Musical Schemata: Modelling Challenges and Pattern Finding (BachBeatles)

Markus Neuwirth, Christoph Finkensiep, and Martin Rohrmeier

Abstract *'From Bach to the Beatles: Exploring Compositional Building Blocks and Musical Style Change with Hermeneutic and Computational Methods' aims at finding voice-leading schemata in digital corpora, analyzing their characteristics, and describing their chronological distributions between ca. 1700 and 1970. Voice-leading schemata are fundamental building blocks of Western music across historical periods and styles, ranging from Renaissance, Baroque, and Classical to modern tonal music. Rather than focusing on concrete findings, the chapter adopts a meta-perspective, explaining and discussing the methodological choices regarding initial hypotheses, modelling options, encodings, pattern-matching methods, and the interpretation frameworks. The final part of the chapter is devoted to a discussion of how computational methods and hermeneutics can interact with one another in a productive manner.*

Computational music theory

Music theory is primarily concerned with characterizing musical structures as they occur across cultures, historical periods, and styles, aiming to reveal the general principles underlying these structures. Previous music-theoretical scholarship has suffered to some extent from a number of core problems including the problem of intuitive statistics, a lack of methodological transparency, and various sampling biases.[1] These shortcomings have been addressed in the more recent field of musical corpus studies. Over the past two decades, music theory has intensified efforts to adopt powerful digital/computational methods in order to empirically scrutinize armchair hypotheses about the nature of musical structures and their historical de-

1 See, for instance, Markus Neuwirth and Martin Rohrmeier, "Wie wissenschaftlich muss Musiktheorie sein? Chancen und Herausforderungen musikalischer Korpusforschung", *Zeitschrift der Gesellschaft für Musiktheorie* 13/2 (2016), 171–193.

velopment, relying on a steadily growing amount of datasets.[2] These datasets can be human- or machine-generated, refer to structural domains such as harmony, counterpoint, melody, rhythm, meter, timbre etc., and involve Western as well as non-Western musical styles.[3]

However, it is only a more recent trend in digital music research to consider more complex musical objects such as voice-leading schemata[4], cadences[5], or sonata form[6]. Successfully dealing with these phenomena not only crucially depends on expert knowledge, but also presents the challenge of how to model and operationalize high-level music-theoretical concepts in ways suitable for digital

---

2 Several societies and interest groups have been launched that testify to this trend, most importantly the Music Encoding Initiative (MEI) in 2005 and the International Society for Music Information Retrieval (ISMIR) in 2008.

3 E.g., Christof Weiß, Frank Zalkow, Vlora Arifi-Müller, Meinard Müller, Hendrik Vincent Koops, Anja Volk and Harald G. Grohganz, "Schubert Winterreise Dataset: A Multimodal Scenario for Music Analysis", *ACM Journal on Computing and Cultural Heritage* (JOCCH), 14/2 (2021), 1–18.

4 See, for instance, James Symons, "A Cognitively Inspired Method for the Statistical Analysis of Eighteenth-Century Music, as Applied in Two Corpus Studies", PhD thesis, Northwestern University, 2017; David R. W. Sears, Marcus T. Pearce, William E. Caplin and Stephen McAdams, "Simulating Melodic and Harmonic Expectations for Tonal Cadences Using Probabilistic Models", *Journal of New Music Research* 47/1 (2018), 29–52; David R. Sears and Gerhard Widmer, "Beneath (or Beyond) the Surface: Discovering Voice-leading Patterns with Skip-grams", *Journal of Mathematics and Music* 15/3 (2020), 1–26; DOI: 10.1080/17459737.2020.1785568; Christoph Finkensiep, Markus Neuwirth and Martin Rohrmeier (2018), "Generalized Skipgrams for Pattern Discovery in Polyphonic Streams", in: E. Benetos, E. Gómez, X. Hu, and E. Humphrey (eds.): *Proceedings of the 19th International Society for Music Information Retrieval Conference (ISMIR)* (Paris 2018) 547–553; Andreas Katsiavalos, Tom Collins and Bret Battey, "An Initial Computational Model for Musical Schemata Theory", in: *Proceedings of the International Society on Music Information Retrieval*, 2019, 166–172; and Christoph Finkensiep, Ken Déguernel, Markus Neuwirth and Martin Rohrmeier, "Voice-Leading Schema Recognition Using Rhythm and Pitch Features", in: *Proceedings of the International Society for Music Information Retrieval* (ISMIR) (Montreal 2020).

5 For instance, Ben Duane, "Melodic Patterns and Tonal Cadences: Bayesian Learning of Cadential Categories from Contrapuntal Information", *Journal of New Music Research* 48/3 (2019), 197–216; and Johannes Hentschel, Markus Neuwirth and Martin Rohrmeier, "The Annotated Mozart Sonatas: Score, Harmony, and Cadence", *Transactions of the International Society for Music Information Retrieval* 4/1 (2021), 67–80; https://doi.org/10.5334/tismir.63

6 Christof Weiß, Stephanie Klauk, Mark Gotham, Meinard Müller and Rainer Kleinertz, "Discourse Not Dualism: An Interdisciplinary Dialogue on Sonata Form in Beethoven's Early Piano Sonatas", in: *Proceedings of the International Society for Music Information Retrieval Conference* (ISMIR) 2020, 199–206; Pierre Allegraud, Louis Bigo, Laurent Feisthauer, Mathieu Giraud, Richard Groult, Emmanuel Leguy and Florence Levé, "Learning Sonata Form Structure on Mozart's String Quartets", *Transactions of the International Society for Music Information Retrieval* 2/1 (2019), 82–96.

and empirical research. Our project titled 'From Bach to the Beatles: Exploring Compositional Building Blocks and Musical Style Change with Hermeneutic and Computational Methods' aimed (1) to address that modelling challenge and (2) to come up with algorithmic solutions to finding voice-leading schemata in digital corpora.

