INTERACTION-BASED ANALYSIS OF FREELY IMPROVISED MUSIC

Stefano Kalonaris
Center for Advanced Intelligence Project (AIP)
RIKEN, Japan
stefano.kalonaris@riken.jp

ABSTRACT

This paper proposes a computational method for the analysis and visualization of structure in freely improvised musical pieces, based on source separation and interaction patterns. A minimal set of descriptive axes is used for eliciting interaction modes, regions and transitions. To this end, a suitable unsupervised segmentation model is selected based on the author's ground truth, and is used to compute and compare event boundaries of the individual audio sources. While still at a prototypal stage of development, this method offers useful insights for evaluating a musical expression that lacks formal rules and protocols, including musical functions (e.g., accompaniment, solo, etc.) and form (e.g., verse, chorus, etc.).

1. INTRODUCTION

When tackling musical structure, it is not uncommon to employ language-based conceptual blends [1]. Some of these focus on a formal and generative grammar approach, while others foreground metapragmatics and conversational metaphors. While the former [2] are conditioned upon a notion of *musical surface* [2] and focus on hierarchical structures of musical phenomena in the context of Western tonal music, the latter have been employed for less formal theories, when dealing with musical improvisation practices, such as *jazz*.

A different perspective is needed when analyzing *free jazz* [3,4] or *free improvisation* [5,6], which are musical expressions that lack an agreed upon representation scheme, and which defy and challenge definitions and categorizations. While recent work has been done in this field to understand how structure is perceived in these musical expressions [7], more research is needed in this regard.

This paper foregrounds the dialogical component of this music, whereby structures are negotiated in real-time, emerge ad-hoc, and cannot be inferred or deduced from a score. A method for the structural segmentation and analysis of musical improvisations of this kind is proposed, inspired by Pelz-Sherman's [8] speculations on interactional listening/music making.

Conditioned upon the analysis of the individual voices (audio sources), this study considers a multi-track recording and implements a distilled version of Pelz-Sherman's scheme, whereby interaction patterns and dynamics are deduced comparing individual boundaries and audio features spaces. By doing so, the method offers itself as a tool for investigation of how structure, in this context, might emerge from the continuous negotiation of musical expectations and demands, how these might be communicated to others, acknowledged or ignored altogether.

2. CONTEXT

Free jazz and free improvisation are not the same musical expression. While the latter is often viewed as the avant-garde European offshoot of the former, they are distinct expressions which can be easily discriminated. However, they also share sufficiently many characteristics, from ideological to musical. For example, they both share the desire to rebel against the *status quo*, to assert freedom from conventions and uniformity, not without political and societal implications.

Generally speaking, no predefined agreement or commitment about the music is made and, according to this paradigm, players negotiate the musical outcome in real-time. In free improvisation it is customary not to abide by musical referents (such as idiom, style, genre, or even tonal keys), while free jazz has a stronger element of idiomatic playing, linked to the broader development and narrative of Afro-American musical expression. Of course, and despite occasional claims of the contrary, there is no such thing as an unbound, *ex nihilo* improvisation since all musicians have an acquired protocol of interaction, based on historical, personal or shared aesthetic and musical preferences. Despite this, free jazz and free improvisation are arguably less formalized than other improvised expressions (e.g., a cadenza in a solo concerto). Paramount to both is the focus on interaction, distributed decision making and lack of predefined musical outcome. Furthermore, "improvisation must be open - that is, open to input, open to contingency - a real-time and (often enough) a real-world mode of production" [9, p.38]. For the sake of simplicity free jazz and free improvisation will be hereinafter referred to as *freely improvised music*, without specific preference for one or the other, unless explicitly stated.

To attempt a computational analysis of this musical expression it is crucial to work from the individual musical

[1] An integration procedure formalized by Fauconnier [1].
[2] A discrete representation of the sounds in a piece, to include pitches, durations and dynamics, intrinsically linked to the concept of music notation.

parts, to see how they relate to one another and how these relations evolve over the course of the piece.

2.1 Source Separation

Attempts to source separate historic and representative freely improvised recordings were made by the author, using non-negative matrix factorization with the Flexible Audio Source Separation Toolbox (FASST) [10] and harmonic-percussive separation [11]. However, results in this respect where not deemed satisfactory and, given that the state of the art in source separation was not the principal focus of this study, improvements on this front were not pursued further. Instead, a multi-track recording from the MedleyDB [12] (see Section 4.2) was used, for lack of a better alternative and despite stronger idiomatic assumptions. Source separation is an open problem and an active topic of research, and it has never, to the author's knowledge, been applied to freely improvised music (although it has been explored for early jazz recordings [13]). This might be due to the low commercial and aesthetic appeal and popularity of this musical expression. Source separation thus far has been more concerned with application to the music industry and to creative music technologies, e.g. automatic mixing [14], automatic transcription [15], orchestration [16] or voice separation [17]. In this context, clearly defined musical functions and registers for the instruments are preferred: drums play rhythm, harmonic instruments play chords, melodic instruments play melody, vocals float on top, and so forth.

