Jahrestagung der Gesellschaft für Musikforschung 2019

Perspectives of Musical Corpus Studies: The Annotated Mozart Sonatas

Markus Neuwirth, Lausanne/Linz,
Johannes Hentschel, Martin Rohrmeier, Lausanne

DOI: 10.25366/2020.96

Zitation: Markus Neuwirth, Johannes Hentschel, Martin Rohrmeier, „Perspectives of Musical Corpus Studies: The Annotated Mozart Sonatas“, in: *Brückenschläge zwischen Musikwissenschaft und Informatik. Theoretische und praktische Aspekte der Kooperation*, in Verbindung mit der Fachgruppe Digitale Musikwissenschaft hrsg. von Stefanie Acquavella-Rauch, Andreas Münzmay und Joachim Veit (= Musikwissenschaft: Aktuelle Perspektiven. Bericht über die Jahrestagung der Gesellschaft für Musikforschung 2019 in Paderborn und Detmold, Bd. 3), Detmold, Musikwissenschaftliches Seminar der Universität Paderborn und der Hochschule für Musik Detmold, 2020, S. 77–81, DOI: 10.25366/2020.96

musiconn
Für vernetzte Musikwissenschaft

Jahrestagung der Gesellschaft für Musikforschung 2019

Perspectives of Musical Corpus Studies: The Annotated Mozart Sonatas

MARKUS NEUWIRTH, LAUSANNE/LINZ, JOHANNES HENTSCHEL, AND MARTIN ROHRMEIER, LAUSANNE

The digital, closely linked with advanced computational and quantitative methods, opens up a potential previously unknown in music research. This concerns digital editions (of scores and writings about music) as well as the analysis of metadata and "the music itself". While much of previous music research (unless psychological in nature) had to confine itself to pondering hypotheses in the armchair, the increased availability of digital scores and data has put musicology in a position to ground hypotheses and theories in empirical corpus studies. Such corpus studies may sometimes produce results that are entirely unexpected and hence lead to a completely revised picture; yet even if a commonly held view is supported, the fact that it could be empirically corroborated by application of rigid statistical methods is a strong merit in itself, as it allows for more precise quantification and follow-up research.

Corpus studies of common-practice harmony nicely illustrate these observations, as these studies corroborate key characteristics of tonal harmony such as "referentiality", "centricity", and "directedness", using unigram and bigram statistics.[1] In our own projects on the use of harmony and voice-leading schemata across history,[2] we have chosen to produce music analytical datasets with the help of human annotators, rather than relying on automated analysis, which is still error-prone and music theoretically less sophisticated. Following up on the Annotated Beethoven Corpus (a publicly available dataset providing digital scores with harmonic analyses of all string quartets by Beethoven[3]), we recently introduced a similar corpus, which features harmonic annotations of chords and cadences as they occur in all keyboard sonatas by Wolfgang Amadé Mozart – another prominent sample of the "Classical Style".[4] These datasets invite

1 E. g., Martin Rohrmeier and Ian Cross, "Statistical Properties of Harmony in Bach's Chorales", in: *Proceedings 10th International Conference Music Perception & Cognition*, ed. by Ken'ichi Miyazaki, Mayumi Adachi, Yuzuru Hiraga, Yoshitaka Nakajima and Minoru Tsuzaki, Sapporo 2008, pp. 619–627; David Temperley, "Composition, Perception, and Schenkerian Theory", in: *Music Theory Spectrum* 33 (2011), pp. 146–168; Dmitri Tymoczko, *A Geometry of Music: Harmony and Counterpoint in the Extended Common Practice*, New York 2011; Fabian C. Moss, Markus Neuwirth, Daniel Harasim and Martin Rohrmeier, "Statistical Characteristics of Tonal Harmony: A Corpus Study of Beethoven's String Quartets", in: *PloS One* 14(6) (2019), <https://doi.org/10.1371/journal.pone.0217242>.

2 These projects are funded by the Swiss National Science Foundation and the Volkswagen Foundation, respectively.

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

4 Johannes Hentschel, Markus Neuwirth and Martin Rohrmeier, "The Annotated Mozart Sonatas: Score, Harmony, and Cadence", article submitted.

comparisons with other classical and non-classical corpora with regard to the use of harmony, cadences, or other aspects (e. g., musical form[5]), thus laying the ground for stylistic comparisons across history and genres.

The Mozart dataset in particular allows researchers to tackle specific questions such as (1) those regarding harmonic syntax in general, (2) the distribution of individual harmonic features across cadence types (e.g., the frequency of V(64) chords, or the use of V vs. V7 chords), or (3) the testing of the cadence typology itself, which in music theory textbooks is based mainly on a cadence's final harmonic and melodic events (e.g., perfect and imperfect authentic cadences (PAC and IAC, respectively), as well as the half cadence (HC) and deceptive and evaded cadences (DC and EC, respectively)).