Why study musical schemata?

Why is the study of voice-leading schemata in particular a worthwhile endeavour? Firstly, most voice-leading schemata are comparatively easy to recognize by both trained and untrained listeners, and they play a pivotal educational role in ear-training classes at music universities. Secondly, voice-leading schemata can be seen as an important musical object from the point of view of diachronic style development. They are frequently used patterns to be found across historical periods, ranging from Renaissance, Baroque, and Classical to modern tonal music. Examples include such well-known and cognitively salient schemata as the Lamento, the Pachelbel, the descending-fifths sequence, and various forms of cadences (i.e., closing formulae). Some schemata even carry with them specific expressive or affective connotations: the Lamento, for instance, is traditionally associated with grief and depression; similarly, the Teufelsmühle (or Omnibus) conveys notions of uncertainty, horror, and imminent death.[7]

Finally, voice-leading schemata constitute a rich musical object in view of both their structure and usage. A schema serves as an abstract contrapuntal template that composers can elaborate in multiple ways by flexibly inserting embellishing notes. Apart from the frequency of use of particular schemata, it is the concrete way in which schemata are instantiated in a composition that helps distinguish musical styles from one another.

Our 'From Bach to the Beatles' project was primarily concerned with finding voice-leading schemata in large digital corpora in order to be able to answer questions about the use of schemata over time and geographical areas. At present, there is only scant quantitative evidence of the frequency and diachronic distribution of

---

7 E.g., William E. Caplin, "Topics and Formal Functions: The Case of the Lament", in: Danuta Mirka (ed.): *The Oxford Handbook of Topic Theory* (New York: Oxford University Press, 2014), 415–452; John A. Rice, "The Morte: A Galant Voice-Leading Schema as Emblem of Lament and Compositional Building-Block", *Eighteenth-Century Music* 12/2 (2015), 157–181; Paula J. Telesco, "Enharmonicism and the Omnibus Progression in Classical-Era Music", *Music Theory Spectrum* 20/2 (1998), 242–279; and Marie-Agnes Dittrich, '"Teufelsmühle" und "Omnibus"', *Zeitschrift der Gesellschaft für Musiktheorie* 4/1–2 (2007), 107–121, https://doi.org/10.31751/247

schemata across history;[8] large-scale, machine-readable datasets on schemata are almost non-existent.[9] For assessing the prevalence of schemata in musical corpora, automated recognition of schema instances can be a time- and cost-efficient alternative to manually labelled data.

There are, however, two key challenges that computational approaches face as they seek to detect note patterns in music: (1) the multidimensional (polyphonic) structure of music as opposed to, for example, the sequential structure of written text; and (2) the highly flexible nature of these patterns, given that the structural notes in the individual voices can be elaborated in a great variety of ways. These problems will be explained in detail further below (see section The challenge of finding schemata).

What are musical schemata?

The schema concept itself comes from cognitive psychology and has subsequently been introduced in the field of music research.[10] Related concepts used in this context are script, frame, prototype, idealtype, archetype, model, meme, idiom (in the sense defined by construction grammar), cognitive map, and associative network.[11] In principle, any music-structural domain (e.g., harmony, form, and counterpoint) can be conceptualized in terms of schemata. Nonetheless, the term 'schema' has cur-

8 Among the exceptions are Robert O. Gjerdingen, *A Classic Turn of Phrase: Music and the Psychology of Convention* (Philadelphia: University of Pennsylvania Press, 1988); Gjerdingen, *Music in the Galant Style* (New York: Oxford University Press, 2007); Vasili Byros, "Towards an 'Archaeology' of Hearing: Schemata and Eighteenth-century Consciousness", *Musica Humana* 1/2 (2009), 235–306; and Byros, "Trazom's Wit: Communicative Strategies in a 'Popular' Yet 'Difficult' Sonata", *Eighteenth-Century Music* 10/2 (2013), 213–252.

9 For a recent contribution to this issue, see Finkensiep et al., "Voice-Leading Schema Recognition".

10 The cognitive notion of schema is described in David E. Rumelhart, "Schemata: The Building Blocks of Cognition", in: Rand J. Spiro, Bertram C. Bruce, and William F. Brewer (eds.): *Theoretical Issues in Reading Comprehension* (Hillsdale, NJ: Lawrence Erlbaum, 1980), 33–58. Arguably the first author to explicitly apply the concept of schema to musical phenomena was Leonard B. Meyer (e.g., Meyer, "Innovation, Choice, and the History of Music", *Critical Inquiry* 9/3 [1983], 517–544); it was picked up by Robert Gjerdingen in 1988.