A scenario of this type is undesirable in freely improvised music. In fact, such compliance with predefined roles and domains is the primary impetus out of which free jazz and free improvisation were developed in the first place. In these musical expressions it is common practice to use extended techniques, whereby the spectral palette of each stream is augmented beyond "normal", shifting the attention from melodic contour to gestural morphologies of sound, such as trajectories, density, functional relations and so forth. The spectral spill-over generated by this mode of playing, along with non exclusive musical roles (e.g., a guitar can be hit with mallets and objects and used as a percussion, etc.) makes it challenging to clearly separate the sources.

2.2 Structure in Freely Improvised Music

The issues linked to spectral spill-over and musical role cross-over go beyond source separation tasks, and can also make the use of standard music information retrieval (MIR) techniques for structural segmentation arduous. Notions commonly used in MIR tasks and cognitive-based approaches to musical surface parsing and segmentation rely on culture-specific axioms. For example, the predicate that major and minor triads form convex subsets [18] (which had already been challenged by Forte's music set theory [19]). Assumptions made in this contexts are difficult to port to musical domains that do not share the same tonal/functional axioms.

Adding to the the difficulties in deducing musical structure in freely improvised music is the issue of representation. Roads [20] posits that one can represent music at three levels: iconic, symbolic and score level. The first would include data relative to an audio waveform (e.g., sequences of values for amplitude and phase) or graphic scores, the second would include the use of signs which would convey syntactical meaning, and the third can be assimilated to what is commonly called music notation. No attempt to define a representation system or scheme for free improvisation has been made up to date, although graphic scores or snippets of musical notation [21] can be used as platforms (sometimes distributed, collaborative and editable on-the-fly [22,23]) for inspiration and suggestive/aleatoric interpretation, especially in free jazz. Nevertheless, these tools cannot fully describe and contain the musical process and product. Audio representations can also be used, but retrospectively. That is, the music is always created on the spot and no musician knows what the outcome will be *a priori*.

The issues outlined above, however, do not imply that these musical expressions lack structure or structural segments [24, 25]. More specifically, it appears that macrostructure in freely improvised music is a surface phenomenon emerging from micro-structures which are sufficiently differentiated at some feature level. The transitional regions between these sections are paramount for the understanding of segmentation boundaries and of their treatment (e.g., gradual, clear-cut, etc.) in real-time. It has been shown that expert improvisers can "generate segmentation in high-level musical structure" [24, p. 235].

3. AN INTERACTION-BASED VIEWPOINT

Improvised music can be challenging in many ways, but can be better understood as a dynamical and distributed decision making process. Structure is thus a by-product of such process, which has been investigated in terms of saliency and coordination [26]. According to this interpretation, the notion of *focal point* is paramount. This can be defined as "a point of convergence for expectations" [26, p.3].

As seen so far, negotiation, coordination and interaction are the key concepts needed to understand and analyze freely improvised music. The focus of this paper is in fact on the real-time interaction of potentially several musical "voices", and the relational nature of the musical result. To this end, it is useful to think of a relationship as the association between two elements/agents which emerges via a specific connection, and of interaction as the events and actions that help (or not) to form or define a relationship [3].

Having established this relational/interactional paradigm, some analogies and metaphors naturally come to the fore, such as the parallels between music-language and improvisation-conversation. While these pairs seem related (two people need a a common language to have a conversation, just as musicians might need to operate within given conventions to musically communicate), the correspondences are subtle, contradictory and often problematic.

[3] For example, an estranged father has a filial relationship with his offspring and a null interaction.

3.1 Linguistic Approaches

Theories inspired by Chomsky's generative grammars [2, 27–30] are common in the context of Western tonal music, and in computational musicology. They have also been proposed for musical improvisation [31], with limited application. Grammar-based perspectives of musical representation fall under the symbolic category, whereby deep structures are inferred, parsed or deduced from the musical surface. These approaches have not been unchallenged. Some have doubted the ability of linguistic-based and Gestalt-based approaches to generalise across listeners at a macro-structural level [32], others argue for a more holistic perspective of sounds [33], others yet note that high-level entities, like beat structure [4], chord simultaneities [5] and voice separation [6] are necessary for the formation of the musical surface [37]. Importantly, freely improvised music has no (or very little) dependency on musical surface, as pieces are not planned, notated and performed accordingly, as discussed in Section 2.2 (although retrospective notation by means of transcription can be done).