Since our dataset combines harmonic and cadence annotations, it can be used to determine potential similarities of cadence instances across the conventional types on the basis of multiple harmonic features, thus enabling scholars to scrutinise the results proposed by Sears and colleagues on the basis of a much larger dataset.[6] More specifically, it can be tested whether PACs on the one hand and IACs, HCs, and failed cadences (DC and EC) on the other differ primarily with respect to their endings[7] or also with regard to their entire harmonic makeup. The latter question becomes particularly acute in view of the (counterintuitive) finding that the PAC is by far the most frequently used type (covering almost 50% of the whole distribution of roughly 1.000 cadence tokens), while the IAC is not on equal footing with the PAC, but a comparatively rare event (ca. 5%). Given that the PAC is also a diverse category, with many of its instances sharing harmonic features with instances from the other types, cadence tokens may be ranked in terms of closural strength (measured by the harmonic content) irrespective of their categorical membership.

The proposed cadence labels might prompt other researchers to add further layers of cadence annotations. This can be achieved either by introducing sub-categories of cadence types, such as subtypes of half cadences in Mozart's works[8], or by drawing on contemporaneous historical categories[9] which may or may not coincide with the cadence labels used in present-day

5 See, for instance, Mark Gotham and Matthew Ireland, "Taking Form: A Representation Standard, Conversion Code, and Example Corpus For Recording, Visualizing, and Studying Analyses of Musical Form", in: *Proceedings of the 20th International Society for Music Information Retrieval Conference, ISMIR*, Delft 2019, pp. 693–699, <http://archives.ismir.net/ismir2019/paper/000085.pdf> (25.06.2020).

6 David R. W. Sears, Marcus T. Pearce, William E. Caplin and Stephen McAdams, "Simulating Melodic and Harmonic Expectations for Tonal Cadences Using Probabilistic Models", in: *Journal of New Music Research* 47/1 (2018), pp. 29–52.

7 William E. Caplin, "The Classical Cadence: Conceptions and Misconceptions", in: *Journal of the American Musicological Society* 57/1 (2004), pp. 51–118.

8 Nathan John Martin and Julie Pedneault-Deslauriers, "The Mozartean Half Cadence", in: *What is a Cadence? Theoretical and Analytical Perspectives on Cadences in the Classical Repertoire*, ed. by Markus Neuwirth and Pieter Bérge, Leuven 2015, pp. 185–213.

9 See, for instance, Thomas Christensen (Ed.), *The Cambridge History of Western Music Theory*, Cambridge 2006.

theories.[10] The comparison and evaluation of such different taxonomies applied to the same corpus, in combination with the automated extraction of note-level features, has the potential of advancing the construction of a formal model of cadence and its constitutive features.

The Mozart dataset has been created using a flexible annotation standard that relies on a formalised vocabulary. Further, it is both machine-readable and user-friendly for the music theory experts who provide the annotations, as the standard builds on the widely shared convention of Roman numeral analysis. The annotation standard in its present form is, however, limited to musical styles that can meaningfully be analysed in terms of third-based sonorities featuring identifiable roots. These roots are required to be located within a local key, which in turn needs to be comprehensible in relation to one overarching (major or minor) key. Inevitably, this poses a challenge when one deals with other types of musical organization, for instance extended tonal music that drops the assumption of one main tonal centre or Tonfeld-based music that assumes octatonic, hexatonic, and stack-of-fifths pitch collections.[11] A further challenge is to find ways to convert the Roman numeral annotations to other encoding schemes. The compatibility of the proposed harmony annotations with more coarse-grained standards (e. g., those providing root information only or those operating with absolute chord labels) offers rich potential for joint research efforts.

There is a lot to be gained from these kinds of datasets. Two aspects in particular will be pointed out by way of conclusion. First, solving questions of authenticity (or authorship attribution) has long been an important concern in historical musicology, which is even more acute if reliable external evidence is missing and the available analytical methods of dealing with internal (stylistic) evidence are too imprecise, arbitrary, and informal. These problems can to some extent be remedied by corpus studies.[12] By training a model based on pieces with known composers, it is possible to discern different musical features (whether low- or high-level) in terms of their discrimination potential. This provides a promising bridge between music analysis and philology. Second, analytical datasets can be linked with written texts (modern or historical) addressing the analysed pieces and hence allow scholars to characterise abstract hermeneutic and aesthetic concepts in terms of analytical findings.[13]

---

10 E. g., Caplin, "The Classical Cadence".

11 E. g., Richard Cohn, *Audacious Euphony: Chromatic Harmony and the Triad's Second Nature*, New York 2012; Michael Polth, "The Individual Tone and Musical Context in Albert Simon's Tonfeldtheorie", in: *Music Theory Online* 24/4 (2018), <https://doi.org/10.30535/mto.24.4.15>.

12 E. g., Andrew Brinkman, Daniel Shanahan and Craig Sapp, "Musical Stylometry, Machine Learning, and Attribution Studies: A Semi-Supervised Approach to the Works of Josquin", in: *Proceedings of the 14th Biennial International Conference on Music Perception and Cognition (ICMPC)*, San Francisco 2016, <https://icmpc.org/icmpc14/files/ICMPC14_Proceedings.pdf> (25.06.2020), pp. 91–97.

