Structural Analysis of Live Coding Performances Through Novelty-based MIR Methodologies

Francesco Ardan Dal Rí
DISI - Department of Information Engineering and Computer Science, University of Trento (IT)
francesco.dalri-2@unitn.it

Francesco Manenti
Conservatory of Music G. Verdi, Milan (IT)
francesco.manenti@hotmail.com

ABSTRACT

In this paper, we assess a pipeline for structural analysis and segmentation on live coding performances. Traditional Music Information Retrieval methodologies are primarily evaluated on Western classical or popular music. However, the practice of live coding poses unique challenges for analytical tasks, due to its complexity and absence of standardized forms. Furthermore, given the extemporaneous/improvisational nature of live coding, the literature lacks shared methodologies and tools. To address these problems, we tailored a novelty-based analysis pipeline on live coding performances, focusing on four musical features commonly used in the literature; such features were employed to compute Similarity Matrices and identify boundaries between musical sections via novelty curves. The effectiveness of the proposed methodology was evaluated with the involvement of six experienced live coders, who provided their live performance recordings to be analyzed by the algorithm. The musicians reflected on the accuracy and relevance of the segmentation through semi-structured interviews, offering insights into how the segmentation related to their personal practices and musical perspectives. Additionally, we applied our pipeline to a large dataset, highlighting structural patterns and general trends across performances.

1 Introduction

In musicological research domains, computer music analysis is still up for debate and poses numerous challenges, largely due to the vast heterogeneity of musical experiences. While several methodologies have been proposed, as recently examined by Lartillot (2022), analytical tasks need to adapt to different compositional processes that define highly intertwined taxonomies and analytical lenses that may even differ from piece to piece. Overall, there is a tendency to focus on the Schaefferian concept of the 'sound object' (Zattra 2005) through its multi-parametric representations, often through sonograms or manually annotated graphic scores, with software such as Ina GRM's *Acousmographe*[1] or IRCAM's *EAnalysis*[2]. This is motivated by the fact that compositional processes themselves are often based on meticulous exploration of the sound material, thus making the hierarchical aggregation of sound objects not trivial and spread across different levels of functional abstraction, typically supported by Gestalt principles (Terhardt 2016). Consequently, the very meaning of 'musical section' becomes blurred and difficult to define holistically: such challenges are mirrored in automatic analysis algorithms. In the field of Music Information Retrieval (MIR), structural analysis and segmentation of music pieces have long been of primary relevance, applied to tasks like automatic recommendation systems and audio thumbnailing (Nieto et al. 2020). Both supervised and unsupervised techniques are usually evaluated on Western classical or popular repertoire, where they prove particularly effective (Buisson et al. 2022; Wang, Hung, and Smith 2022). However, these musical experiences are often characterized by recognizable patterns in harmony, melody, and rhythm, which tend to repeat throughout the pieces, constituting well-defined sections like verses, choruses, or movements, according to genre-specific functional taxonomies. Moreover, high-level structures retrieving in such contexts also benefits from a long-established tradition that facilitates refining and evaluating such segmentation algorithms (Müller and Kurth 2006).

On the contrary, computer music - and by extension live coding - rarely adheres to such strict rules and exhibits more fluid and unconventional forms, where pieces may develop through subtle modification creating intricate functional hierarchies or textural layers (Couprie 2022). Research in this area is further hampered by the subjective nature of structural

[1] https://inagrm.com/en/showcase/news/203/acousmographe
[2] https://forum.ircam.fr/projects/detail/eanalysis

perception, which involves cognitive processes influenced by experience, habit, and cultural factors, thus resulting in possible discrepancies, e.g. between performers and listeners (Popescu, Widdess, and Rohrmeier 2021). Moreover, the improvisational and extemporaneous nature of live coding tends to shift the research focus onto the performative act, where MIR techniques are usually deployed for creative purposes - e.g. (Xambó, Lerch, and Freeman 2018), rather than for high-level analytical ones. Indeed, only a few recent works addressed *a posteriori* considerations oriented towards musical form and structural development - e.g. (Magnusson 2014b; Diapoulis and Carlé 2023).

In this paper, we're interested in investigating to which extent traditional segmentation techniques adapt to this practice (Section 2). As such, we propose a pipeline for structural segmentation and similarity analysis, exploiting novelty-based algorithms in order to retrieve sections within recorded performances. According to the literature, e.g. (Vatolkin, Koch, and Müller 2021; Paulus and Klapuri 2008), we focus on four common hand-crafted features, namely chroma, Mel-Frequency Cepstral Coefficients (MFCCs), Root-Mean-Square (RMS), and tempo, then computing self-similarity matrices and novelty curves whose peaks define boundaries between adjacent musical parts (Section 3). After fine-tuning over a selection of pieces of different genres (Section 4), we tested our pipeline on live coding performances provided by six experienced live coders. Besides the algorithm itself, which we open-sourced, the core of this contribution lies in the reflections that emerged during the interviews with the live coders involved in the evaluation. Practitioners helped us in validating our methodology and provided insights on how the structural analysis aligns with and diverges from their usual creative process, highlighting relevances and limitations of the proposed audio-driven approach (Section 5). Finally, we also applied our pipeline to a broader dataset collected from a recent streaming event, uncovering structural patterns and trends across live coding performances (Section 6). Results are discussed in Section 7.

2 Background

2.1 Traditional Techniques for Music Structure Segmentation

Within the field of MIR, Music Structure Segmentation tasks concern the automatic subdivision of a musical piece into meaningful sections, such as verses, choruses, or bridges. Such procedures are critical in a variety of applications, such as music recommendation systems, automatic music transcription, audio thumbnailing, and music structure analysis (Murthy and Koolagudi 2018). Music segmentation typically involves high-level hierarchical clustering of musical features around the three principles of segmentation: repetition, homogeneity, and novelty (Maezawa 2019). The majority of segmentation pipelines still operate in an unsupervised fashion, relying on hand-crafted features and a series of algorithms retrieving some sort of similarity metrics between adjacent segments. For instance, Miotto et al. (2007) applied Hidden Markov Models on incipits of tonal classical music; McFee and Ellis (2014) make use of spectral clustering and Laplacians to aggregate structure classes; Marmoret et al. (2023) evaluated different similarity matrices from 80 mel coefficient as input of a correlation block-matching algorithm.

Overall, novelty functions are a common method for retrieving boundaries between adjacent sections, with several segmentation algorithms being indeed based on the seminal work by Foote (2000). Examples include the work by Serra et al. (2014), where novelty curves are extracted from lag matrices of Harmonic Pitch Class Profile (HPCP) features; Hartmann et al. (2017) explored how different features impact novelty-based segmentation; and McMcallum (2019) adopted a hybrid approach, retrieving novelty functions on compressed and meaningful feature embedding extracted by a Convolutional Neural Network, thus directly avoiding features selection.

