's changes throughout history. Horn and Huron (2015) used this technique to examine how mode's expressive relationships with timing, dynamics, and articulation changed between the Classical and Romantic eras—periods reputed for notable stylistic differences. In a cluster analysis of 750 works composed by 330 composers, they identified pertinent shifts in the major and minor modes' relationships to prosodic cues in periods spanning 1750-99 and 1850-99.[7] In pieces sampled from the Classical era, mode played a prominent role in separating clusters, followed by combinations of loudness and timing. In contrast, mode's role in separating clusters in Romantic era pieces was overshadowed by combinations of prosodic cues such as loudness and timing.[8] The authors identified a doubling of major pieces demarcated by soft/quiet/legato passages,[9] arguing the category's increasing prevalence strayed from conventional descriptions of the major and minor modes expressing "happiness" and "sadness", respectively.

[1.7] Other research using cluster analysis has explored mode from a lexical perspective (Harasim et al. 2021), building on previous work examining pitch class distributions in historically diverse music (Albrecht and Shanahan 2013; Albrecht and Huron 2014; White 2014). The authors used cluster analysis in tandem with other statistical techniques to examine how well the pitch class distribution of pieces from four different eras characterize the major and minor modes. Their findings contrast with general accounts that mode is clearly dichotomized into "major" and "minor" categories in the Romantic era, instead highlighting decreased tonal

[7] The authors' analysis examined mode's relationship with loudness, timing, and articulation.
[8] Note that while the output of a cluster analysis provides an agnostic means for comparing high level clusters, interpretations of the meaning of the clustering is ultimately subjective.
[9] Our team's recent work explores the psychological implications of music's shifting cue relationships (Anderson and Schutz, in press; Kelly, Anderson, and Schutz, under review).

clarity compared to music from the Baroque and Classical eras. Consistent with the observed reduction in modal clarity, application of a key-finding algorithm to music by J.S. Bach and F. Chopin revealed worse performance in Chopin's music (Schmuckler and Tomovski 2005). These examples highlight how exploring large-scale shifts in musical phenomena along with corpora of individual composers can shed new light on music's complex historical changes.

The Case for Examining Composers' Corpora

[1.8] Here we provide a connecting link between qualitative and quantitative studies of music's changes throughout history by using cluster analysis to explore Bach's *Well-Tempered Clavier Book 1* (WTC; Bach, 1883 [1722]) and Chopin's *Preludes* (Chopin 1839)—musical sets our team has explored extensively from empirical (Poon and Schutz 2015; Schutz 2017) and psychological perspectives (Anderson and Schutz, in press; Kelly, Anderson, and Schutz, in press; Battcock and Schutz 2021). Our approach complements empiricists' bottom-up inferences from studies of large corpora, along with historians' top-down inferences from analyses of individual scores by providing focused analyses of comprehensive corpora. These corpora enable assessing musicological questions with a high degree of precision. Both sets contain one piece in each major and minor key, allowing comparison of expressive differences between modes; they are written for keyboard instruments, enabling reasonably consistent analyses and interpretations of findings; [10] finally, they have been explored extensively from historical and empirical perspectives, providing opportunity to compare and contrast results with qualitative and quantitative studies past (Battcock and Schutz 2019; Beuerman 2003; Bisesi, Macritchie, and Parncutt 2012; Ledbetter 2002).

[10] Although both sets were composed for keyboard instruments, the composers worked with different musical constraints relating to pitch availability, dynamic range, and instrument tuning; the compositional context for Chopin's set is still disputed (Temperley 2009; Frederick 1979).

[1.9] Applying empirical techniques traditionally reserved for big data questions to the corpora of individual composers can provide insight into how trends examined at a macro level relate to composers' expressive choices. This approach avoids issues associated with large-scale studies concerning classifying composers into eras and deciding which pieces (and how many) should be sampled. Consequently, focused studies of individual composers' musical sets can enable insights with fewer caveats for understanding their significance in history classrooms and performance spaces. We attempt to focus our descriptions on quantifiable musical phenomena, relying on statistical algorithms to assess optimal clustering, detect cue patterns, and measure cues' importance within and between composers. Among these is a novel statistical technique capable of clarifying cues' varied influence and importance within and between clusters (Apley and Zhu 2020). These efforts aim to paint a comprehensive picture of musical patterns in Bach and Chopin's sets that can be cast in the light of historical observations from qualitative and quantitative research alike.

[2.0] Method and Analysis

Cluster Analysis

[2.1] To examine how composers' pieces group according to similarities in cue structure, we performed hierarchical cluster analyses using the normalized attack rate, pitch height, mode, and root mean squared (RMS) amplitude values in each set. We codified attack rate, mode, and pitch height of eight-measure excerpts of the pieces (including pick-up measures) using a method previously outlined by Poon and Schutz (2017). We then prepared 16-bit waveform audio of each eight-measure excerpt from recordings by Vladimir Ashkenazy (Ashkenazy 1993) and Pietro De Maria (De Maria 2015). This enabled us to codify the average RMS amplitude value of each excerpt from mono-converted versions of the excerpts in Audacity.

[2.2] In hierarchical clustering, each observation is first assigned to its own unique cluster; then clusters are iteratively fused together until reaching a certain (prespecified) number. To measure dissimilarity between pieces, we created Euclidean distance matrices for each composer—subtracting the mean values from each observation and dividing by the standard deviation of each cue. We used Ward's linkage method to sequentially fuse similar clusters (see Ward 1961 for methodological details).

Optimal Clustering

[2.3] The optimal number of clusters to use in cluster analysis is not fixed and can be decided using a priori knowledge or diagnostic visualizations. Approaches are numerous, and two popular ones include elbow plots—which quantify how much additional variance each added cluster explains; and silhouette plots—which guide appropriate clustering by quantifying the distance of an observation in one cluster to those in other clusters. Although useful, these techniques rely on visual inspection of graphs which can introduce interpretive biases.

[2.4] Recent analytical tools enable assessing multiple quantitative metrics using agnostic statistical procedures. These methods not only offer additional guards against researcher biases, but also perform a wide array of analyses to assess optimal clustering. Using them, we evaluated 30 algorithms with solutions ranging between two and six clusters with the *NbClust* package in *R* (Charrad et al. 2014). Because different algorithms can yield differences in optimal clustering, a common strategy is to use a majority vote to determine the optimal number of clusters to use. For Bach, most algorithms indicated three clusters are optimal, whereas they indicated four to six clusters are optimal for Chopin. To provide the clearest comparisons between composers within these constraints, we settled on using three clusters for Bach, and four for Chopin.

Clustering Solution

[2.5] *Figure* 1 depicts dendrograms of cluster analyses of Bach (left) and Chopin's (right) prelude sets. An online version of this visualization in the companion app to this article offers interactivity with these data, including the option to hear the actual excerpts analyzed. Dendrograms are useful for visualizing clustering patterns, measuring dissimilarity between clusters along the *x* axis. Although they provide a bird's eye view of the similarities and differences between clusters, they do not reveal how pieces within a cluster are similar. Consequently, we prepared a series of additional visualizations to gain deeper insight into the expressive groupings within each prelude set.