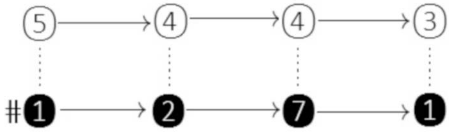
11 The notion of 'script' is introduced in Roger Schank and Robert P. Abelson, *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures* (New York: Halsted, 1977). For a Neo-Darwinian, memetic approach to schemata, see Steven Jan, "Using Galant Schemata as Evidence for Universal Darwinism", *Interdisciplinary Science Reviews* 38/2 (2013), 149–168. On the remaining terms frame (Minsky), prototype (Rosch), idealtype/archetype (Meyer), and model (Stachowiak), see Gjerdingen 1988 and 2007.

rently become almost a synonym of standardized *voice-leading* (or contrapuntal) patterns[12]; and this is how we also employ the term in this paper.

Voice-leading schemata are usually defined as configurations of two or more voices that move together through a sequence of stages, forming specific patterns of successive vertical intervals that occur within a specific tonal context.

Consider the example of the Fonte (a concept introduced as early as the mid-eighteenth century): The Fonte is usually defined as a four-stage pattern involving at least two voices. The bass moves through the scale degrees #1-2-7-1 of a major scale, while the soprano follows the pattern 5-4-4-3, thus producing the following sequence of vertical intervals: tritone → minor third → tritone → major third (see Example 1 for an abstract schema representation). In actual compositions, the schema prototype can be elaborated in multiple ways. For instance, the notes belonging to one stage can be displaced in time, as long as their belonging to a particular stage is indisputable. Also, structural notes can be flexibly ornamented by inserting additional notes. A real-world example illustrating the surface realization of a Fonte is given in Example 2.

Example 1: An abstract representation of a Fonte schema. The arrows denote the temporal sequence of vertical intervals (with dashed lines connecting the scale degrees in the upper voice and the bass voice).



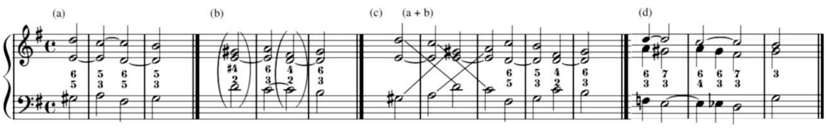
12 In German, these patterns are commonly referred to as 'Satzmodelle'; see, for instance, the special issue in the *Zeitschrift der Gesellschaft für Musiktheorie* 4/1–2 (2007), https://www.gmth.de/zeitschrift/ausgabe-4-1-2007/inhalt.aspx.

Example 2: A Fonte instance in Wolfgang Amadé Mozart's Piano Sonata in B-flat major, K 281, opening bars of final movement. The two diagonal dashed lines show note displacement within a stage. The notes occurring between the highlighted structural notes represent embellishments (or elaborations). (The figure is taken from Finkensiep et al. 2018.)

While the correct interval pattern is a central property of any schema instance, it is not a sufficient one. The selected notes must also provide the contrapuntal template for its context, such that all the notes contained in the time span covered by the schema instance can be meaningfully interpreted as ornamentations of the selected notes.

Although the Fonte seems to be a comparatively simple object, a variety of subcategories can be produced (and have in fact been used throughout music history) by applying standard operations such as voice shifting, voice exchange, note insertion, and manifold combinations thereof (see Example 3).

Example 3: The Fonte in its prototypical version (a) and variants produced by voice shifting (b), voice exchange (c), and note insertion (d).



Challenges of modelling and operationalization

To enable music theory and computer-assisted corpus research to efficiently interact, theories need to be sufficiently precise and formalizable, and the abstract concepts used need to be operationalized, that is, translated into observable entities. Since our project is concerned with identifying in digital corpora one particular class of musical objects such as voice-leading schemata, the very need to operationalize and formalize these objects requires one to reconsider and refine traditional descriptions and definitions. Also, it may involve straightening out incompatible aspects between historical and modern definitions of particular schemata. The flipside of such strict definitions is that one may be forced to abstract from an object's fuzzy, but rich aspects in terms of contextual connotation. This is why operationalization is a notoriously intricate issue.

Even for such a well-known schema as the Fonte, several (slightly divergent) definitions exist in the music theoretical literature. Starting with the historical literature, Joseph Riepel (who coined the term Fonte) describes the schema as follows: "[…] gleich wie das *Monte* einen Ton hinauf – also steigt das *Fonte*, wie du weißt, einen Ton herab; das erste Glied wendet sich nämlich in die Secund *Tertz minor*, und das zweyte Glied in den Hauptton *Tertz major*".[13] Riepel thus emphasizes the sequential component: the first part as a whole is moved down by whole-step, thus producing a shift from minor to major (and therefore, in synaesthetic terms, from darkness to brightness).

A much broader definition is proposed by Markus Schwenkreis, who basically equates a Fonte with a descending-fifth sequence; as a result, there is no intrinsic extension of the schema in terms of the number of stages involved.[14] Finally, Robert Gjerdingen defines the Fonte primarily in terms of scale degrees in the soprano and bass: "The 7–1 ascent in the bass is matched by a 4–3 descent in the melody, often terminating a larger 6–5–4–3 descent […] All these features together suggest a Fonte prototype of four events arranged into two pairs […]."[15] In other words, the resulting vertical intervals are seen as a product of the contrapuntal combination of the two outer voices. This definition cannot capture those instances, however, in which no seventh (the local 4) is involved in the dominant (or V) chords launching each pair of events.

Describing schemata only in terms of a specific combination of voices and, therefore, as a succession of vertical intervals may be seen as insufficient, as schemata typically also feature a particular harmonic signature which can be expressed in terms of Roman numerals (which in turn assume chords, their roots, and their relation to a given key). The harmonic signature for the main variant of

13 Joseph Riepel, *Erläuterung der betrüglichen Tonordnung* (Augsburg 1765), 24 ('as the Monte goes up a step– so, as you know, the Fonte goes down a step; namely, the first member turns to the minor-mode second scale degree, and the second member to the main key tonic' (translation by the authors)).