Albeit all the aforementioned methodologies proved to be particularly accurate, their evaluation has been only conducted on Western popular (in the broad sense) songs or classical music excerpts: to the best of our knowledge, apart from few exceptions - e.g. (Tian and Sandler 2016), repertoire belonging to other musical cultures and experiences are currently overlooked (X. Serra 2011).

2.2 Musical Form in Live Coding Performances

Live coding practice is traditionally characterized by a strong improvisational approach, where the musical discourse develops through a series of operations applied by the performer on his/her code based on extemporaneous ideas (Collins et al. 2003; Brown and Sorensen 2009). The literature strongly emphasizes this attention to the liveness of performative acts, in which formal thinking is implicitly linked to the use of performative strategies or development ideas focused on single patterns in the moment or the near future (Brown and Sorensen 2007; McLean 2014; Dal Rì, Zanghellini, and Masu 2023). Therefore, the musical structure is inseparably related to the code structure, and emerges from processes of juxtaposition, addition, removal, and modification of individual code blocks that organize events into patterns, whose development generally occurs in a linear manner (McLean and Wiggins 2010; McLean 2014). At the current state, reflections on the functional character of higher-level sections with respect to the economy of the entire performance still remain an overall under-explored area.

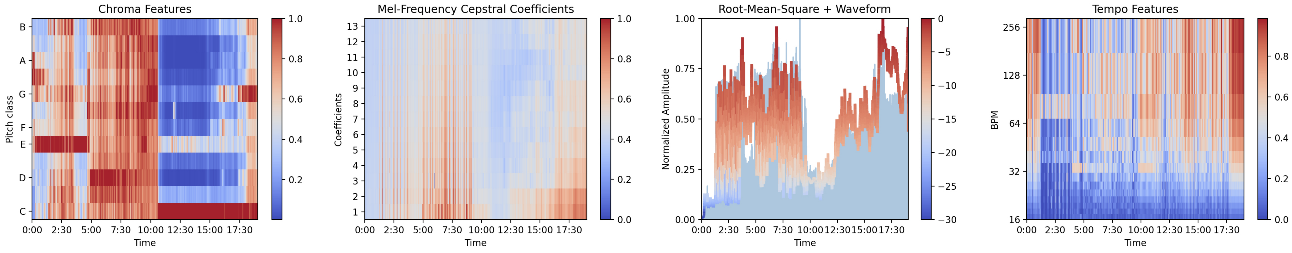

Figure 1: *Examples of the extracted feature. a) Chroma; b) MFCCs; c) RMS; and d) Tempo*

Recently, some works provided broader reasoning on musical form in live coding performances, e.g. through visual representations for fostering its comprehension both in real-time and offline. Magnusson (2011) first tackled the problem from the point of view of the notation, which being intrinsic to the code itself already constitutes an archetypal score of the piece. He further developed such reflections around the Threnoscope (Magnusson 2014b), a system which integrates textual and graphical interface allowing the performer to visualize the form of his/her piece. Dal Rí and Masu (2022) explored two complementary visualization systems allowing live coders to visually recall their performances, thus helping in raising awareness on their own formal choices with respect to the evolution of the piece; similarly, Manesh et al. (2024) developed a version control plugin which inherently facilitates incorporating structural reprises into performances; Diapoulis and Carlé (2023) proposed a framework for reproducible, *a posteriori* analysis of live coding performances, in which onsets and spectral centroids are employed to graphically represent the global development on a circular graph.

3 Analysis Pipeline

We focused our analysis on mid-length performances (~ 10 to 30 minutes), retrieving high-level information by calculating peaks in novelty functions with respect to four different musical features. By doing so, we aimed to identify meaningful differences throughout each piece that could serve as boundaries between distinct musical sections, expressed in terms of novelty peaks. The considered musical pieces are preprocessed by resampling at 16kHz, 16-bit PCM, normalizing at -1dB, and removing possible silence at the beginning and at the end. Then, we extracted the spectrograms via Short-Time Fourier Transform with $n\_fft = 8192$ and $hop\_size = 4096$: therefore, each frame represents 256ms of audio. In addition, we also retrieve harmonic and percussive spectrograms $\mathcal{S}_H$ and $\mathcal{S}_P$, via median-filtering source separation (Fitzgerald 2010).

According to the literature, we choose the most common features which can be representative of multiple aspects of music: chroma features for the pitch, MFCCs for the timbre, RMS for the loudness, and tempo features for the speed. All features have been extracted using librosa[3]; examples are shown in Figure 1. For the sake of consistency, each figure in this section refers to the same test performance by the first author of this paper. For details on implementation, please refer to the code: https://github.com/return-nihil/LC_SEG

3.1 Feature Extraction

3.1.1 Chroma Features

Chroma features allow to represent a musical signal by aggregating each frequency bin of its spectrogram into a feature vector of 12 chroma coefficients, corresponding to the 12 discrete pitches of the Western Chromatic Scale, in an octave-agnostic fashion (Ellis 2007). Such features cannot capture small, microtonal shifts, but provide a good approximation of the harmonic relationships within a musical piece. As such, they have been used extensively for a variety of MIR tasks, e.g. (Ewert, Muller, and Grosche 2009; Bartsch and Wakefield 2005).

A chroma coefficient $\mathcal{C}$ at frame t for a pitch $p \in \mathbf{P} = \{C, C\#, D, \ldots, B\}$ is defined as:

$$\mathcal{C}(t,p) = \sum_{f \in \mathcal{F}_p} |\mathcal{S}_H(t,f)| \tag{1}$$

where $\mathcal{F}_p$ is the set of frequencies associated with p across all the octaves, and $\mathcal{S}_H(t,f)$ is a pitch-based log-frequency spectrogram derived, in our case, from the harmonic one. The entire chromagram is thereby computed by iterating over all $p \in \mathbf{P}$ and $t \in \mathbf{T}$, producing a matrix of shape $P \times T$.

[3] https://librosa.org/

3.1.2 Mel-Frequency Cepstral Coefficients

MFCCs provide an efficient representation of the most significative spectral characteristics of an audio signal (Zheng, Zhang, and Song 2001). They are often referred to as 'spectrum of a spectrum', as they indeed are derived from the Mel-frequency cepstrum, which decorrelates the log-power spectrum of a signal on the Mel scale via a Discrete Cosine Transform. By applying these transformations, the most significant information is concentrated in the first few coefficients. Due to their compactness and robustness, they are the most common feature in audio retrieval and speech processing (Ittichaichareon, Suksri, and Yingthawornsuk 2012; Nagawade and Ratnaparkhe 2017). MFCCs $\mathcal{C}_i$ at frame t are computed as follows:

$$\mathcal{C}_i(t) = \sum_{n=1}^{N_f} \log(\mathcal{S}_n(t)) \cdot \cos\left(i\,(n-0.5)\,\frac{\pi}{N_f}\right), \quad i = 1, \dots, L \tag{2}$$

where N_f is the number of Mel filters, $\mathcal{S}_n(t)$ is the energy of the n-th Mel-spectrum at frame t, and L is the number of cepstral coefficients (in our case, according to most literature, we considered the first 13).