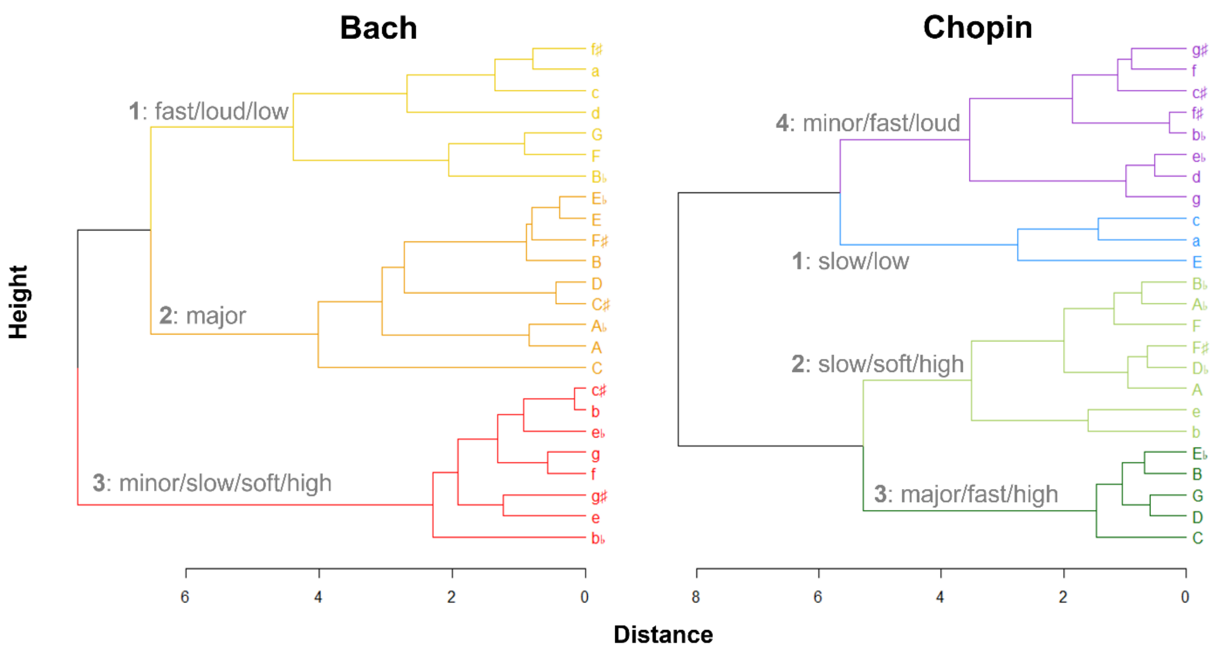


Figure 1. Dendrogram output from cluster analyses of Bach (left) and Chopin (right). Annotations indicate post-hoc summaries of individual cue clusters shown in *Figure* 2. Colors have been arbitrarily assigned for clarity.

Cue Patterns

Mode

[2.6] To clarify cues' relative influence on the observed clustering patterns, we prepared plots summarizing the distribution of pieces in each cluster. *Figure* 2 visualizes each cluster using the average attack rate, pitch height, and RMS amplitude of the pieces contained within, along with the proportion of major and minor pieces in each cluster. Examining mode reveals that Bach's first cluster is nearly evenly split between major and minor pieces (with slightly more minor than major pieces), whereas the second (major) and third (minor) clusters contain pieces of only one mode. For Chopin the first (66% minor) and second (75% major) clusters feature major and minor pieces, whereas the third (major) and fourth (minor) are characterized by only one mode.

Expressive (Prosodic) Cues

[2.7] Our analyses suggest attack rate plays a critical role in characterizing both composers' clusters. For Bach, clusters one to three exhibit fast, moderate, and slow attack rates, respectively; whereas Chopin's first two clusters exhibit slow attack rates and his latter two fast attack rates. The relationship between timing and loudness is rather clear for Bach: the fastest cluster has the highest RMS amplitude; the slowest has the lowest. For Chopin, timing's relation to loudness is less direct: clusters two (slow) and four (fast) are distinguished by soft and loud RMS amplitude, respectively; clusters one (slow) and three (fast) exhibit moderate amplitude values.

[2.8] Of the analyzed cues, pitch's role is least consistent in explaining clustering. For Bach, cluster one sits at an average pitch height of ~C4, whereas his second and third are about a major fourth above at ~F4. For Chopin, pitch height is more varied amongst clusters, reaching much lower pitches—likely reflecting (in part) the larger range of pitches available during his lifetime. His first cluster exhibits a low average pitch of E3; his fourth sits about a fifth above (~B3); and his second and third clusters are highest (~Eb4).

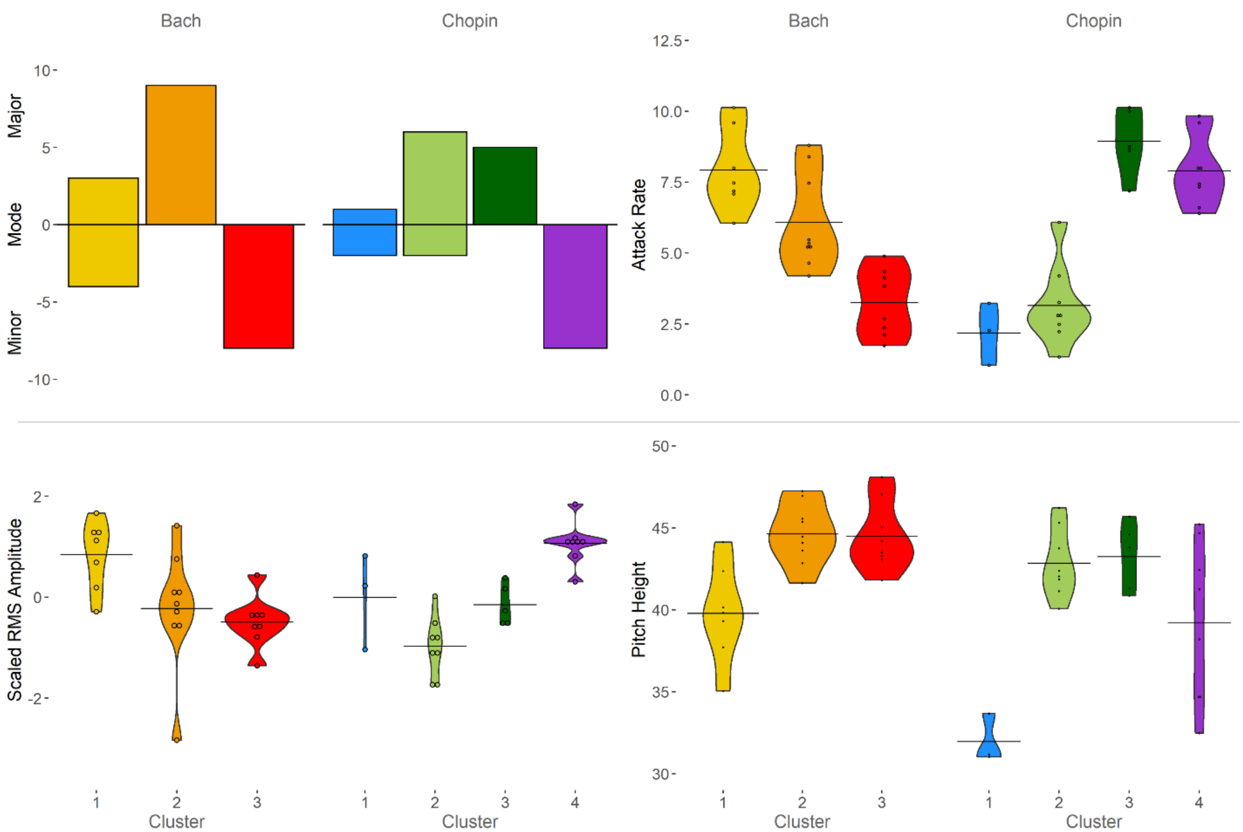


Figure 2. Bar plots indicating distribution of mode within each cluster (top left); violin plots with points indicating density and distribution of attack rate (top right), RMS amplitude (bottom left), and pitch height (bottom right). Points within violins denote individual observations. Clusters are demarcated with the same colors as in Fig 1. Width of violins scaled by number of observations per cluster.

