DYNAMIC REGIMES IN FREE IMPROVISED JAZZ MUSIC

Matt Setzler
Indiana University
mattsetz@gmail.com

Minje Kim
Indiana University
minje@indiana.edu

Tyler Marghetis
Indiana University
tyler.marghetis@gmail.com

ABSTRACT

In ‘free jazz’ ensembles, musicians collectively generate new improvised music without a score, preconceived song structure, or explicit advance planning. Free jazz is structured quite differently from more extensively studied genres—such as pop, classical, and even straight-ahead jazz—as it lacks recurring song forms and stable tonal structure. Instead, free jazz music operates more like an open-ended conversation, in which the topic of discourse changes over time as a function of ongoing interactions of the conversants. Inspired by theoretical developments in cognitive science, we conceive of free jazz ensembles as distributed cognitive systems — systems in which the cognitive function of a group emerges out of the coordinated activity of its constituent members. Such distributed cognitive systems have been shown to support the emergence of stable patterns of activity (regimes), systematic drift from one regime to another, and sudden transitions between distinct regimes. In the present study we describe unsupervised methods for identifying and quantifying these emergent dynamics in a corpus of free jazz audio recordings. In so doing, we empirically demonstrate ways in which free jazz ensembles resemble other distributed cognitive systems — and illustrate the benefits to MIR of leveraging existing theory in cognitive and complexity science.

1. INTRODUCTION

In most musical genres, both composition and performance are constrained by a set of recurring structures. Both contemporary pop music and baroque cantatas, for example, follow recurring song forms and stable tonal structure; a typical pop song consists of a series of verses separated by a recurring chorus, along with an intro, a bridge, and an outro. This recurring structure is especially useful in the context of MIR, where it can place *a priori* constraints on search, segmentation, and labeling. But this becomes more complicated when there is no such standard structure – such as in the genre of jazz music know as ’free jazz.’ In free improvised jazz (in contrast with more traditional ’straight ahead’ jazz), musicians collectively improvise new music without a score, preconceived song structure, or explicit advance planning. Performances can vary wildly in sound, structure, and length. How, then, should we make sense of the structure of such unstructured music?

While free jazz is not constrained by standard forms, the experience at a free jazz performance is nevertheless not just of random noise. Instead, free jazz music operates more like an open-ended conversation, in which the topic of discourse changes over time as a function of ongoing interactions of the conversants. As a result, the audience may experience structure at multiple levels, from brief shifts in the dynamics of the performance, to a cohesive structure that describes the piece as a whole. The present study draws inspiration from this analogy – and from associated research in cognitive science and complex systems theory – to inform an approach to that is both theoretically informed and data-driven.

1.1 Conversations and complex systems

Ensemble music performance, and free jazz performance in particular, are examples of *complex systems* [2]. Complex systems are systems that consist of multiple interacting parts, in which the behavior of the system as a whole (e.g., a music ensemble) is determined primarily by the interactions among parts rather (than the parts in isolation) [3]. This is certainly true of free improvised jazz, where the musical performance is driven by interactions among musicians rather than by some form of centralized control (e.g., a musical score, a conductor, etc.).

It is useful to compare the complex system of jazz improvisation to another instance of coordinated human activity: multi-person conversation. Like free jazz performances, conversations are typically unplanned. Nevertheless, reliable structure emerges at multiple timescales [9]. This structure ranges from the brief dynamics of turn-taking to longer-timescale regularities in the particular topics being discussed [1, 7]. This structure is "emergent," in the sense that it does not reflect the goals or intentions of any individual person in the conversation, but rather it exists at the level of the entire distributed system.

Within conversations in particular and complex systems more generally, qualitatively different regimes of activity can emerge. Many complex systems will exhibit multiple "attractors," states of activity towards which the system will converge (citation). When a system enters the ’basin of attraction’ for an attractor, the system’s activity can gradually move towards the attractor. For instance, for a particular group of people in conversation, Canadian politics may be an attractor; if the topic of their conversation gets close to this attractor (e.g., somebody brings up a political issue), then they may gradually shift the topic of conversation until they are explicitly discussing Canadian politics.