3.1.3 Root-Mean-Square

RMS is a simple metric to measure the loudness of an audio signal (Panagiotakis and Tziritas 2005), typically employed in signal processing. The RMS of a signal amplitude y over a frame t is calculated as follows:

$$\text{RMS}(y_t) = \sqrt{\frac{1}{N}\sum_{i=0}^{N-1} y_t[i]^2} \tag{3}$$

where N is the number of samples in the current frame. Due to the many fluctuations observed, which are less informative from a high-level point of view, we opted for slightly smoothing the curve using a moving average.

3.1.4 Tempo Features

Tempo features represent the magnitudes of different tempis over a given window. They are usually aggregated in tempo bins to create a tempogram, similar to a chromagram. Several ways to compute tempograms exist, here we used the most standard one, which consists of computing the local auto-correlation of the onset envelopes.

Onset envelopes $o(t)$ are retrieved by maximum-filtering a spectral flux of a spectrogram by mean of a half-wave rectifier function - see (Böck and Widmer 2013) for a detailed demonstration; in our case, we processed the percussive spectrogram $\mathcal{S}_P$ to enhance precision and feature separation. The auto-correlation of the onset envelopes R_o at frame t with respect of a lag value τ is then computed over a window of length M:

$$R_o(t,\tau) = \sum_{m=0}^{M-1} \Theta(m) \cdot o(t+m) \cdot o(t+m+\tau) \tag{4}$$

where Θ is a Hann window (Grosche, Müller, and Kurth 2010). A tempogram is then produced by iterating over each time step t producing a lag matrix, and displayed by converting the lag values into BPM. In our case, we set $M = 15$, therefore every frame of the tempogram represents 14 correlation bins over 3.84 seconds.

3.2 Self-Similarity Matrices

Self-Similarity Matrices (SSMs) are often used in MIR to express similarities between feature vectors and highlight recursive patterns and relationships across musical pieces (Foote 1999). Several methodologies for computing the similarity between feature vector exist. In this work, we measured vector-wise Euclidean distance in order to better deal with magnitudes in low-dimensional feature spaces. First, for each feature, we memory-stacked the feature matrix $\mathbf{F}$, incorporating information from the previous 10 steps, thus providing context and improving smoothing. Then, we applied z-score normalization method to obtain $\hat{\mathbf{F}}$, and computed the pairwise Euclidean distance matrix for the transposed feature matrix $\mathbf{D}_{ij} = \sum_k(\hat{\mathbf{F}}_{ki} - \hat{\mathbf{F}}_{kj})^2$ with $k \in [1, n]$ representing the feature index. Finally, we constructed the SSM $\mathbf{S}_{ij}$ by applying a Gaussian Radial Basis Function as a kernel to convert distances into similarities:

$$\mathbf{S}_{ij} = \exp\left(-\frac{\mathbf{D}_{ij}^2}{2\sigma^2}\right) \tag{5}$$

where σ is the median of the pairwise distances between all feature vectors. Examples of the SSMs relative to each of the four features considered are shown in Figure 2.

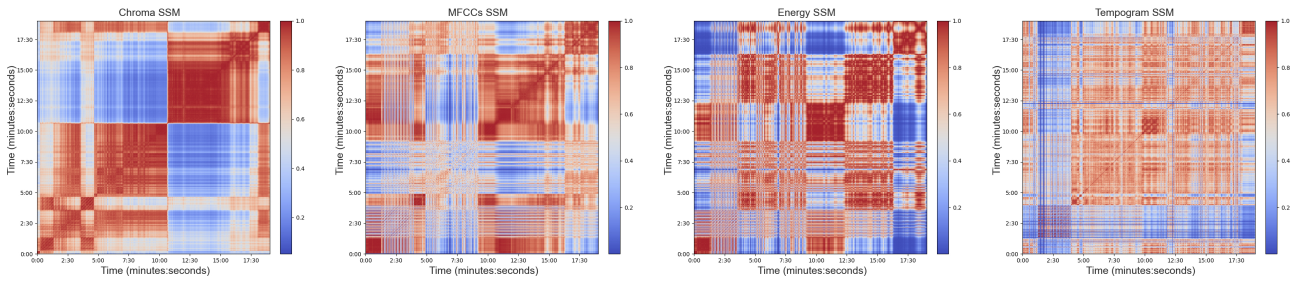

Figure 2: *Examples of the SSMs computed frame-wise from each feature. a) Chroma; b) MFCCs; c) RMS; d) Tempo*

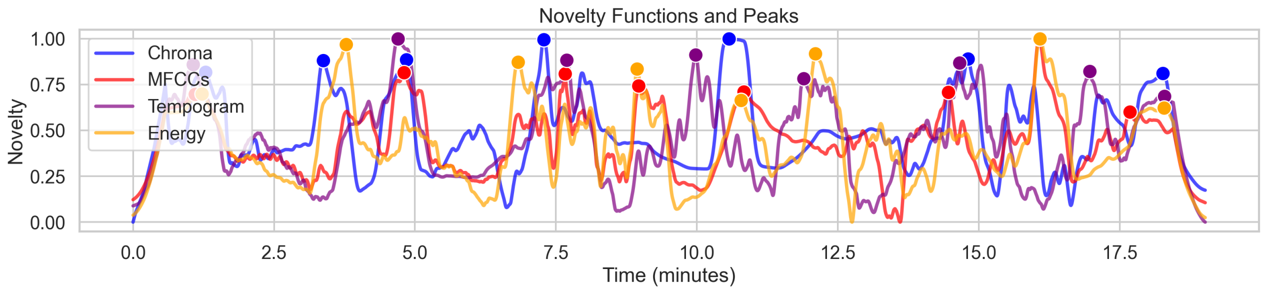

Figure 3: *Examples of the novelty functions and respective peaks retrieved from each feature SSM.*

3.3 Boundary Detection with Novelty Curves

Novelty functions constitute an optimal method for detecting feature changes in time, as they activate in the presence of significant discrepancies within a given window. Usually, they are computed by sliding large kernels over the diagonal of a SSM or, as in our case, by Gaussian-smoothing. Indeed, following the methodology in Serra et al. (2014), we first converted every $\mathbf{S}_{ij}$ into a lag matrix $\mathbf{L}_{ij}$ and created two Gaussian kernels s_1 and s_2 of different size. A smoothed matrix $\hat{\mathbf{L}}_{ij}$ is computed by sequentially colvolving a padded $\mathbf{L'}_{ij}$ with s_1 and s_2. By applying two kernels, we thereby take into account both local and contextual information. Finally, we retrieved the novelty functions $\mathcal{N}(t)$ by computing the norm between pairs of consecutive columns:

$$\mathcal{N}(t) = \left\| \hat{\mathbf{L}}_{ij}[:, t+1] - \hat{\mathbf{L}}_{ij}[:, t] \right\| \tag{6}$$

Each $\mathcal{N}(t)$ is then normalized and its peaks, representing section boundaries, are extracted by thresholding, considering a minimum value of 0.6 and a minimum distance of 100 frames (~30”). The novelty functions and relative peaks for all the features are shown in Figure 3.

