Computational Strategies for Measuring and Visualizing Tonal Structures: Towards a Large-Scale Analysis of Wagner's Ring Cycle

CHRISTOF WEIß, Center for Artificial Intelligence and Data Science (CAIDAS), Universität Würzburg, Würzburg, Germany
RAINER KLEINERTZ, Institut für Musikwissenschaft, Universität des Saarlandes, Saarbrücken, Germany and Department of Musicology, Hochschule für Musik Saar, Saarbrücken, Germany
STEPHANIE KLAUK, Institut für Musikwissenschaft, Universität des Saarlandes, Saarbrücken, Germany
MEINARD MÜLLER, International Audio Laboratories Erlangen, Erlangen, Germany

Within Western classical music, Richard Wagner's *Der Ring des Nibelungen* stands out as a monumental work, comprising nearly 22,000 measures and 15 h of music with a continuous dramatic narrative. Structuring such a large-scale work comprises an enormous challenge, which particularly affects the work's tonal organization. While other compositional techniques such as Wagner's use of leitmotifs have been extensively studied, his approach to structuring the tonal framework of the *Ring* remains subject to musicological debate. In this article, we present a computer-assisted approach for exploring the tonal organization of the *Ring* combining two traditional analytical perspectives based on sheet music with a novel computational strategy based on audio recordings. The first traditional perspective is the idealized tonal analysis by Alfred Lorenz from the 1920s, the second perspective examines key signature regions derived from Wagner's original score. We complement and compare these two views with a third, computational perspective: an audio-based strategy for measuring diatonic scale content over time, which is based on deep-learning techniques introduced in our previous work. To integrate these three perspectives, we propose a visual overlay technique using a dedicated color scheme, which offers a comprehensive and comparative view of Wagner's tonal conception. This integrated approach not only provides new insights into the *Ring*'s large-scale tonal organization but also demonstrates the broader potential of computational methods for musicological research regarding their ability to address challenges such as scalability, reproducibility, and objectivity of analysis. Audio-based strategies, in particular, offer a promising alternative to using symbolic (machine-readable) scores, which are often unavailable for large-scale works and tedious to create. Overall, we demonstrate that by providing automated measurements of tonal structures, computational tools can significantly enhance and expedite the tonal analysis of large-scale musical compositions such as Wagner's *Ring*.

Christof Weiß conducted this research together with the co-authors while working at the International Audio Laboratories Erlangen. Since November 2022, he has been Professor for Computational Humanities with the Center for Artificial Intelligence and Data Science (CAIDAS) at University of Würzburg.
This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Grant No. 252013209 (DFG MU 2686/7-2, DFG KL 864/4-2).
Authors' Contact Information: Christof Weiß (corresponding author), Center for Artificial Intelligence and Data Science (CAIDAS), Universität Würzburg, Würzburg, Germany; e-mail: christof.weiss@uni-wuerzburg.de; Rainer Kleinertz, Institut für Musikwissenschaft, Universität des Saarlandes, Saarbrücken, Germany and Department of Musicology, Hochschule für Musik Saar, Saarbrücken, Germany; e-mail: r.kleinertz@hfm.saarland.de; Stephanie Klauk, Institut für Musikwissenschaft, Universität des Saarlandes, Saarbrücken, Germany; e-mail: s.klauk@mx.uni-saarland.de; Meinard Müller, International Audio Laboratories Erlangen, Erlangen, Germany; e-mail: meinard.mueller@audiolabs-erlangen.de.

ACM 1556-4711/2026/9-ART38
https://doi.org/10.1145/3801964

CCS Concepts: • **Information systems** → **Music retrieval**; *Content analysis and feature selection*; • **Applied computing** → **Sound and music computing**; *Digital libraries and archives*; *Document metadata*; • **Human-centered computing** → *Information visualization*; • **Hardware** → *Digital signal processing*;

Additional Key Words and Phrases: Music information retrieval, harmony analysis, computational musicology, corpus analysis

ACM Reference format:
Christof Weiß, Rainer Kleinertz, Stephanie Klauk, and Meinard Müller. 2026. Computational Strategies for Measuring and Visualizing Tonal Structures: Towards a Large-Scale Analysis of Wagner's Ring Cycle. *ACM J. Comput. Cult. Herit.* 19, 3, Article 38 (September 2026), 23 pages.
https://doi.org/10.1145/3801964

1 Introduction

Richard Wagner's cycle *Der Ring des Nibelungen* constitutes an outstanding example of Western classical music, not least due to its unparalleled temporal extent spanning four operas (also called "evenings" or music dramas) with more than 14 h of music, 3,244 pages of orchestral score,[1] and 21,939 musical measures. The *Ring*'s parts are interwoven by a continuous dramatic plot, whose complex mesh of relationships is also reflected in the composition, emphasized by a comprehensive set of so-called leitmotifs. Structuring such a long work poses an enormous compositional challenge. This particularly affects the large-scale *tonal* organization since tonality marked the most important musical parameter for structuring compositions across time, at least throughout the common-practice period. While Wagner's usage of leitmotifs has been extensively studied [3, 10, 29, 35], there is still an open debate among musicologists regarding his approach to the tonal organization of the *Ring* [16, 30]. A first large-scale tonal analysis was performed in 1924 by conductor and musicologist Alfred Lorenz [26] in an attempt to structure the *Ring* into a set of so-called *poetico-musical periods* (German: *dichterisch-musikalische Perioden*)—a notion taken from Wagner's own literary works [44]. According to this idea, a period ideally begins and ends in the same key (principal tonality), thus achieving a sense of closure. This principal tonality (a kind of local key) is abandoned for dramatic reasons (e.g., a new emotion appearing) and later revisited (when the former emotion reappears). While the results of Lorenz's analyses have been heavily criticized [6, 7, 27], his annotations can still be considered an impressive early attempt towards a coarse-scale tonal analysis of the full *Ring*. Their role is even more important in light of the absence of any follow-up overall analyses, partly caused by musicologists' focus on other notions in the *Ring* (predominantly, the leitmotifs) or the long-standing belief that the tonal organization of Wagner's works is rather unstructured or "wandering" [7]—a view that music theorists recently questioned by analyzing extended tonality according to "tone fields" [36].

For conducting tonal analyses of such large-scale works, computational strategies provide promising opportunities for complementing manual analyses. In the field of music information retrieval, researchers are developing automated methods for estimating local keys from digitized music data [4, 15, 25, 33, 55]. Recently, substantial progress has been made regarding this task using **deep-learning (DL)** systems [21, 39, 50], usually based on supervised end-to-end training. As a main drawback, this methodology requires annotated training data, which introduces various biases, for example, towards the repertoire covered by the training examples and their specific annotations. A recent study [50] showed that the cross-annotator consistency across three human experts is as low as 75% for local keys in Franz Schubert's Winterreise—a cycle of 24 songs, which are much shorter in length and significantly less ambiguous regarding tonal organization than Wagner's operas. As demonstrated in [50], DL systems inevitably adapt to these subjective decisions of annotators during training, thus undermining an objective and independent analysis. To address this issue, we follow in this article a *multi-step strategy*, which we introduced in an earlier work [48] that constitutes the conceptual basis for this study. Instead of *predicting*

[1]Following the complete edition by Schott [8].

local keys with an end-to-end DL approach (involving hard decisions and expert annotations), we *measure* the presence of the 12 *diatonic scales* over the course of the *Ring*. For this, we rely on a simple template-matching approach [47] that does not require annotated training data. Rather than explicitly detecting diatonic scales, we then visualize scale probabilities over a circle-of-fifths axis, thus being capable of highlighting ambiguities, which play an important role in Wagner's tonal language and make them apparent to both expert and non-expert readers.[2]

As the starting point for this computational approach, a suitable music representation is needed. Typically, machine-readable scores are considered, which are of limited availability (in high quality) and tedious to create for complex and large-scale works such as entire operas. Audio recordings of performances thus provide an interesting alternative for the computational analysis. This, however, demands for specialized signal processing techniques to extract semantic information from the acoustic waveforms. As shown in our recent work [48, 52], novel processing techniques based on deep neural networks, which are trained on synchronized score–audio pairs to estimate pitch-class activations, lead to a substantial improvement, almost closing the gap between score- and audio-based strategies when being used for measuring tonal structures. We therefore consider such DL-based pitch-class estimates as the underlying representation for the subsequent tonal measurements.

Measuring tonal structures enables an objective view (not dependent on annotator's choices), being able to provide a new insightful perspective on the *Ring*'s tonal organization. Computing such measurements based on audio recordings is practical and scalable due to the abundance of audio data. In this article, we utilize such audio-based measurements to complement two traditional approaches, one based on the idealized tonal analyses by Alfred Lorenz mentioned above, and another one based on the key signature regions extracted from the score, which are only roughly corresponding to a tonal analysis but originate from Wagner's own notation.[3] We propose an overlay technique for integrating the three approaches by *jointly visualizing* their results with a dedicated color scheme. While all approaches realize an either simplified or idealized notion of local key, their combined view allows for comparative studies that can shed new light on the tonal conception of the *Ring*.

The main contributions of this article can be summarized as follows. First, we consider a recently proposed multi-step strategy [48] using deep pitch-class representations of audio for measuring and visualizing large-scale tonal structures and apply this to Wagner's *Ring* cycle. Second, we compare the resulting measurements of diatonic scales with the analyses by Alfred Lorenz as well as with key signature regions as written in Wagner's score. Third, to obtain a combined view of these three perspectives, we use an overlay technique for jointly visualizing the results, thereby exploiting a dedicated color scheme. Based on these comparative visualizations of the full *Ring* cycle, we discuss possible implications for understanding its large-scale tonal conception. Finally, we quantify the mutual similarity between the three perspectives. To this end, we introduce a suitable distance measure that reflects the musical similarity between scales (using a circular variant of the Earth Mover's Distance). Together with another measure from previous work, we apply this similarity measure for investigating the influence of visualization parameters such as the window size and the confidence (softmax temperature). Our results indicate that, with a suitable parameter choice, the combined visualizations of traditional analyses and computational measurements yield new insights into the *Ring*'s structure, thus serving as an example of how computer-assisted strategies can be applied for musicological research in a meaningful way.

The remainder of this article is organized as follows. Section 2 introduces related work on analyzing Wagner's *Ring* and its accessibility for computational methods. In Section 3, we present the three perspectives (Lorenz's analysis, Wagner's key signatures, and our computational measurements of diatonic scales), introduce our visualizing technique, and apply this to the full *Ring* while discussing the musicological implications. Finally, in Section 4, we introduce our similarity measures and investigate the influence of parameters. Section 5 concludes the article.

[2] For details on this computational analysis procedure, we refer the reader to the previous article [48], which is the basis for this study.

[3] Please note that with the simplification to the 12 diatonic scales (instead of 24 major and minor keys), we not only avoid frequent but often uninformative mode confusions (major vs. relative minor) but also allow for a direct comparison to KS regions.

2 Music Scenario: Wagner's *Ring*

This section introduces relevant background on Wagner's *Ring* cycle regarding its structure (following [45]) and related analysis (Section 2.1) as well as available data for computational processing (Section 2.2).

2.1 Related Work

Since the origins of the genre in the 17th century, operas have usually been structured into a series of so-called "numbers," that is, individual movements such as arias connected by recitatives or spoken dialogues. One of the revolutions Wagner introduced to this genre in the 19th century was a novel approach to structuring operas, which he theoretically described in *Opera and Drama* [44]. Overcoming the number opera, Wagner postulated a steady musical flow (denoted as *through-composed style* or *endless melody*), abandoning structural means such as repetitions or interruptions. This lack of evident caesuras not only makes the structural analysis of Wagner's operas particularly difficult, it also led Wagner scholars to the conclusion that there "is no structure" in his works, making any structural analysis obsolete. However, this conclusion is in sharp contrast to Wagner's own awareness for the structure of his compositions, stating himself the "unity of the symphony movement" to be an ideal for opera's structure [43].

In this vein, it is important to emphasize that harmony and tonality play a central role for Wagner to achieve such structure. An important hint to this is the notion of the *poetico-musical period*. Introduced in *Opera and Drama* [44], such a period should involve a close relationship between poetry and drama on the one hand and the music on the other hand. In particular, he considers dramatic elements such as the appearance of persons or sentiments as a cause to trigger a modulation, that is, change of the local key. Such a change usually reverts along with a dramatic change, thus achieving a sense of closure as implied by the term "period."

