DISCOURSE NOT DUALISM: AN INTERDISCIPLINARY DIALOGUE ON SONATA FORM IN BEETHOVEN'S EARLY PIANO SONATAS

Christof Weiß[1], Stephanie Klauk[2], Mark Gotham[2], Meinard Müller[1], Rainer Kleinertz[2]

[1] International Audio Laboratories Erlangen, Germany
[2] Institut für Musikwissenschaft, Saarland University, Germany
`christof.weiss@audiolabs-erlangen.de`

ABSTRACT

The computational analysis of music has traditionally seen a sharp divide between the "audio approach" relying on signal processing and the "symbolic approach" based on scores. Likewise, there has also been an unfortunate gap between any such computational endeavour and more traditional approaches as used in historical musicology. In this paper, we take a step towards ameliorating this situation through the application of a computational method for visualizing local key characteristics in audio recordings. We exploit these visualizations of diatonic scale content by discussing their musicological implications, being aware of methodological limitations as for the case of minor keys. As a proof of concept, we use this method for investigating differences between the traditional sonata-form model and selected Beethoven piano sonatas in the context of sonata theory from the end of the 18th century. We consider this scenario as an example for a rewarding dialogue between computer science and historical musicology.

1. INTRODUCTION

The analysis of musical works in terms of their compositional style and context is at the core of historical musicology. Scholars engage with a reasonably large body of works over the course of their career and make observations about these works through manual analysis. As valuable as this is, it is a time-consuming and individuated approach that poses difficulties for making meaningful observations at scale. In this paper, we seek to demonstrate how a question of musicological relevance could be complemented by using computational analysis methods. While such methods can never achieve the flexible and interconnected consideration of the human mind nor capture the compositional intricacies of a specific work, we aim to show that suitable visualizations can assist in the process of expert interpretation in a rewarding way.

Numerous computational methods for harmony analysis have been developed over the past decades, centered on global key detection [1–4], local key estimation [5–7], chordal analysis [8–10], and their combination into functional harmony analysis systems [11–13]. Many of these are based on *symbolic* encodings of music notation such as MIDI or MusicXML. However, symbolic music datasets are rarely available, in particular when requiring symbolic encodings of high quality covering an entire corpus of music (not just individual pieces). Optical Music Recognition (OMR) software for automatically converting graphical formats into symbolic data does not yet offer reliable results, meaning that time-consuming manual postprocessing is often required [14, 15].

Beyond such practical problems, we should remember that Western music notation is essentially "prescriptive": a set of instructions for performance that requires reading and interpretation. As an alternative to the processing of sheet music, analyses can be carried out on the basis of audio recordings [4, 7, 10]. In this paper, we make use of an existing audio-based method [16] for visualizing the diatonic scale content of a music recording over time in order to complement and facilitate the close analytical reading by human experts. We address this paper to both musicologists and computer scientists alike and thus explain the relevant background from both domains.

We first set out the musicological context (Section 2), describe the computational method (Section 3), and explain its musical implications through the example of three piano sonatas by L. v. Beethoven (Section 4), which have been subject to automatic analysis [9, 12, 17]. Using our visualizations, we discuss the implications for sonata form theories (Section 5). Finally, we summarize our findings and the rewards of an interdisciplinary dialogue between computer science and musicology (Section 6).

2. MUSICOLOGICAL BACKGROUND

Our focus in this paper is on large-scale tonal structures that are attested to play a crucial role in sonata form. This section provides a short, simplified introduction to sonata form for readers unfamiliar with it and serves to introduce the main musicological question addressed in this paper.

Sonata form is a central model for describing the first movement of most multi-movement works in classical music from about 1770 far into the 19th century. The term is applicable not only to *sonatas* but also to symphonies, string quartets, and other genres, and thus has a wide explanatory potential for the music of the time, with a focus on these works' first movements.

The definition of sonata form widely adopted by musicians and musicologists seems to stem primarily from the writings of A. B. Marx [18], which were driven mainly by an effort to understand the music of then recently deceased Ludwig van Beethoven (1770–1827). According to this model, the sonata is divided into three main sections: a first section, later denoted as *exposition* (in two or more keys, often repeated), a central *development*, and a *recapitulation* of the exposition (set mainly in one key). This is the core structure which may be framed by an introduction at the start, and/or a coda at the end [19]. Automatically detecting these large-scale segments and their tonal relations has been approached both for Mozart's string quartets [20, 21] and Beethoven's piano sonatas [17].

The exposition is tasked with setting out the melodic material (e.g., themes or motives) and the main key relationships. It is usually divided into a first subject area, followed by a transition into a second subject area, and a short cadential passage (*Schlussgruppe*). Crucially, the two focal areas are defined by contrasts both in theme (melody) and key (tonality). The typical tonal pairing is of a major key and its dominant key (the major key one fifth higher, e.g., C major and G major), or of a minor key with its relative major key (e.g., A minor and C major).