In addition, we created a global novelty function $\mathcal{N}_G(t)$ by summing all the $\mathcal{N}(t)^2$ (exponentiation is meant to weight higher values over averaging) relative to the four features, and extracted peaks. We used these peaks to perform a global segmentation that takes into account the overall variations with respect to all the novelty curves. The algorithm also stores a copy of the original audio file, overlaying an audio click at each boundary.

3.4 Segment-Similarity Matrices

Finally, we employed the novelty peaks to chunk the original features accordingly, and performed dimensionality reduction by simple column-wise averaging over entire segments. Per-feature Segment-Similarity Matrices (SegSMs) are then computed following the same procedure and algorithms as in 3.2 (only without memory-stacking) to retrieve the degree of similarity with respect to entire sections, and displayed maintaining the relative temporal proportion for easier consultation and comparison. The final output of this process is shown in Figure 4.

4 Preliminary Evaluation

We tested the proposed pipeline on a selection of cherry-picked musical pieces, belonging to different genres, to ensure that the algorithm is sufficiently robust to different scenarios. While an in-depth discussion of the algorithm’s behaviour

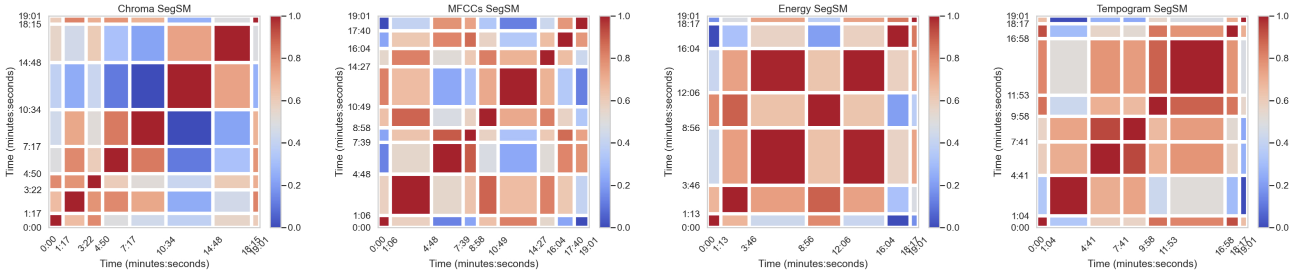

Figure 4: *Examples of the SegSMs for every feature according to a global subdivision in sections. a) Chroma; b) MFCCs; c) RMS; d) Tempo. Non-homogeneous grid sizes represent the actual section length.*

in such contexts is beyond the scope of this paper, we still provide some general considerations that emerged during our preliminary evaluation.

In line with the literature, the pipeline proved overall robust in correctly segmenting tonal, beat-based music. The combination of local moving smoothings and averages along with the different features considered allow the algorithm to adapt to a variety of different musical structures, identifying typical verse+chorus structures in pop/rock music, different themes within classical movements, breaks in danceable electronic pieces, up to short instrumental segments in progressive music. Boundaries for the most part occur within the typical time tolerance of 3 seconds (Ong and Herrera 2005). We found that local contexts present the advantage of both tracking subtle novelties (e.g. in music which is overall very similar) and dealing with internal variations (e.g. pieces characterized by substantial outliers), which can be easily overlooked by averaging globally. However, such a solution shows also a noticeable drawback: indeed, adaptive algorithms exhibit the tendency to forcefully look for novelties even in perceptually flat sections, e.g. in drone-ish music. This could possibly result in different analysis levels, in which the algorithm's lens moves from high-level to mid-level hierarchies within the course of the same piece. While this fact could still be valuable for analytical purposes, it is worth noticing that this may not always reflect the overall perception of the listener.

5 Evaluation with Live Coders

The proposed pipeline has been evaluated in-depth with six live coders with the aim of understanding its viability, strengths, and weaknesses. The participants, involved on a voluntary basis, were experienced performers; five or them are male and one female, all of them are between 30 and 45 years old and based in Europe. We acknowledge that involving more people, possibly from different communities and cultural backgrounds, could have led to more differentiated outcomes. However, we argue that the variety of the approaches and the high degree of system personalization by the participants are still sufficient to initiate a discourse on the topic, which - we hope - could stimulate further reflections.

5.1 Methodology

We asked each participant to provide us with the audio recording of one single live performance which they consider significant with respect to their usual pratcice. Each file has been analyzed by the algorithm, extracting feature-wise SSMs and SegSMs, along with audio files containing markers (audio clicks) in correspondence with every global boundary. Then, we interviewed each participant individually, following a semi-structure interview as follows: 1. General questions - personal approaches to musical form, how they are used to structure the development of their pieces (this mostly served us as a guideline to tailor the the interview on their practice); 2. Global segmentation - agnostic review of the provided audio file and comparison against the retrieved markers; 3. Similarity matrices - review of the provided per-feature similarity matrices; 4. Overall remarks - comments and suggestions on the algorithm.

We analyzed the transcriptions of the recorded interviews via thematic analysis (Braun and Clarke 2006), progressively harmonised the text, and extracted the most relevant themes which emerged - highlighted in bold. Interviews were conducted in Italian and directly translated by the authors - direct quotes are reported in quotation marks.

5.2 Emerged Reflections

All the pieces submitted by the participants were characterized by a strong improvisational component. However, only one was performed completely from-scratch, with the others being to a certain extent prepared beforehand, in terms of general outlines that help them maintain a musical direction, then elaborated according to extemporaneous ideas and sensations.

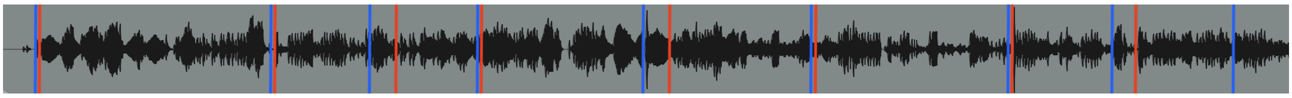

Figure 5: *One of the audio tracks analyzed, highlighting the boundaries retrieved by the system (in blue) and the boundaries identified by the live coder (in red)*

Although the participants underlined some differences in terms of musical complexity and aesthetics, they perceived their cognitive processes as substantially identical to the from-scratch approach.