On the other hand, once a complex system reaches an attractor (i.e., is near the center of the attractor’s "basic of attraction"), the system can exhibit striking stability — despite the lack of a central control. A group conversation can get stuck on a particular topic for an extended period of time, if that topic is attractor for the system.

Finally, after a system becomes established in a stable regime, it can sometimes undergo a sudden jump to an entirely different regime – an event known as a 'critical transition' [8]. A conversation that has been stuck talking about Canadian politics might suddenly shift, for instance, to a discussion of the best new neighborhood restaurants. The theory of complex systems has identified a number of 'early warning signals' of such transitions [8]; past work has applied these to free jazz improvisation [10].

1.2 The emergent dynamics of free jazz

Improvising free jazz ensembles are an excellent example of a complex system. The spontaneously generated musical structures in free jazz are emergent products of a distributed cognitive system of mutually interacting improvisers. This is qualitatively different from most musical genres, in which musical structure is specified ahead of time in the form of a written score, or template which constrains improvised activity (in the case of straight-ahead jazz). Furthermore, free jazz ensembles often eschew or deviate from the use of tonal harmony or pulse-based rhythm. How then might we extract musical information from such a seemingly unstructured musical enterprise?

In the present study we draw inspiration from complex systems theory and research in distributed cognition to identify emergent structure in free improvised jazz music. Because we can't rely on ordered tonal and rhythmic structure, we analyze how "acoustic textures" change over improvised musical pieces. The acoustic textures displayed by free jazz ensembles vary in terms of volume (from very loud, to prolonged periods of complete silence), instrumentation (from many instruments playing together, to only a single instrument) and pitch (e.g. whether instruments are playing all high pitches, low pitches or any combination therein).

We propose three dynamic regimes that identify different ways in which an acoustic texture can change (or persist) over time: stability, directed drift and transitions. "Stability" refers to an extended musical window in which the ensemble maintains more of less the same acoustic texture. "Directed drift" refers to the degree to which an ensemble's playing changes in a systematic, directed way, starting with one acoustic texture and drifting systematically towards a different acoustic texture. "Transitions" refer to sudden changes from one stable acoustic texture to a very different acoustic texture. For instance, a piece might start with a sustained drum solo, and then very quickly transition to a prolonged period of silence. Or, a section might start with a period where multiple instruments are playing very quietly, and then very quickly transition to a period where a single instrument is producing a loud, high-pitched sound. These would each be examples of sudden transitions. We present unsupervised methods for automatically detecting these three dynamic regimes (in parallel) in a corpus of free jazz audio recordings.

1.3 Current study

The current study draws on theories of distributed cognition and complex systems to develop unsupervised methods for quantifying dynamical regimes within free jazz performances. We describe a new corpus of free jazz recordings. We then describe three measures: a simple measure of the acoustic *stability* of a musical performance; a novel measure of systematic, directed *drift* in a musical performance; and a measure of whether the performance is undergoing a sudden transition from one stable regime to another, based on the novelty function introduced by Foote [5]. We then use these three measures to quantify the emergent structure of free jazz – including the temporal distribution of stability, drift, and sudden transitions over the course of free jazz pieces, as well as the interrelationships of these three dynamical aspects. We also verify that our measures align with subjective human experiences. In the Discussion, we unpack the promise of our approach for quantifying the dynamics of un-scored, fully improvised music more generally, and discuss the value of combining unsupervised methods with theoretical insights from cognitive science and complex systems theory.