Our primary focus will be on key, which we approximate in the broader sense with our visualizations of diatonic pitch class content. With this computational method, we elucidate the tonal relations in specific early Beethoven piano sonatas. In this paper, we focus on works in major keys overall and on the exposition sections of those works.

3. COMPUTATIONAL APPROACH

In this work, we employ a computational method for visualizing local keys or, more precisely, the *diatonic pitch class content* over the course of a piece, closely following the approach proposed in [16]. The method operates on audio recordings, i.e., *performances* of the pieces, thus allowing for scalability to a wide repertoire (see Section 1).

3.1 Chroma-based Scale Estimation

Our method is based on the measurement of spectral energies over time. These energies are summarized into twelve *chroma bands* irrespective of their octave, according to the pitch classes of the twelve-tone equal-tempered scale. The resulting *chroma features* can be represented as twelve-dimensional vectors whose entries refer to the pitch classes C, C♯, ..., B in chromatic order. For chroma extraction, we use the filter-bank approach provided by the chroma toolbox [22] with a feature rate of 10 Hz (i.e., ten chroma vectors per second). For each frame, we match the chroma

vector with binary diatonic scale templates using the inner product (cosine similarity). For instance, the template for the "0 diatonic scale" (corresponding to the pitch classes of the C major and A natural minor scales) is given by

$$\mathbf{t}_0 = (1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1)^\top . \qquad (1)$$

Assuming enharmonic equivalence (C♯=D♭), there are a total of 12 diatonic scales, whose templates are obtained by circularly shifting the template $\mathbf{t}_0$ shown above. Thus, we obtain for each frame an analysis given by a twelve-dimensional vector. We will discuss our choice of using binary diatonic templates in Section 3.4.

3.2 Pre-processing

Before the template matching step described above, we apply several pre-processing steps. Since local keys or scales refer to the pitch class content of larger sections of music, we smooth the chromagrams (with an initial feature rate of 10 Hz) using a window of size $w \in \mathbb{N}$ in frames and a hopsize of 10 frames (one second). The musical implications of the window size parameter will be discussed in Section 4.3. Additionally, we normalize the smoothed chroma features according to the ℓ_2-norm.

3.3 Post-processing and Visualisation

In this interdisciplinary work, we do not aim for a fully automatic "key detection," which locally decides on the most likely key or scale. Instead, following [16], we propose the use of suitable visualization techniques allowing for a direct interpretation of the continuous-valued diatonic scale probabilities in the musicological discussion. For generating this visualization, we obtain re-scaled local analyses by using the *softmax* function, thus suppressing weak components and enhancing large ones. Using a normalization with respect to the ℓ_1-norm, we can interpret the analysis as pseudo-probabilities of diatonic scales, which we then visualize in grayscale, where darker gray corresponds to higher probabilities (see Figure 1 for an example).

We adopt a musical criterion for arranging the order of scales in this visualisation. There are two main options of ordering the scales, either chromatically (C, C♯, ...) or according to the circle of fifths (C, G, ...). Motivated by [23, 24], we prefer the latter arrangement accounting for the similarity of fifth-related scales, which have six out of seven pitch classes in common. We set the diatonic scale corresponding to the piece's global key in the center of the visualization, with upper-fifth-related scales (more sharps) above and lower-fifth-related scales (more flats) below that center scale. This "global key normalization" facilitates the comparison between movements in different keys.

3.4 Scale versus Local Key

Our analytical approach directly maps the locally predominant pitch class content (as measured from the audio recordings) to probabilities for diatonic scales. While there have been many such template-based approaches based on psychological and empirical studies [25, 26], our choice
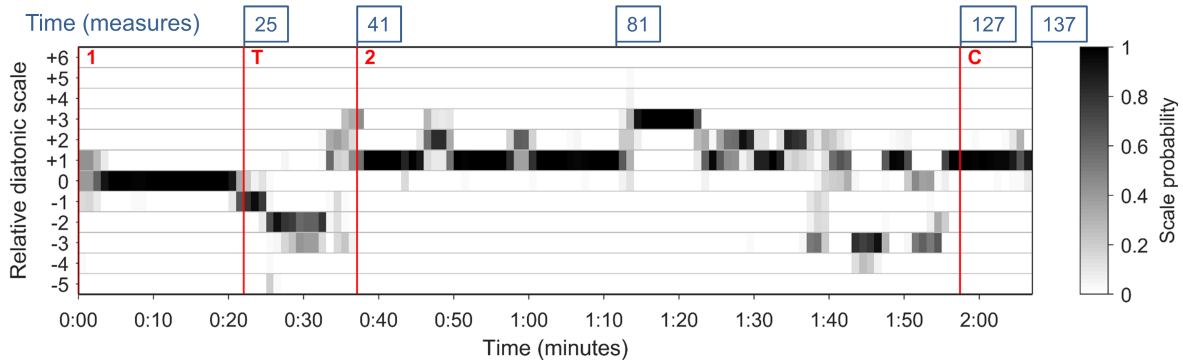
Figure 1. L. v. Beethoven, piano sonata Op. 7 in E♭ major, 1st mvmt. *Allegro molto e con brio*, exposition. Computational tonal analysis with a window size of $w = 4$ seconds.

of straightforward diatonic templates allows for an interpretable and objective investigation of the tonal content.