Furthermore, a grammar-based representation of musical deep structure is purely functional, and fails to account for sociological, emotional, moral, aesthetic, and cultural aspects involved in musical expression. The main difficulty encountered when employing grammar-based approaches to music representation is that of modelling context. This is particularly problematic in improvised music, where phrasing and context are often interrupted and reinstantiated. To this end, Roads [20] suggests that more research should be undertaken in "interrupt-driven" grammars.

3.2 Conversational Approaches

Musical interaction occurring in improvisation has been often viewed under the paradigm of verbal communication. Drawing from *metapragmatics* [38], both Sawyer [39] and Monson [40] develop their frameworks for understanding jazz improvisation though a conversational metaphor. Sawyer, in the larger context of improvisational studies, notes that "improvisational interaction can be mediated by both linguistic and musical symbols" [41, p. 150]. Improvisation is thus associated to different voices in dialogue with one another, a real-time conversation (without a predefined topic). It is arguable that a successful conversation relies on effective communication. While more idiomatic forms of musical improvisation such as *jazz* focus more on narrative [42] and story telling [43], actively engaging with tradition and lineage, freely improvised music is less preoccupied with linear accounts and more focused on real-time distribution of agency and the dialogical aspect of communication [44].

[4] Beat structure comprises *beat induction* (finding an appropriate relative clock) and *beat tracking* (a dynamically changing clock).

[5] Chord simultaneity presupposes culture-specific knowledge stored in long-term memory [34] and it is an *emergent quality* [35].

[6] The auditory system can decompose spectral fusions [35, p.64] into separate streams, based on pattern analysis. This process was described in [36] as "auditory stream segregation".

3.3 Interactional Music-making

Despite the many similarities between music and language or musical improvisation and conversation, there remain sufficiently many fundamental differences to warrant caution when blending domains. Beyond the inadequacy of formal grammars or the lack of formal theories in conversational approaches, the relational nature of the musical interactions occurring in freely improvised music is paramount, and it is the principal motivation for the method proposed in this paper. A dialogical perspective foregrounds such interaction between the musical constituent parts, which can be assimilated to the audio sources (streams) in a recording. Stream segregation with respect to music improvisation has been investigated in [45], where the concept of interactional listening is developed. Interaction is also the pivot of Peltz-Shermans work, which is reviewed in the next section to investigate boundary localization and segmentation of freely improvised musical pieces.

4. METHOD

4.1 Overview

According to Pelz-Sherman's [46] distinction between monoriginal and heteroriginal musical expressions, freely improvised music classes among the latter. He posits [8] that performers are, at any given time, either in a state of transmitting and/or receiving musical signals, with *i-events* representing the mutual response to a musical request, called a *cue*. Pelz-Sherman does not offer an exhaustive list of i-events, but he lists *imitation*, *question and answer*, *completion/punctuation* and *interruption*. Furthermore, he discriminates between static and dynamic modes of interaction, whereby the former (*sharing*, *solo/accompaniment*, *not sharing*) are fundamental states at which players operate at any given time, and are associated with levels of i-event density (high, medium, low, respectively). Dynamic modes (*emerging/withdrawing*, *merging/accepting*, *interjecting/supporting*, *initiating/responding*) instead can be thought of as the types of transitions between any two static modes. To further clarify, static modes of interaction can be assimilated to the inter-boundary regions in the context of structural segmentation, whereas dynamic modes can be considered the intra-boundary segments. Other frameworks have been proposed in this context, some of which have a more extensive taxonomy of transitions and/or relational functions [47], however, these were not considered in the current study for the sake of simplicity.

In this paper Pelz-Sherman's scheme is reduced to comprise two essential descriptors: *static mode* (spanning from not sharing to sharing) and *dynamic mode* (either morphing or clear-cut). Concretely, the former is a measure of similarity between the musical features of the audio sources in-between the segment boundaries, whereas the latter is a level of agreement between the boundary placement/detection over them. To this end, the author posits that if a boundary is detected in most parts in a given time window (thus, a cue was responded to within this threshold), then a (more or less) clear-cut transition is assumed.

Algorithm	Accuracy	Precision	Recall	F1 score
vmo (mfcc)	0.805	0.182	0.133	0.154
foote (mfcc)	0.8	0.2	0.056	0.088
cnmf (mfcc)	0.8	0.2	0.056	0.088
olda (mfcc)	0.81	0.375	0.167	0.231
scluster (mfcc)	0.762	0.182	0.111	0.138
sf (mfcc)	0.79	0.167	0.056	0.084

Table 1: Metrics for the different (MFCC-based) algorithms used for segmentation

Conversely, if inter-part segment boundaries do not agree, a morphing transition is assumed. This might be a scenario whereby one player sends a cue which is not followed by a significant change in the musical feature space of the other player(s) (either as the result of a deliberate musical choice/strategy, or simply because they missed it).