14 Markus Schwenkreis (ed.), *Compendium Improvisation – Fantasieren nach historischen Quellen des 17. und 18. Jahrhunderts* (Basel: Schwabe, 2018), 87. A similar definition can be found in Hubert Mossburger, *Ästhetische Harmonielehre: Quellen, Analysen, Aufgaben*, vol. II (Wilhelmshaven: Noetzel, 2012), 770ff.

15 Gjerdingen, *Music in the Galant Style*.

the Fonte, for instance, would read as V65/ii–ii–V65–I. This harmonic layout is sometimes seen as a mere by-product of the voice combination described above.

Overall, the schema descriptions found in the scholarly literature are often somewhat imprecise and leave room for interpretation, which means they cannot readily be used for the purposes of computer-assisted music analysis. Therefore, in our project we had to come up with a (formal) model that translated the ideas described above into observable entities.

Generally, a formal model is a description of a particular segment of the world in terms of *entities* and their *relationships*, using the language of mathematics.[16] Since models aim to specify which aspects of the world are included and which are to be ignored, they inevitably involve abstraction from and simplification of the world under study.[17] As long as the researcher keeps in mind that the model is not to be confused with the richer real-world object, the model-based approach comes with the obvious advantage that otherwise diverse objects can now be compared with one another, thus enabling more general insights (see Conclusion).

From the perspective of model-based music theory, voice-leading schemata define the relationships between such entities as structural (schematic) and ornamental (non-schematic) notes and, in so doing, generalize over a wide variety of note configurations on the musical surface. In other words, schematic models provide an 'explanation' of the musical surface. The determination of how exactly the schematic core notes are elaborated in approaching the musical surface is the task of a comprehensive theory of counterpoint and historical styles, which has yet to be developed, though.

Our model of schemata

A notionally *comprehensive* model of musical schemata would no doubt require including a wealth of structural information about keys, harmonies, metrical weights, and distinct voices. In our model, however, we do not make any assumptions about these components in order not to considerably increase the complexity of the model.

As for the voices involved in the actual polyphonic structure of a piece, we do not make any assumption about their number and nature. Also, we do not aim to define

16 Christoph Finkensiep, Markus Neuwirth and Martin Rohrmeier, "Music Theory and Model-Driven Corpus Research", in: Daniel Shanahan, Ashley Burgoyne and Ian Quinn (eds.): *Oxford Handbook of Music and Corpus Studies* (New York: Oxford University Press, forthcoming). Generally on model-based digital humanities, see Julia Flanders and Fotis Jannidis (eds.), *The Shape of Data in Digital Humanities: Modeling Texts and Text-based Resources* (London: Routledge, 2018).

17 In contrast to 'approximation', which trades a simpler model for slight inconsistencies with the real world, 'abstraction' does not introduce any world-model contradictions.

rules that would help us to determine which note events on the musical surface belong to which voices.[18] Further, note events are not assigned scale-degrees within a particular key context, as this would require an analysis of the musical input in terms of tonal keys; nor does our model assign metrical weights to schema events. Instead, we opt for an underspecified and somewhat simplified model: voice-leading schemata are modelled merely as a specific succession of vertical intervals—a series of 'stages' based on a fixed number of voices. Each vertical interval in the model structure constitutes a 'stage' that contains one note per voice; and the number of stages is defined as fixed for each schema type.[19] Further, we assume that on the musical surface the structural notes belonging to a stage may be displaced in time, without making any restrictions on the number of intervening note events. Heuristically, however, we define a temporal limit where the distance is constrained between the displaced note events forming a stage.

The prototype for each schema variant (or subtype) is specified by using a formal notation. For instance, the prototype of the two-voice Fonte is encoded as: 'fonte.2': [['a1', 'P5'], ['M2', 'P4'], ['M7', 'P4'], ['P1', 'M3']], where '2' indicates the two-voice variant of the Fonte. Each note is given as an interval to some arbitrary reference point (for instance, P5 stands for 'perfect fifth', M2 for major second etc.). Since this comprises all possible transpositions of the schema, it is not necessary to know the reference key.

The challenge of finding schemata by using gram-based methods

The fact that (polyphonic) music consists of multiple simultaneous voices, and that the schematic core tones can be greatly embellished, makes it challenging to find voice-leading schemata in digitally encoded corpora. (Local) gram-based methods, which are standard in fields such as computational linguistics, constituted the first option that we considered. Gram-based methods extract short sequences from a longer stream of entities (e.g., letters or words in language; notes or chords in music). The most basic gram model, the n-gram, is just a *consecutive* subsequence in the input stream that has n elements.

18 Voice separation is itself a demanding theoretical and computational problem; see, for instance, Tillman Weyde and Reinier de Valk, "Chord- and Note-based Approaches to Voice Separation", in: David Meredith (ed.): *Computational Music Analysis* (Cham: Springer, 2016), 137–154.

19 This is another simplification: schemata such as the 5–6 progression or the descending-fifths sequence do not have a fixed size in terms of the number of stages involved.