Assuming the poetico-musical period to be foundational for understanding the *Ring*'s structure, Alfred Lorenz presented a first large-scale tonal analysis of Wagner's *Ring* cycle [26], where he segmented the *Ring* into 146 such periods and assigned a principal tonality (key) to each of them. While Lorenz's analyses represented a kind of "state of the art" for analyzing the *Ring* until the 1960s, his analyses were criticized by various musicologists later on [2, 6, 7] for several reasons including the overstretching of the notion of periods to explain the full *Ring* [6] as well as Lorenz's problematic political role [27]. Nowadays, many musicologists agree that Wagner at least partially realized the notion of periods in his works [16] and that tonality in Wagner's works is not random or "wandering" [7] but an essential compositional aspect, meticulously planned by the composer [30]. In this article (see Section 3.4), we therefore abstract from the individual decisions Lorenz made in his annotations but use them at large as one perspective on the tonal structure of the *Ring* (see Section 3.1).

2.2 Raw Data and Accessibility

As mentioned above, the tetralogy *Der Ring des Nibelungen*, created 1848–1874 in two phases, constitutes one of the most remarkable manifestations of Wagner's conceptual ideas on composing "music dramas." The *Ring* serves as an ideal subject for studying Wagner's use of tonality. Moreover, its significance is emphasized by the many artists, composers, and genres (e.g., film music) that have drawn inspiration from it. This strong influence has also spurred the *Ring*'s study with computational methods, addressing various aspects such as leitmotifs [19, 24, 29, 35], measure boundaries [46, 54], singing voice activity [22, 23], or listeners' experience [32, 35].

Some of these studies were conducted within a long-term interdisciplinary project,[4] which resulted in the recently released **Wagner Ring Dataset**[5] **(`WRD`)** [45], a multi-modal and multi-version dataset comprising sheet music, annotations, and 16 different performances of the *Ring*, each lasting roughly 14–15 h (232 h of audio in total). Following [45], the `WRD` is structured into 4 operas, 11 acts, and 21,939 measures, where we use the *11 acts*, each of which exhibits a contiguous measure count, as the units of analysis (see Figure 1 for an overview).

[4] *Computer-Assisted Analysis of Harmonic Structures (CAS)*, see https://www.audiolabs-erlangen.de/fau/professor/mueller/projects/cas.
[5] https://zenodo.org/record/7672157#.Y_fxDXbMJaR.

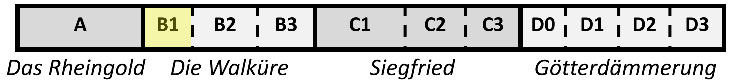


Fig. 1. Overview of the musical parts (operas/acts) of the *Ring*. Figure from [48]; for an overview including available versions see [45]. The first act of *Die Walküre* (WWV 86 B1) serves as running example for explaining our methodologies.

Table 1. Statistics of the Different Segmentations and Annotations of the WRD

			Segment Duration (Measures)		
Segmentation/Annotation	# Segments	# Segm. per Act	Mean ± Std. Dev.	Min–Max	Median
Acts	11	1.0	1,994 ± 764.0	892–3,897	1,789.0
Scenes	38	3.5	577 ± 294.4	59–1,290	567.5
Time signature regions	555	50.5	40 ± 67.5	1–486	14.0
KS regions (KS)	**528**	**48.0**	**42 ± 43.6**	**2–261**	**24.0**
Lorenz segments (AL)	**219**	**19.9**	**100 ± 89.1**	**3–489**	**73.0**
Lorenz periods	166	15.1	120 ± 92.1	8–489	87.5

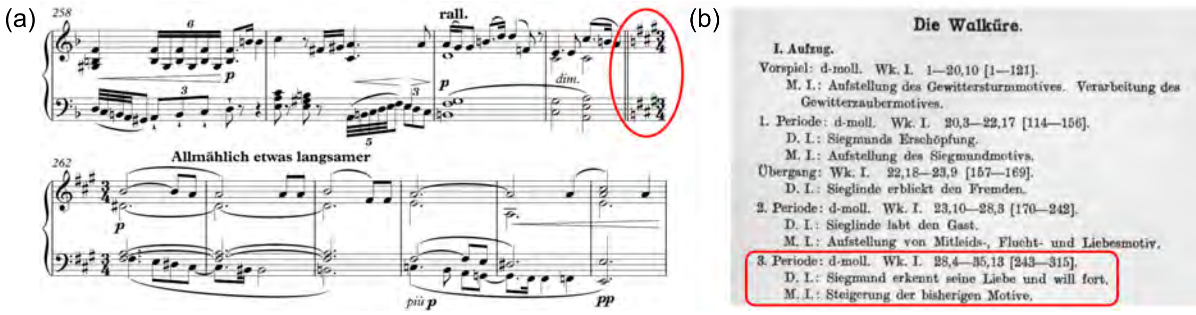
All durations are given in musical measures. For the bottom row, we only count Lorenz period regions ignoring introductions, transitions, and the like. Also, we individually count sub-periods like "12Aa," "12Ab," "12Ba," thus arriving at a higher number (166) as the one mentioned in Section 1 (146). Bold lines denote the annotations used in the subsequent analyses.

Within each act, there are various fine-level segmentations such as scenes (see Table 1 for some statistics). Regarding sheet music, the WRD does not include a full orchestral score in machine-readable format (which would be extremely tedious to create) but only a piano reduction in symbolic format. Instead of using this data as basis for our computational analysis, we rely on the audio recordings. As shown in our previous study [48], processing such audio recordings with DL methods leads to a high similarity between audio- and score-based analyses. Section 3 discusses this choice in more detail.

The manual analysis of tonality in the *Ring* is not only tedious but also highly subjective and, therefore, prone to a low inter-annotator agreement (as found for shorter and simpler compositions in [50]). As a result, there are no reliable tonal analyses for the *Ring*. Nevertheless, the WRD comes with structure annotations on different levels, which we summarize in Table 1 together with statistics on the number and duration of segments. Beyond the acts and scenes discussed above, this also encompasses finer-level segmentations based on the time and key signature regions, which we extracted from the symbolic piano scores that follow the full scores of the complete edition [8]. While key signatures do not provide an actual key information or tonal analysis but rather indicate a suitable diatonic scale to notate the tonal content of a specific passage, this information still provides an interesting perspective on the *Ring*'s tonal organization—not least due to the fact that the decision on the key signature and its changes over time was taken by Wagner himself. Going beyond the WRD, we also consider two segmentations based on by Alfred Lorenz's analyses [26], which are of intermediate duration ranging between the scenes and the time/key signature regions. We hereby discriminate between all segments (including introductions and transitions) and the pure periods. In the following, we discuss these segmentations in more detail.

3 Measuring and Visualizing Tonal Structures

In this section, we describe the mentioned analyses and annotations in more detail (Section 3.1) and introduce our computational strategy (Section 3.2). We then present our approach of jointly visualizing the three perspectives (Section 3.3) and finally discuss these visualizations for the full *Ring* cycle (Section 3.4).



Die Walküre.

I. Aufzug.

Vorspiel: d-moll. Wk. I. 1—20,10 [1—121].
M. I.: Aufstellung des Gewittersturmmotives. Verarbeitung des Gewitterzaubermotives.

1. Periode: d-moll. Wk. I. 20,3—22,17 [114—156].
D. I.: Siegmunds Erschöpfung.
M. I.: Aufstellung des Siegmundmotivs.

Übergang: Wk. I. 22,18—23,9 [157—169].
D. I.: Sieglinde erblickt den Fremden.

2. Periode: d-moll. Wk. I. 23,10—28,3 [170—242].
D. I.: Sieglinde labt den Gast.
M. I.: Aufstellung von Mitleids-, Flucht- und Liebesmotiv.

3. Periode: d-moll. Wk. I. 28,4—35,13 [243—315].
D. I.: Siegmund erkennt seine Liebe und will fort.
M. I.: Steigerung der bisherigen Motive.

Fig. 2. Examples for tonal structure annotations derived from different sources. (a) KS example: An excerpt from the first act of *Die Walküre* (WWV 86 B1, measures 258–267) in the piano reduction from the WRD, with a key signature change (KS) in measure 262. (b) AL example: An excerpt from Alfred Lorenz's book [26, p. 28] indicating the first periods of the same act (WWV 86 B1) with the third period ranging from measures 243–315, featuring D minor as principal tonality.

3.1 Analyses and Annotations

As mentioned in the previous section, we consider several perspectives on the *Ring*'s large-scale tonal structure. To enable a quantitative (and visual) comparison, we relate each of these perspectives to the 12 diatonic scales (see Section 3.2 for details). While Wagner's tonal language often extends far beyond diatonic pitch-class constellations, there remains a substantial number of diatonic passages in the *Ring*. The prominence of diatonic scales closely corresponds to tonally stable regions. Moreover, analyzing on the scale level helps to avoid confusions between relative keys, which often occur across different annotators [50], and facilitates the visualization of tonal relationships of scales on the circle of fifth. We indicate the presence or likelihood of these scales for each of the M musical measures of an act, resulting in a matrix $\mathcal{A} \in [0,1]^{12 \times M}$ for each of the eleven acts.

The first perspective, denoted by KS, relies on basic information extracted from the score, namely the key signatures that Wagner chose for notating a specific musical section. A key signature directly corresponds to one diatonic scale (e.g., in Figure 2(a), 1♭ correspond to the −1 diatonic and 3♯ to the +3 diatonic), and we assume enharmonic equivalence ($-6 \,\widehat{=}\, +6$) obtaining 12 possible scales. This numerical scheme ("Quintenmessung," see [13]) entails a simple musical distance measure (neighbors on the circle of fifths, which share 6/7 accidentals, obtain a distance of 1), while we acknowledge that other notions of tonal distance [41] also play an important role in Wagner's music. Please note that there is no specification of the tonic or mode (e.g., the +2 diatonic might relate to D major or B minor equally well) involved. Moreover, the decision on the key signature is sometimes rather pragmatic since its purpose is to ensure good readability and comprehension by the musicians.[6] In this sense, the key signature information is clearly not an analysis nor corresponds to local key regions. Nevertheless, the choice of a key signature provides at least a weak indication of a tonal region or scale, since it marks the pitch-class content that can be notated without further accidentals. We therefore consider it a relevant perspective on how Wagner structured the *Ring* tonally in his score. To obtain the diatonic scale information, we automatically extract key signatures from the symbolic piano scores in the WRD (see Figure 2(a) for an example) and convert the key signature information to a diatonic scale, thus obtaining a matrix $\mathcal{A}_{\mathrm{KS}} \in [0,1]^{12 \times M}$.

For the second perspective, denoted by AL, we consider Alfred Lorenz's book on the "musical structure" (German: *musikalischer Aufbau*) of the *Ring* published in 1924 [26]. As mentioned above, Lorenz mainly focuses on the notion of the poetico-musical periods for his analysis, extending those with additional segments such as an introduction (*Vorspiel*) or transition (*Übergang*). Figure 2(b) shows an example for his descriptions, which include the measure regions as well as—for most segments—the principal tonality (key). Most importantly, we

[6] For example, in the first opera (*Das Rheingold* WWV 86 A), measures 3827–3842 are notated with 4♭ for the female singers (Rheintöchter) and one of the harps (on stage) while the remaining singers and instruments stay with 5♭. Here, we follow the latter for our annotation.

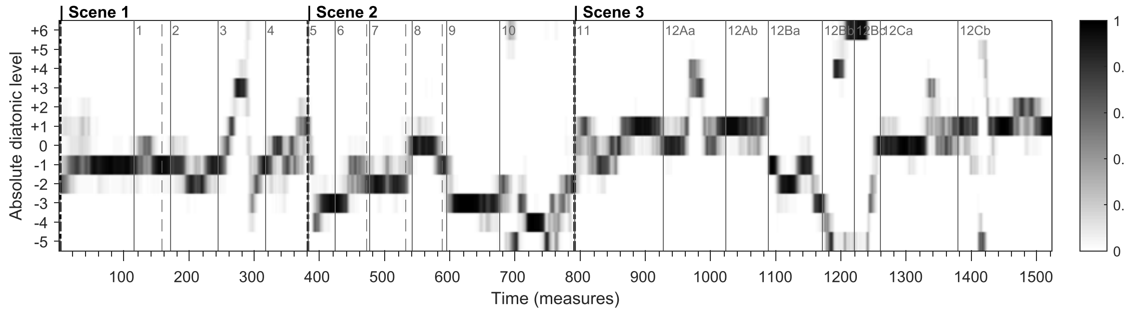


Fig. 3. Visualization of diatonic scale measurements for the act WWV 86 B-1 in the version conducted by H. v. Karajan. Parameters: window size $w = 32$, softmax parameter $\beta = 50$. Lorenz's periods are indicated by solid lines, transitions and introductions by dashed lines.

emphasize that this principal tonality serves as the reference (frame key) for the period, while the whole idea of the period is to depart from this frame key and return to it later. Intermediate keys (targets of the modulations) are not explicitly indicated in the descriptions.[7] For our purposes, we consider all annotated segments including introductions, transitions, or the like[8] (see Figure 3 for an example). Analogously to the previous perspective, we end up with a matrix $\mathcal{A}_{\mathrm{AL}} \in [0,1]^{12 \times M}$. We publish the initial annotations for both KS and AL as supplementary material to this article.