4.2 Procedure

To test the method, the multi-track recording *FreeJazz* by MusicDelta, from the MedleyDB [12] was used. This trio (clarinet, double bass and drums) recording was chosen for several reasons. Firstly, and as discussed in Section 2, free jazz shares many of the broader concerns of freely improvised music (such as the desire to break regular tempos, tones, and chord changes conventions). Secondly, the track was deemed by the author, a domain expert and practitioner, sufficiently apt to investigate multi-part interaction. Thirdly, and after extensive search, it was not possible to source an historical example of either free jazz or free improvisation in multi-track format. Several constraints (e.g., time, location, musicians' network) at the time of writing did not allow for a bespoke recording of a multi-track piece.

The raw individual audio sources and the audio mix were segmented using the MSAF Python package [48], based on several of the available algorithms (e.g., variable Markov oracle [49], audio novelty [50], convex non-negative matrix factorization [51], ordinal linear discriminant analysis [52], spectral clustering [53]), each in turn based either on Mel-frequency cepstrum coefficients [7] (MFCCs) or tempogram features. The former are shown in Figure 1, for comparison's sake.

To choose one of these algorithms, the results on the audio mix were compared to the author's analysis of the same file. This analysis was used as the ground truth for computing the F-score, shown in Table 1.

The ordinal linear discriminant analysis (OLDA) [52] algorithm was selected, based on its score. In Figure 2 the ground truth and the detected bounds are plotted for comparison. Using the best performing model, the boundaries obtained on the audio sources were compared. Figure 3 illustrates these boundaries, and reveals several salient events in the piece. Saliency is inferred because the boundaries feature in both the audio mix and the individual sources.

[7] Relating to a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear scale of frequency.

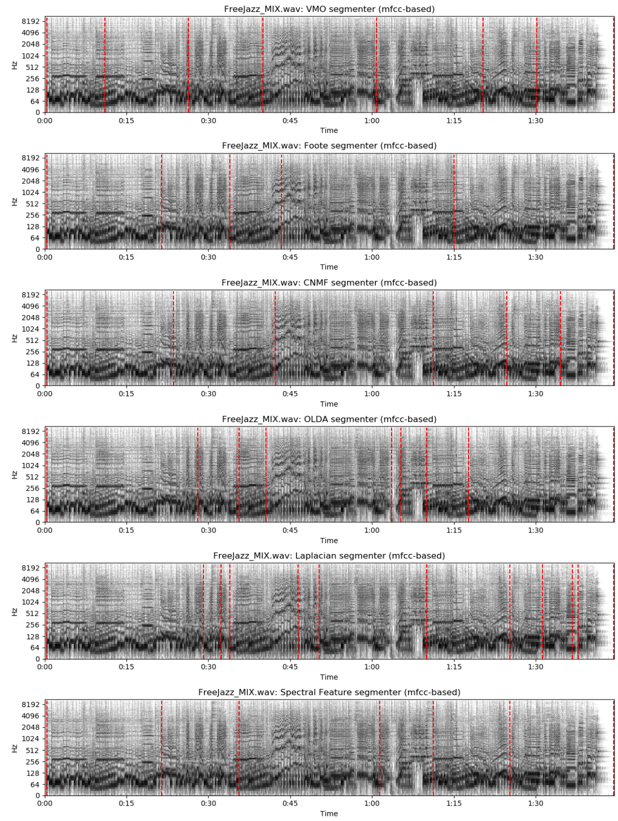

Figure 1: Segmentation boundaries on the audio mix, using several algorithms and based on MFCC features.

Based on this consideration, clear-cut and morphing transitions were identified, according to whether the individual audio sources' boundaries agreed (+ or - a 2 seconds [8] buffer, factored in to account for the reaction time needed by one player to respond to a musical cue originated from the other players) or not, respectively. This procedure formalizes what was called dynamic mode in Section 4.1 and it is shown in Figure 4.

Static mode, on the other hand, is concerned with the sharing of the musical feature space at a given time. To this end, regions in between the boundaries were used to compute the similarity over given audio features. Figure 5 shows the inter-regions zero-crossing rate [9] dynamics for all three audio sources. A total of 27 such features were initially computed and are available for inspection, although they are omitted here for the sake of brevity.

In the example shown in Figure 6, 6 features (root-mean-square energy, spectral centroid, spectral bandwidth, spectral flatness, spectral roll-off, and zero crossing rate) were used to calculate the (average) inter-region similarity between the audio sources, using cosine similarity (Pearson correlation or other metrics are also possible). The similarity values so obtained were used as the color gradient.

[8] This time window is heuristically determined, and used to divide the audio buffer into bins of this length.

[9] The rate at which the signal changes from positive to zero to negative or vice versa.