Assessing Cue Importance

[2.9] Understanding which cues most strongly influence pieces' organization into clusters requires disentangling their collinear relationships and quantifying their relative contributions to the cluster analysis. To accomplish this, we employed Random Forests (RFs) to assess how specific cues influence cluster classification (for a technical overview of RFs, see Breiman, 2001). RFs apply random sampling and ensemble learning (i.e., training multiple decision tree algorithms) to improve generalizability and prediction accuracy when predicting the class of observations. To classify pieces into clusters, we trained the algorithm to predict the clusters of six pieces (i.e., the test set) using information about how the other 18 pieces cluster based on their cue structure (i.e., the training set). For both composers, a RF algorithm successfully classified all pieces into their corresponding cluster after controlling for random error.

[2.10] Next, we used accumulated local effects (ALE) to provide insight into how the RF algorithm used pieces' cue structure to inform classifications (Apley and Zhu 2020). ALEs estimate changes in cue effects while accounting for multicollinearity by estimating differences within a "window" of the cue values along a grid (see Molnar, 2019 for summary and implementation details). *Figure* 3 reveals how cues differ in describing clusters between composers. Although attack rate influences clustering most for both composers, RMS amplitude and mode differ between them—with mode more important for Bach, and RMS amplitude for Chopin. For both composers, pitch information influences clustering least.

[2.11] *Figure* 4 shows ALE variable importance within clusters, highlighting how cue importance varies depending on the cluster in question. As mode often receives the most attention in studies on musical expression, its varied influence in each cluster is of particular interest. For Bach, mode characterizes clustering most in cluster two, whereas attack rate distinguishes clusters one and three more distinctly. However, mode is least important to cluster one—here attack rate is most influential, followed closely by pitch and amplitude information. Similarly, although mode distinguishes Chopin's third and fourth clusters most, it distinguishes his first two clusters least, characterized by attack rate and pitch, and attack rate and RMS amplitude, respectively.

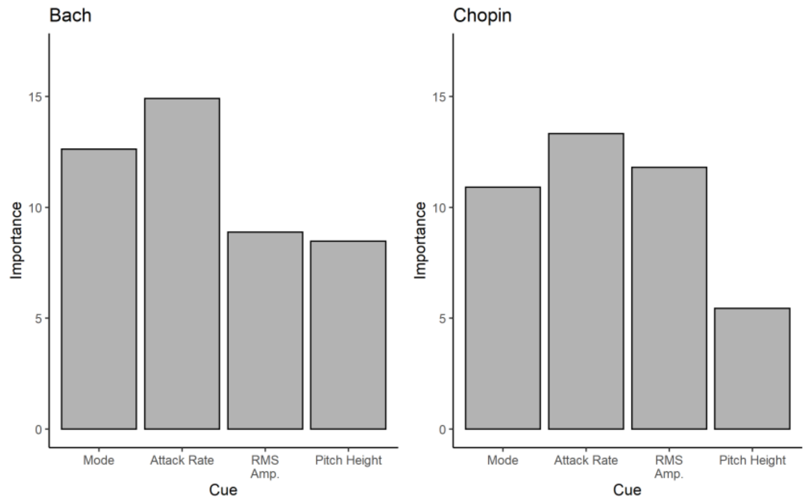


Figure 3. ALE variable importance quantifying the standard deviation of each cue's (mode, attack rate, RMS amplitude, pitch height) influence on pieces' clustering.

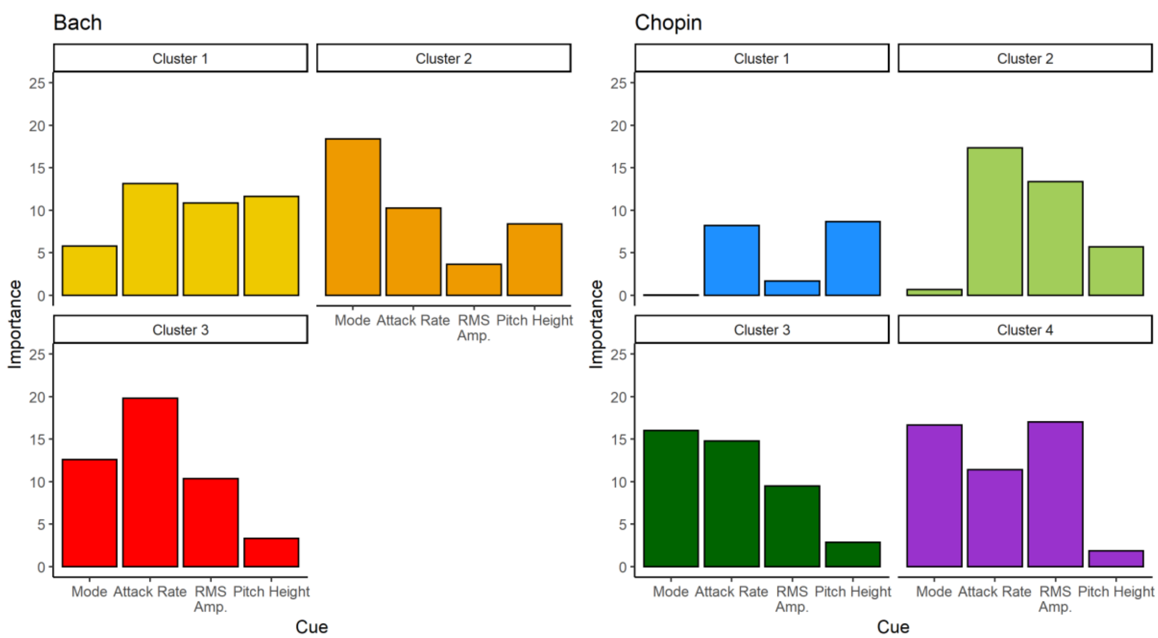


Figure 4. ALE variable importance quantifying the standard deviation of each cue's (mode, attack rate, RMS amplitude, pitch height) influence on individual clusters.

[2.12] *Figure* 5 illustrates how changing values of mode, attack rate, RMS amplitude, and pitch height influence the probability of clustering classification—pinpointing how specific cue values lead to differences in clustering probability for Bach and Chopin. The two middle panels show circle plots comparing the magnitude of the optimal values for clustering, with upward and downward arrows indicating whether the optimal value is higher or lower than the middle value for that cue across the full dataset. To highlight how each cue plays a distinct role in defining the musically expressive profile of each cluster, we briefly summarize their contributions below.