While n-gram approaches are useful for uncovering contrapuntal patterns that consist of elements appearing adjacent on the musical surface,[20] skip-gram approaches allow one to capture non-adjacent structural elements.[21] By extending the n-gram idea, skip-grams produce subsequences by leaving out (or 'skipping') up to k elements.[22] Thus, skip-grams may reveal particular patterns that are otherwise 'disguised' by intervening events.

However, using a conventional skip-gram approach did not prove sufficient for our purpose. Both n-grams and their skip-gram extensions assume the distance between entities to be determined by their positions in the stream. While this assumption might be reasonable for text, monophonic melodies, or chord sequences, it is problematic for other applications that involve multiple (simultaneous) temporal streams, in particular streams of musical events such as notes. Therefore, it is desirable to measure the distance between events (notes) based on their *timing* information (that is onset, offset, and duration). Second, while notes might be simultaneous in a score, they occur sequentially in a stream or list-of-notes representation, which becomes problematic if distance is measured by index.

As a novel approach to the problem of finding polyphonic patterns in music, we developed an algorithm that extended previous n-gram and skip-gram approaches by using an arbitrary distance measure (e.g., temporal distance).[23] This allows skip-grams to be applied to non-sequential structures such as polyphonic music. Furthermore, the ability to operate on individual notes enables us to group non-simultaneous notes into stages. The stages can then be grouped into schema candidates in a second pass. Note, however, that due to the exhaustive search and a great number of possible note combinations, our resulting dataset is extremely unbalanced, containing many more accidental occurrences of the pre-defined interval pattern (false positives) than true schema instances. To further reduce the number of schema candidates, guided by prior music-theoretical intuition, we restricted the window size

---

20 E.g., Mathieu Bergeron and Darrell Conklin, "Subsumption of Vertical Viewpoint Patterns", in *International Conference on Mathematics and Computation in Music* (Springer: Berlin, Heidelberg, 2011), 1–12; and Christopher Antila and Julie Cumming, "The VIS Framework. Analyzing Counterpoint in Large Datasets", in: Hsin-Min Wang, Yi-Hsuan Yang and Jin Ha Lee (eds.) *Proceedings of the 15th International Society for Music Information Retrieval Conference* ISMIR (Taipei, Taiwan, 2014), 71–76.

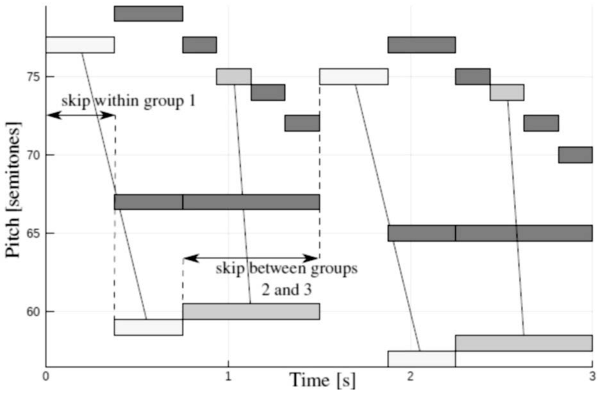
21 David R. Sears, Andreas Arzt, Harald Frostel, Reinhard Sonnleitner and Gerhard Widmer, "Modeling Harmony with Skip-Grams", in: *Proceedings of the International Society on Music Information Retrieval* (2017), 332–338.

22 D. Guthrie, B. Allison, W. Liu, L. Guthrie and Y. Wilks, "A Closer Look at Skip-Gram Modelling", in: *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'06)* (Genoa: European Language Resources Association (ELRA), 2006), 1222–1225.

23 See Finkensiep et al., "Generalized Skipgrams for Pattern Discovery"; see also Sears and Widmer, "Beneath (or Beyond) the Surface".

to a maximal note displacement of one bar per stage and a maximal distance of one bar between the onsets of two adjacent stages (for an example of schema-matching using our skip-gram approach, see Example 4).

Example 4: The figure shows an example of the application of skip-grams to polyphonic music displayed in piano-roll visualization. The highlighted notes are members of the skip-gram; the stages are indicated by solid lines between notes belonging to the same stage. The skip-gram pattern refers to the schema instance shown in Example 2. (The figure is taken from Finkensiep et al. 2018.).



The dataset

The data created in the project and used for schema matching consist of digital musical scores and expert-labelled schemata added to them.[24] Our dataset is based on the full set of Mozart's piano sonatas encoded in MusicXML format. These 18 sonatas with three movements each (thus 54 movements in total) were composed between 1774 and 1789 and constitute a prominent sample of the classical style. The pieces in the dataset contain 103,829 notes in total distributed over 7,500 measures, with 244 hand-annotated true schema instances. They are complemented by 190,994 false instances (99.87% vs. 0.13% true instances) which were additionally found by the skip-gram matcher for the selected schema types and subtypes. In Finkensiep et al.

24 The underlying annotation standard has been developed in the course of this project. The dataset is stored for version control in a GitHub repository. For historical sources of classical music, there are no copyright issues. See Finkensiep et al., "Voice-Leading Schema Recognition".

2020, we selected 10 schema types and 20 subtypes from a comprehensive lexicon of schemata. Keeping the high combinatorial complexity of multi-voice structures in mind, we decided to restrict this study to two-voice schema variants. Furthermore, we reduced the number of candidates to at most 25 per group of temporally overlapping candidates.

Schema instances are encoded as nested arrays of notes in the same form as the corresponding prototypes. Instances may deviate from the shape of the prototype if (a) a note that would repeat its predecessor (e.g. the second 4 in the Fonte) is held over or missing, or (b) two adjacent voices merge and are represented by a single note on the surface.