While reviewing their pieces against the markers retrieved by the algorithm, participants initially focused on the **accuracy** of the segmentation, compared to how it was conceived in the course of the performance. We tracked such evaluation by clustering the markers as coincident (corresponding to the live coders' intention), non-coincident (where a new section is correctly identified, but in the wrong place), missing (new section not identified), and exceeding (incorrect new section). Overall, they positively evaluated the algorithm's accuracy. This is particularly valuable, especially in relation to several ambiguous transitions, such as crossfades or incremental parameter changes, for which the precision of the segmentation was particularly appreciated. Furthermore, the number of identified segments mostly matched the number of sections intended by the artists, with only one missed, and four exceeding. In three cases out of four where the algorithm erroneously identified an additional section, performers declared that they were slowly increasing some sort of distortion on the master bus. For all of them, this constitutes a semantic continuum: "In my head, it's the same slow and linear development" - P1; "It's a crescendo of noise, variations at that point are mostly random and irrelevant" - P3. Table 1 summarizes the perceived accuracy of the algorithm, while Figure 5 shows an example of a segmentation as perceived by the performer and as retrieved by the algorithm.

Table 1: Evaluation of the retrieved peaks with respect to the segmentations defined by the participants.

Total	Coincident	Non-coincident	Missing	Exceeding
44	33	9	1	4

Discussing the segmentation discrepancies in the case of boundaries close to those identified by the algorithm, differences emerged in terms of **auditory vs performative perception**. In fact, in relative proximity to a new section, performers motivated the algorithm's choices at a perceptual level: "It correctly identified the reprise [of a melodic line], but the sounds underneath are quite connoted and continue for a few seconds more, the section for me ends when I remove them" - P2. It is evident how the perception of formal variations may be strongly influenced by the actions on the code, rather than by the actual musical outcome, as well as by possible randomness typically often intrinsic to live coding systems: "I understand why the boundary was identified here, but I would not consider it as synchronized to my actions" - P4; "This is simply the result of an involuntary outlier that I then immediately corrected, in fact the new section started a little earlier" - P6; "There is actually a modulation that I hadn't thought of at the time, so I understand that it was interpreted as the main change point, but for me it was still linked to the previous part" - P1.

A common trend regarding the **structural windowing lens** emerged from the interviews, and all performers confirmed our arbitrary choice of a minimum distance of 30 seconds between consecutive structures. Although we believe that the calibration of such tolerance windows should still be assessed on a case-by-case basis, participants repeatedly mentioned how the practice of live coding imposes a practical constraint due to the time needed to elaborate a new idea in the form of code, which tends to slow down development processes and rarely allows for sudden and close changes in structure. We argue that this factor radically influences both their performative strategies and their perception of formal development, which is in fact based on broader structures: "The piece is based on macroforms with a certain degree of tolerance, in compliance with an arched structure" - P5; "I work with wide structures, processing a single sample until I get bored, usually between 2 and 7 minutes" - P1; "Formally, I think in terms of continuous flow of small events, whose consequentiality and overlapping creates for me a unitary musical section" - P2. Indeed, small variations are often delegated to the machine and tend not to be considered as influential in defining structures: "The piece is conceived in an organic way" - P5; "I think about building a system, for me a new section begins when I make radical changes to the system itself or to its behaviour" - P2.

In the third phase of the interview, we showed participants the SegSSMs related to each feature, allowing them to comment on the retrieved **similarity** between each segment. Overall, performers rated them positively: "It reflects my formal intentions" - P5; "The difference between the first part and the ending [of the piece] is very noticeable, in line with my idea of development" - P2. The similarities of the segments were particularly appreciated in the case of voluntary recovery of previously exposed materials or structures: "You can see that these parts are similar: I use the same sample, in fact it's a sort of reprise" - P1; "It clearly identifies the similarities between the most chaotic moments, differentiating them from the more stable ones" - P3. In relation to the global segmentation of the audio file, by seeing the different segmentation on each feature the performers declared to better understand the reason for some discrepancies between their perception and the subdivision operated by the system: "The difference here is reasonable, but I personally never noticed it, it was probably somehow unconscious" - P1; "The algorithm must have identified some minimal elements that I wasn't thinking of, which actually makes these sections similar to each other" - P4.

Further elaborating on this point, it emerged how each performer evaluates differently the **features functional significance**, in relation to his/her practice and musical style. Indeed, some of them stated to not consider one or more features at all while performing, or conceived them as almost irrelevant: "There was no intention in the change of intonation, for me it was just a matter of timbral variations" - P1; "The structural element is catalysed in a timbre and loudness factor. The harmonic and rhythmic fields are static, and their variations are peripheral and do not influence my idea of development: - P5;"I avoided using the temporal parameter, ensuring that it was not perceived" - P3; "The segmentation on the conceptual level can be very different from piece to piece" - P2. In other cases, in accordance with what highlighted in the previous paragraphs, it emerged that structural intentions were mainly expressed through variations of specific features, while others possibly varied accordingly: "The chromagram resonates with my intention, although I was actually thinking in terms of sample categories, which actually had repercussions on harmonic variations." - P2. Moreover, in two cases, the global segmentation algorithm delayed the identification of a new section because it had been based on a single feature, while the others remained identical, causing a peak on the relative novelty curve which was averaged on the global one: "Here I only change the root note, while everything else remains unchanged. For me, however, it is a very significant change" - P6.

Furthermore, participants also commented on the **features effectiveness**. In general, some doubts were expressed regarding the use of tempograms, sometimes perceived as not particularly informative, also in relation to the fact that none of the analyzed pieces was clearly beat-based: "It is difficult to extract temporal information in this kind of performance, it can be misleading because there is no clear rhythmic part" - P1; "It could at most provide a general idea of rhythmic density, but I guess it also depends on the relevance of the attack onsets" - P2. Some participants have reflected on the fact that all the features can somehow be effective: "They describe well the main musical categories, and what they do not indicate is implicit in the absence of variation" - P4; "Even if I would not consider all the features, they can still represent a benchmark on my effort to deny some of them" - P3. No participant suggested the inclusion of additional features: "I think they are sufficient to identify all the useful information, even in the case of very different musical performances" - P4.

Finally, participants provided comments on the system and its possible **practical use**. Although they strongly base their practice on improvisation and prefer a more on-the-fly approach to structural development, they confirmed the validity of the proposed system and expressed positive comments: "It's interesting being able to see the segmentation of my piece" - P1; "This kind of feedback it's very interesting and allows me to ask myself questions about how I tend to develop musical forms." - P2. Interviewees often remarked on their close connection with the performative act which cannot be deprived of the live context, thus expressing little interest in an *a posteriori* analytical approach of a musicological/ethnographic nature. However, all suggested the usefulness of the proposed system during rehearsals in preparation for a concert, as an analytical tool oriented towards supporting awareness of their cognitive processes: "It is important to do a minimal analysis and understand how to improve your own code flow" - P2; "It certainly speeds up the analysis times, I would use it in a creative way as a support for the construction of my automatisms" - P4; "I think it allows me to better understand if I was going in the right direction with respect to what I was trying to express" - P3.