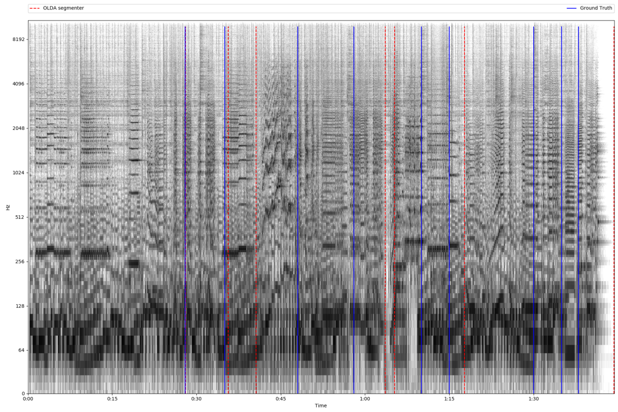

Figure 2: Comparing the ground truth (blue) and the OLDA algorithm (red) based on MFCC features.

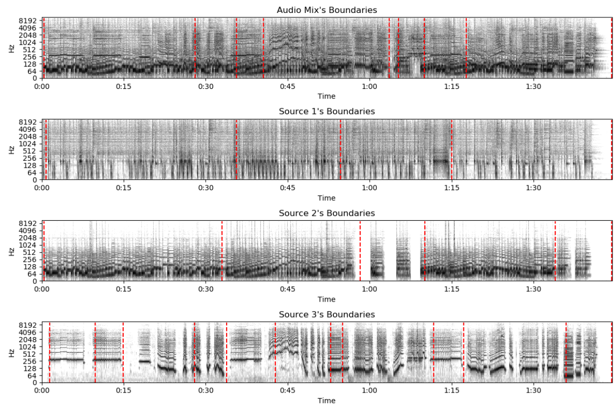

Figure 3: Segmentation boundaries on the audio mix and the audio sources, using the OLDA algorithm based on MFCC features.

5. DISCUSSION

The objective of this study was not to claim a methodology able to improve the current state of the art in segmentation of audio tracks, but rather to offer a perspective foregrounding the interaction of the musical voices, their contribution, and how these negotiate structure in real-time during freely improvised pieces. Neither is the objective truth used to be considered as a target for the optimization of the algorithm. Instead, it is used as an initial pruning and approximation, to choose a segmenter for the exploration described in Section 4.2. The i-events, the transitions, and the 'sharing' quality of the sections outlined by the method are offering an opportunity for a re-evaluation of the interpretative and cognitive process occurring when trying to infer structural dynamics in a freely improvised piece. In this sense, the method contributes suggestions, hints and viewpoints. Thus, it can be considered under the same dialogical paradigm as the music that it analyzes. For the user, this might be akin to having a conversation with another musicologist or practitioner, who would present her opinion about how the musical parts interact with one another, and how the piece emerges from such dynamics.

Inspecting the boundaries in the dynamic mode (see Figure 4), and ignoring the first clear-cut transition (which is not very informative, since it is the point where the instruments start playing, at the beginning of the recording) one can consider the activation times given in Table 2.

The method so far operates with an arbitrary time kernel (or stride), whereby the total length of the piece is divided into equally spaced windows of such length. While this is heuristically determined and can be changed to one's liking, it nevertheless produces a lower accuracy as for the boundaries' activation times. Rather than considering this as an handicap, it is useful to be reminded that the objective of the study is to identify patterns of interaction rather than accurate segmentation boundaries. Furthermore, this allows to factor in reaction times for the musicians who, while making decisions in real-time, might change their musical behavior in the order of a few seconds. Combining the information gathered from the dynamic mode analysis, it is thus possible to infer more pronounced regions of musical interaction between the following times in seconds: 32-34, 52-58, 70-76, and 92-94. These are obtained by combining activation times that are sufficiently close together (in both the clear-cut and the morphing class). 'Sufficiently' is taken to mean as up to twice the time kernel, on either side.

Interestingly, these sections exhibit a lower feature similarity (see Figure 6), which might suggest that when cues are acknowledged and responded to this corresponds to a higher interaction. Conversely, in the regions in between, the individual players might adopt more consistent, static and function-oriented musical behaviors. This finding seems in accordance with the framework proposed by Pelz-Sherman, as well as with the notions of *transitions* and *relational functions/composites* in Nunn's work [47]. To test the validity of this approach it would be appropriate to conduct further studies accounting for the opinion of