Table 1 summarizes statistical properties of these two perspectives (the two middle rows in boldface). For reference, the table also shows statistics on the acts, scenes, and time signature regions. Comparing time and key signature, we observe a similar behavior, with time signatures changing slightly more often on average. As for tonal segments, we immediately observe that both key signatures (KS) and Lorenz segments (AL) are on a substantially finer level than the scenes within an act, which is also visually evident (for AL) from Figure 3. A typical scene (median) lasts for 567.5 measures, while a KS region only lasts 24.0 measures. AL segments are clearly longer (median 73.0 measures) than KS regions. On average, an act comprises more than twice the amount of KS regions (48.0) than AL segments (19.9). This already shows that AL rather denotes the (often idealized) frame key of a period while KS tends to follow the modulations within the periods on a finer temporal level. In addition to the more detailed Lorenz sections (AL), which yield full coverage of an act, we also provide statistics that only account for the actual periods (bottom line in Table 1). For all types of segmentation, we note that the segments' duration varies substantially, with the standard deviation sometimes being even greater than the mean. This suggests a high degree of variability, ranging from nearly measure-wise changes to entire scenes remaining in a certain time/key signature or principal tonality.[9] Nevertheless, we consider these statistics valuable for gaining insights into the structure of the *Ring* and its time scales, which we will relate to our computational measurements described in the following.

3.2 Measuring Diatonic Scale Content

As the third perspective, which we denote by CM, we consider a previously proposed strategy for measuring tonal structures from audio recordings [48]. For convenience, we now summarize both related work in tonal analysis and the main processing steps of our strategy, closely following our previous work [48], which we refer to for further details on the methodology.

[7] Sometimes, Lorenz mentions such intermediate keys but never specifies the corresponding measure region, e.g., "g-Moll (Mittelsatz in f-Moll[])" [26, p. 31].
[8] For nested period structures with sub-periods such as 12Aa, 12Ab, 12Ba, and so on, we opt for the finest level in our annotations.
[9] Concerning the Lorenz periods, this inhomogeneity posed a main argument of Dahlhaus criticizing Lorenz's analyses [6].

As analyzing tonal structures is central to computational musicology, various automated methods for tonal analysis have been proposed over decades (see, e.g., [5, 9, 28, 40, 53]). Regarding different time scales, typical tasks refer to global keys [42], local keys [25], and chords [20]. While local key or chord analysis is often done as a framewise classification task (involving an explicit segmentation and hard decision on labels), this strategy cannot account for gradual modulation processes, tonal ambiguities, or different music-theoretical interpretations. This frequently leads to a high subjectivity in the labels, which has been shown for chord [18, 31] and local key estimation [50] and becomes particularly prevalent when using deep neural networks trained for these tasks.

To address these problems, we consider as our third perspective a multi-step strategy: We rely on the efficacy of DL only for a first processing step, which aims for detecting the activity of the 12 chromatic pitch classes over time as proposed in [52]. To compute such pitch-class representations from the audio recordings, we use a recently proposed model [49]—a deep residual neural network of medium size (4.8M parameters) applied to a music-specific spectral audio representation (harmonic constant-Q transform, HCQT [1], with a resolution of roughly 43 frames per second). We trained this network on several datasets comprising various instrumentations and styles including orchestral, choir, and chamber music (see [48], Table 1 for a detailed description of the training datasets). The pre-trained model is fully available online.[10] Applying the trained network to the audio data of the `WRD` (corresponding to the cross-dataset experiment in [48]), we obtain a 12-dimensional vector of pitch-class activations for each frame. We aggregate this frame-wise representation to an average pitch-class distribution for each musical measure, using the measure annotations provided in the `WRD`, which were created through an extensive manual annotation process [45, 46].

As the second step, we use the resulting measure-wise pitch-class representations to visualize the presence of specific tonal structures. To this end, we consider a model-based approach relying on template matching, which can be realized for chords [12], scales [47], or other tonal structures defined as pitch-class sets. In this article, we focus on the measurement of the 12 *diatonic scales*, which are suitable for highlighting the overall modulation process while avoiding the frequent ambiguity between relative keys (e.g., C major vs. A minor). We are aware of this diatonic scale measurement being a simplification of local key estimation. The missing specification of the referential chord (tonic) and, thus, the missing discrimination between major and relative minor is certainly a lack of information, which has downsides (such as the non-consideration of altered scale degrees in harmonic or melodic minor) but at the same time also advantages. Being a pure measurement of pitch-class content over time rather than a sort of traditional "harmonic analysis," it involves less interpretation by trained experts, with perceptual aspects as the implied "center of gravity" or "feeling of arrival" necessary for establishing a local key becoming less relevant. This makes the measurements less prone to annotator subjectivity and disagreement, which frequently occurs in local key annotations as shown in [45], while still being explicitly interpretable for musicologists.

To perform this diatonic scale measurement, we proceed as follows. Since the pitch-class content of a diatonic scale usually occurs within longer segments, we first smooth the measure-wise representations using a window of size $w \in \mathbb{N}$ measures and a hop size of one measure, followed by a normalization using the ℓ_2-norm. This results in a vector $\mathbf{c} = (c_0, c_1, \ldots, c_{11})^\top \in \mathbb{R}^{12}$ for each window (with $c_0, c_1, \ldots, c_{11}$ referring to pitch classes C, C♯, . . ., B). To measure the presence of each diatonic scale, we match each pitch-class vector $\mathbf{c}$ with binary scale templates $\mathbf{t} \in \mathbb{R}^{12}$ using the inner product $\langle \mathbf{t}, \mathbf{c} \rangle$ of ℓ_2-normalized vectors (also known as cosine similarity). The template $\mathbf{t}_0$ for the "0 diatonic scale" (corresponding to C major or A natural minor, or the white keys on the piano) is given by:

$$\mathbf{t}_0 = \frac{1}{\sqrt{7}} \cdot (1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1)^\top , \tag{1}$$

[10] Source code (PyTorch), pre-trained models, and estimated pitch-class activations for the full *Ring* published at https://github.com/christofw/pitchclass_tonalstructures. For this study, we utilized the model "DeepResWide" specified in `exp302c`.

where 1 denotes an active (in-scale) pitch class and 0 an inactive (out-of-scale) one. We obtain templates for the other scales by performing cyclic shifts of $\mathbf{t}_0$. Due to enharmonic equivalence (identifying C♯ and D♭, etc.), we distinguish 12 diatonic scales. For each window, this results in a measurement that we specify as a 12-dimensional vector $\hat{\mathbf{a}} \in \mathbb{R}^{12}$, with $\hat{a}_i = \langle \mathbf{t}_i, \mathbf{c} \rangle$, $i \in \{-5, -4, \ldots, +5, +6\}$. As argued in [48], instead of deciding on the best matching scale, we propose to use suitable visualizations [47], which allow for directly interpreting the continuous-valued diatonic scale measurements. For this purpose, we compute rescaled local analyses $\mathbf{a} \in \mathbb{R}^{12}$ utilizing the softmax function, which suppresses weaker components and emphasizes stronger ones:

$$a_i = \exp(\beta \hat{a}_i) \Big/ \sum_{j=-5}^{6} \exp(\beta \hat{a}_j), \tag{2}$$

with $i \in \{-5, \ldots, +6\}$. The parameter $\beta \in \mathbb{R}$ (inverse softmax temperature) serves to adjust the level of enhancement, controlling how strongly the prominent pitch classes in the representation are emphasized. Due to the ℓ_1-normalization inherent to the softmax, we can interpret the values a_i as pseudo-probabilities of diatonic scales. We combine the local analysis vectors $\mathbf{a} \in \mathbb{R}^{12}$ for all musical measures to a matrix $\mathcal{A}_{\mathrm{CM}} \in [0, 1]^{12 \times M}$. When comparing the resulting matrices for different underlying pitch-class representations, as shown in the comprehensive investigation in [48], we found that using our DL-based representations results in a high similarity between audio-based measurements of diatonic scales and those based on symbolic scores. Therefore, in this article, we only use the deep pitch-class representations as basis for computing $\mathcal{A}_{\mathrm{CM}}$ in this article.

As shown in Figure 3, one can visualize this analysis $\mathcal{A}_{\mathrm{CM}}$ using a grayscale plot. To account for the tonal proximity of fifth-related scales, we order the scales along a circle-of-fifths axis, thus achieving a good overview of the modulation process, in particular for large-scale works such as symphonies or operas (compare [47]). Figure 3 shows such a plot for the first act of *Die Walküre* (WWV 86 B-1) as recorded under H. v. Karajan in 1966 (Deutsche Grammophon). The parameters used for this example are a window size of $w = 32$ measures and an inverse softmax temperature of $\beta = 50$ (we investigate this parameter choice in Section 4). We observe clearly stable sections remaining mostly in one scale such as for measures (mm.) 20–200 (−1 scale), mm. 600–680 (−3), or mm. 1260–1330 (0), as well as transitional regions such as mm. 1140–1190 (gradually modulating from −1 to −5). For better orientation, we additionally show the boundaries for different segmentations (described in Table 1) such as scenes (thick dotted-dashed vertical lines) and Lorenz segments (thin straight lines and numbers for periods, thin dashed lines for additional segments such as introductions or transitions). We observe that scene boundaries may correspond to strong tonal changes (a new scene starting in a new key). Moreover, for some of the Lorenz periods (such as periods 2, 3, or 12Aa), our visualization of scales suggests a modulation structure corresponding to Wagner's idea with one scale modulating another one and then returning back.

Summarizing our observations and the ones from [48], this computational strategy leverages the strengths of DL (in contrast to traditional signal-processing approaches) while providing an objective and interpretable measurement of the tonal structures. We again want to emphasize that the simplification to diatonic scales has downsides, such as the missing discrimination between major and relative minor. Moreover, the presence of additional altered pitch classes occurring in harmonic and melodic minor scales (scale degrees ♯6 and ♯7) poses challenges.[11] Simply including such scale degrees into the template, however, leads to poorer results due to confusions with other scales. In contrast, the proposed straightforward mapping—while not being an actual "analysis" that is able to resolve complex pitch-class relationships—can be directly interpreted as what it is: The measurement of a certain *diatonic pitch-class set* (not a major or minor scale, nor any key) being present in a given temporal segment. As such, it provides certain benefits by demanding less interpretation by experts, therefore being less subjective and avoiding biases related to training data and annotations that may arise in an end-to-end

[11] In practice, we found that these additional pitch classes lead to ambiguity especially when using more noisy pitch-class representations based on signal processing. Pitch-class features based on deep learning, in contrast, have less energy in inactive pitch-class bands, which reduces confusions with other scales. An example for this behavior can be observed in several passages of *Die Walküre* (WWV 86 B1, mm. 1–100 or mm. 476–586, see Figure 8 in [48]), where the signal-processing approach generates artifacts towards scales with more sharps while the deep-learning approach, also used for our study, suffers considerably less from such artifacts.

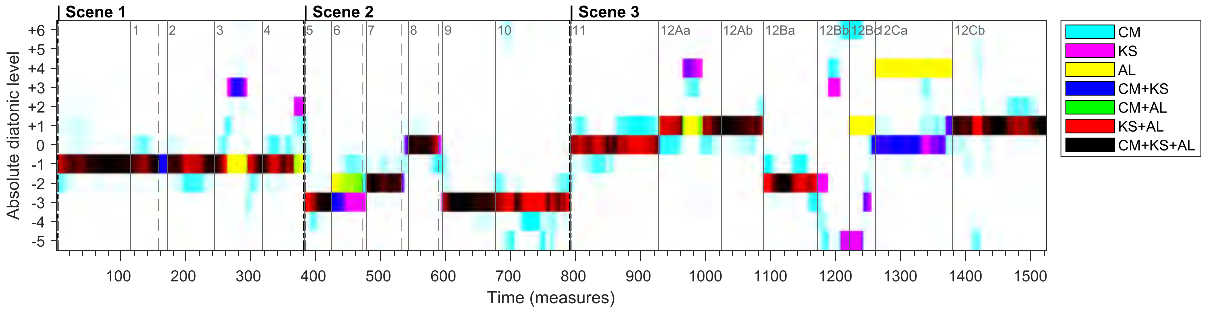


Fig. 4. Comparative visualization of diatonic scale measurements for the act WWV 86 B-1 in the version conducted by H. v. Karajan. Visualization parameters are $w = 32$, $\beta = 50$.

local key estimation approach as shown in [50]. In the following sections, we will contrast this computational measurement with the analyses described in Section 3.1, which also helps to better understand the behavior of this method.