[2.13] *Mode.* For Bach, the minor mode most strongly distinguishes pieces in clusters one and three. However, the lack of major pieces in cluster three leads to a stronger influence on clustering than cluster one; conversely, major pieces most accurately characterize cluster two, reaching the highest clustering probability at ~55%. For Chopin, the major and minor modes strongly distinguish clusters three and four, respectively. They also distinguish clusters one (minor) and two (major) but to a much lesser extent due to a combination of major and minor pieces in both clusters. Mode appears most influential for clustering pieces in Bach's second cluster; of Chopin's clusters, mode is most influential for clusters three and four.

[2.14] *Attack Rate.* Fast attack rates distinguish Bach's first cluster, peaking at eight attacks per second. Moderate attack rates near five attacks per second (APS) distinguish pieces in his second cluster. Finally, slower attack rates near four APS best distinguish pieces in his third cluster. For Chopin, slow attack rates near 2.5 and 2 APS best classify clusters one and two, whereas fast attack rates near 10 and 7.5 APS characterize clusters three and four, respectively. Comparing clusters within each composer's analysis indicates attack rate influences both composers' third cluster most strongly.

[2.15] *RMS Amplitude.* For Bach, louder RMS amplitudes distinguish cluster one, peaking at -26 dB, whereas clusters two and three are progressively softer at ~ -31 dB and ~ -33 dB, respectively. For Chopin, high RMS amplitude values strongly distinguish clusters one and four, reaching ~ -24 dB and ~ -25 dB, respectively. Clusters two and three exhibit softer RMS amplitudes at ~ -35 dB and ~ -29 dB, respectively. RMS amplitude appears most influential on Bach's third cluster and on Chopin's second and fourth clusters.

[2.16] Bach's first cluster peaks at lower pitches (near C4) than his other clusters. In contrast, clusters two and three peak a tritone and perfect fifth above at ~ F#4 and ~ G4, respectively (however, pitch is less influential on clustering probability for both). For Chopin, low pitches strongly demarcate clusters one and four, peaking lower than Bach at ~ G3 and Ab3, respectively, whereas clusters two and three peak at higher pitches (~D4 and F4, respectively). Pitch height most strongly influences both composers' first cluster.

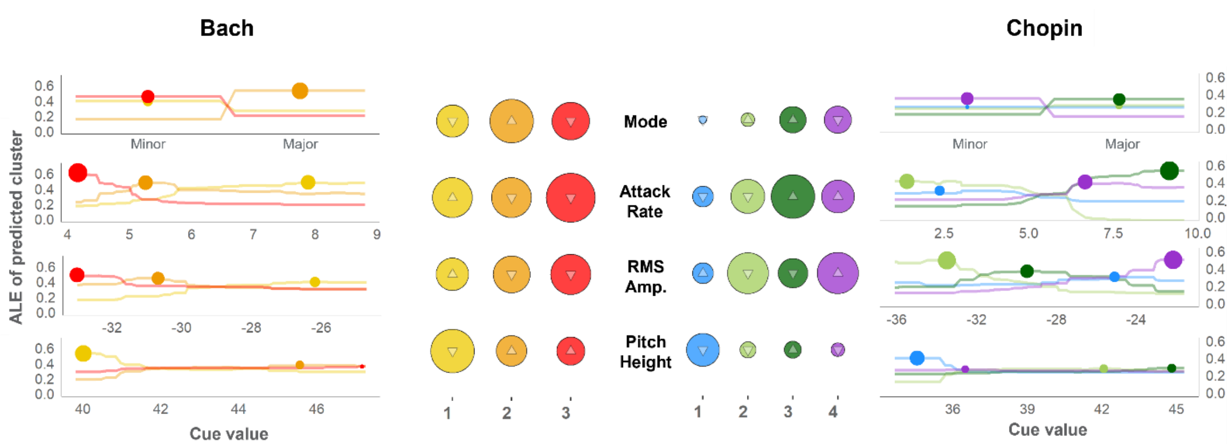


Figure 5. Accumulated local effects (ALE) indicating values' influence on clustering probability for each cluster (indicated by color) for each composer (Bach, left; Chopin, right). Circles indicate values where the probability is highest for pieces belonging to a cluster, with the size of the circle corresponding to the magnitude of the probability. Circle plots summarize the magnitude of these effects, with upward and downward arrows showing whether these values are "high" or "low" based on a median split of the dataset. Cues (top to bottom): mode, attack rate, RMS amplitude, pitch height.

Cluster Analysis Explorer

[2.17] Presenting the cluster analysis findings in an interactive format offers greater interpretability and exploratory value. *Figure* 6 depicts a 3-dimensional scatter plot akin to those found in the interactive tool. These visualizations provide an interface for deeper insights by enabling musicians and researchers alike to gain a sense of the complex cue relationships underpinning clustering decisions. They also serve as an interface for comparing the algorithm's clustering decisions with findings from empirical and historical research exploring changes over music history. Music theorists and performers alike can gain insight into which cues drive expression most strongly for pieces within each composer's set.

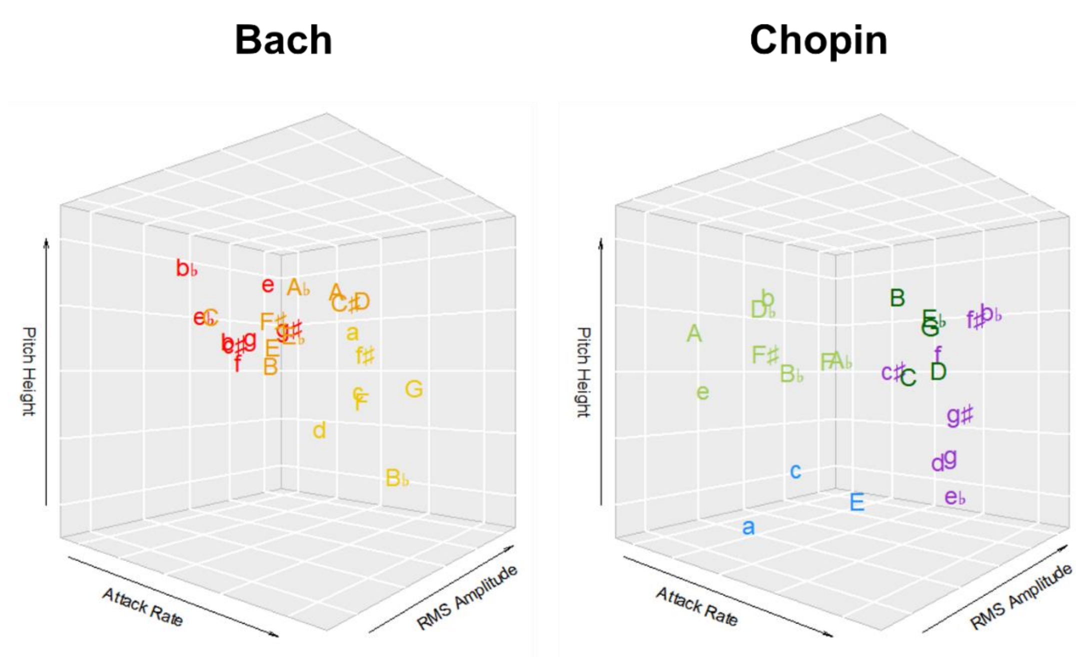


Figure 6. 3-dimensional scatter plot of Bach (left) and Chopin's (right) preludes, color-coded according to the cluster analysis output. X, Y, and Z (inward to page) axes correspond to attack rate, pitch height, and RMS amplitude. Cue values scaled using both composers' sets.