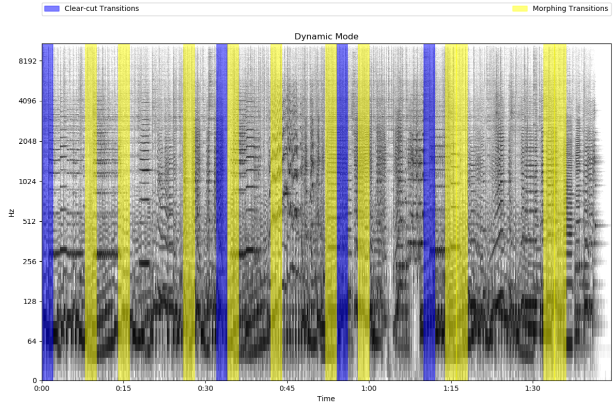

Figure 4: Dynamic Mode: clear-cut (blue) and morphing (yellow) transitions shown over the audio mix's spectrogram.

clear-cut	32, 54, 70
morphing	8, 14, 26, 34, 42, 52, 58, 74, 76, 92, 94

Table 2: Transitions types and their activation times, in seconds

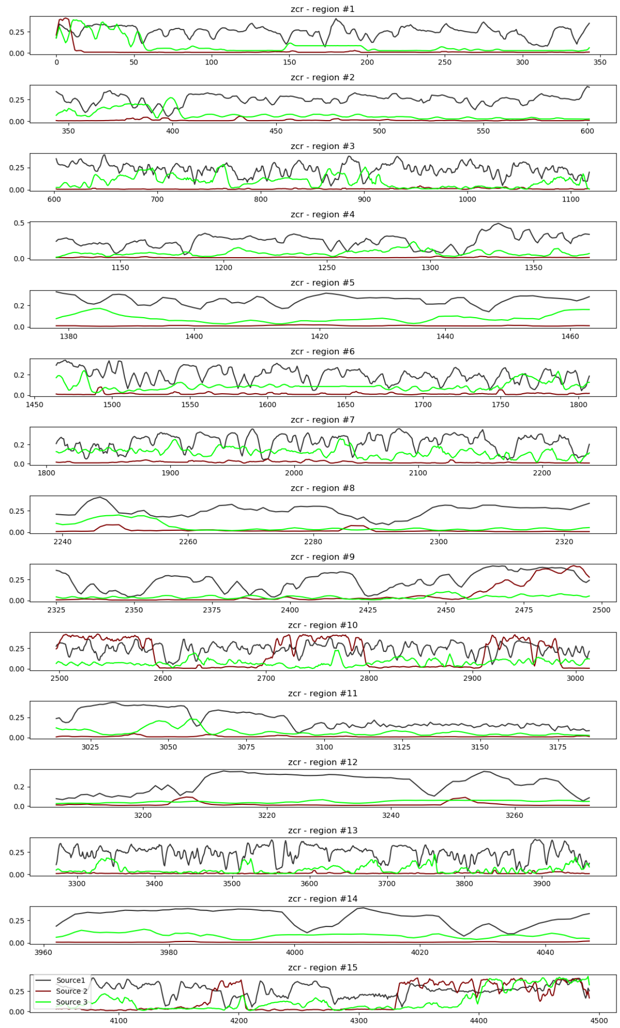

Figure 5: Inter-regions zero-crossing rate dynamics.

a wider pool of practitioners and using more recordings, since one piece cannot offer general insights. These endeavors are left for future work.

6. CONCLUSION

To analyze freely improvised music, it is paramount to consider the interaction between players and how their choices help shape the piece in real-time, without a predetermined plan or specific musical goals other than those arising as contingencies of the creative process. In keeping with the distributed and dialogical nature of this musical expression, this paper explores a relational stance and implements an essential method for visualizing such interactions. While mainstream musical expressions, styles and genres are more conducive to elicit well defined segments and regions (by virtue of intrinsic structural assumptions. i.e.: chorus, verse, etc.), freely improvised music requires bespoke treatment and a focus shift from surface to functional level. The current study represent but a small step in this direction, albeit partial and with a reduced scope of analysis (limited to two basic modes of interaction). Future work would benefit from implementing more discrimination within these levels, as formalized in [8, 46, 47], as

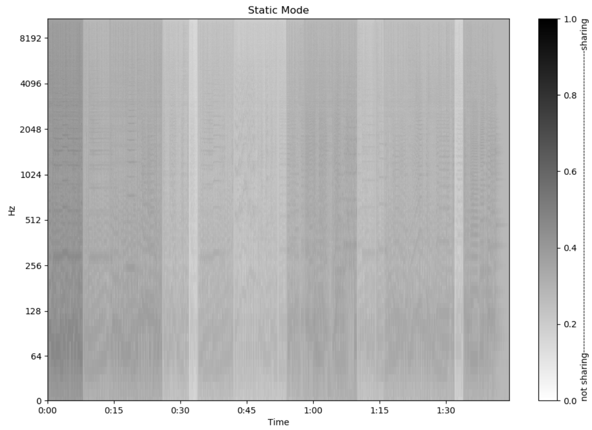

Figure 6: Static Mode: inter-regions feature space sharing.

well as from recording a dataset of multi-track freely improvised pieces. To this end, the author plans to record several duets comprising a wide range of instrumentation, with active practitioners pooled from the freely improvised music scene in UK and Japan.