3.3 Comparative Visualization

As described above, a central benefit of the proposed measurement of tonal structures is its accessibility by means of suitable visualizations. Inspired by this fact, we now introduce a strategy for comparing different analytical perspectives in a visual way. To this end, we re-consider a color overlay technique introduced in [48]. This technique (or related ones) may serve to compare various different notions such as measurements computed for *different versions* (cross-versions visualization, see [17, 51]) or based on *different representations* [11, 48]. In this article, we apply the technique for comparing our three perspectives CM, KS, and AL, opting for the additive color scheme introduced in [48]. Next, we briefly summarize the computation steps, following [48].

We start with the three analyses $\mathcal{A}_{\mathrm{KS}}$, $\mathcal{A}_{\mathrm{AL}}$ (Section 3.1), and $\mathcal{A}_{\mathrm{CM}}$ (Section 3.2). To compute the comparative visualization, we represent the three measurement results in the RGB space in the following way:

$$\begin{pmatrix} \mathrm{R} \\ \mathrm{G} \\ \mathrm{B} \end{pmatrix} = \begin{pmatrix} 1 - \mathcal{A}_{\mathrm{CM}}(n,m) \\ 1 - \mathcal{A}_{\mathrm{KS}}(n,m) \\ 1 - \mathcal{A}_{\mathrm{AL}}(n,m) \end{pmatrix}, \tag{3}$$

where $n \in \{-5, \ldots, 6\}$, $m \in \{1, \ldots, M\}$. Plotting this color array (as in Figure 4) shows the three matrices overlaid in a transparent way. The colors resulting from this assignment are shown in the legend of Figure 4. When all three measurements are consistent, we obtain black. In cases of disagreement, different colors arise: A CM measurement that does not align with any other perspective results in cyan, while a KS measurement that stands alone yields magenta. An isolated AL measurement results in yellow. When two measurements match (but the third differs), mixed colors appear: Blue for CM matching KS, green for CM matching AL, and red for KS matching AL.

Continuing our example from Figure 3, Figure 4 shows this comparative visualization for the first act of *Die Walküre* (WWV 86 B1) in the Karajan version. Regarding the CM, this visualization exactly corresponds to Figure 3, with the grayscale values of Figure 3 now shown in degrees of cyan. The other two measurements are discrete (only values 0 or 1 possible). Consequently, these two perspectives (KS and AL) are visible as solid bars in either magenta (KS), yellow (AL), or their color mix red (KS + AL). As soon as CM is matching one of the others, these colors (gradually) change towards another color mix (magenta→blue, yellow→green). For example, in mm. 262–292, KS indicates +3 (magenta) and CM changes to +3 a bit later (magenta changing to blue). If CM approaches the already consistent measurement of KS and AL, their overlay color (red) gradually changes to black. For example, in the last Period 12Cb (mm. 1378–1523), KS and AL consistently indicate the +1 scale (in red) and, for most of the time but with some interruptions, CM also yields the highest probability for +1, turning red into black.

We now discuss the visualization in more detail to understand potential inconsistencies between the different perspectives (CM, KS, and AL). While the visualization helps to identify areas of agreement, it also highlights discrepancies. By analyzing these differences, we can better assess the strengths/weaknesses or the different focus of each approach, thus gaining deeper insights into the *Ring*'s tonal organization as viewed from these three perspectives.

We observe several long regions of high agreement among all measurements (black) such as the prelude as well as periods 1, 2, 7, 8, 9, 12Aa, and 12Cb. On the contrary, all three perspectives sometimes disagree, as in Period 12Bc (mm. 1219–1258). We also find sections where the two reference perspectives (KS and AL) agree but the CM yields mostly other (yet closely related) scales (periods 10 and 11). For Period 12Bb (mm. 1170–1218), Lorenz does not specify a key but marks this period as "strongly modulating" [26, p. 30]. Concerning the period structure, we make a highly interesting observation: In both Period 3 (mm. 243–315) and Period 12Aa (mm. 926–1021), CM and KS consistently show a modulation from a first scale to a second one and back (with CM sometimes slightly deviating from KS or indicating a smooth transition between two KS regions). In Period 3, this is a modulation from D minor to A major (fifth parallel) and in Period 12Aa, from E minor to E major (parallel; tonic root stays). In contrast, AL does not indicate these modulations but exactly specifies the first scale as the "principal tonality" of each period, turning the beginning and end of both sections into black. In the middle, we observe two colors: blue (mix between KS and CM) for the second scale, and yellow (AL) remaining in the first scale. These visual structures almost perfectly reveal the prototype of Wagner's poetico-musical periods regarding their tonal implications (we do not discuss the textual or dramatic part here, see [16, 30]).

Another interesting situation arises in Period 6 (mm. 423–410), where the three perspectives appear to be one fifth apart from each other. While Wagner notates the whole period in 3♭ (−3 diatonic in KS), Lorenz annotates it as B♭ major (−2 diatonic in AL). Looking at the score, we observe the period to start in C minor (−3) in m. 423. This results in CM and KS matching (blue color) at the start of Period 6. After several modulations, the music reaches a clear B♭ major region in mm. 446–449. Here, CM changes to −2, now coinciding with KS (green). In the following, CM highlights a region between −2 and −1, the latter being exemplified through, e. g., an F major cadence in mm. 458–450. This F major can be interpreted as a tonicization, still with dominant tension towards the −2 scale of the following Period 7, which all three perspectives agree on again.

We now turn to the discussion of Period 12Ca (mm. 1259–1377), where we observe a strong contrast between AL (indicating +4) and the other two perspectives (mostly agreeing on the 0 scale). As for KS, almost the full period is notated without accidentals (0 diatonic), with a change to +1 in m. 1358. Lorenz's annotation (AL), in contrast, marks this period to be in E major (+4) [26, p. 30]. Upon close investigation, we indeed find three clears—though quite short—E major segments (mm. 1269–1276, mm. 1327–1334, mm. 1362–1365) within a longer passage revolving around C major. Here, the A minor triad (as the resolution of the E major triad, interpreted as a dominant) serves as a pivot chord facilitating the switch between 0 and +4. As shown in Figure 4, the CM analysis mostly follows the predominant 0 scale but changes between the two conflicting analysis of KS and AL towards the end of the period.

Overall, the comparative visualization proves to be a valuable tool for quickly grasping commonalities and differences between the three perspectives on the tonal structure of this act. We observe stable sections where all perspectives coincide as well as contrasting passages where the perspectives focus on different levels of tonal organization, with AL focusing on the "principal tonality," KS on the medium-scale tonal regions, and CM on the more detailed modulation process. When directly comparing CM and KS, they are often identical or tonally close (differing by only one or two scales) while AL tends to diverge more from the two other perspectives. In Section 4, we will explore how adjusting CM's window size parameter w allows us to shift between different levels of granularity. While the agreement between these perspectives does not imply a certain quality or correctness of

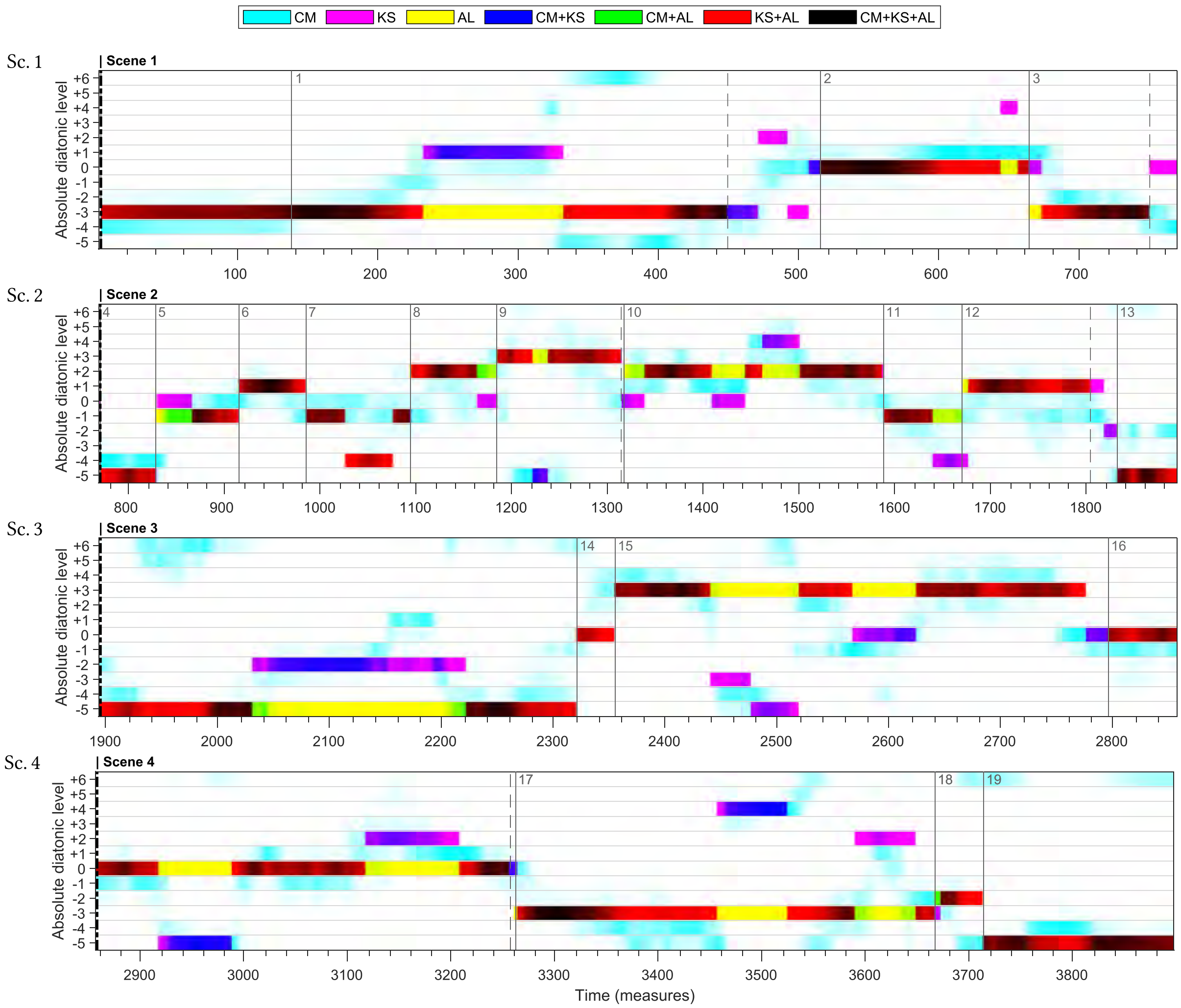


Fig. 5. *Das Rheingold* WWV 86 A, structured into four scenes. Visualization parameters are $w = 64$ and $\beta = 50$.

the analysis, we emphasize that—beyond the choice of temporal hierarchy levels—the signal-driven CM is able to highlight similar structures than other, human-based readings of tonal structures.