6 Overall Trends over Large Corpora

As a marginal contribution, we provide some brief resumes of structural segmentation over a large number of live performances. In light of what discussed in the previous section, we recognize the mere indicative value of such an operation, since every performance should be evaluated *per se*. However, we argue that the following graphs can still be useful to outline some recurrent patterns and to provide a general overview of current trends.
As such, we first created a corpora of audio files retrieving all the performances from the recent TOPLAP Solstice Stream from December 2023[4], where every set is ~15 minutes long. The files have been stored in .mp3 format and the music, although publicly available, has been used exclusively for analysis purposes. We systematically applied the analysis pipeline illustrated in Section 3 to the whole corpora, storing structural boundaries retrieved for every feature in every file in .json format. Then, we aggregated all such data.

The histograms on top of Figure 6 show 1. the number of retrieved novelty peaks in each file, with respect of every feature, and 2. - similarly - the duration of every segment. According to the proposed analysis methodology, the vast majority of the performances contain 6 or 7 segments, with very few exhibiting 4 or less, independently of the feature considered. In any case, none of the analyzed performances returns more than 9 segments. The second graph shows that only a small part of the segments last less than 80 seconds; this is overall in line with what our participants considered as a general "rule of

[4]https://tinyurl.com/solsticeLC2023

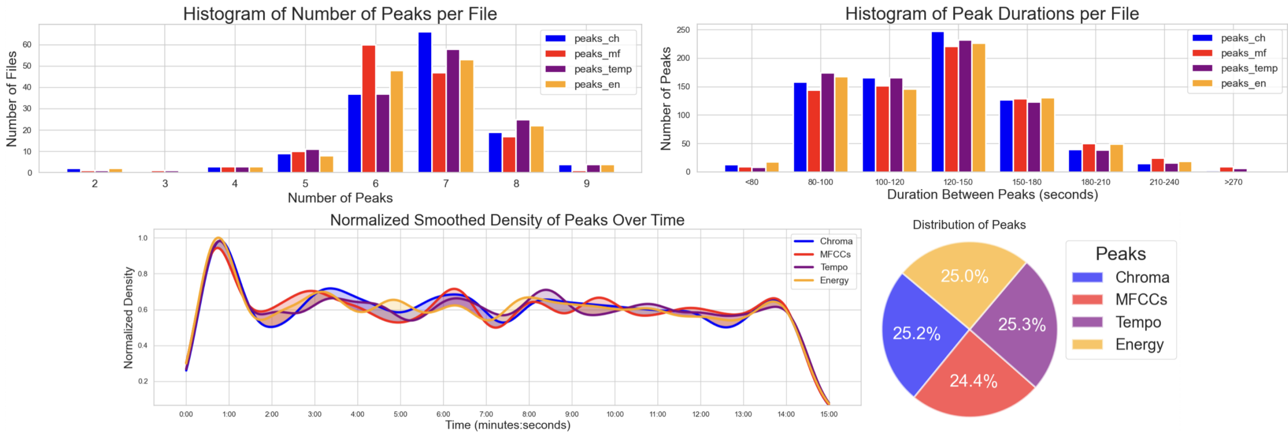

Figure 6: *Top: Per-feature density distribution (normalized) over the average duration of 15 minutes (left) and distribution of the peaks in percentages (right). Bottom: Per-feature density distribution (normalized) over the average duration of 15 minutes (left) and distribution of the peaks in percentages (right).*

thumb" for single sections, especially in from-scratch performances where the actual act of coding usually takes more time. Sections mostly last from 80 to 180 seconds, and few more than 210; in general, durations are evenly distributed featurewise.

The plots on bottom of Figure 6 shows 1. the density distribution of the retrieved novelty peaks over the average duration of 15 minutes, and 2. the percentage distribution of peaks. The first graph shows that peaks are overall distributed throughout the entire duration of the performances, with a clear spike at around 40 seconds. This shows that performances tend to exhibit clear differences in their initial part: we can speculate on such fact being due to most pieces starting out of silence or very simple musical cells, gradually layering sound events after a while, or to the use of intro sections. This is also evident by inspecting individual recurrence matrices, the majority of which clearly show similar trends in the beginning. Finally, the pie chart shows that no feature prevails overall. At first, this is in contrast with what discussed by the live coders involved in our evaluation, as they stated to concentrate on specific features during their performances. While acknowledging that the small number of participants prevents us from drawing general assessments, we argue that the great variety of performative approaches can reduce the weight of individual features in large numbers. Looking at the individual similarity matrices, it is possible to notice how there is variance in the number of segments identified for each feature within a single performance; yet each performance is different in this sense.

7 Discussions

All the interviews shared common traits that provide a wider perspective on the musical form in the context of live-coding practice. First, a deep connection emerged between the perception of formal structures and the individual practice of each musician. This resonates with the live coding community being overall oriented towards the performative moment which cannot be separated from the 'here and now' (Magnusson 2014a), implying that the interest in analytical practice is mostly aligned with enhancing personal awareness and refining musiking processes, rather than pursuing musicological objectives.

Segmentation via novelty curves was positively evaluated by participants, who declared their propensity to consider high-level structures as defined by significant variations, in line with the literature, e.g. (Müller and Kurth 2006). Also, they overall favoured a holistic approach across multiple audio descriptors, but expressed the need for control over peak thresholding and segmentation tailored to individual preferences. This once again confirms the interest in a rehearsing-oriented use of such tools, as already highlighted by Masu and Dal Rí (2023) and Manesh et al. (2024). In addition, we argue that the minimal prevalence of any single feature in the aggregated analysis, initially in contrast with participants' emphasis on prioritizing specific attributes, is due to the great variety of performative approaches observed, which may dilute the prominence of individual features over large datasets.

Participants also focused on how the unique characteristics of live coding impact their way of creating formal development, underlining strong relationships between interfaces and musical outcomes, as pointed out by Di Scipio (2003). All practitioners agreed that their development techniques are strongly characterized by a procedural approach based on the definition of constraints within a code block, where low-level decisions are often entrusted to the machine, while the performer focuses on high-level ones. Live coding environments then become systems of 'high symbolic pertinence' (Magnusson 2009) where musicians tend to offload part of their cognitive processes (Sayer 2016). Such peculiar interaction has repercussions on the musical outcomes: indeed, interviewees positively evaluated the use of a minimum distance parameter between adjacent boundaries as they rarely identify short musical sections. We argue that, especially in relation to from-scratch oriented approaches, this is due to the actual time needed by live coders to formalize their musical ideas in the form of the code - 'idea-to-code latency' (McLean and Wiggins 2009). The aggregated analysis reinforces such claims, as the majority of pieces in the database exhibit 6 - 7 segments between 80 and 180 seconds.