There are certainly limitations to this approach. Most importantly, it relates to the notion of local key only in a loose way since there is a methodological gap between "scale" and "key." In particular, pitch class content is not the only determinant of musical key. Furthermore, the pitch class content is rarely exclusively diatonic, which is especially the case for pieces in minor, where scale degrees outside the natural minor scale (♯$\hat{6}$ and ♯$\hat{7}$) play a crucial role. Moreover, our method cannot resolve relative key differences such as C major – A minor, which would be relevant particularly for minor key movements. [1] On the other hand, it provides an easier overview (only 12 scales) and is not susceptible to relative key confusions.

Leaving aside these methodological gaps, even identifying a scale can be problematic. For example, certain chords such as the (relatively rare) augmented triad and the (not at all rare) diminished seventh chord pose challenges because they cannot be unambiguously assigned to a diatonic scale. Moreover, frequent modulations constitute a problem for assigning a single scale to a specified time window. That being said, this computational issue reflects a genuine musical problem: hearing a tonal moment in isolation would leave the allocation of a key highly ambiguous—and composers (Beethoven very much included) exploit this potential for ambiguity. It is only through context that we are able to make clear assertions. We remain mindful of these limitations as we proceed to consider how these visualizations may yet be useful in support of a better understanding of sonata form in general and, more precisely, in Beethoven's early piano sonatas.

4. APPLICATION TO SONATA EXPOSITIONS

Consequently, we apply the method described above to several exposition sections of Beethoven's early sonatas (first movement, respectively). The time stamps are given in MM:SS and refer to Daniel Barenboim's 1984 set of recordings for Deutsche Grammophon.

---

[1] We provide an overview of all 28 first movements in sonata form (including minor key examples) on `www.audiolabs-erlangen.de/resources/MIR/BeethovenSonataAnalyses`

4.1 Sonata Op. 7 in E♭ major

Figure 1 provides an example of our visualization method for the exposition from Beethoven's piano sonata Op. 7 in E♭ major, first movement. As E♭ major is the global key of this movement, the relative diatonic level 0 on the y-axis of this figure refers to an absolute scale of –3, corresponding to the key signature of E♭ major and its relative minor (C minor, both with a key signature of three ♭s). Likewise, +1 refers to the key signature of B♭ major (and G minor), –1 stands for A♭ major (and F minor), and so on.

In addition to the time stamps, the figure (and all subsequent figures) also provide vertical lines (in red) to divide the sections on the basis of Donald F. Tovey's iconic guide to Beethoven's piano sonatas [27]. These lines add additional score-related timestamps given in measure numbers (in blue) and are paired with the formal labels that Tovey used, following the standard terms for the sonata form's main parts discussed above. The vertical lines on Figure 1 use the following abbreviations: "1" stands for the *first group*, "T" for the *transition*, "2" for the *second group*, and "C" for the *cadence group*. The following section takes a closer look at the example with a view to both the visualization method and the implications for understanding sonata form in this specific repertoire.

4.2 A Closer Look

The visualization in Figure 1 provides a straightforward, easily readable overview of the diatonic scale content in the exposition of the first movement. The movement begins with a moment of ambiguity, centered on the central (0) scale (corresponding to E♭ major), and quickly settles much more emphatically into that tonal region for the first 22 seconds (until m. 24). A phase of tonal instability follows, befitting the transition phase of a sonata form exposition, which traverses at least the scales –1 (4♭ or A♭ major / F minor) to –2 (5♭), and then +1 (2♭) to +3 (no accidentals).

At 0:37 (m. 41), a longer section starts in the +1 area, neatly corresponding to what is traditionally called the "Second Theme" or "Second Group." At 1:11 (m. 79), the proportions (with a second subject equal in length to the first) might lead us to expect the exposition to close. Instead, about a minute more music follows, with only the last approx. 10 seconds being displayed in the anticipated

+1 area (mm. 127–136, largely corresponding to the final group). The course of the movement from 1:11 to 1:57 (mm. 79–127), by contrast, seems to be characterized by greater harmonic mobility, centered on the +1 axis, but not restricted to it. Regardless of any designation of the formal parts, the graphic shows a clear distinction in the course of the exposition from 0:38 (m. 41) into a tonally stable part until 1:11 (mm. 78–79), a more flexible part until 1:57 (m. 127), and again a stable one until the end.