3.4 Three Perspectives on the *Ring*

We now want to discuss these comparative visualizations for the full *Ring* cycle, showing them for the four different operas *Das Rheingold* WWV 86 A (Figure 5), *Die Walküre* WWV 86 B (Figure 6), *Siegfried* WWV 86 C (Figure 7), and *Götterdämmerung* WWV 86 D (Figure 8). To limit the scope of this work, we focus on the tonal structures here. For discussing their interplay with the operas' plot, we refer to our related work [38]. As visualization parameters, we chose $w = 64$ and $\beta = 50$ and refer to Section 4 for a detailed discussion of this choice. For a better overview, all visualizations are split into acts except for Figure 5 (split into scenes). This figure, referring to the first opera (or "Vorabend") *Das Rheingold*, shows a slightly less complex structure as compared to the other operas. This is partly due to its extent—in Figure 5, each subplot spans one scene of roughly 1,000 measures as

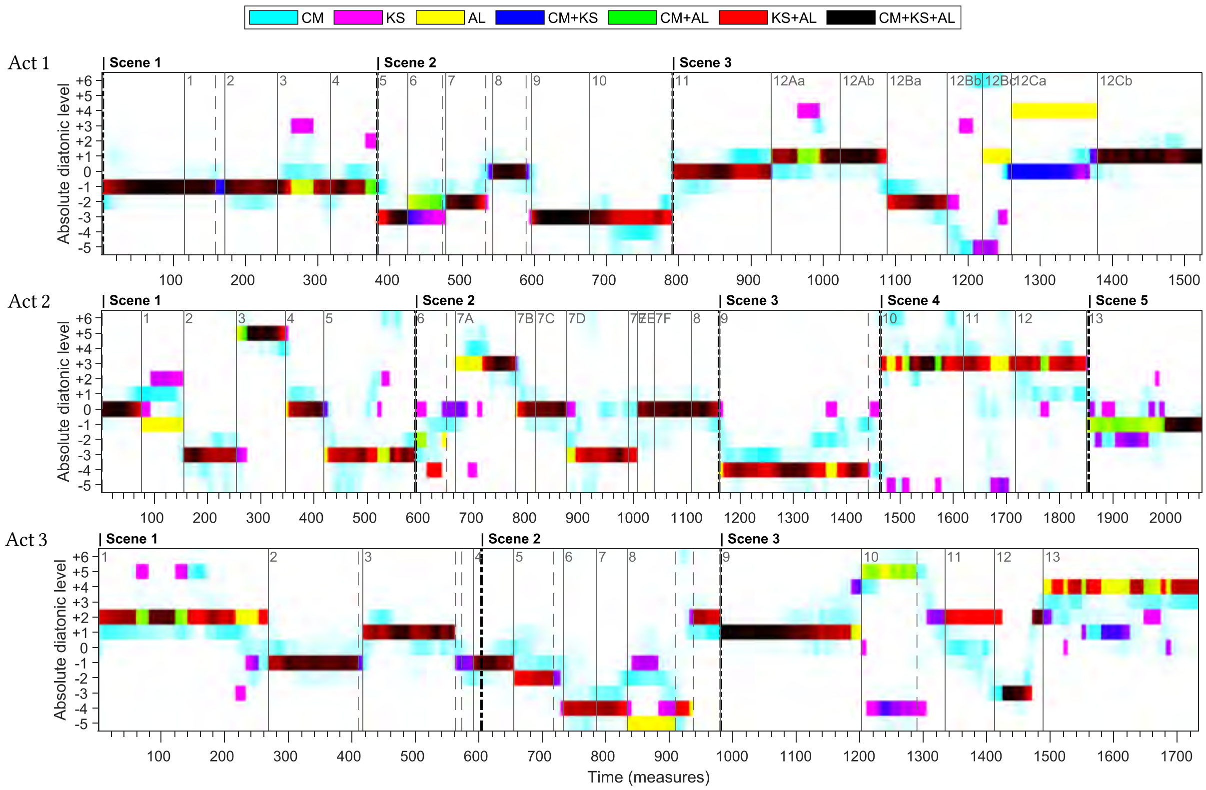


Fig. 6. *Die Walküre* WWV 86 B, structured into three acts (*Aufzüge*). Visualization parameters are $w = 64$ and $\beta = 50$.

opposed to the acts spanning 1,500–3,000 measures in Figures 6–8—but also indicates a simpler tonal structure with less modulations. In *Das Rheingold*, KS and AL are often consistent and only deviate in case of typical "period structures," where KS leaves to an intermediate key while AL stays in the frame key (see for instance periods 1, 10, 13, 16, or 17 for such cases). The computed measurements (CM) closely follow the key signatures (KS) leading to blue or black regions or are only one or two scales apart from KS. Sometimes, CM indicates interesting transitions such as the one in Period 17, where C♯ minor is gradually left from m. 3523 onward via C♭ major (−7, corresponding to +5) in mm. 3529–3534 and E♭ minor (−6, corresponding to +6) in mm. 3543–3555 towards E♭ major (−3), which is finally reached in m. 3560.

Compared to *Das Rheingold* and the first act of *Die Walküre* (discussed in Section 3.3), the other two acts of *Die Walküre* already show a higher degree of modulations and tonal complexity (Figure 6). We still find passages of high consistency such as Periods 7B-C and the end of Act 2, as well as Periods 2, 3, and 9 in Act 3. However, there are more frequent discrepancies between KS and AL, some of which cannot be easily explained by the notions of periods (see Period 1 in Act 2, or Periods 1, 8, and 10 in Act 3). We also find several passages where all three perspectives disagree such as the beginning of Period 7A in Act 2, or the full Period 8 in Act 3. Moreover, we observe regions where KS and AL agree (red), but CM indicates a step-wise transition (cyan), such as the one in Period 11 of Act 3 starting from B minor (+2) and gradually modulating via G major (mm. 1355–1357), A minor (mm. 1368–1375), and G minor (mm. 1396–1400) towards −3 (reached in m. 1427).

Next, we have a look at the third opera, *Siegfried* WWV 86 C (Figure 7), whose first act spans 2,983 measures being by far the longest act of the *Ring*. We still observe passages of high consistency among all perspectives,

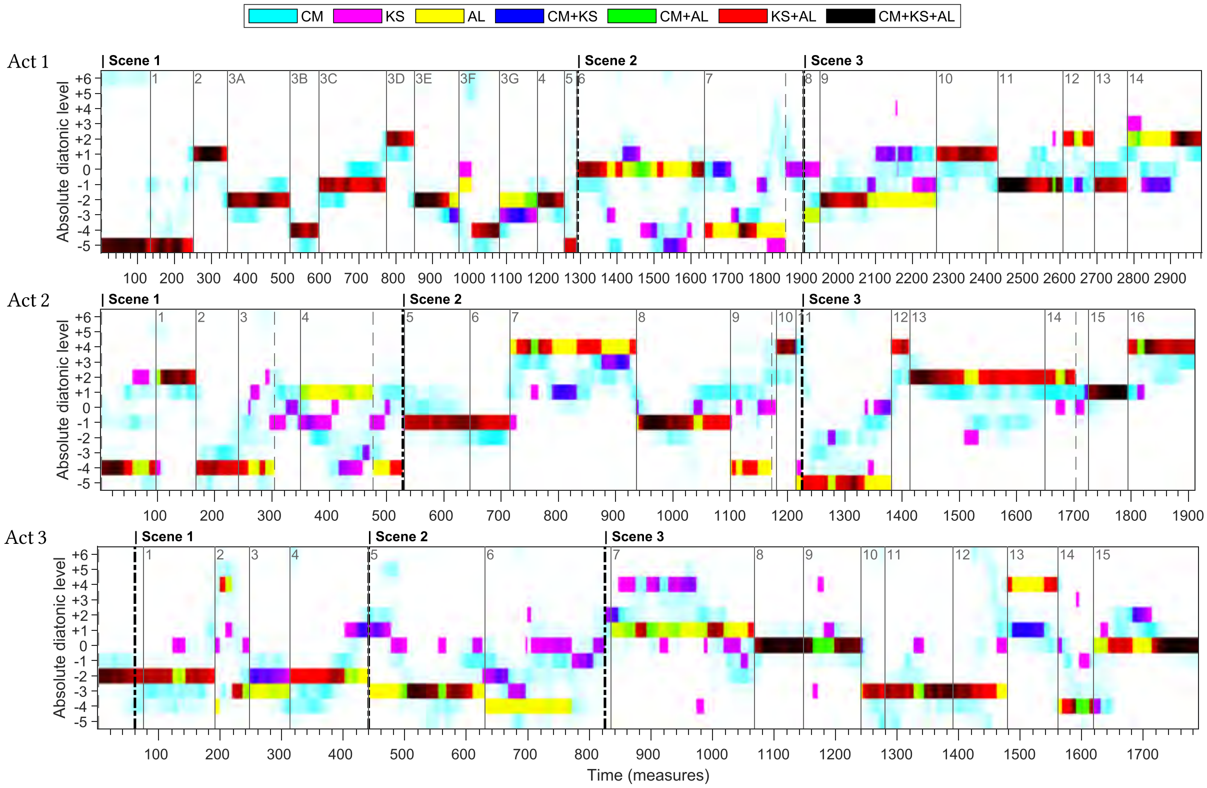


Fig. 7. *Siegfried* WWV 86 C, structured into three acts. Visualization parameters are $w = 64$ and $\beta = 50$.

especially within the first scene of Act 1 and the end (m. 1067 onwards) of Act 3, thus delivering a "stable frame" for this opera. In between—starting from Scene 2 of Act 1—we find far more complex structures. This is already indicated by the substantially larger number of key signature changes. For instance, Period 6 of Act 1 (mm. 1289–1634) comprises nine different key signatures revolving around the constant "principal tonality" C major as indicated by AL. Similar or even more extreme situations can be found in Period 7 of Act 1 or Period 7 of Act 2. Looking at CM, we find considerably more passages of unclear tonality, pointing towards more chromatic harmony or fast modulations. For instance, in Period 1 in Act 3, CM indicates a tonally vague region from −4 to −2. Similarly, Period 11 in Act 2, is characterized by a continuum of diatonic scales ranging from −5 to 0. Only rarely, we find the typical structure of Lorenz's periods, for example, in Periods 9 and 15 in Act 3 (both with frame key C major). Other segments, such as Period 15 in Act 2 and Period 8 in Act 3, show high consistency but lack the typical modulation for a period.

Let us finally turn to the last and longest opera *Götterdämmerung* WWV 86 D (Figure 8). Already in the Prologue (which has a separate measure counting, therefore considered as an individual act), we observe a clear discrepancy between Lorenz's idealized frame keys (E♭ minor, +6, and E♭ major, −3) and the key signatures used by Wagner (revealing many changes). CM often follows KS (e.g., in mm. 109–153 or mm. 670–786), but sometimes differs substantially from both KS and AL. One such instance occurs in mm. 470–500, where KS alternates between −3 and 0 while CM highlights +5 and +4. A manual inspection of this passage reveals a clear and stable section (mm. 488–513) in +5 (initially notated as C♭ major with accidentals within key signature −3, then enharmonically changed to B major within key signature 0), framed by complex chromatic modulations. This is an example

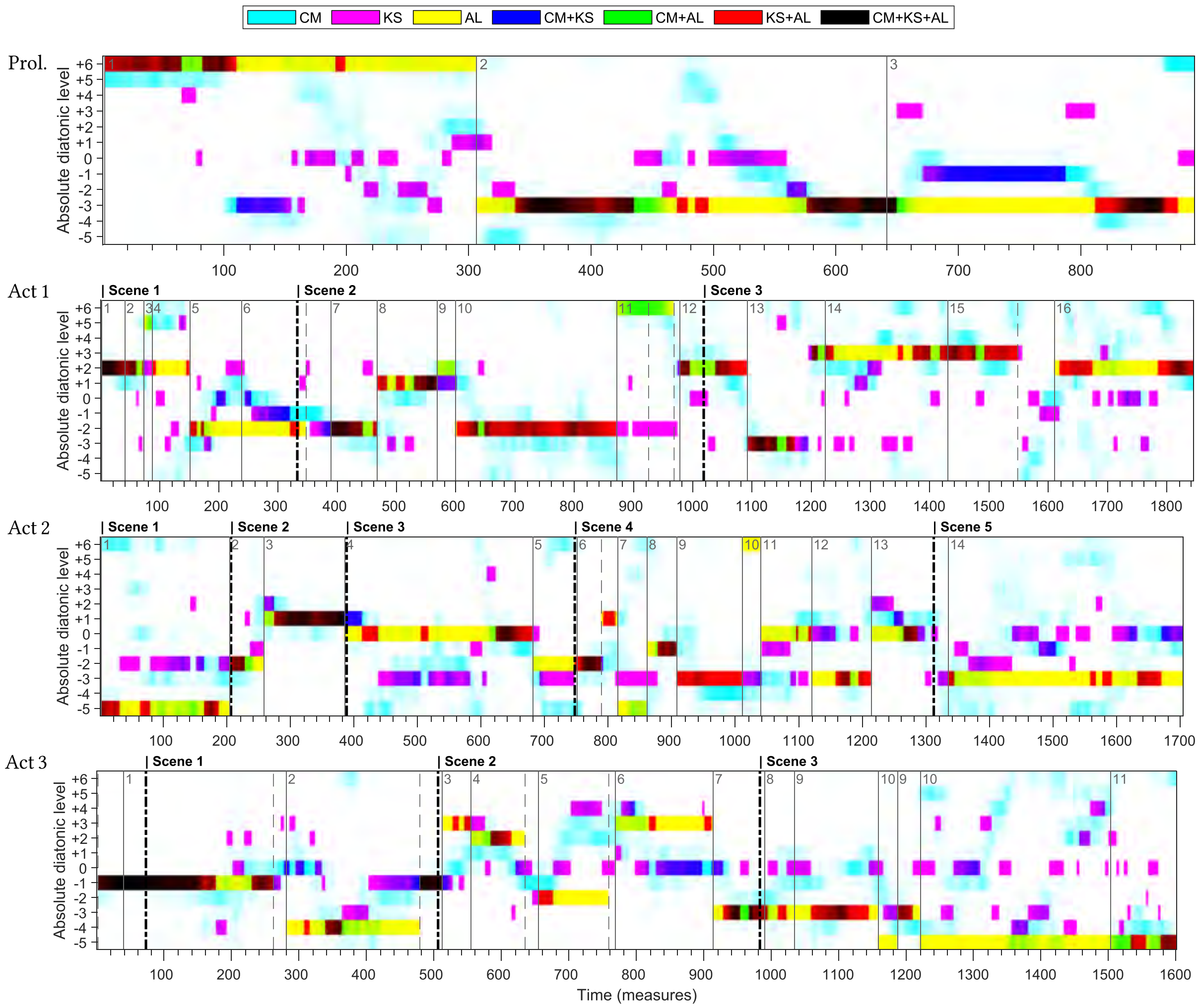


Fig. 8. *Götterdämmerung* WWV 86 D, structured into a *Prologue* and three acts. Visualization parameters are $w = 64$ and $\beta = 50$.