In the domain of Human-Computer Interaction and musical ecologies, research investigates how live coding systems foster embodiment mechanisms, through which the perception of music - and its formal structures - is inextricably linked to the performative action (Leman and Maes 2014). Live coding lies in an in-between space between improvisation and composition, where ideas and actions taken in the 'now' can impact musical outcomes in the 'future' (McLean and Wiggins 2009). While the relationship between gesture and musical structures is still up to debate (Baalman 2015; Armitage 2016), interviews indicated that performers often perceive structures through their actual actions on the code, which are not necessarily synchronized with an immediate auditory outcome. From their perspective, this disconnect can lead to ambiguity, making it difficult to discern intended structural changes based solely on the musical outcome - a limitation for audio-driven approaches.

Moreover, such changes are often carried out gradually, and developed through moments of listening to the machine's output in a constant 'performative loop' (McLean 2014): this validates our choice to compute local averages through different analysis windows, aimed at capturing broad contextual information. However, there does not seem to be a univocal answer regarding the duration of such a window. We believe that, as reported for example by Marmoret et al. (2023), a segmentation based on bars (or cycles) can be a good starting point, but the temporal fluidity expressed in the considered pieces hampers the extraction of this parameter in a fully audio-driven fashion.

Finally, other works provided an analytical perspective on musical form in live coding performance. Specifically, the visualization system 'Time_X' by Dal Rí and Masu (2022) plots geometric objects based on OSC messages, creating a graphic score which can be consulted afterwards; Diapoulis and Carlé (2023) provided audio-driven, post-performance analysis tool, displaying onset strength and spectral centroids on a circular graph. With the former system providing higher details on single events and the latter a global, linear representation of the whole performance, our approach aims at retrieving structural boundaries based on several musical features, with SegSM providing insights on the similarity relationship between different structural blocks. In this sense, we can consider such approaches complementary.

8 Conclusions

In this paper, we presented an algorithm for structural segmentation and similarity analysis specifically tested on live coding performances. By involving six experienced live coders in the evaluation process, we highlighted the strengths and weaknesses of the proposed methodology, in relation to the personal practice of the participants. The outcome of the interview with the performers helped us to point out some reflections that emerged, especially with regard to their practice, allowing us to validate the effectiveness of our algorithm in direct comparison with the artists' structural ideas. We also applied the algorithm over a larger pool of performances, providing general trends around current live coding performances. Through this work, we hope to contribute to the development of robust and adaptable analytical tools for live coding, fostering further exploration and retrospectives on how performers develop musical meaning.

8.1 Acknowledgments

We would like to thank TOPLAP Italia for the support and appreciation of our work.

References

Armitage, Joanne. 2016. "Revealing Timelines: Live Coding and Its Gestures." *Proceedings of ICLC*.

Baalman, Marije. 2015. "Embodiment of Code." In *Proceedings of the First International Conference on Live Coding*, 35–40.

Bartsch, Mark A, and Gregory H Wakefield. 2005. "Audio Thumbnailing of Popular Music Using Chroma-Based Representations." *IEEE Transactions on Multimedia* 7 (1): 96–104.

Böck, Sebastian, and Gerhard Widmer. 2013. "Maximum Filter Vibrato Suppression for Onset Detection." In *Proc. Of the 16th Int. Conf. On Digital Audio Effects (DAFx). Maynooth, Ireland (Sept 2013)*, 7:4. Citeseer.

Braun, Virginia, and Victoria Clarke. 2006. "Using Thematic Analysis in Psychology." *Qualitative Research in Psychology* 3 (2): 77–101.

Brown, Andrew R, and Andrew Sorensen. 2009. "Interacting with Generative Music Through Live Coding." *Contemporary Music Review* 28 (1): 17–29.

Brown, Andrew R, and Andrew C Sorensen. 2007. "Aa-Cell in Practice: An Approach to Musical Live Coding." In *International Computer Music Conference*, 292–99.

Buisson, Morgan, Brian Mcfee, Slim Essid, and Helene-Camille Crayencour. 2022. "Learning Multi-Level Representations for Hierarchical Music Structure Analysis." In *International Society for Music Information Retrieval (ISMIR)*.

Collins, Nick, Alex McLean, Julian Rohrhuber, and Adrian Ward. 2003. "Live Coding in Laptop Performance." *Organised Sound* 8 (3): 321–30.

Couprie, Pierre. 2022. "Analytical Approaches to Electroacoustic Music Improvisation." *Organised Sound* 27 (2): 117–30.

Dal Rì, Francesco Ardan, and Raul Masu. 2022. "Exploring Musical Form: Digital Scores to Support Live Coding Practice." In *NIME 2022*.

Dal Rì, Francesco Ardan, Francesca Zanghellini, and Raul Masu. 2023. "Sharing the Same Sound: Reflecting on Interactions Between a Live Coder and a Violinist." In *Proceedings of the International Conference on New Interfaces for Musical Expression*.

Di Scipio, Agostino. 2003. "'Sound Is the Interface': From Interactive to Ecosystemic Signal Processing." *Organised Sound* 8 (3): 269–77.

Diapoulis, Georgios, and Martin Carlé. 2023. "Reproducible Musical Analysis of Live Coding Performances Using Information Retrieval: A Case Study on the Algorave 10th Anniversary." In *Proceedings of the 7th International Conference on Live Coding (ICLC 2023)*.

Ellis, Daniel PW. 2007. "Classifying Music Audio with Timbral and Chroma Features."

Ewert, Sebastian, Meinard Muller, and Peter Grosche. 2009. "High Resolution Audio Synchronization Using Chroma Onset Features." In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, 1869–72. IEEE.

Fitzgerald, Derry. 2010. "Harmonic/Percussive Separation Using Median Filtering." In *13th International Conference on Digital Audio Effects (DAFX10)*. Graz, Austria.

Foote, Jonathan. 1999. "Visualizing Music and Audio Using Self-Similarity." In *Proceedings of the Seventh ACM International Conference on Multimedia (Part 1)*, 77–80.

———. 2000. "Automatic Audio Segmentation Using a Measure of Audio Novelty." In *2000 Ieee International Conference on Multimedia and Expo. Icme2000. Proceedings. Latest Advances in the Fast Changing World of Multimedia (Cat. No. 00th8532)*, 1:452–55. IEEE.

Grosche, Peter, Meinard Müller, and Frank Kurth. 2010. "Cyclic Tempogram — a Mid-Level Tempo Representation for Musicsignals." In *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, 5522–25. IEEE.

Hartmann, Martin, Olivier Lartillot, and Petri Toiviainen. 2017. "Musical Feature and Novelty Curve Characterizations as Predictors of Segmentation Accuracy." In *Proceedings of the Sound and Music Computing Conferences*. Aalto-yliopisto.

Ittichaichareon, Chadawan, Siwat Suksri, and Thaweesak Yingthawornsuk. 2012. "Speech Recognition Using MFCC." In *International Conference on Computer Graphics, Simulation and Modeling*. Vol. 9.