7. REFERENCES

[1] G. Fauconnier and M. Turner, "Conceptual integration networks," *Cognitive Science*, vol. 22, no. 2, pp. 133 – 187, 1998.

[2] F. Lerdahl and R. Jackendoff, *A generative theory of tonal music*. Cambridge, MA: MIT Press, 1983.

[3] E. Jost, *Free jazz*. Universal Edition Graz, 1974.

[4] J. Schwartz, *Free Jazz: A Research and Information Guide*. Routledge, 2018.

[5] D. Bailey, *Improvisation: Its Nature And Practice In Music*. Da Capo Press, 1993.

[6] E. Prévost, *No Sound is Innocent: AMM and the Practice of Self-invention, Meta-musical Narratives, Essays*. Copula, 1995.

[7] C. Canonne and N. B. Garnier, "Cognition and Segmentation In Collective Free Improvisation: An Exploratory Study," in *International Conference on Music Perception and Cognition*, Thessaloniki, Greece, 2012.

[8] M. Pelz-Sherman, "Suggested Applications of Musical Improvisation Analysis to Sonification," in *International Workshop on Interactive Sonification*, York, UK, 2007.

[9] G. E. Lewis, "Too many notes: Computers, complexity and culture in voyager," *Leonardo Music Journal*, vol. 10, pp. 33–39, 2000.

[10] A. Ozerov, E. Vincent, and F. Bimbot, "A general flexible framework for the handling of prior information in

audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 4, pp. 1118–1133, May 2012.

[11] D. Fitzgerald, "Harmonic/percussive separation using median filtering," in *International Conference on Digital Audio Effects (DAFx)*, 2010.

[12] R. Bittner, J. Salamon, M. Tierney, M. Mauch, J. Bello, and et al., *MedleyDB: A Multitrack Dataset for Annotation-Intensive MIR Research*, 10 2014.

[13] S. Mimilakis, E. Cano, J. Abeber, and G. Schuller, "New sonorities for jazz recordings: Separation and mixing using deep neural networks," in *International Conference on Music Information Retrieval*, 09 2016.

[14] E. Perez-Gonzalez and J. Reiss, "Automatic equalization of multichannel audio using cross-adaptive methods," in *Audio Engineering Society Convention 127*, Oct 2009.

[15] M. D. Plumbley, S. A. Abdallah, J. P. Bello, M. E. Davies, G. Monti, and M. B. Sandler, "Automatic music transcription and audio source separation," *Cybernetics and Systems*, vol. 33, pp. 603–627, 2002.

[16] S. Ewert, B. Pardo, M. Mueller, and M. D. Plumbley, "Score-informed source separation for musical audio recordings: An overview," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 116–124, May 2014.

[17] E. M. Grais, D. Ward, and M. D. Plumbley, "Raw multi-channel audio source separation using multi-resolution convolutional auto-encoders," *CoRR*, vol. abs/1803.00702, 2018.

[18] A. Honingh and R. Bod, "Convexity and the well-formedness of musical objects," *Journal of New Music Research*, vol. 34, no. 3, pp. 293–303, 2005.

[19] A. Forte, *The Structure of Atonal Music*. Yale University Press, 1973.

[20] C. Roads and P. Wieneke, "Grammars as representations for music," *Computer Music Journal*, vol. 3, no. 1, pp. 48–55, 1979.

[21] E. Neeman, "Free Improvisation as a Performance Technique: Group Creativity and Interpreting Graphic Scores," Ph.D. dissertation, The Julliard School, New York, 5 2014.

[22] V. Goudard, "John, the semi-conductor : a tool for comprovisation," *CoRR*, vol. abs/1811.06858, 2018.

[23] R. Dudas, "Comprovisation: The Various Facets of Composed Improvisation within Interactive Performance Systems," *Leonardo Music Journal*, vol. 20, pp. 29–31, 2010.

[24] R. T. Dean, F. Bailes, and J. Drummond, "Generative structures in improvisation: Computational segmentation of keyboard performances," *Journal of New Music Research*, vol. 43, no. 2, pp. 224–236, 2014.

[25] J. Pressing, "The micro- and macrostructural design of improvised music," *Music Perception: An Interdisciplinary Journal*, vol. 5, no. 2, pp. 133–172, 1987.

[26] C. Canonne, "Focal Points in Collective Free Improvisation," *Perspectives of New Music*, vol. 2, 2013.

[27] J. Nattiez, "Fondements d'une semiologie de la musique," *Journal of Aesthetics and Art Criticism*, vol. 35, no. 2, pp. 239–242, 1976.