where CM indicates the prevalent and clearly perceivable tonality, while KS and AL (for different reasons) do not explicitly indicate this tonality. Reviewing the whole Prologue, we repeatedly find consistent regions for +6 (enharmonically equivalent to −6 as indicated by KS and AL (E♭ minor)), and −3 (E♭ major), underscoring the importance of a frame tonality (or "frame tonic," E♭) in structuring this whole part. A remarkable situation exists at the Prologue's conclusion, where all perspectives differ substantially: After a consistent −3 passage (E♭ major, mm. 840–860), CM changes to +6 (focusing on the E♭ minor prevalent in mm. 864–879) while KS changes to 0, facilitating the transition to B minor (+2, the cadence to the following Act 1). In contrast, AL does not indicate a change for this transitional passage, remaining in the frame tonality −3.

In contrast to the Prologue, Acts 1–3 are clearly longer and more complex. Overall, we find much less agreement among the three perspectives as compared to the previous operas. For instance, the second half of Act 2 and the final scene of Act 3 almost completely lack consistent regions, except for the very end in D♭ major (−5). Similar to *Siegfried*, we again find passages with many KS changes (e.g., Periods 4, 14, and 16 in Act 1) and regions of rather vague tonality in CM (e.g., Period 10 in Act 1 or the final scene of Act 2). Moreover, we observe interesting transitions from CM (Act 3, mm. 1320), leaving A major via E major and a complex chromatic modulation process to reach D♭ major (−5) in m. 1363. Summarizing some of the discussed passages, where KS was actually quite far from the analyzed tonality, we hypothesize that—after the long break from 1857–1869 during composing Act 2 of *Siegfried*—Wagner started to use KS 0 more often not because these passages feature the actual 0 diatonic (C major or A minor). Instead, they might be as complex (going beyond diatonic tonality) that there is no actually suitable key signature, thus going for the simplest one and incorporating all alterations as in-measure accidentals.

In summary, we found several indications that Wagner's tonal language gets more complex over the course of the *Ring*, an observation that might relate to the long break (1857–1869) within its composition (pausing at the second act of *Siegfried*). While there are stable and consistent passages throughout all operas (such as the first 160 measures of the final act), the inconsistent, vague, and transitional passages occur more often towards the end of the cycle. Concerning the realization of the poetico-musical period idea, we find some clear examples for such periods. However, by far not all the periods indicated by Lorenz in [26] show the clear tonal structure implied by this concept. This suggests that Wagner may have had the concept in mind but did not strictly adhere to it throughout the entire *Ring*, exposing Lorenz's attempt to explain the whole work through such periods as overambitious.

Concluding on the methodology, we summarize that our combined view yields reliable insights into the tonal organization of the *Ring* and the discrepancy of the different perspectives, thus enabling a multi-faceted and differentiated musicological discourse.

3.5 Musicological Discourse: An Exemplary Discussion

As pointed out above, the focus of our contribution is on the comparative visualization strategy. Nevertheless, we want to exemplify how these visualizations can be utilized for musicological discourse. Choosing the last scene of *Siegfried* (Act 3, Scene 3), which plays a central role in the drama, as an example (Figure 7), we want to highlight how the three perspectives can help understanding Wagner's shaping of form through tonality (partly following [38]). According to Lorenz [27], this scene is approached with a "transformational music" (mm. 770–833). It is followed by Period 7 (mm. 834–1066), which is an extreme case for the discrepancy of the perspectives as mentioned above: Within only one AL segment, denoting E minor as principal tonality (yet also indicating intermediate segments in E major without further specifying their location), no less than 12 key signature regions occur. In this period, an average KS region lasts less than 20 measures as opposed to 42 measures averaged over the full *Ring* (see Table 1). The predominant scales measured by CM oscillate between these two annotations (KS and AL), which rarely coincide. For instance, CM agrees with AL around mm. 900–920 and with KS around mm. 940–970. Nevertheless, both CM and KS circulate around the +1 scale annotated by AL, which therefore may be indeed a "center," yet with only few regions actually *in* that tonality. These observations are in line with the dramatic development of the scene: Its first part (Period 7) corresponds to a monologue by Siegfried, describing the process of discovering Brünnhilde. While the outer process of this scene is stable and coherent (matching the principal tonality by AL), there is an intensive inner process of Siegfried experiencing fear for the first time when facing a woman (Brünnhilde). This inner conflict between love and fear is expressed by the many intermediate keys and transitions reflected by both KS and CM. The second, considerably longer part of Scene 3 (mm. 1067ff.) unfolds in three sub-parts: The outer two are framed by a clear and stable 0 diatonic regions (C major), where all perspectives agree. Both are duets between Brünnhilde and Siegfried, corresponding in terms of affect and structure: the first (mm. 1067ff.) reflects Brünnhilde's passionate yet restrained turn towards Siegfried, while the second (mm. 1719ff.) is the moment of their "ultimate affirmation" [37]. Between these two passages lies a transitional section, which we can interpret as a process of overcoming fear. Accordingly, the dramatic turning points of the scene are shaped by fear-related affects: first, Siegfried's fear of sexual intimacy (mm. 930ff.) and his uncertainty or lack of understanding (mm. 1151–1278); then Brünnhilde's experience of sexual fear (mm. 1279–1369), followed by her fear of losing her divinity and knowledge (mm. 1361–1478) [14]. The most important phase of these negative feelings are Periods 10–12, where both protagonists mourn the loss of their former identities. This phase stands out from the surrounding 0 scale by being centered around −3, confirmed by all three perspectives yet with slightly less stability, as indicated by KS and CM. The final transition towards reaching the "ultimate affirmation" is then described differently by the three perspectives; yet, their interplay shows the procedural nature of how Wagner shapes this transition until reaching the final C major.

While conducting such a detailed musicological discourse for the entire *Ring* is beyond the scope of this article, the example of this final scene of *Siegfried* showed how the interplay of the different perspectives unveils Wagner's shaping of form through tonality. Stable phases of the drama are coined by an agreement of all perspectives, both the more idealized ones (AL) and the more descriptive ones (KS, CM). In contrast, phases of development are characterized by discrepancy between the "principal tonality" highlighted by AL and the actual transition process comprising modulations and a variety of shorter tonal phases (KS, CM). The exact assignment of a key or scale to each segment is less relevant (and hard to do even for experts). However, the overall picture arising from these perspectives clearly supports Newcomb's reading that musical form in Wagner's works is best understood as a "dynamic musical process" shaped by dramatic development [30]. These readings are supported by the results of a recent study [38], which takes further musical parameters (tempo and singers' activity) into account for a multi-dimensional visualization.

4 Quantitative Comparison

As discussed in Section 3, we consider the three analyses as different *perspectives* on the *Ring*, which have been created under different assumptions and with different goals. However, they are still comparable and should be similar in the way that all of them can be directly projected to the 12 diatonic scales. For this reason, we expect them to at least coincide for long and rather clear tonal regions. Therefore, while we compared these analyses in a visual way in Section 3, we now complement this with a quantitative comparison of the three analyses (Section 4.1). We emphasize that such measurements are not an "evaluation" (of the analyses' quality) but rather an attempt to quantify their mutual similarity, highlighting in which way they differ from each other and how we would have to select the parameters of CM to get closer to one of those. On this basis, we finally use both the visual and the quantitative approach for investigating the influence of the visualization parameters for CM in detail (Section 4.2).

4.1 Musical Similarity Measures

For a quantitative comparison of the different perspectives, we interpret their result for a given measure as a histogram over a 12-bin circular axis (musically corresponding to the circle of fifths). In [48], we already proposed a measure for comparing such distributions based on the **Jensen–Shannon (JS)** divergence $\mathcal{J}$—a symmetrized version of the Kullback–Leibler divergence $\mathcal{K}$. Following [48], we calculate this metric by interpreting the vectors $\mathbf{a} \in \mathbb{R}^{12}$ (columns of $\mathcal{A}$) as probability distributions. We compute the JS divergence between two such vectors $\mathbf{a}, \mathbf{a}' \in \mathbb{R}^{12}$ with $\|\mathbf{a}\|_1 = \|\mathbf{a}'\|_1 = 1$ as follows:

$$\mathcal{J}(\mathbf{a}, \mathbf{a}') = \frac{1}{2}\left(\mathcal{K}(\mathbf{a}, \mathbf{m}) + \mathcal{K}(\mathbf{a}', \mathbf{m})\right), \tag{4}$$

with $\mathbf{m} = 1/2(\mathbf{a} + \mathbf{a}')$. We invert this to define the JS *similarity* $\mathcal{S}$ as:

$$\mathcal{S}(\mathbf{a}, \mathbf{a}') = 1 - \mathcal{J}(\mathbf{a}, \mathbf{a}'), \tag{5}$$

yielding a value $\mathcal{S}(\mathbf{a}, \mathbf{a}') \in [0, 1]$ with $\mathcal{S}(\mathbf{a}, \mathbf{a}') = 1$ if $\mathbf{a} = \mathbf{a}'$. For comparing two measurements of a whole act or the full *Ring*, we calculate the pairwise JS similarity for each musical measure and finally average over all measures.

In [48], the JS similarity turned out useful to compare analyses based on different representations. However, there is a drawback regarding its musical meaning. While the JS similarity is invariant to permutations of the histograms (i.e., to shuffling the order of histogram bins in the same way for both $\mathbf{a}$ and $\mathbf{a}'$), the order of bins crucially matters for capturing similarity relations between diatonic scales over the circle of fifths. Here, neighboring (fifth-related) scales such as +1 and +2 are musically similar (sharing 6 of 7 pitch classes) while distant (e.g., chromatically shifted) scales such as +1 and +6 are dissimilar (sharing only 2 of 7 pitch classes). The JS similarity, however, yields the same value for both relationships. To account for this, we propose another similarity measure specifically suitable for the circle-of-fifth relationships. To this end, we consider a circular variant of the earth mover's distance [34], which is based on the idea of "optimal transport," that is, estimating the amount of "probability mass" to be moved when turning one distribution into another. For the circular case, this transport may occur in both directions around the circle, with the optimality criterion always favoring the "shorter" direction. Adapting the definitions from [34] to our scenario—that is, two normalized histograms $\mathbf{a}, \mathbf{a}' \in \mathbb{R}^{12}$—, we compute this circular similarity based on the earth mover's distance (CEMD similarity) $C(\mathbf{a}, \mathbf{a}') \in [0, 1]$ as follows. First, we define the cumulative histogram $\mathbf{h}_k(\mathbf{a}) \in \mathbb{R}^{12}$ starting at bin k as:

$$h_{k,i}(\mathbf{a}) = \begin{cases} \sum_{j=1}^{i} a_j & \text{if } i \geq k \\ \sum_{j=k}^{12} a_j + \sum_{j=1}^{i} a_j & \text{if } i < k \end{cases}, \tag{6}$$

for $i, k \in \{0, \dots, 11\}$. Similarly, we compute $\mathbf{h}_k(\mathbf{a}') \in \mathbb{R}^{12}$. With this, we calculate our similarity measure as:

$$C(\mathbf{a}, \mathbf{a}') = 1 - 2 \cdot \min_{k \in \{0,\dots,11\}} \frac{1}{12} \sum_{i=0}^{11} |h_{k,i}(\mathbf{a}) - h_{k,i}(\mathbf{a}')|. \tag{7}$$

One can show that, with this definition, we obtain $C(\mathbf{a}, \mathbf{a}') = 1$ if and only if $\mathbf{a} = \mathbf{a}'$. At the other extreme, one obtains $C(\mathbf{a}, \mathbf{a}') = 0$ in the case that $\mathbf{a}$ and $\mathbf{a}'$ are two one-hot distributions on exactly opposite positions on the circle, e.g., $\mathbf{a} = (0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0)^\top$ and $\mathbf{a}' = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1)^\top$, which musically corresponds to scales 0 and +6, respectively. For any other pair of sparse distributions (with a probability of one at exactly one scale), we obtain a similarity of $C = 1 - d/6$ with d being the shortest possible distance of the two peaks on the circle of fifths. For instance, comparing such sparse distributions for scales +1 and +2 ($d = 1$) yields $C = 5/6$, comparing scales −3 and +1 ($d = 4$) yields $C = 2/6$, and comparing scales −5 and +6 ($d = 1$ due to the cyclic definition) yields $C = 5/6$. For less sparse distributions, indicating non-zero probabilities for several (usually neighboring) scales, the similarities increase since less "probability mass" has to be moved. With these properties, the CEMD similarity measure well reflects the musical similarity of diatonic scale probabilities.