4.3 Effects of the Window Size

As one of the essential properties of our methodology, we have to specify the window size parameter w, which defines the temporal context of the local scale analysis. The use of a window size of four seconds in Figure 1 affects the result we see in that visualization. In order to demonstrate the effect of this parameter, Figure 2a–e shows five different visualizations of the same example with window sizes varying from 2–20 seconds.

Naturally, shorter time windows portray a more "atomized" harmonic course, emphasizing the moment-to-moment details, while longer windows diminish the individual moments in favor of the "bigger picture." Any gain in uniformity and clarity comes at the cost of a corresponding loss in detail. In this respect, a computational analysis does not differ from a manual one, in which the analyst also has to decide whether to focus on fine-grained details or to favour better readability and clarity. This speaks to music theory's attention to the "level" of reductive analysis [28]—an approach to which the variable window size on offer here may be highly suitable. Multi-scale approaches for simultaneously using several window sizes suggest an alternative [29]. However, these visualizations require a third dimension (usually color-coded), which complicates readability in our application scenario. Moreover, interactive visualizations with a flexible adjustment of the window size can be realized with user interfaces or websites.

To better illustrate the behaviour of our method, we now proceed to discuss the effects of the window size at the example of several specific passages. With a window size of four seconds (Figure 1), the dominant seventh chord G-B♭-D♭-E♭ of A♭ major in m. 10 (at 0:08) does not lead to any deviation from the level 0. With a resolution of two seconds (Figure 2a), it causes a much stronger gray coloration, which reaches down to the –3 level (corresponding to G♭ major / E♭ minor). Similarly, the diminished seventh chord F♯-A-C-E♭ in mm. 79–89 (1:11–1:22) triggers a whole range of possible interpretations. First, the chord eludes the assignment to a single diatonic scale, and second, through the pitch class F♯, it does not fit into the previous B♭ major context (–2), nor through F♯ and E♭ into the subsequent C major context (0). In the case of longer windows, these uncertainties are smoothed out and therefore no longer catch the eye (see Figure 2d–e).

An even larger window size, e. g. of 20 seconds (Figure 2e), shows that only certain resolutions make sense for a specific work. Such large windows neither increase clarity nor do all the gray shadings of ambiguity disappear. In
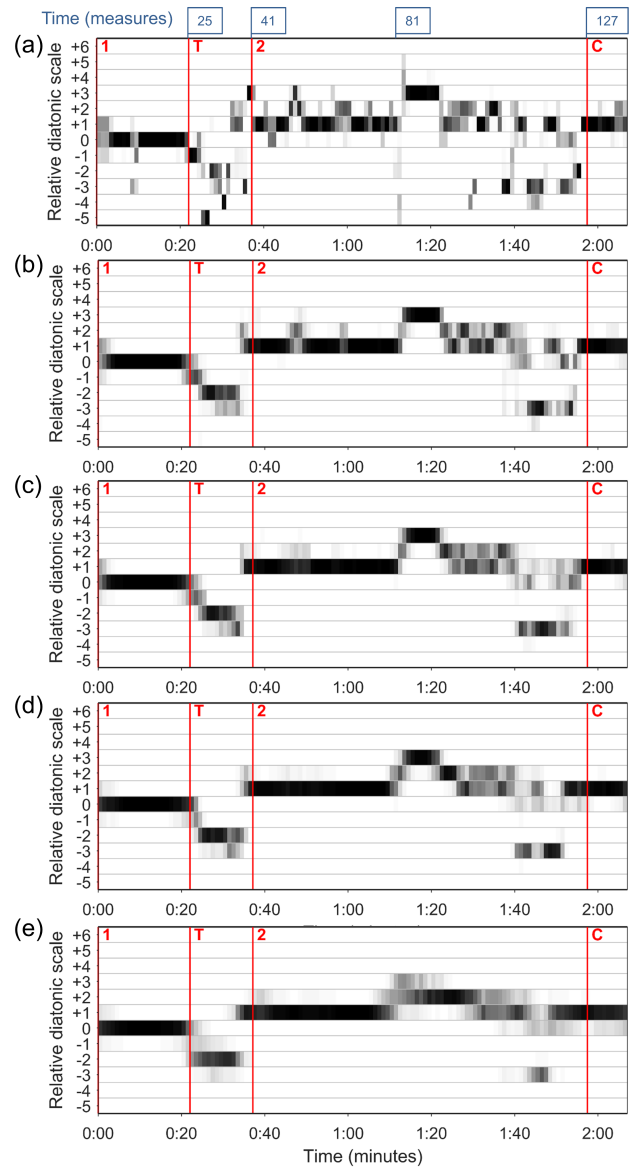

Figure 2. L. v. Beethoven, piano sonata Op. 7 in E♭ major, 1st mvmt. *Allegro molto e con brio*, exposition. Computational tonal analysis with a window size w of (a) 2 sec., (b) 6 sec., (c) 8 sec., (d) 12 sec., (e) 20 sec.