[3.0] Discussion

[3.1] Using empirical techniques to analyze cue patterns provides a deeper understanding of compositional choices. Our algorithm-aided approach offers useful new insight into sets of pieces widely studied using qualitative and quantitative methods. This enables bridging epistemological divides between different fields of music study by consolidating research on music's historical changes. By investigating clusters empirically, we can pinpoint cues' influence for demarcating expressive differences. Using quantitative techniques to assess cues' importance within clusters sheds new light on expressive meaning in Bach and Chopin's sets.

[3.2] Our analyses reveal three notable insights about the expressive differences between Bach and Chopin's prelude sets. First, whereas three clusters best demarcate Bach's set, four best demarcate Chopin's. This expansion in expressive groupings may indicate Chopin employed a greater range of varied cue combinations than Bach, possibly reflecting the greater expressive nuance possible on instruments developed during the 1800s (Burkholder, Grout, and Palisca 2019; Frederick 1979).

[3.3] Second, our findings support and contextualize accounts of mode's changing relationships to timing and loudness by relating insights from large-scale studies to corpora composed by individuals. For example, selectively comparing clusters containing only major pieces with those containing only minor pieces reveals parallels to past research on mode's changing relationships to loudness and timing in the Romantic era. Despite certain interpretive challenges posed by a mixture of compositional and performance cues in our analyses,[11] these

[11] The lack of score-defined dynamic information from both composers necessitated using values from performances recorded since the 1990s. Consequently, the degree to which performance cues reflect composers' intentions can confound patterns the clustering algorithm detects.

contrasting patterns suggest differences in the composers' cue use partly reflect those observed between musical eras.

[3.4] Finally, to enable a better understanding of cues' expressive roles, we can evaluate their varied importance between composers as well as clusters. Rank ordering cue importance between composers indicates that mode demarcates clusters less prominently in Chopin's set than Bach's. Additionally, patterns in cue importance vary markedly between clusters. These differences suggest cues' importance to expressive meaning depends on musical context and composers' expressive intentions.

Cue Characteristics Within Clusters

[3.5] Historical music research describes broad shifts in mode's relationship to prosodic cues during the Romantic era, where highly contrastive cue relationships replace the fast/loud/major and slow/soft/minor compositional patterns of epochs past (Horn and Huron 2015). Analyzing corpora by individual composers enables examining macroscopic phenomena at a microscopic level, contextualizing musical differences observed between eras. Do Bach and Chopin's prelude sets reflect broader expressive differences between Baroque and Romantic composers? If so, to what extent do they hold to other composers of those eras? These important questions can only be explored in the context of music that has stood the test of time.

[3.6] Investigating how timing, dynamics, and pitch height relate to mode within each cluster provides opportunity to compare Bach and Chopin's cue use to findings from larger corpora. In Bach's (third) cluster comprising only minor pieces, the musical excerpts are lower and slower than those in his (second) fully major cluster. These contrastive patterns in loudness and timing support conventional descriptions of minor pieces being slower and softer than their major counterparts during Bach's time. In contrast, Chopin's fully minor (fourth) cluster is similarly fast, but louder and lower than his fully major (third) cluster, providing some support (albeit weak) for the rising prominence of fast/loud/minor and slow/soft/major categories arising in the Romantic era (Horn and Huron 2015; Ladinig and Huron 2010).

[3.7] Mode's shifting relationship to timing in the Romantic era evinces in the cue patterns of Chopin's second cluster. Its (predominately major) pieces are slower and softer than his fourth cluster's minor pieces. However, its concurrence with the fast/loud/major pieces of cluster three suggests mode's relationship to prosodic cues broadened rather than shifted. Could access to a wider expressive palette during the Romantic era enable Chopin to move beyond mode's traditionally dichotomous expressive role? To gain a better sense of how nuanced cue relationships influence clustering, we trained an algorithm to classify pieces into their respective clusters, enabling assessment of cues' importance for clustering.

Exploring the Importance of Cues Within and Between Sets

[3.8] Mode often receives more attention than low level cues in explorations of expressive differences—likely due to its importance in the compositional organization of a piece. However, our exploration of cues' relative importance suggests mode does not always strongly demarcate clusters relative to other analyzed cues. In our analyses, attack rate provided the best explanation of pieces' separation into clusters. For Chopin's set, RMS amplitude also ranked higher in importance than mode.[12] Consequently, although mode figures prominently in composers' organizational decisions, timing information appears to distinguish their organization into clusters more strongly. The greater importance of timing over mode in our analysis supports perceptual accounts ordering their relative contributions to music's perceived emotional meaning

[12] Despite the importance of RMS amplitude for Chopin's pieces, the meaning of the compositional importance of dynamics cannot be assessed in the present study due to the use of dynamics from recorded performances.

(Juslin and Lindström 2010; but see Eerola, Friberg, and Bresin 2013). Comparing our findings to earlier investigations of expressive differences across eras, mode's lower rank-order importance in Chopin's set may reflect evidence for its increasing complexity in the Romantic era (Harasim et al. 2021).

[3.9] Exploring the significance of specific cues in demarcating clusters provides deeper insight into preludes' musical structure. The inconsistency of mode's importance between clusters can clarify how composers emphasize specific cues to create expressive contrast. Whereas mode most strongly distinguishes pieces in Bach's second cluster, timing beats mode in cluster three, where low attack rates best explaining clustering. Curiously, mode is least important in cluster one, where timing and pitch play a comparably strong role (followed closely by RMS amplitude). For Chopin, mode is most important to clusters three and four, but least important to clusters one and two. Consequently, its influence appears to vary between composers and clusters.

Pinpointing Cues' Clustering Influence Through Accumulated Local Effects

[3.10] Assessing which cue values influence how pieces cluster together enables more precise descriptions of their musical characteristics. *Figure* 5 summarizes the value most strongly influencing clustering for each cue. Bach's first cluster is strongly distinguished by fast timing, average pitch close to middle C, and high amplitude values. It is only weakly distinguished by the minor mode. Timing is also influential for Bach's third cluster, with slow attack rates (near 4 attacks per second [APS]) and soft RMS amplitude (~-32dB) predicting clustering. In contrast, mode plays a greater role for his second cluster, where major pieces with attack rates near five APS prominently affect clustering.

[3.11] Analyzing Chopin's clusters reveals RMS amplitude and timing distinguish clusters two and four—however, in contrasting ways: whereas soft RMS amplitude and slow attack rates distinguish cluster two, loud RMS amplitude and fast attack rates distinguish cluster four. Unlike his second and fourth clusters, Chopin's first and third clusters are less similar in terms of how cues influence classification. Slow timing and low pitches mostly strongly distinguish cluster one, and fast timing and major mode distinguish cluster three.

[3.12] The varied sizes of circles in *Figure* 5 highlight how much each cue's most predictive value characterizes each cluster. The attack rate value most influential in Bach's third cluster (4 APS) influences clustering accuracy more strongly than the most influential value for his first cluster (8 APS). Consequently, although his third cluster is "slow" and his first is "fast", timing as a cue appears to be more influential for cluster three overall. Examining the different sizes of circles between composers reveals cues' predictive roles are more uniform for Bach than for Chopin, indicating individual cues more often take the lead in distinguishing how Chopin's pieces cluster together. For example, pitch height is much more influential than any other cue in Chopin's first cluster, whereas RMS amplitude and attack rate are most predictive in distinguishing his second cluster's pieces. This greater variability in how prosodic cues distinguish clusters for Chopin may also reflect a greater ability to use compositional cues to emulate speech-like expression—a possibility propelled by advancements in instrument design (Dolge 1972). Although these analyses are exploratory, contextualizing large-scale observations within corpora by specific composers introduces novel questions to explore through quantitative and qualitative inquiry.