2. METHODS

2.1 Corpus

Our corpus consisted of 56 audio recordings of free jazz music. Altogether there was 6 hours and 41 minutes of improvised material in total. Each recording contained one complete improvisation, from beginning to end. All recordings were taken from professional improvising ensembles. These ensembles ranged in terms of size and instrumentation, although we only included recordings that had at least two performers collaborating together. The average performance duration was 7 minutes 10 seconds, although there was considerable variability (min = 56.seconds, max = 2426 seconds, standard deviation = 391 seconds). Some of the recordings were recorded in professional studios, while others were recorded at live performances and subsequently released as full albums. Most of the recordings in our corpus are publicly available albums, but a small fraction are private and were given to us directly by the artists for the purpose of this project. We find the variability in our recordings to be a strength which always us to address general questions about collective improvisation without being limited to specific musical settings such as ensemble size.

2.2 Pre-Processing

All of our measures are based off the self-dissimilarity matrix. For each recording, we first extracted time series of 12 MFCC components in 2 second frame sizes with 0.1 second hop size. The large frame size was used to account for sections in which instruments left spaces of silence between successive phrases. We wanted to contributions from these musicians to be accounted in as many frames as possible, so long as they were consistently playing. After extracting MFCC time series we computed the dissimilarity matrix with .5 seconds between successive indices, using Euclidean distance as a measure of dissimilarity.

2.3 Stability

Stability refers to the extent to which a window of music maintains more or less the same acoustic texture. There is a direct negative correlation between stability and the mean dissimilarity between frames in a given window. To formalize this notion, we first obtained time series of 'instability' by computing average dissimilarity in sliding windows of length 10, 30 and 60 seconds, with a hop size of 0.5 seconds. We then measured stability by subtracting instability values from the maximum similarity out of all the recordings. Thus 0 similarity is not an absolute measure, but represents the highest level of dissimilarity in our dataset, and the maximum similarity value is also relative to the recordings in our corpus.

2.4 Directed Drift

Directed drift refers to a gradual, systematic transition from one acoustic texture to another. Therefore, in a window with high directed drift, frames should successively become more similar to the final frame of the window as they approach it. In other words, we should expect that dissimilarity between successive frames in a window of high drift should monotonically decrease as they approach the final frame in the window is approached. To assess such monotonic decrease in dissimilarity, we measured the rank-based correlation between dissimilarity and the (negative) rank-based temporal distance to end of window using kendall's tau. This gave us a measure of drift which ranged between [-1,1], with 1 representing the highest possible degree of directed drift. This was performed on all recordings with sliding windows of length 10, 30 and 60 seconds, and a hop size of 0.5 seconds.

2.5 Transitions

To detect transitions, we used Jonathan Foote's novelty function with window sizes 30 and 60 seconds [5]. For time efficiency we computed this on subset of dissimilarity matrix, with 2 seconds between successive indices. We used the method in its simplest form – without any kernal smoothing (i.e. all indices in the dissimilarity matrix window were weighted equally).

3. RESULTS

3.1 Distribution of Dynamic Regimes

3.1.1 Stability and Drift

How are dynamic regimes distributed throughout improvised performances? What is the relationship between regimes at different timescales? Figure 1 displays the distribution of stability (A) and drift (B) for all time indices and all pieces in the corpus. It is easy to observed that on average stability increases with smaller window sizes. This is to be expected, as larger windows comprised more points in the dissimilarity matrix, as well as longer time window in which musical textures may change. As described previously, the maximum value of stability is relative to our data-set and window size used.

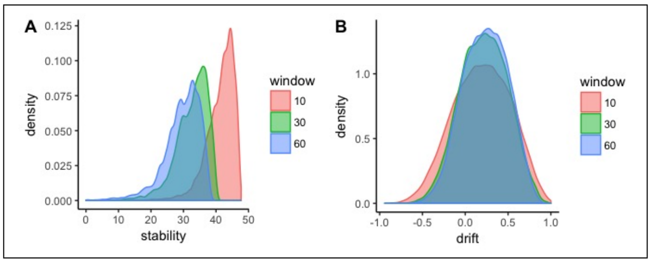

Figure 1. Distribution of stability and drift across all pieces.