Figure 2e, the C major section in mm. 81ff. (1:12–1:22) is no longer visible as +3 (C major / A minor); instead, it blurs with the +2 level. At the same time, the tonally stable final group (+1), which only lasts about 10 seconds (mm. 127–136), shows considerably more shades of gray on the 0 level than visualizations with shorter windows, which is caused by the previous measures and the subsequent repetition of the exposition. The choice of very large windows not only leads to an increased smoothing of the visualization but also reaches limits beyond which the results are no longer meaningful.

In the present sonata, window sizes of 4–12 seconds have proven to be useful, illuminating most of the relevant phenomena. In particular, these visualizations clearly highlight phases of tonal stability and instability as well as diatonic regions, which are of high importance for the for-
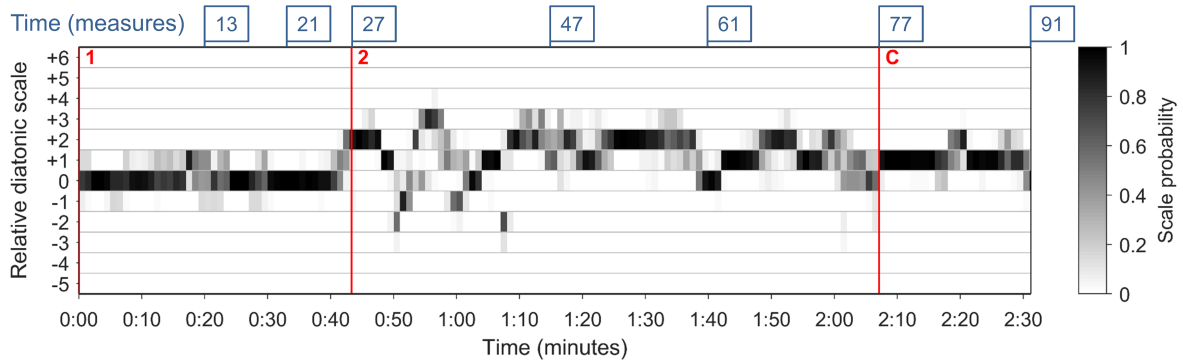
Figure 3. L. v. Beethoven, piano sonata Op. 2 No. 3 in C major, 1st mvmt. *Allegro con brio*, exposition. Computational tonal analysis with a window size of $w = 4$ seconds.

Figure 4. L. v. Beethoven, piano sonata Op. 2 No. 3 in C major, 1st mvmt. *Allegro con brio*, mm. 25–28.

mal organization of expositions. Based on these observations, we now examine two further sonata expositions that have received frequent attention in the literature: sonatas Op. 2 No. 3 in C major and Op. 10 No. 3 in D major.

4.4 Sonata Op. 2 No. 3 in C major

Spanning 90 measures, the exposition of the sonata Op. 2 No. 3 in C major is even more extensive than that of Op. 7. The computational analysis (Figure 3) shows that the piece remains in a C major context for an unusually long time. These first 26 measures include the main subject (mm. 1–13) followed by playful figurations, which surprisingly do not modulate (mm. 13–21), and a cadence passage (mm. 21–26), which ends after a G major scale on the single tone G without having modulated (see Figure 4). [2] From m. 27 (0:43) on, this is followed by the melodic motif in G minor mentioned above, which is repeated in D minor (m. 33) and continued towards A minor (m. 39).

The visualization makes it clear that this G minor passage is by no means the beginning of the "Second Group (or Transition and Second Group) in Dominant"—as marked in Figure 3 after Tovey—but the beginning of the modulation to the upper-fifth key, i. e., the "transition." First of all, the abrupt tonal change in m. 27 (0:43) is visible. [3] However, the visualization does not show G minor (−2) but D major (+2), which may be caused by the frequent occurrence of the leading notes F♯ and C♯. In the following modulation, we observe level 0 at 1:03 (mm. 39ff., pointing to in A minor), then again level +2 at 1:09-–1:14 (mm. 43–45, pointing to D major), thereby terminating the transition to the second group (mm. 47–61; 1:15–1:40). The second group starts at level +1, then from m. 53 (1:26) on mainly represented as +2 due to several neighbor notes

C♯ and the secondary dominant chord A major sounding for a whole measure. This is followed by a longer cadence section, initially at the 0 level (C major) from 1:40 (mm. 61ff.) and then confirms level +1 with smaller swings to the levels +2 and 0. The largely stable level +1 in the final group is then clearly visible (mm. 77–90).