General Discussion

Where Qualitative Insights Clarify Quantitative Findings

[3.13] Our analysis focuses on prelude sets by two widely studied composers. The prevalence of these sets in empirical and qualitative work enables opportunity to consolidate knowledge gained from distinct epistemologies. Whereas qualitative studies inform these sets' musical context and performance conventions (Beuerman 2003; Kirkpatrick 1984), quantitative approaches explore technical and pragmatic issues such as quantifying tonality (Purwins et al. 2004; Schmuckler and Tomovski 2005) and performance expression (Bisesi, Macritchie, and Parncutt 2012; Dodson 2011; Rector 2021). Consolidating these perspectives can create a robust epistemological framework accounting for biographic and historical musical considerations, enabling further musical insights through data visualizations and exploratory tools.

[3.14] Engaging with both perspectives clarifies the historic importance of these sets. We can relate our findings to empirical studies examining Bach and Chopin's prelude sets. For example, mode's greater rank-order importance for Bach's clusters compared to Chopin's is consistent with studies using these sets in key-finding algorithms (Schmuckler and Tomovski 2005). Additionally, our findings can complement and extend more traditional musicological analyses focused on individual compositions. For example, a high degree of chromaticism demarcates Chopin's A-minor prelude (Kramer 1985). Interestingly, this piece belongs to his first cluster, where attack rate and pitch contribute most to clustering (*Figure* 5). Mode's minimal influence in this cluster may suggest Chopin compensated for low modal clarity by emphasizing other cues. Similarly, we can explore the link between compositional features and performance expression. Performers often use dynamic contrasts to accent the musical structure of Bach's C-Minor Prelude (Martens 2012). The prevalence of loud RMS amplitude in the prelude's corresponding cluster (cluster one) suggests loud dynamics are important to the cluster's organizational structure. These comparisons (though merely exploratory and by no means exhaustive) provide a useful starting point for further investigation into the link between historical, compositional, and performance-related features.

[3.15] Beyond contextualizing findings using evidence from past studies, listening to clusters' distinctive auditory characteristics can contextualize what these differences mean for perception by clarifying how they sound. For this purpose, we developed an exploratory, interactive application to summarize the analysis findings. Past small-scale and large-scale approaches have implemented similar demonstrative tools to clarify perceived emotion in Beethoven's Sonata *Pathetique* (Albrecht 2018) and emotive pieces selected by American and Chinese participant groups (Cowen et al. 2020) to clarify complex analysis findings. Readers can explore the auditory characteristics of the clustered pieces in interactive versions of *Figures* 1 and 6. Clicking on coordinates triggers audio for each excerpt; individual clusters can also be shown and hidden by clicking the corresponding checkboxes.

Limitations

[3.16] This novel analytical process exploring differences in musical corpora holds benefits for consolidating knowledge from qualitative and quantitative examinations of musical expression. Exploring how specific cue values influence pieces' clustering helps to overcome subjectivity in interpreting cluster analysis results by offering a deeper look of cues' roles within clusters. Applying this technique to the corpora of individual composers enables drawing on the unique strengths of large- and small-scale studies by offering the interpretive breadth characteristic of data-driven analyses and the analytical depth of score-based approaches.

[3.17] To inform future research exploring expressive differences within the musical works of individual composers, we highlight three limitations to the present study. First, aside from the obvious limitation to generalizability by focusing on only two composers, relying on only two recorded performances for assessing dynamics can confound compositional and performance features in the cluster analysis. Part of this problem stems from the unavailability of notated dynamic information in Bach's prelude set. Although our approach adequately disentangles these factors in our quantification of variable importance, including performance cues may have unintended effects on the clustering algorithms' decisions. Consequently, in attempts to mitigate potential confounds, we selected performances using objective evaluation of their prevalence in widely used classical music databases (Kelly, Anderson, and Schutz, in press).

[3.18] Second, average values of pitch, timing, and RMS amplitude overlook nuanced smeasure-to-measure changes within a composition. Investigations of interpretative differences suggest this nuance can provide a better understanding of historic differences in musical expression (Rector 2021). Finally, our analysis establishes variable importance using a novel process combining algorithms to cluster data, assess accuracy, and clarify effects while decomposing cues' intercorrelated structure. Although some theoretical aspects of this approach are not completely unprecedented,[13] the extent of its applicability should be further explored by music researchers and data scientists alike. Nonetheless, this process shows promise for elucidating the patterns emerging from compositional and performance cues.

[13] See Badih et al. (2019) for a similar technique implementing decision trees using cluster analysis.

Conclusion

[3.19] With the growing availability and accessibility of complex analytical toolkits, quantitative analyses on extensive musical corpora will shape the future for understanding music's changing expressive properties. This exploratory research connects perspectives from historical and empirical findings by making use of a data-driven approach comprising novel statistical techniques. Developing techniques to enable a consistent and comprehensive framework for understanding changes throughout music history is increasingly important as researchers discover incompatibilities in the theories of knowledge driving qualitative and quantitative research (Kang and Evans 2020).

[3.20] In addition to clarifying connections between these approaches, focusing on composer-crafted corpora suggests broad generalizations about musical style sometimes overlook the subtle nuances which make music so appealing to listeners. For an accurate understanding of composers' distinctive compositional approaches, future research might compare large- and small-scale musical phenomena to analyses of well-organized musical collections.[14] This holds potential to elucidate composers' complex musical choices. Applying these techniques to multiple performances of historic renditions also enables exploring questions of how musical training and contemporary performance influences may affect cue weights in music from specific eras.

[3.21] In a recent collection of interviews with 65 composers, the late music critic Bálint András Varga reported several composers described deliberately avoiding self-repetition, using music's expressive elements to embark on new creative avenues (Varga, 2011). Consequently,

[14] The development and exploration of the Annotated Beethoven Corpus reveal the power of these approaches (see Moss et al. 2019; Neuwirth et al. 2018).

as music research continues to parse the complex musical vocabulary of epochs past using theoretical and empirical methods, composers' complex and highly idiosyncratic musical choices should receive special attention in musical analyses. This in turn will enable a better understanding of music's evolution and interest throughout history.

Works Cited

Adler, Guido. 1885. *Umfang, Methode Und Ziel Der Musikwissenschaft*. Leipzig, Germany: Breitkopf und Härtel.

Albrecht, Joshua D. 2018. “Expressive Meaning and the Empirical Analysis of Musical Gesture: The Progressive Exposure Method and the Second Movement of Beethoven’s Pathetique Sonata.” *Music Theory Online* 24 (3). https://mtosmt.org/ojs/index.php/mto/article/view/45.

Albrecht, Joshua D., and David Huron. 2014. “A Statistical Approach to Tracing the Historical Development of Major and Minor Pitch Distributions, 1400-1750.” *Music Perception*. University of California Press. https://doi.org/10.1525/MP.2014.31.3.223.