4.2 Investigating Visualization Parameters

We now want to utilize the two proposed similarity measures for quantitatively comparing the different analytical perspectives (CM, KS, AL) discussed in Section 3. For a pair of analyses, we therefore compute the respective similarity (JS or CEMD) for each musical measure and finally average over all measures of the full *Ring* cycle. In particular, this quantification allows us to systematically investigate the role of the two visualization parameters for the CM, that is, the window size w and the inverse softmax temperature β.

Figure 9(a) shows the pairwise JS similarities for two comparisons (CM-KS, CM-AL) as a function of the window size (left) and the inverse softmax temperature (right), respectively. Regarding the window size (left plot, with $\beta = 50$), we observe a peak for both curves yet at different positions: The CM shows the highest similarity with the key signature regions (KS) at $w = 48$. In contrast, the similarity of CM with Lorenz's annotations (AL) is maximal for $w = 96$ and slowly decreases for larger windows. These "optimal" window sizes closely correspond to the typical segment lengths indicated in Table 1 (mean of 42 measures for KS, 100 measures for AL). Using the CEMD similarity (Figure 9(b)), one can observe a similar trend. Interestingly, however, the maximal position for CM-KS is at a smaller window size of $w = 32$ (closer to the median length for KS in Table 1) while for CM-AL,

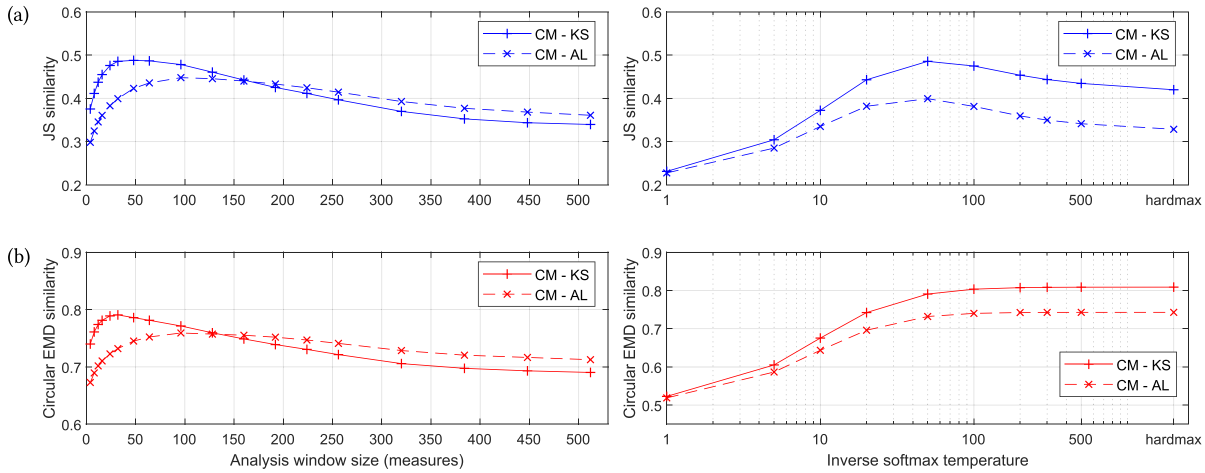


Fig. 9. Pairwise comparison of scale probabilities based on CMs with different annotations (key signature regions KS, Lorenz annotations AL). The figure shows sweeps over different window sizes (left) and over different softmax temperature resp. "hard" argmax (right). (a) Evaluation using the JS similarity measure. (b) Evaluation using the CEMD similarity measure based on the circular earth mover's distance. Measure-wise computation averaged over all measures of the *Ring* cycle.

the maximizing position is at a similar window size as for JS. Hence, with the CEMD similarity, we find an even stronger difference between KS and AL regarding their temporal context. From both curves, we conclude that KS describes the pitch class content of medium-sized sections, around 30–50 measures, while AL assumes an idealized perspective focusing on sections of 100–200 measures. By suitable choosing the window size parameter w, we can adjust CM's primary focus to align with either of these perspectives.

We now want to investigate how the choice of w affects the visualizations. To this end, we show in Figure 10 our comparative visualization for four different window sizes ranging from $w = 12$ to $w = 224$, using the first act of *Die Walküre* as an example. For the short window size of $w = 12$ (Figure 10(a)), CM clearly focuses on the fine-scale tonal regions and transitions, often deviating from KS and AL. Increasing the window size to $w = 32$ (Figure 10(b)), many of these short passages disappear with CM resembling KS more closely as indicated by black and blue colors (e.g., in Periods 3 or 12Ca). This begins to change when we further increase the window size to $w = 64$ (Figure 10(c)). Here, we repeatedly observe regions where CM and AL coincide (green) while KS differs (magenta), see for instance Periods 6 and 12Aa. For a very large window size of $w = 224$ (Figure 10(d)), this trend is further reinforced, with CM now often displaying an overly smooth behavior that substantially obscures local information. From these visualizations, as Figure 9 also indicates, the window size of 64 measures chosen as default for CM in Section 3 turns out to be a meaningful compromise between the temporal resolution of the other two perspectives, KS and AL.

Finally, let us discuss the influence of the softmax parameter β, which adjusts the level of enhancement (focus on an individual scale). The pairwise similarities in the two right plots of Figure 9 (with a logarithmic axis and the hard maximum as the extreme case $\beta \to \infty$), where we use $w = 32$ throughout. Both similarity measures (JS and CEMD) clearly indicate that an inverse softmax temperature of at least $\beta = 20$ is necessary to obtain meaningful results. Towards larger β, however, the two measures differ substantially. For JS, we observe a peak at $\beta = 50$ after which the similarity of CM with both other perspectives (KS and AL) decreases. In contrast, the CEMD similarity (lower right plot) for both pairs further increase and then plateaus at $\beta = 100$. This discrepancy can be explained by the increased flatness of CM, which becomes smoothed across neighbor scales as β grows. This "smoothing along the circle of fifths" affects the JS similarity since less probability mass is assigned to the

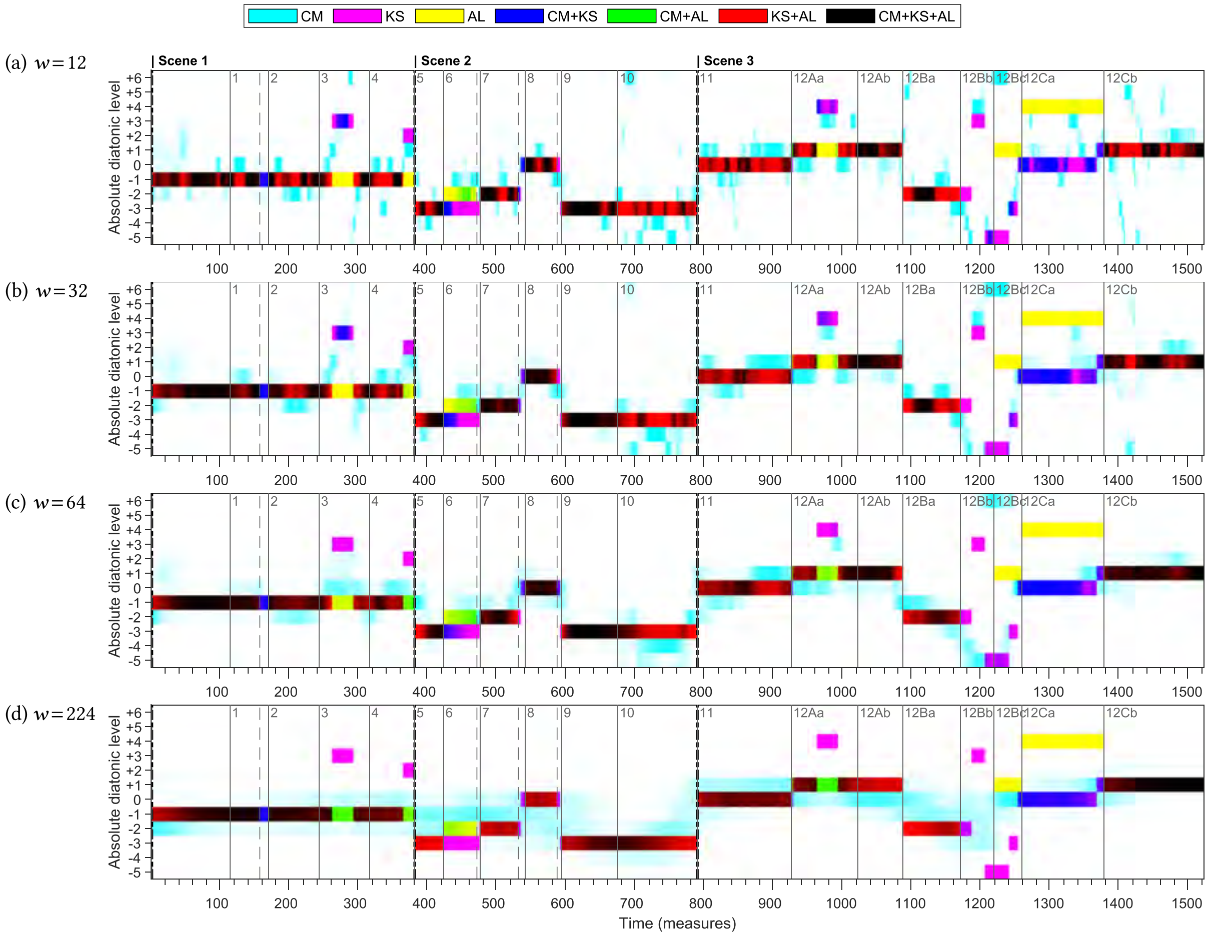


Fig. 10. Comparative visualization of diatonic scale measurements for the act WWV 86 B-1 for Karajan version using a window size w of (a) 12 measures, (b) 32 measures, (c) 64 measures, (d) 224 measures. Softmax parameter $\beta = 50$.

"correct" scales (indicated by KS or AL). On the other hand, the CEMD similarity also considers the close musical relationships between neighboring scales and is thereby less affected by the strong softmax rescaling. From the two experiments, we conclude that $\beta = 50$, our default choice for CM in Section 3, constitutes a good compromise.

Overall, our quantitative comparisons have underlined the suitability of the computational strategy (CM) for highlighting interesting tonal structures in large-scale musical works such as the *Ring*. Moreover, choosing the visualization parameters offers high flexibility, allowing the focus to shift between different temporal granularities from detailed modulations up to the coarse-scale tonal structure. This makes it possible to retrace, examine, and contextualize different analytical perspectives, as demonstrated for two specific examples (KS and AL).

5 Conclusions

In this article, we utilized a recently proposed computational method for measuring tonal structures and demonstrated its application to musical works of large scale. As a concrete example, we analyzed the large-scale tonal organization of Richard Wagner's *Ring* cycle. To this end, we compared three perspectives on the work: a computational measurement of diatonic scales from audio recordings, the key signature regions chosen by

Wagner, and an idealized analysis by musicologist Alfred Lorenz. We visually compared these perspectives using a color-based overlay technique and discussed the visualizations for the entire *Ring* cycle. Moreover, we conducted a quantitative comparison where we considered a novel, musically inspired similarity measure. Our results show that combining the three perspectives contributes to a better understanding of the *Ring*'s tonal structure. They complement each other by focusing on different temporal granularities, from a few up to hundreds of measures. We conclude that a joint visualization as conducted in this article provides a unique overview of large-scale works. From a broader perspective, the potential of such computer-assisted strategies highlights how an interdisciplinary dialogue between computer science and musicology can be beneficial to both disciplines.