4.5 Sonata Op. 10 No. 3 in D major

The visualization of the sonata Op. 10 No. 3 gives a completely different picture (Figure 5). After about 20 seconds (m. 22), the initial tonality (D major, level 0) is left, followed by a longer section on the level +1 (mm. 23–45). In fact, we find here a lyrical motif in B minor, which actually corresponds to level 0. Due to the frequent occurrence of the leading note A♯, however, there seems to be a kind of "statistical averaging" between B minor (0) and B major (+3). The occurrence of C♯ major and F♯ minor in mm. 31–34 leads to a small shade of gray in the +4 level at 0:31. The second group in A major is then reached at 0:48 (m. 54). The repetition of the second subject in −2 (pointing to A minor) is briefly visible at 0:54 (mm. 60–63). It is then striking that the second group again ends more or less in the middle of the movement and is followed by a longer developmental passage at 1:00–1:18 (mm. 67–87).

5. MUSICOLOGICAL DISCUSSION

Section 4 examined the use of this visualization method for illuminating the details of three exemplary cases. We now broaden the scope to consider its potential contribution to the wider question of sonata form itself.

The visualizations discussed above have shown a multi-faceted structure within the tonal course of the exposition. The plots point out that the musical course cannot be unproblematically reconciled with the traditional sonata form schema of *first group – transition – second group –*
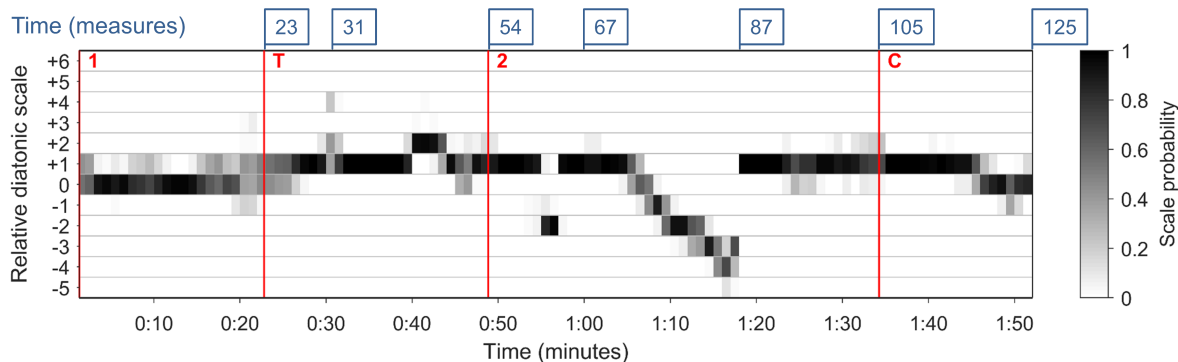
---

[2] This ending constitutes a clear example of a so-called "bifocal close" [30].

[3] See also [17] for a method to highlight such changes.

Figure 5. L. v. Beethoven, piano sonata Op. 10 No. 3 in D major, 1st mvmt. *Presto*, exposition. Computational tonal analysis with a window size of $w = 4$ seconds.

cadence group [19], a problem raised by Carl Dahlhaus decades ago [31, p. 101–103]. This leads us to reconsider other, earlier theories of sonata form from the late 18th century which emphasize the idea of a musical *discourse* as opposed to a thematic *dualism*.

A detailed description of such an 18th-century form model is given by Francesco Galeazzi in the second volume of his *Elementi teorico-pratici di musica* (Rome 1796) [32]. Galeazzi's model differs from the theory more familiar today in two principal respects: the presence of a contrasting *second motif* before or at the beginning of the transition and a *cadential period* ahead of what he calls the *coda* (i. e., a codetta or final group).

For the section we now denote as *exposition* (which Galeazzi simply calls *prima parte*), Galeazzi sets out an alternative schema of seven elements [32, p. 324]:

1. Prelude (*preludio*)
2. Principal motive (*motivo*)
3. Second motive (*secondo motivo*)
4. Departure to [...] related keys (*uscita di tono*)
5. Characteristic passage / middle passage (*passo caratteristico / passo di mezzo*)
6. Cadential period (*periodo di cadenza*)
7. Codetta (*coda*)

Of these seven parts, according to Galeazzi [32], parts 2, 4, and 6 (in modern terms: *first key subject*, *transition*, and *cadential confirmation of the second key*) are compulsory, the remainder are optional. Accordingly, Galeazzi bases his model on a main motif that dominates the musical discourse like the topic of a speech and that can be followed by a whole series of new, partly related thoughts:

> "The motive [...] must be very conspicuous and perceptible because inasmuch as it is the theme of the discourse, if it is not well understood, neither will the consecutive discourse be understood." [32, pp. 326–327] [4]

If one starts from such a discursive form idea, the occurrence of a new thought before or in the transition (e. g.

---

[4] Galeazzi does not at all favor a monothematic structure as found sometimes in the works of Joseph Haydn. The model described by him is open to new motives in any part of the exposition, not just in the sense of two contrasting themes as modelled by Marx [18].