The schema annotation was performed in a hybrid fashion. First, we developed our skip-gram-based schema matcher that located potential instances of schemata in a digital score and helped the user visualize and explore these candidates.[25] Second, we developed a schema annotation app that allowed users to annotate musical schemata. The app interacts with the schema matcher so that algorithmically computed schema instances can be taken into account or modified in the human annotation process.

Regarding the human-machine interaction, we observe an effect of mutual interaction: the human expert assesses schema instances as classified by the algorithm while the algorithm also identifies schema instances which might surprise the expert or which the expert has missed. We begin to observe a phenomenon that was previously seen in other developments of Artificial Intelligence: as evident in the research on games such as chess, backgammon, poker, or Go from the 80s onwards, human expertise is not the only infallible source of evidence and computational models (begin to) infer patterns that may partly extend expert knowledge.[26]

Scrutinizing domain expert intuitions

The dataset compiled by computer-assisted annotation has been explored by using statistical methods. The results help scrutinize theoretical intuitions and hypotheses about the use of individual schemata and their frequency of occurrence.

Since the aim of our project was not to solve an engineering task, we chose to design a formal model of schemata that explicitly reflected their structural properties

25 Note that schema visualization is an integral part of schema matching and annotation apps.

26 On AI and music as well as related challenges, see, for instance, Martin Rohrmeier, "On Creativity, Music's AI Completeness, and Four Challenges for Artificial Musical Creativity", in: *Transactions of the International Society for Music Information Retrieval* 5/1 (2022).

(rather than a neural network, for instance). This way, the results and model properties remain interpretable for the human theorist and also provide feedback for music theory by usefully defining schemata and its related set of features. The very process of developing the different versions of the computational models (leading to the current, but not final, model) is itself a rich resource, where more traditional music analysis and formal modelling enrich each other in various ways (see Conclusion).[27]

While classifying musical building blocks as instances of an underlying schema, human listeners and analysts rely, often unconsciously, on a variety of features. Whether or not this is done in a consistent fashion is open to debate and empirical scrutiny. In our project, the uncertainty associated with schema identification and classification played an important role. There are numerous borderline cases which do not neatly fit into one of the available categories, or which combine elements from two or more categories. In this regard, computational approaches prove powerful in testing the criteria on which human judgment implicitly relies. In our case of schema detection, these (pitch- and rhythm-related) criteria are 'complexity', 'regularity', and 'salience'. While complexity involves the issue of whether two structural events of one stage occur at the same time or at different onsets, regularity concerns, for instance, the rhythmic similarity across stages;[28] and salience is related to duration and metrical weight.

In Finkensiep et al. (2020), we have evaluated the above features for schema recognition by using a binary classifier. Here, we have shown that parallelism (i.e. rhythmic regularity aligned to the metrical grid) has indeed a strong positive influence, thus indicating a preference for regular temporal organization. Generally, true schema instances exhibit higher internal regularity and lower complexity (e.g. in terms of note displacement) than non-instances do. In contrast, properties related to intrinsic salience (e.g. duration or metric weight of the matched notes) are less important.

As the results in Finkensiep et al. (2020) show, distinguishing between incidental and structural note configurations based on a small number of musically and cognitively motivated heuristics works well in the vast majority of cases. Even if a number of misclassifications remain, a closer look at these cases provides valuable insights into the problem at hand. First, the main limitation of our approach is that the model

---

27 The proximity of formal modelling in computer science to hermeneutic techniques as practised in music analysis is discussed in Michael Piotrowski and Markus Neuwirth (2020), "Prospects for Computational Hermeneutics", in: Cristina Marras and Marco Passarotti (eds.): *Proceedings of the 9th Annual Conference of the AIUCD* (Milano, Jan. 15–17, 2020), Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD), 204–209.

28 See, for instance, Mathieu Giraud, Ken Déguernel and Emilios Cambouropoulos, "Fragmentations with Pitch, Rhythm and Parallelism Constraints for Variation Matching", in: *International Symposium on Computer Music Multidisciplinary Research (CMMR)*, 2013, 298–312.

assesses suggested schema instances individually without considering or comparing them to alternative interpretations. In many cases, the main reason for human experts to reject a candidate does not seem to be a lack of plausibility of the match itself, but the unavailability of a 'better explanation', that is, an alternative analysis of the context of the match that identifies a more plausible contrapuntal scaffold.

A second insight concerns the idea of schema itself and its relation to a classification task. From a cognitive perspective, a schema does not need to be instantiated unambiguously or even completely. It is sufficient if listeners recognize the schema as the template for the surface events, or if they understand the composer's intention to evoke the schema. In this regard, discrete binary classification into instances and non-instances may be as unattainable as it is undesirable, falling short of the complexity that the relationship between schema and realization can exhibit.

Conclusion: The interactive potential between close and distant reading in music research

By way of conclusion, we would like to consider the relationship between traditional music theory (as a humanities discipline) and digital/quantitative research on a more general level, pondering the various objections raised against their interaction.

The frequently voiced accusation of positivism (in its various forms) that the generalizing impulse of quantitative research is not capable of capturing the 'qualities' realized in each individual (historical) case is often a reflexive reaction to corpus-based computational studies. The widespread scepticism towards quantitative approaches in historical or cultural music research is based on a number of concerns.