13 See, for instance, Michael Piotrowski and Markus Neuwirth, "Prospects for Computational Hermeneutics", in: *Proceedings of the 9th Annual Conference of the AIUCD (Milan, Jan. 15–17, 2020)*, ed. by Cristina Marras and Marco Passarotti, Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD), Milan 2020, pp. 204–209, <https://aiucd2020.unicatt.it/aiucd-Piotrowski_Neuwirth.pdf> (25.06.2020).

Zitation: Markus Neuwirth, Johannes Hentschel, Martin Rohrmeier, „Perspectives of Musical Corpus Studies: The Annotated Mozart Sonatas", in: *Brückenschläge zwischen Musikwissenschaft und Informatik. Theoretische und praktische Aspekte der Kooperation*, in Verbindung mit der Fachgruppe Digitale Musikwissenschaft hrsg. von Stefanie Acquavella-Rauch, Andreas Münzmay und Joachim Veit (= Musikwissenschaft: Aktuelle Perspektiven. Bericht über die Jahrestagung der Gesellschaft für Musikforschung 2019 in Paderborn und Detmold, Bd. 3), Detmold, Musikwissenschaftliches Seminar der Universität Paderborn und der Hochschule für Musik Detmold, 2020, S. 77–81, DOI: 10.25366/2020.96

Abstract

This paper discusses the potential of musical corpus studies, taking research on common-practice tonal harmony as a case in point. Based on a brief depiction of a project carried out at the École polytechnique fédérale de Lausanne (EPFL) and two novel datasets of harmonic analyses of music in the classical style, we elaborate on research questions, applications, and the need of extending the annotation standard used.

Short bios

Markus Neuwirth is a Professor of Music Analysis at the Anton Bruckner University Linz and a Researcher at the Digital and Cognitive Musicology Lab of the École polytechnique fédérale de Lausanne (EPFL). At EPFL he is conducting a research project titled "From Bach to the Beatles – Exploring compositional building blocks and musical style change with hermeneutic and computational methods" funded by the *Volkswagen Foundation* (2018–2020; together with Martin Rohrmeier). Previously he held a Postdoctoral Fellowship (2013–2016) from the Research Foundation Flanders (FWO) at Leuven University, where he obtained his PhD in musicology in 2013 (again funded by the FWO). Neuwirth is editor-in-chief of the journal *Music Theory and Analysis*.

Johannes Hentschel studied music education, music theory, and Romance studies in Freiburg i. Br., Lübeck, and Helsinki. Proficient as an accordionist, singer, and conductor, he has been a lecturer for music theory at music universities. In 2018, he joined the École polytechnique fédérale de Lausanne (EPFL). Supervised by Martin Rohrmeier at the Digital and Cognitive Musicology Lab (DCML), Johannes Hentschel is preparing a doctoral dissertation on diachronic style change in music while being highly active in the domain of corpus building and metadata organization.

Martin Rohrmeier is the director of the Digital and Cognitive Musicology Lab (DCML) at the École Polytechnique Fédéral de Lausanne (EPFL). He studied philosophy, mathematics, and musicology in Bonn, and received his master and doctoral degrees from Cambridge University at the Centre for Music & Science under the supervision of Ian Cross. He subsequently worked at Microsoft Research, FU Berlin, and MIT Linguistics with a fellowship from the MIT Intelligence Initiative. In 2014, he was appointed one of the Open Topic Professorships in Systematic Musicology and Music Cognition at TU Dresden. In 2017, he joined the EPFL as associate professor for Digital Musicology.

Brückenschläge zwischen Musikwissenschaft und Informatik

Theoretische und praktische Aspekte der Kooperation

Herausgegeben von Stefanie Acquavella-Rauch, Andreas Münzmay und Joachim Veit

Musikwissenschaft:
Aktuelle Perspektiven 3

musiconn
Für vernetzte Musikwissenschaft

Brückenschläge zwischen Musikwissenschaft und Informatik

Musikwissenschaft: Aktuelle Perspektiven

Bericht über die Jahrestagung der Gesellschaft für Musikforschung 2019 in Paderborn und Detmold

Herausgegeben von Rebecca Grotjahn und Nina Jaeschke

Band 3

Brückenschläge

zwischen Musikwissenschaft und Informatik

Theoretische und praktische Aspekte der Kooperation

Beiträge der Symposien zur Digitalen Musikwissenschaft
Osnabrück 2018 und Paderborn 2019
im Rahmen der Jahrestagungen der Gesellschaft für Musikforschung

In Verbindung mit der Fachgruppe Digitale Musikwissenschaft
herausgegeben von
Stefanie Acquavella-Rauch, Andreas Münzmay und Joachim Veit

Detmold: Musikwissenschaftliches Seminar der Universität Paderborn
und der Hochschule für Musik Detmold
2020

musiconn
Für vernetzte Musikwissenschaft

DOI: 10.25366/2020.87

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Impressum

Redaktion: Stefanie Acquavella-Rauch, Andreas Münzmay und Joachim Veit
Satz: Nina Jaeschke und Joachim Veit

INHALT

Gabriele Buschmeier in memoriam