Albrecht, Joshua D., and Daniel Shanahan. 2013. “The Use of Large Corpora to Train a New Type of Key-Finding Algorithm: An Improved Treatment of the Minor Mode.” *Music Perception* 31 (1): 59–67. https://doi.org/10.1525/MP.2013.31.1.59.

Aldwell, Edward, Carl Schachter, and Allen Cadwallader. 2002. *Harmony & Voice Leading*. 3rd ed. Boston, MA: Schirmer.

Anderson, Cameron J., and Michael Schutz. In press. “Exploring Historic Changes in Musical Communication: Deconstructing Emotional Cues in Preludes by Bach and Chopin.” *Psychology of Music*.

Apley, Daniel W., and Jingyu Zhu. 2020. “Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models.” *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 82 (4): 1059–86. https://doi.org/10.1111/rssb.12377.

Bach, Johann Sebastian. 1883. “The Well-Tempered Clavier Book I Ed. Hans Bischoff.” Berlin: Kalmus & Co. (Music Publisher).

Badih, Ghattas, Michel Pierre, and Boyer Laurent. 2019. “Assessing Variable Importance in

Clustering: A New Method Based on Unsupervised Binary Decision Trees.” *Computational Statistics* 34 (1). https://doi.org/10.1007/s00180-018-0857-0ï.

Battcock, Aimee, and Michael Schutz. 2019. “Acoustically Expressing Affect.” *Music Perception* 37 (1): 66–91.

———. 2021. “Emotion and Expertise : How Listeners with Formal Music Training Use Cues to Perceive Emotion.” *Psychological Research*.

Beuerman, Eric Gilbert. 2003. “The Evolution of the Twenty-Four Prelude Set for Piano.” University of Arizona.

Bisesi, Erica, Jennifer Macritchie, and Richard Parncutt. 2012. “Recorded Interpretations of Chopin Preludes: Performer’s Choice of Score Events for Emphasis and Emotional Communication.”

Blashfield, Roger K., and Mark S. Aldenderfer. 1988. “The Methods and Problems of Cluster Analysis.” In *Handbook of Multivariate Experimental Psychology*, 447–73. Springer US. https://doi.org/10.1007/978-1-4613-0893-5_14.

Breiman, Leo. 2001. “Random Forests.” *Machine Learning* 45 (1): 5–32. https://doi.org/10.1023/A:1010933404324.

Broze, Yuri, and Daniel Shanahan. 2013. “Diachronic Changes in Jazz Harmony: A Cognitive Perspective.” *Music Perception: An Interdisciplinary Journal* 3 (1): 32–45.

Burkholder, J. P., D. J. Grout, and C. V. Palisca. 2019. *A History of Western Music: Tenth International Student Edition*.

Caleon, Imelda, and Subramaniam Ramanathan. 2008. “From Music to Physics: The Undervalued Legacy of Pythagoras.” *Science and Education* 17 (4): 449–56. https://doi.org/10.1007/s11191-007-9090-x.

Charrad, Malika, Nadia Ghazzali, Véronique Boiteau, and Azam Niknafs. 2014. "NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set." *Journal of Statistical Software* 61 (1): 1–36. https://www.jstatsoft.org/index.php/jss/article/view/v061i06.

Chopin, Frederic. 1839. *Preludes Op. 28*. Edited by Raoul Pugno. Berlin: Schlesinger'sche Buch-und Musikhandlung.

Condit-Schultz, Nathaniel. 2019. "Deconstructing the NPVI: A Methodological Critique of the Normalized Pairwise Variability Index as Applied to Music." *Music Perception* 36 (3): 300–313. https://doi.org/10.1525/MP.2019.36.3.300.

Cowen, Alan S, Xia Fang, Disa Sauter, and Dacher J. Keltner. 2020. "What Music Makes Us Feel: At Least 13 Dimensions Organize Subjective Experiences Associated with Music across Different Cultures" 117 (4). https://doi.org/10.1073/pnas.1910704117/-/DCSupplemental.

Cruder, Cinzia, Marco Barbero, Emiliano Soldini, and Nigel Gleeson. 2021. "Patterns of Pain Location in Music Students: A Cluster Analysis." *BMC Musculoskeletal Disorders* 22 (1): 1–13. https://doi.org/10.1186/s12891-021-04046-6.

Daniele, Joseph R., and Aniruddh D. Patel. 2013. "An Empirical Study of Historical Patterns in Musical Rhythm: Analysis of German & Italian Classical Music Using the NPVI Equation." *Music Perception: An Interdisciplinary Journal* 31 (1): 10–18.

Dodson, Alan. 2011. "Expressive Timing in Expanded Phrases: An Empirical Study of Recordings of Three Chopin Preludes." *Royal Northern College of Music* 4: 2–29.

Dolge, A. 1972. *Pianos and Their Makers: A Comprehensive History of the Development of the Piano from the Monochord to the Concert Grand Player Piano*.

Duckles, Vincent, Jann Pasler, Glenn Stanley, Thomas Christensen, Barbara H. Haggh, Robert Balchin, Laurence Libin, et al. 2001. “Musicology.” In *Oxford Music Online*. Oxford University Press. https://doi.org/10.1093/gmo/9781561592630.article.46710.

Eerola, Tuomas, Anders Friberg, and Roberto Bresin. 2013. “Emotional Expression in Music: Contribution, Linearity, and Additivity of Primary Musical Cues.” *Frontiers in Psychology* 4 (July): 1–12. https://doi.org/10.3389/fpsyg.2013.00487.

Frederick, EM. 1979. “The" Romantic" Sound in Four Pianos of Chopin’s Era.” https://www.jstor.org/stable/746286?casa_token=gH2niX5ujzIAAAAA:wkXdwVWOObSQYAck98zGor6XmyRfUnVGiQBr5DqzqNORgyPrxztJoC3JG4zx2ceYswwYjTZHzfZooFRXVx3AHFk0zRlPXSlK90cFIV7R4HRrGE4p0YeX.

Ge, Ren Shan, Qiang Dong, Chantai M. Sottas, Haolin Chen, Barry R. Zirkin, and Matthew P. Hardy. 2005. “Gene Expression in Rat Leydig Cells during Development from the Progenitor to Adult Stage: A Cluster Analysis.” *Biology of Reproduction* 72 (6): 1405–15. https://doi.org/10.1095/biolreprod.104.037499.

Hansen, Niels Chr, Makiko Sadakata, and Marcus T. Pearce. 2016. “Nonlinear Changes in the Rhythm of European Art Music: Quantitative Support for Historical Musicology.” *Music Perception*. University of California Press. https://doi.org/10.1525/MP.2016.33.4.414.

Harasim, Daniel, Fabian C. Moss, Matthias Ramirez, and Martin Rohrmeier. 2021. “Exploring the Foundations of Tonality: Statistical Cognitive Modeling of Modes in the History of Western Classical Music.” *Humanities and Social Sciences Communications* 8 (1): 1–11. https://doi.org/10.1057/s41599-020-00678-6.

Hatten, Robert S. 2004. *Musical Meaning in Beethoven: Markedness, Correlation, and Interpretation*. Bloomington, IN: Indiana University Press.

Horn, Katelyn, and David Huron. 2015. "On the Changing Use of the Major and Minor Modes 1750–1900." *Music Theory Online* 21 (1): 1–11. http://www.mtosmt.org/issues/mto.15.21.1/mto.15.21.1.horn_huron.html.