Acknowledgments

We thank Vlora Arifi-Müller for help in preparing data and annotations. The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer Institute for Integrated Circuits IIS.

References

[1] Rachel M. Bittner, Brian McFee, Justin Salamon, Peter Li, and Juan P. Bello. 2017. Deep salience representations for F0 tracking in polyphonic music. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 63–70. DOI: https://doi.org/10.5281/zenodo.1417937

[2] Werner Breig. 2002. Wagners begriff der "dichterisch-musikalischen periode." In "*Schlagen Sie Die Kraft Der Reflexion Nicht zu Gering an." Beiträge zu Richard Wagners Denken, Werk Und Wirken*. Klaus Döge, Christa Jost, and Peter Jost (Eds.), Mainz, 158–172.

[3] Matthew Bribitzer-Stull. 2015. *Understanding the Leitmotif* . Cambridge University Press.

[4] Wei Chai and Barry Vercoe. 2005. Detection of key change in classical piano music. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 468–474.

[5] Tsung-Ping Chen and Li Su. 2018. Functional harmony recognition of symbolic music data with multi-task recurrent neural networks. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 90–97.

[6] Carl Dahlhaus. 1965. Wagners Begriff der "dichterisch-musikalischen Periode." In *Beiträge Zur Geschichte Der Musikanschauung im 19. Jahrhundert*. Walter Salmen (Ed.), Bosse, 179–194.

[7] Carl Dahlhaus. 1969. Formprinzipien in Wagners "Ring des Nibelungen." In *Beiträge zur Geschichte der Oper*. Heinz Becker (Ed.), Bosse, 95–129.

[8] Carl Dahlhaus, Klaus Döge, and Egon Voss (Eds.). 1970–2020. *Richard Wagner–Gesamtausgabe*. Schott Music.

[9] W. Bas de Haas, José Pedro Magalhães, Remco C. Veltkamp, and Frans Wiering. 2011. HARMTRACE: Improving harmonic similarity estimation using functional harmony analysis. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 67–72.

[10] Laurence Dreyfus and Carolin Rindfleisch. 2014. Using digital libraries in the research of the reception and interpretation of Richard Wagner's leitmotifs. In *Proceedings of the International Workshop on Digital Libraries for Musicology*, 1–3.

[11] Sebastian Ewert, Meinard Müller, Verena Konz, Daniel Müllensiefen, and Gerraint A. Wiggins. 2012. Towards cross-version harmonic analysis of music. *IEEE Transactions on Multimedia* 14, 3–2 (2012), 770–782.

[12] Takuya Fujishima. 1999. Realtime chord recognition of musical sound: A system using common lisp music. In *Proceedings of the International Computer Music Conference (ICMC)*, 464–467.

[13] Zsolt Gárdonyi and Hubert Nordhoff. 2002. *Harmonik* (2nd ed.). Möseler.

[14] Katharina Hottmann. 2011. Fürchten oder Nicht-Fürchten? Zur Dramaturgie der Schluss-Szene des 'Siegfried.' In *Wagners Siegfried Und Die (Post-)Heroische Moderne*. Tobias Janz (Ed.), Königshausen & Neumann, 291–324.

[15] Özgür İzmirli. 2007. Localized key finding from audio using nonnegative matrix factorization for segmentation. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 195–200.

[16] Rainer Kleinertz. 2014. Richard Wagners Begriff der "dichterisch-musikalischen Periode." *Die Musikforschung* 67 (2014), 26–47.

[17] Verena Konz, Meinard Müller, and Rainer Kleinertz. 2013. A cross-version chord labelling approach for exploring harmonic structures—A case study on Beethoven's Appassionata. *Journal of New Music Research* 42, 1 (2013), 61–77.

[18] Hendrik Vincent Koops, W. Bas de Haas, John Ashley Burgoyne, Jeroen Bransen, Anna Kent-Muller, and Anja Volk. 2019. Annotator subjectivity in harmony annotations of popular music. *Journal of New Music Research* 48, 3 (2019), 232–252. DOI: https://doi.org/10.1080/09298215.2019.1613436

[19] Andreas Kornstädt. 2001. The JRing system for computer-assisted musicological analysis. In *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*, 93–98.

[20] Filip Korzeniowski and Gerhard Widmer. 2016. Feature learning for chord recognition: The deep chroma extractor. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 37–43. DOI: https://doi.org/10.5281/zenodo.1416314

[21] Filip Korzeniowski and Gerhard Widmer. 2018. Genre-agnostic key classification with convolutional neural networks. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 264–270.

[22] Michael Krause and Meinard Müller. 2022. Hierarchical classification of singing activity, gender, and type in complex music recordings. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 406–410.

[23] Michael Krause, Meinard Müller, and Christof Weiß. 2021. Singing voice detection in opera recordings: A case study on robustness and generalization. *Electronics* 10, 10 (2021), 1–14. DOI: https://doi.org/10.3390/electronics10101214

[24] Michael Krause, Meinard Müller, and Christof Weiß. 2021. Towards leitmotif activity detection in opera recordings. *Transactions of the International Society for Music Information Retrieval (TISMIR)* 4, 1 (2021), 127–140. DOI: https://doi.org/10.5334/tismir.116

[25] Néstor Nápoles López, Laurent Feisthauer, Florence Levé, and Ichiro Fujinaga. 2020. On local keys, modulations, and tonicizations: A dataset and methodology for evaluating changes of key. In *Proceedings of the International Conference on Digital Libraries for Musicology (DLfM)*, 18–26.

[26] Alfred O. Lorenz. 1924. *Der musikalische Aufbau des Bühnenfestspiels Der Ring des Nibelungen*. Das Geheimnis Der Form Bei Richard Wagner, Vol. 1, M. Hesse, Berlin.

[27] Stephen McClatchie. 1998. *Analyzing Wagner's Operas: Alfred Lorenz and German Nationalist Ideology*. University of Rochester Press.

[28] Lesley Mearns. 2013. *The Computational Analysis of Harmony in Western Art Music*. Ph.D. Dissertation. Queen Mary University of London, UK.

[29] Daniel Müllensiefen, David Baker, Christophe Rhodes, Tim Crawford, and Laurence Dreyfus. 2016. Recognition of leitmotives in Richard Wagner's music: An item response theory approach. In *Analysis of Large and Complex Data*. Springer, 473–483. DOI: https://doi.org/10.1007/978-3-319-25226-1

[30] Anthony Newcomb. 1981. The birth of music out of the spirit of drama: An essay in Wagnerian formal analysis. *19th-Century Music* 5 (1981), 38–66.

[31] Yizhao Ni, Matt McVicar, Raúl Santos-Rodríguez, and Tijl De Bie. 2013. Understanding effects of subjectivity in measuring chord estimation accuracy. *IEEE Transactions on Audio, Speech, and Language Processing* 21, 12 (2013), 2607–2615.

[32] Kevin R. Page, Terhi Nurmikko-Fuller, Carolin Rindfleisch, David M. Weigl, Richard Lewis, Laurence Dreyfus, and David De Roure. 2015. A toolkit for live annotation of opera performance: Experiences capturing Wagner's Ring cycle. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 211–217.

[33] Hélène Papadopoulos and Geoffroy Peeters. 2012. Local key estimation from an audio signal relying on harmonic and metrical structures. *IEEE Transactions on Audio, Speech, and Language Processing* 20, 4 (2012), 1297–1312.

[34] Julien Rabin, Julie Delon, and Yann Gousseau. 2008. Circular earth mover's distance for the comparison of local features. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*.

[35] Carolin Rindfleisch. 2016. Modelling music reception: An ontology for representing interpretations of Richard Wagner's leitmotifs. In *Proceedings of the International Conference of the Alliance of Digital Humanities Organizations (DH)*, 333–336.

[36] Johannes Schild. 2010. "Zum Raum wird hier die Zeit"–Tonfelder in Wagners Parsifal. In *Funktionale Analyse: Musik–Malerei–Antike Literatur*. Bernhard Haas and Bruno Haas (Eds.), Olms, 313–373.

[37] Jürgen Schläder. 1995. *Das Opernduett: ein Szenentypus des 19. Jahrhunderts und seine Vorgeschichte*. Studien Zur Geschichte Und Theorie Der Dramatischen Künste, Vol. 6. Max Niemeyer.

[38] Pascal Schmolenzky, Stephanie Klauk, Rainer Kleinertz, Christof Weiß, and Meinard Müller. 2025. A multidimensional approach to opera analysis: Harmony, tempo, and dramatic interaction in Wagner's Siegfried. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*.

[39] Hendrik Schreiber and Meinard Müller. 2019. Musical tempo and key estimation using convolutional neural networks with directional filters. In *Proceedings of the Sound and Music Computing Conference (SMC)*, 47–54.

[40] David Temperley. 1997. An algorithm for harmonic analysis. *Music Perception* 15, 1 (1997), 31–68.

[41] Dmitri Tymoczko. 2009. Three conceptions of musical distance. In *Proceedings of Bridges 2009: Mathematics, Music, Art, Architecture, Culture*. Tarquin Publications, 29–38.

[42] Steven van de Par, Martin F. McKinney, and André Redert. 2006. Musical key extraction from audio using profile training. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 328–329.

[43] Richard Wagner. 1892–1899. *On the Application of Music to the Drama. Prose Works*, Vol. 6. Kegan Paul, Trench, Trübner & Co., London, UK, 173–192.

[44] Richard Wagner. 1995. *Opera and Drama*. University of Nebraska Press. Translation of the original edition from 1851.

[45] Christof Weiß, Vlora Arifi-Müller, Michael Krause, Frank Zalkow, Stephanie Klauk, Rainer Kleinertz, and Meinard Müller. 2023. Wagner Ring dataset: A complex opera scenario for music processing and computational musicology. *Transactions of the International Society for Music Information Retrieval* 6, 1 (2023), 135–149. DOI: https://doi.org/10.5334/tismir.161

[46] Christof Weiß, Vlora Arifi-Müller, Thomas Prätzlich, Rainer Kleinertz, and Meinard Müller. 2016. Analyzing measure annotations for Western classical music recordings. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 517–523.

[47] Christof Weiß and Julian Habryka. 2014. Chroma-based scale matching for audio tonality analysis. In *Proceedings of the Conference on Interdisciplinary Musicology (CIM)*, 168–173.

[48] Christof Weiß and Meinard Müller. 2024. From music scores to audio recordings: Deep pitch-class representations for measuring tonal structures. *Journal on Computing and Cultural Heritage* 17, 3, Article 45 (2024), 1–19.

[49] Christof Weiß and Geoffroy Peeters. 2022. Comparing deep models and evaluation strategies for multi-pitch estimation in music recordings. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 30 (2022), 2814–2827. DOI: https://doi.org/10.1109/TASLP.2022.3200547

[50] Christof Weiß, Hendrik Schreiber, and Meinard Müller. 2020. Local key estimation in music recordings: A case study across songs, versions, and annotators. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 28 (2020), 2919–2932. DOI: https://doi.org/10.1109/TASLP.2020.3030485

[51] Christof Weiß, Frank Zalkow, Meinard Müller, Stephanie Klauk, and Rainer Kleinertz. 2017. Versionsübergreifende visualisierung harmonischer verläufe: Eine fallstudie zu wagners Ring-Zyklus. In *Proceedings of the Jahrestagung Der Gesellschaft Für Informatik (GI)*, 205–217. DOI: https://doi.org/10.18420/in2017_14

[52] Christof Weiß, Johannes Zeitler, Tim Zunner, Florian Schuberth, and Meinard Müller. 2021. Learning pitch-class representations from score–audio pairs of classical music. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 746–753.

[53] Terry Winograd. 1968. Linguistics and the computer analysis of tonal harmony. *Journal of Music Theory* 12 (1968), 2–49.

[54] Frank Zalkow, Christof Weiß, Thomas Prätzlich, Vlora Arifi-Müller, and Meinard Müller. 2017. A multi-version approach for transferring measure annotations between music recordings. In *Proceedings of the AES International Conference on Semantic Audio*, 148–155.

[55] Yongwei Zhu and Mohan S. Kankanhalli. 2003. Music scale modeling for melody matching. In *Proceedings of the ACM International Conference on Multimedia*, 359–362.

Received 31 October 2024; revised 21 November 2025; accepted 27 February 2026