Op. 2 No. 3 mm. 27–39, Op. 10 No. 3 mm. 23–30) and of a longer form part after the "middle sentence" (Op. 2 No. 3 mm. 61–77, Op. 7 mm. 81–127, Op. 10 No. 3 mm. 67–105) in no way leads into "insoluble theoretical difficulties" [31]. Even though some detailed structures might differ from a human analysis, the computational visualizations show something that was taken for granted for composers and audiences of the late 18th century: the individual formal parts are not opposed in a *dualistic* tension, but rather formed a series, where uniformity and diversity, tonal stability and modulation are combined into a living, *discursive* whole. Beethoven's early piano sonatas—differing considerably from the traditional model as codified by Marx—fit perfectly in this context [33]. Beethoven as well as Mozart before him [30] seem to have been influenced by Italian music. Galeazzi's account of sonata form—based on Italian composers of the 1770s and 1780s (Mozart and Beethoven were unknown to him)—points towards such an understanding as reflected in our visualizations.

6. CONCLUSIONS AND OUTLOOK

In this paper, we have demonstrated the use of a computational analysis system for shedding new light on a research question at the heart of historical musicology. The method relies on audio recordings and visualizes the diatonic scale content of a piece in an objective and interpretable way, providing an easy, at-a-glance insight into the phases of stability, instability, and tonal transition. Being aware of several alternative analysis strategies, we plan to work on a closer interrogation of the method, its relation to local key analysis, and the comparison of the output to systematic human analyses as provided by [12, 34, 35].

Even though any type of automated approach can never achieve the flexibility of human analysis, we have shown that it can provide an overview of large-scale structures, thus aiding the research process of historical musicology. Since this approach can be scaled up easily without requiring human annotations [36], it allows for corpus studies in a novel order of magnitude, which can enrich musicological research. In future work, we thus intend to apply this method to a wider range of musical contexts involving extensive corpora and individual large-scale works, both of which would benefit from these "at-a-glance" reductions.

7. REFERENCES

[1] S. R. Holtzman, "A program for key determination," *Interface*, vol. 6, no. 1, pp. 29–56, 1977.

[2] D. Temperley, "An algorithm for harmonic analysis," *Music Perception*, vol. 15, no. 1, pp. 31–68, 1997.

[3] S. T. Madsen and G. Widmer, "Key-finding with interval profiles," in *Proceedings of the International Computer Music Conference (ICMC)*, Copenhagen, Denmark, 2007, pp. 212–215.

[4] B. Schuller and B. Gollan, "Music theoretic and perception-based features for audio key determination," *Journal of New Music Research*, vol. 41, no. 2, pp. 175–193, 2012.

[5] C. L. Krumhansl and E. J. Kessler, "Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys," *Psychological Review*, vol. 89, no. 4, pp. 334–368, 1982.

[6] T. Rocher, M. Robine, P. Hanna, and L. Oudre, "Concurrent estimation of chords and keys from audio," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Utrecht, The Netherlands, 2010, pp. 141–146.

[7] H. Papadopoulos and G. Peeters, "Local key estimation from an audio signal relying on harmonic and metrical structures," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 4, pp. 1297–1312, 2012.

[8] T. Fujishima, "Realtime chord recognition of musical sound: A system using common lisp music," in *Proceedings of the International Computer Music Conference (ICMC)*, Beijing, China, 1999, pp. 464–467.

[9] V. Konz, M. Müller, and R. Kleinertz, "A cross-version chord labelling approach for exploring harmonic structures—a case study on Beethoven's Appassionata," *Journal of New Music Research*, vol. 42, no. 1, pp. 61–77, 2013.

[10] J. Pauwels, K. O'Hanlon, E. Gómez, and M. B. Sandler, "20 years of automatic chord recognition from audio," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Delft, The Netherlands, 2019, pp. 54–63.

[11] P. R. Illescas, D. Rizo, and J. M. Iñesta, "Harmonic, melodic, and functional automatic analysis," in *Proceedings of the International Computer Music Conference (ICMC)*, Copenhagen, Denmark, 2007, pp. 165–168.

[12] T. Chen and L. Su, "Functional harmony recognition of symbolic music data with multi-task recurrent neural networks," in *Proceedings of the 19th International Society for Music Information Retrieval Conference (ISMIR)*, Paris, France, 2018, pp. 90–97.

[13] G. Micchi, M. Gotham, and M. Giraud, "Not all roads lead to rome: Pitch representation and model architecture for automatic harmonic analysis," *Transactions of the International Society for Music Information Retrieval (TISMIR)*, vol. 3, no. 1, pp. 42–54, 2020.