[28] S. W. Smoliar, "Music programs: An approach to music theory through computational linguistics," *Journal of Music Theory*, vol. 20, no. 1, pp. 105–131, 1976.

[29] J. A. Moorer, "Music and computer composition," *Commun. ACM*, vol. 15, no. 2, pp. 104–113, Feb. 1972.

[30] T. Winograd, "Linguistics and the computer analysis of tonal harmony," *Journal of Music Theory*, vol. 12, no. 1, pp. 2–49, 1968.

[31] A. M. Perlman and D. Greenblatt, "Noam Chomsky Meets Miles Davis: Some Observations on Jazz Improvisation and Language Structure," in *The sign in music and literature*, W. Steiner, Ed. Austin, TX: University of Texas Press, 1981.

[32] B. L. J. Smith, I. Schankler, and E. Chew, "Listening as a creative act: Meaningful differences in structural annotations of improvised performances," *Music Theory Online*, vol. 20, no. 3, 2014.

[33] E. D. Scheirer, "Music-Listening Systems," Ph.D. dissertation, Massachusetts Institute of Technology, Boston, MA, 6 2000.

[34] D.-J. Povel and E. Jansen, "Perceptual mechanisms in music processing," *Music Perception: An Interdisciplinary Journal*, vol. 19, no. 2, pp. 169–197, 2001.

[35] J. K. Wright and A. S. Bregman, "Auditory stream segregation and the control of dissonance in polyphonic music," *Contemporary Music Review*, vol. 2, no. 1, pp. 63–92, 1987.

[36] A. S. Bregman and J. Campbell, "Primary auditory stream segregation and perception of order in rapid sequences of tones." *Journal of Experimental Psychology*, vol. 89, no. 2, pp. 244–249, 1971.

[37] E. Cambouropoulos, "The musical surface: Challenging basic assumptions," *Musicae Scientiae*, vol. 14, no. 2_suppl, pp. 131–147, 2010.

[38] M. Silverstein, "Metapragmatic discourse and metapragmatic function," in *Reflexive Language: Reported Speech and Metapragmatics*, J. A. Lucy, Ed. Cambridge University Press, 1993, pp. 33–58.

[39] R. K. Sawyer, *Group creativity: Music, theater, collaboration.* Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, 2003.

[40] I. T. Monson, *Saying Something: Jazz Improvisation and Interaction*. Chicago: University of Chicago Press, 1996, winner of the Sonneck Society's 1998 Irving Lowens Prize for best monograph in American music published in 1996.

[41] R. K. Sawyer, "Improvisation and the creative process: Dewey, collingwood, and the aesthetics of spontaneity," *The Journal of Aesthetics and Art Criticism*, vol. 58, no. 2, pp. 149–161, 2000.

[42] V. Iyer, "Exploding the narrative in jazz improvisation," in *Uptown Conversation: The New Jazz Studies*, R. O'meally, B. Edwards, and F. Griffin, Eds. New York: Columbia University Press, 2004.

[43] F. Okiji, "Storytelling in jazz work as retrospective collaboration," *Journal of the Society for American Music*, vol. 11, no. 1, p. 7092, 2017.

[44] G. B. Wilson and R. A. R. MacDonald, "The sign of silence: Negotiating musical identities in an improvising ensemble," *Psychology of Music*, vol. 40, no. 5, pp. 558–573, 2012.

[45] G. Michaelsen, "Analyzing musical Interactionin Jazz improvisations of the 1960s," Ph.D. dissertation, Indiana University, Bloomington, IN, 2013.

[46] M. Pelz-Sherman, "A Framework for the Analysis of Performer Interactions in Western Improvised Contemporary Art Music," Ph.D. dissertation, University of California, San Diego, CA, 1998.

[47] T. E. Nunn, "Wisdom of the Impulse: On the Nature of Musical Free Improvisation," 1998.

[48] O. Nieto and J. Bello, "Systematic exploration of computational music structure research," in *Proceedings of the 17th International Society for Music Information Retrieval Conference (ISMIR)*, 8 2016.

[49] C. i Wang and G. J. Mysore, "Structural segmentation with the variable markov oracle and boundary adjustment," *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 291–295, 2016.

[50] J. Foote, "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)*, vol. 1, July 2000, pp. 452–455 vol.1.

[51] O. Nieto and T. Jehan, "Convex non-negative matrix factorization for automatic music structure identification," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 236–240.

[52] B. McFee and D. P. W. Ellis, "Learning to segment songs with ordinal linear discriminant analysis," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 5197–5201.

[53] ——, "Analyzing song structure with spectral clustering," in *Proceedings of the 17th International Society for Music Information Retrieval Conference (ISMIR)*, 2014.