Lartillot, Olivier. 2022. "Musicological and Technological Perspectives on Computational Analysis of Electroacoustic Music." In *International Seminar on Sonic Design*, 271–97. Springer.

Leman, Marc, and Pieter-Jan Maes. 2014. "The Role of Embodiment in the Perception of Music." *Empirical Musicology Review* 9 (3-4): 236–46.

Maezawa, Akira. 2019. "Music Boundary Detection Based on a Hybrid Deep Model of Novelty, Homogeneity, Repetition and Duration." In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 206–10. IEEE.

Magnusson, Thor. 2009. "Of Epistemic Tools: Musical Instruments as Cognitive Extensions." *Organised Sound* 14 (2): 168–76.

———. 2011. "Algorithms as Scores: Coding Live Music." *Leonardo Music Journal* 21: 19–23.

———. 2014a. "Herding Cats: Observing Live Coding in the Wild." *Computer Music Journal* 38 (1): 8–16.

———. 2014b. "Scoring with Code: Composing with Algorithmic Notation." *Organised Sound* 19 (3): 268–75.

Manesh, Daniel, Douglas Bowman Jr, and Sang Won Lee. 2024. "SHARP: Exploring Version Control Systems in Live Coding Music." In *Proceedings of the 16th Conference on Creativity & Cognition*, 426–37.

Marmoret, Axel, Jérémy E Cohen, and Frédéric Bimbot. 2023. "Barwise Music Structure Analysis with the Correlation Block-Matching Segmentation Algorithm." *Transactions of the International Society for Music Information Retrieval (TISMIR)* 6 (1): 167–85.

Masu, Raul, and Francesco Ardan Dal Rì. 2023. "Visual Representations to Stimulate New Musicking Strategies in Live Coding." *Organised Sound* 28 (2): 218–30.

McCallum, Matthew C. 2019. "Unsupervised Learning of Deep Features for Music Segmentation." In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 346–50. IEEE.

McFee, Brian, and Dan Ellis. 2014. "Analyzing Song Structure with Spectral Clustering." In *ISMIR*, 405–10. Citeseer.

McLean, Alex. 2014. "Making Programming Languages to Dance to: Live Coding with Tidal." In *Proceedings of the 2nd ACM SIGPLAN International Workshop on Functional Art, Music, Modeling & Design*, 63–70.

McLean, Alex, and Geraint Wiggins. 2009. "Patterns of Movement in Live Languages."

———. 2010. "Tidal–Pattern Language for the Live Coding of Music." In *Proceedings of the 7th Sound and Music Computing Conference*, 331–34.

Miotto, Riccardo, Nicola Orio, et al. 2007. "A Methodology for the Segmentation and Identification of Music Works." In

ISMIR, 7:239–44.
Müller, Meinard, and Frank Kurth. 2006. "Towards Structural Analysis of Audio Recordings in the Presence of Musical Variations." *EURASIP Journal on Advances in Signal Processing* 2007: 1–18.
Murthy, Y. V. Srinivasa, and Shashidhar G. Koolagudi. 2018. "Content-Based Music Information Retrieval (CB-MIR) and Its Applications Toward the Music Industry: A Review" 51 (3).
Nagawade, Monica S, and Varsha R Ratnaparkhe. 2017. "Musical Instrument Identification Using MFCC." In *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, 2198–2202. IEEE.
Nieto, Oriol, Gautham J Mysore, Cheng-i Wang, Jordan BL Smith, Jan Schlüter, Thomas Grill, and Brian McFee. 2020. "Audio-Based Music Structure Analysis: Current Trends, Open Challenges, and Applications." *Trans. Int. Soc. Music. Inf. Retr.* 3 (1): 246–63.
Ong, Bee Suan, and Perfecto Herrera. 2005. "Semantic Segmentation of Music Audio." In *Proceedings of the... International Computer Music Conference*, 61. Computer Music Association.
Panagiotakis, Costas, and Georgios Tziritas. 2005. "A Speech/Music Discriminator Based on RMS and Zero-Crossings." *IEEE Transactions on Multimedia* 7 (1): 155–66.
Paulus, Jouni, and Anssi Klapuri. 2008. "Acoustic Features for Music Piece Structure Analysis." In *Proc. Of 11th International Conference on Digital Audio Effects*, 309–12. Citeseer.
Popescu, Tudor, Richard Widdess, and Martin Rohrmeier. 2021. "Western Listeners Detect Boundary Hierarchy in Indian Music: A Segmentation Study." *Scientific Reports* 11 (1): 3112.
Sayer, Tim. 2016. "Cognitive Load and Live Coding: A Comparison with Improvisation Using Traditional Instruments." *International Journal of Performance Arts and Digital Media* 12 (2): 129–38.
Serra, Joan, Meinard Müller, Peter Grosche, and Josep Ll Arcos. 2014. "Unsupervised Music Structure Annotation by Time Series Structure Features and Segment Similarity." *IEEE Transactions on Multimedia* 16 (5): 1229–40.
Serra, Xavier. 2011. "A Multicultural Approach in Music Information Research." In *Klapuri a, Leider c, Editors. ISMIR 2011: Proceedings of the 12th International Society for Music Information Retrieval Conference; 2011 October 24-28; Miami, Florida (USA). Miami: University of Miami; 2011.* International Society for Music Information Retrieval (ISMIR).
Terhardt, Ernst. 2016. "Gestalt Principles and Music Perception." In *Auditory Processing of Complex Sounds*, 157–66. Routledge.
Tian, Mi, and Mark B Sandler. 2016. "Towards Music Structural Segmentation Across Genres: Features, Structural Hypotheses, and Annotation Principles." *ACM Transactions on Intelligent Systems and Technology (TIST)* 8 (2): 1–19.
Vatolkin, Igor, Marcel Koch, and Meinard Müller. 2021. "A Multi-Objective Evolutionary Approach to Identify Relevant Audio Features for Music Segmentation." In *International Conference on Computational Intelligence in Music, Sound, Art and Design (Part of EvoStar)*, 327–43. Springer.
Wang, Ju-Chiang, Yun-Ning Hung, and Jordan BL Smith. 2022. "To Catch a Chorus, Verse, Intro, or Anything Else: Analyzing a Song with Structural Functions." In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 416–20. IEEE.
Xambó, Anna, Alexander Lerch, and Jason Freeman. 2018. "Music Information Retrieval in Live Coding: A Theoretical Framework." *Computer Music Journal* 42 (4): 9–25.
Zattra, Laura. 2005. "Analysis and Analyses of Electroacoustic Music." *Sound and Music Computing (SMC05), Salerno, Italy* 36.
Zheng, Fang, Guoliang Zhang, and Zhanjiang Song. 2001. "Comparison of Different Implementations of MFCC." *Journal of Computer Science and Technology* 16: 582–89.