[14] D. Byrd and J. G. Simonsen, "Towards a standard testbed for optical music recognition: Definitions, metrics, and page images," *Journal of New Music Research*, vol. 44, no. 3, pp. 169–195, 2015.

[15] I. Fujinaga, A. Hankinson, and L. Pugin, "Automatic score extraction with Optical Music Recognition (OMR)," in *Springer Handbook of Systematic Musicology*. Springer, 2018, pp. 299–311.

[16] C. Weiß and J. Habryka, "Chroma-based scale matching for audio tonality analysis," in *Proceedings of the Conference on Interdisciplinary Musicology (CIM)*, Berlin, Germany, 2014, pp. 168–173.

[17] N. Jiang and M. Müller, "Automated methods for analyzing music recordings in sonata form," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Curitiba, Brazil, 2013, pp. 595–600.

[18] A. B. Marx, *Die Lehre von der musikalischen Komposition [The School of Musical Composition]*. Leipzig, Germany: Breitkopf und Härtel, 1837–1847.

[19] J. Hepokoski and W. Darcy, *Elements of Sonata Theory. Norms, Types, and Deformations in the Late-Eighteenth-Century Sonata*. Oxford University Press, 2006.

[20] L. Bigo, M. Giraud, R. Groult, N. Guiomard-Kagan, and F. Levé, "Sketching sonata form structure in selected classical string quartets," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Suzhou, China, 2017, pp. 752–759.

[21] P. Allegraud, L. Bigo, L. Feisthauer, M. Giraud, R. Groult, E. Leguy, and F. Levé, "Learning sonata form structure on Mozart's string quartets," *Transactions of the International Society for Music Information Retrieval (TISMIR)*, vol. 2, no. 1, pp. 82–96, 2019.

[22] M. Müller and S. Ewert, "Chroma Toolbox: MATLAB implementations for extracting variants of chroma-based audio features," in *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Miami, Florida, USA, 2011, pp. 215–220.

[23] Z. Gárdonyi and H. Nordhoff, *Harmonik*, 2nd ed. Wolfenbüttel: Möseler, 2002.

[24] C. Weiß, "Computational methods for tonality-based style analysis of classical music audio recordings," Ph.D. dissertation, Ilmenau University of Technology, Ilmenau, Germany, 2017.

[25] C. L. Krumhansl, *Cognitive Foundations of Musical Pitch*. Oxford, UK: Oxford University Press, 1990.

[26] D. Temperley, "What's key for key? The Krumhansl-Schmuckler key-finding algorithm reconsidered," *Music Perception*, vol. 17, no. 1, pp. 65–100, 1999.

[27] D. F. Tovey, *A Companion to Beethoven's Pianoforte Sonatas*. The Associated Board of the Royal Schools of Music, 1931.

[28] F. Lerdahl and R. Jackendoff, *A Generative Theory of Tonal Music*. MIT Press, 1983.

[29] C. S. Sapp, "Visual hierarchical key analysis," *ACM Computers in Entertainment*, vol. 3, no. 4, pp. 1–19, 2005.

[30] S. Klauk and R. Kleinertz, "Mozart's italianate response to Haydn's opus 33," *Music & Letters*, vol. 97, pp. 575–621, 2016.

[31] C. Dahlhaus, *Ludwig van Beethoven. Approaches to his music*. Oxford University Press, 1991.

[32] F. Galeazzi, *Theoretical-Practical Elements of Music, Parts III & IV, a translation of Elementi teorico-pratici di musica*, D. Burton and G. W. Harwood, Eds. Edited by Deborah Burton and Gregory W. Harwood, University of Illinois Press, 2012.

[33] S. Klauk, "Über Ausdruckscharakter und Programmmusik: Ludwig van Beethovens opp. 13 und 18 im Kontext italienischer Kammermusik," in *Nineteenth-Century Programme Music. Creation, Negotiations, Reception*, J. Kregor, Ed. Turnhout: Speculum musicae 32, Brepols, 2018, pp. 3–23.

[34] M. Neuwirth, D. Harasim, F. C. Moss, and M. Rohrmeier, "The Annotated Beethoven Corpus (ABC): A dataset of harmonic analyses of all Beethoven string quartets," *Frontiers in Digital Humanities*, vol. 5, p. 16, 2018.

[35] M. Gotham and M. Ireland, "Taking form: A representation standard, conversion code, and example corpora for recording, visualizing, and studying analyses of musical form," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Delft, The Netherlands, 2019, pp. 693–699.

[36] C. Weiß and M. Müller, "Computergestützte Visualisierung von Tonalitätsverläufen in Musikaufnahmen. Möglichkeiten für die Korpusanalyse," in *Instrumentalmusik neben Haydn und Mozart. Analyse, Aufführungspraxis und Edition*, S. Klauk, Ed. Saarbrücker Studien zur Musikwissenschaft 20, Königshausen & Neumann, 2020, pp. 107–130.