Firstly, one may argue that historically differentiating, context-sensitive 'close reading' and (much) more coarse-grained 'distant reading' are mutually exclusive modes of research. The observation of patterns in a collection of objects by necessity requires abstraction from details in order to make the objects comparable (see above on modelling and its challenges). The individuality-driven analysis as practised in hermeneutically oriented studies has a different purpose and research interest than the observation of common features or trends in a large corpus.

Secondly, qualitative hermeneutic methods usually operate with a larger number of categories; and they are furthermore characterized by a higher degree of granularity in the sense of a 'thick description' (Clifford Geertz) than the ideally few and more large-grained categories of corpus research. This is precisely why a corpus study cannot be obtained from the accumulation of many small hermeneutic analyses.

Thirdly, there is the fear that in the long run computational and quantitative methods would eventually replace qualitative hermeneutic approaches and make human analysts expendable.

Nonetheless, without casting doubt on the principle validity of points 1 and 2, we want to argue that close and distant readings can be bridged such that they complement each other in productive ways. To begin with, qualitative insights—or pre-empirical intuitions—serve as a starting point that allows us to come up with musically and historically sensitive formal definitions of our research objects (i.e., schemata); and hypotheses concerning their historical use are to be tested on a broader empirical basis. Furthermore, hermeneutics and computer-assisted music analysis share a common concern for uncertainty which arises at all stages of the research agenda including data production, analysis, and interpretation. In our project, uncertainty arises in terms of schema identification and classification.

Finally, 'distant reading' has the potential of providing important insights on which detailed individual case studies can additionally be based. The digital representation of musical objects allows scholars to systematically scrutinize and compare different degrees of granularity and their information content in relation to some prior hypotheses and research aims. Qualitative hermeneutic questions can thus be raised to a new scientific level by corpus research without losing its own legitimacy.

More generally, we are currently in the midst of a digital revolution that may prove to be a major turning point in human society. The amount of music available and the historical scale that can now be explored are much higher than in previous, more traditional research. Computational modelling brings musicological research to an entirely new and unprecedented level. But computational musicology, being still in its early phase of development (despite the use of electronic and digital tools for music for a long time), has yet to unfold its enormous potential. At present, one of the biggest challenges in the field is the synthesis between traditional, analytical, and historical methods, not to mention the new methods and tools coming from digital research in order to reshape the discipline of musicology for the future. Exploration of the many possibilities of interaction between quantitative and qualitative methods in an integrative 'mixed methods' framework remains one of the urgent tasks of future research.

Bibliography

Allegraud, Pierre, Louis Bigo, Laurent Feisthauer, Mathieu Giraud, Richard Groult, Emmanuel Leguy and Florence Levé. "Learning Sonata Form Structure on Mozart's String Quartets", In *Transactions of the International Society for Music Information Retrieval* 2, 1 (2019): 82–96.

Antila, Christopher and Julie Cumming. "The VIS Framework. Analyzing Counterpoint in Large Datasets". In *Proceedings of the 15th International Society for Music Information Retrieval Conference*, ISMIR, edited by Hsin-Min Wang, Yi-Hsuan Yang and Jin Ha Lee, 71–76, Taipei, Taiwan, 2014.

Bergeron, Mathieu and Darrell Conklin. "Subsumption of Vertical Viewpoint Patterns". In *International Conference on Mathematics and Computation in Music*, 1–12, Springer: Berlin/Heidelberg, 2011.

Byros, Vasili. "Towards an 'Archaeology' of Hearing: Schemata and Eighteenth-century Consciousness". In *Musica Humana 1*, 2 (2009): 235–306.

Byros, Vasili. "Trazom's Wit: Communicative Strategies in a 'Popular' Yet 'Difficult' Sonata". In *Eighteenth-Century Music 10*, 2 (2013): 213–252.

Caplin, William E. "Topics and Formal Functions: The Case of the Lament". In *The Oxford Handbook of Topic Theory*, edited by Danuta Mirka, 415–452. Oxford: Oxford University Press, 2014.

Conklin, Daniel and Mathieu Bergeron. "Discovery of Contrapuntal Patterns". In *Proceedings of the International Society on Music Information Retrieval*, 2010, 201–206.

Dittrich, Marie-Agnes. "'Teufelsmühle' und 'Omnibus'". In *Zeitschrift der Gesellschaft für Musiktheorie* 4, 1–2 (2007): 107–121. https://doi.org/10.31751/247

Duane, Ben. "Melodic Patterns and Tonal Cadences: Bayesian Learning of Cadential Categories from Contrapuntal Information". Journal of New Music Research 48, 3 (2019): 197–216.

Finkensiep, Christoph, Markus Neuwirth and Martin Rohrmeier 2018. "Generalized Skipgrams for Pattern Discovery in Polyphonic Streams". In *Proceedings of the 19th International Society for Music Information Retrieval Conference* (ISMIR), Paris 2018, edited by Emmanouil Benetos, Emilia Gómez, Xiao Hu and Eric Humphrey, 547–553.

Finkensiep, Christoph, Markus Neuwirth and Martin Rohrmeier. "Music Theory and Model-Driven Corpus Research". In *Oxford Handbook of Music and Corpus Studies*, edited by Daniel Shanahan, Ashley Burgoyne and Ian Quinn, New York: Oxford University Press. (forthcoming)

Finkensiep, Christoph, Ken Déguernel, Markus Neuwirth and Martin Rohrmeier. "Voice-Leading Schema Recognition Using Rhythm and Pitch Features". In *Proceedings of the International Society for Music Information Retrieval* (ISMIR), Montreal 2020, edited by Julie Cumming, Jin Ha Lee, Brian McFee et al., 520–526.