Howard, James H., and Eugene B. Silverman. 1976. "A Multidimensional Scaling Analysis of 16 Complex Sounds." *Perception & Psychophysics* 19 (2): 193–200. https://doi.org/10.3758/BF03204226.

Juslin, Patrik N., and Erik Lindström. 2010. "Musical Expression of Emotions: Modelling Listeners' Judgements of Composed and Performed Features." In *Music Analysis*, 29:334–64. https://doi.org/10.1111/j.1468-2249.2011.00323.x.

Kabók, József, Slobodan Radišić, and Bogdan Kuzmanović. 2017. "Cluster Analysis of Higher-Education Competitiveness in Selected European Countries." *Economic Research-Ekonomska Istraživanja* 30 (1): 845–57. https://doi.org/10.1080/1331677X.2017.1305783.

Kang, Donghyun, and James Evans. 2020. "Against Method: Exploding the Boundary between Qualitative and Quantitative Studies of Science." *Quantitative Science Studies* 1 (3): 930–44. https://doi.org/10.1162/qss_a_00056.

Kelly, Benjamin O., Cameron J. Anderson, and Michael Schutz. Under review. "Perceptual Categorization of Music between Eras."

Kirkpatrick, R. 1984. *Interpreting Bach's Well-Tempered Clavier: A Performer's Discourse of Method*.

Kramer, Lawrence. 1985. "Romantic Meaning in Chopin's Prelude in A Minor." *19th-Century Music*. Vol. 9. Autumn.

Ladinig, Olivia, and David Huron. 2010. "Dynamic Levels in Classical and Romantic Keyboard Music: Effect of Musical Mode." *Empirical Musicology Review* 5 (2): 51–56.

Ledbetter, D. 2002. *Bach's Well-Tempered Clavier: The 48 Preludes and Fugues*.

Martens, Daniel Barolsky and Peter. 2012. "Rendering the Prosaic Persuasive: Gould and the Performance of Bach's C-Minor Prelude (WTC I)." *Music Theory Online* 18 (1).

Molnar, Christoph. 2019. "Interpretable Machine Learning." 2019. https://christophm.github.io/interpretable-ml-book/.

Moss, Fabian C., Markus Neuwirth, Daniel Harasim, and Martin Rohrmeier. 2019. "Statistical Characteristics of Tonal Harmony: A Corpus Study of Beethoven's String Quartets." *PLoS ONE* 14 (6). https://doi.org/10.1371/journal.pone.0217242.

Neuwirth, Markus, Daniel Harasim, Fabian C. Moss, and Martin Rohrmeier. 2018. "The Annotated Beethoven Corpus (ABC): A Dataset of Harmonic Analyses of All Beethoven String Quartets." *Frontiers in Digital Humanities* 5 (July): 16. https://doi.org/10.3389/fdigh.2018.00016.

Patel, Aniruddh D., and Joseph R. Daniele. 2003. "An Empirical Comparison of Rhythm in Language and Music." *Cognition* 87 (1): B35–B45.

Poon, Matthew, and Michael Schutz. 2015. "Cueing Musical Emotions: An Empirical Analysis of 24-Piece Sets by Bach and Chopin Documents Parallels with Emotional Speech." *Frontiers in Psychology* 6 (November): 1–13. https://doi.org/10.3389/fpsyg.2015.01419.

Post, Olaf, and David Huron. 2009. "Western Classical Music in the Minor Mode Is Slower (Except in the Romantic Period)." *Empirical Musicology Review* 4 (1): 2–10.

Purwins, Hendrik, Benjamin Blankertz, Guido Dornhege, and Klaus Obermayer. 2004. "Scale Degree Profiles from Audio Investigated with Machine Learning." *Audio Engineering Society Convention 116.* www.aes.org.

Rector, Michael. 2021. “Historical Trends in Expressive Timing Strategies: Chopin’s Etude, Op. 25 No. 1.” https://doi.org/10.18061/emr.v15i3-4.7338.

Rehding, Alexander. 2016. “Instruments of Music Theory.” *Society for Music Theory* 22 (4): 22.

Savage, Patrick E., and Steven Brown. 2014. “Mapping Music: Cluster Analysis of Song-Type Frequencies within and between Cultures.” *Ethnomusicology* 58 (1): 133–55. https://doi.org/10.5406/ethnomusicology.58.1.0133.

Schmuckler, MA, and R. Tomovski. 2005. “Perceptual Tests of an Algorithm for Musical Key-Finding.” *Journal of Experimental Psychology: Human Perception and Performance* 31 (5): 1124. https://doi.org/10.1037/0096-1523.31.5.1124.

Schutz, Michael. 2017. “Acoustic Constraints and Musical Consequences: Exploring Composers’ Use of Cues for Musical Emotion.” *Frontiers in Psychology* 8: 1–10. https://doi.org/10.3389/fpsyg.2017.01402.

Sharma, Asuda, Hesham Ali, and Dario Ghersi. 2018. “Cluster Analysis of Biological Networks.” In *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, 1–3:1036–46. Elsevier. https://doi.org/10.1016/B978-0-12-809633-8.20494-9.

Shensa, Ariel, Jaime E. Sidani, Mary Amanda Dew, César G. Escobar-Viera, and Brian A. Primack. 2018. “Social Media Use and Depression and Anxiety Symptoms: A Cluster Analysis.” *American Journal of Health Behavior* 42 (2): 116–28. https://doi.org/10.5993/AJHB.42.2.11.

Sneath, Peter HA, and Robert R. Sokal. 1973. “Numerical Taxonomy. The Principles and Practice of Numerical Classification.,” June. https://doi.org/10.2307/1217562.

Sokal, Robert R. 1958. "A Statistical Method for Evaluating Systematic Relationships." *Univ. Kansas, Sci. Bull.* 38: 1409–38. https://doi.org/10.24517/00055549.

Stenlund, Tova, Per Erik Lyrén, and Hanna Eklöf. 2018. "The Successful Test Taker: Exploring Test-Taking Behavior Profiles through Cluster Analysis." *European Journal of Psychology of Education* 33 (2): 403–17. https://doi.org/10.1007/s10212-017-0332-2.

Stylidis, Dimitrios. 2018. "Journal of Sustainable Tourism Residents' Place Image: A Cluster Analysis and Its Links to Place Attachment and Support for Tourism." *Taylor & Francis* 26 (6): 1007–26. https://doi.org/10.1080/09669582.2018.1435668.

Temperley, Nicholas. 2009. "Preluding at the Piano." In *Musical Improvisation: Art, Education, and Society*, 323–41.

Varga, B. A. 2011. *Three Questions for Sixty-Five Composers*. United Kingdom: University of Rochester Press.

Ward, Joe Jr H. 1961. "Hierarchical Grouping to Maximize Payoff."

Weiß, Christof, Matthias Mauch, Simon Dixon, and Meinard Müller. 2019. "Investigating Style Evolution of Western Classical Music: A Computational Approach." *Musicae Scientiae* 23 (4): 486–507. https://doi.org/10.1177/1029864918757595.

White, Christopher W. M. 2014. "Changing Styles, Changing Corpora, Changing Tonal Models." *Music Perception* 31 (3): 244–53. https://doi.org/10.1525/MP.2014.31.3.244.

Zhukov, Katie. 2007. "Student Learning Styles in Advanced Instrumental Music Lessons." *Music Education Research* 9 (1): 111–27. https://doi.org/10.1080/14613800601127585.