Flanders, Julia and Fotis Jannidis (eds.). *The Shape of Data in Digital Humanities: Modeling Texts and Text-based Resources*. London: Routledge, 2018.

Giraud, Mathieu, Ken Déguernel and Emilios Cambouropoulos. "Fragmentations with Pitch, Rhythm and Parallelism Constraints for Variation Matching". In *International Symposium on Computer Music Multidisciplinary Research* (CMMR), 2013, 298–312.

Gjerdingen, Robert O. *Music in the Galant Style*, New York: Oxford University Press, 2007.

Hentschel, Johannes, Markus Neuwirth and Martin Rohrmeier. "The Annotated Mozart Sonatas: Score, Harmony and Cadence". In *Transactions of the International Society for Music Information Retrieval* 4, 1 (2021): 67–80. http://doi.org/10.5334/tismir.63.

Jan, Steven. "Using Galant Schemata as Evidence for Universal Darwinism". In *Interdisciplinary Science Reviews* 38, 2 (2013): 149–168.

Katsiavalos, Andreas, Tom Collins and Bret Battey. "An Initial Computational Model for Musical Schemata Theory". In *Proceedings of the International Society on Music Information Retrieval*, 2019, 166–172.

Meredith, David, Kjell Lemström and Geraint A. Wiggins. "Algorithms for Discovering Repeated Patterns in Multidimensional Representations of Polyphonic Music". In *Journal of New Music Research 31*, 4 (2002): 321–345. DOI:10.1076/jnmr.31.4.321.14162.

Meyer, Leonard B. "Innovation, Choice and the History of Music". In *Critical Inquiry* 9, 3 (1983): 517–544.

Moss, Fabian C., 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), e0217242.

Mossburger, Hubert. Ästhetische Harmonielehre: Quellen, Analysen, Aufgaben, vol. II, Wilhelmshaven: Noetzel, 2012.

Neuwirth, Markus and Martin Rohrmeier. "Wie wissenschaftlich muss Musiktheorie sein? Chancen und Herausforderungen musikalischer Korpusforschung". In *Zeitschrift der Gesellschaft für Musiktheorie 13*, 2 (2016): 171–193.

Piotrowski, Michael and Markus Neuwirth (2020). "Prospects for Computational Hermeneutics". In *Proceedings of the 9th Annual Conference of the AIUCD* (Milano, Jan. 15–17, 2020), edited by Cristina Marras and Marco Passarotti, Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD), 204–209.

Rice, John. "The Morte: A Galant Voice-Leading Schema as Emblem of Lament and Compositional Building-Block". In *Eighteenth-Century Music 12*, 2 (2015): 157–181.

Rohrmeier, Martin. "On Creativity, Music's AI Completeness, and Four Challenges for Artificial Musical Creativity". In *Transactions of the International Society for Music Information Retrieval 5*, 1 (2022): 50–66. DOI: http://doi.org/10.5334/tismir.104

Rumelhart, David E. “Schemata: The Building Blocks of Cognition”. In *Theoretical Issues in Reading Comprehension*, edited by Rand J. Spiro, Bertram C. Bruce and William F. Brewer, 33–58, Hillsdale, NJ: Lawrence Erlbaum, 1980.

Schank, Roger and Robert P. Abelson. *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures*, New York: Halsted, 1977.

Schwenkreis, Markus (ed.), *Compendium Improvisation—Fantasieren nach historischen Quellen des 17. und 18. Jahrhunderts*, Basel: Schwabe, 2018.

Sears, David R., Andreas Arzt, Harald Frostel, Reinhard Sonnleitner and Gerhard Widmer. “Modeling Harmony with Skip-Grams”. In *Proceedings of the International Society on Music Information Retrieval*, 2017, 332–338.

Sears, David R., Marcus T. Pearce, William E. Caplin and Steven McAdams. “Simulating Melodic and Harmonic Expectations for Tonal Cadences Using Probabilistic Models”. In *Journal of New Music Research* 47, 1 (2018): 29–52.

Sears, David R. and Gerhard Widmer. “Beneath (or Beyond) the Surface: Discovering Voice-leading Patterns with Skip-grams”. In *Journal of Mathematics and Music* 15, 3 (2020): 1–26. DOI: 10.1080/17459737.2020.1785568.

Symons, James. “A Cognitively Inspired Method for the Statistical Analysis of Eighteenth-Century Music, as Applied in Two Corpus Studies”. PhD Thesis, Northwestern University, 2017.

Telesco, Paula J. “Enharmonicism and the Omnibus Progression in Classical-Era Music”. In *Music Theory Spectrum* 20, 2 (1998): 242–279.

Weiß, Christof, Frank Zalkow, Vlora Arifi-Müller, Meinard Müller, Hendrik Vincent Koops, Anja Volk and Harald G. Grohganz. “Schubert Winterreise Dataset: A Multimodal Scenario for Music Analysis”. In *ACM Journal on Computing and Cultural Heritage (JOCCH) 14*, no. 2 (2021): 1–18.

Weyde Tillman and Reinier de Valk. “Chord- and Note-based Approaches to Voice Separation”. In *Computational Music Analysis*, edited by David Meredith, Cham: Springer, 2016, 137–154.