Figure 1 [B] depicts the distribution of directed drift across all pieces. Drift is bounded between [-1,1], with positive values indicating a large degree of systematic, directed drift. Unlike stability, mean drift is roughly same at the three window sizes assessed. Interestingly, overall drift is positive, as opposed to being centered at 0, which would indicate no drift (mean=0.21; t(133,350)=267, $p < 0.001$). This indicates that there is generally some "forward momentum" characterizing the improvisations in our corpus.

What is the relationship between stability and drift? Intuitively it may seem like they should be mutually exclusive. But its worth noting that a section can be characterized by low stability and low drift, in the event that improvisers are deviating widely from any given texture without moving towards any particular texture ("random walk"). It could also be the case that musicians drift systematically, but very slowly from one texture to another, which would result in a high degree of directed drift, while maintaining relatively high stability. To observed the interplay between stability and drift in the improvised pieces in our performance, we fit a linear model between the two measures for every frame in each piece and window, which demonstrated a negative relationship between stability and drift (m=-1.182).

3.1.2 Transitions

Next we sought to understand how "novelty" is distributed across improvised performances. How often do musicians transition from one stable texture to a qualitatively different texture? And what is the magnitude of such transitions? To explore these questions, we looked at the level of novelty at all local maxima in our novelty time series for each piece. Local maxima with novelty less than zero were filtered out.

As depicted in Figure 2 (A), there is a long-tailed distribution of novelty in these pieces. A large number of small-magnitude transitions were identified, but the frequency quickly drops such that there are a small number of very large-magnitude transitions. Figure 2 (B) depicts

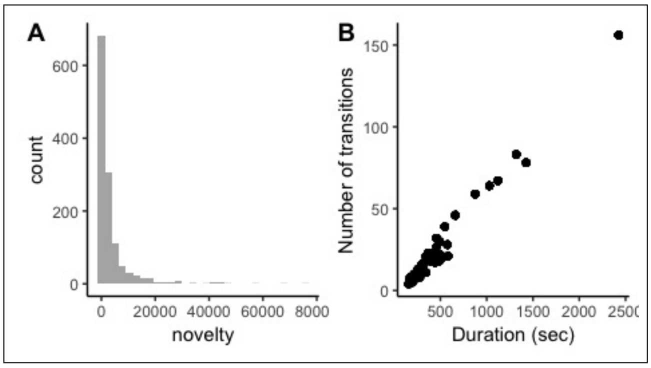

Figure 2. Distribution of novelty across corpus (A) and number of transitions per piece plotted against piece duration (B).

the number of transitions that occur in a particular piece as a function of its duration. There is a striking positive linear trend between number of transitions and duration – improvising ensembles were proportionately more likely to transition the longer they improvised for. This suggests that whatever the mechanism that determines rate of transition in improvising ensembles, it appears to be constant despite variable-length performances.

3.2 Dynamic Regimes Over Time

Next we examine how dynamic regimes, specifically stability and directed drift, evolve over the course of improvised performances. To control for variability in piece length, we looked at how these regimes changed over the "normalized" time course of performances, with 0 representing the beginning of a piece and 1 representing the end.

Figure 3 depicts the mean stability (A) and mean drift (B) for all pieces over normalized time. There are several interesting trends to notice. Overall, improvisations begin with a low level of stability, which gradually increases for most of the performance, until about 80%, at which point there is a sharp decrease in stability which continues into the end of the performance. This trend is robust in the sense that it can be observed at all timescales.

Conversely, there is no robust trend in drift consistent across timescales in the first half of pieces. That being said, drift begins to gradually increase across all timescales roughly halfway through improvised pieces. This trend continues until the piece is about 80% complete, at which point there is a dramatic ramp-up of drift until the end of pieces. This sharp increase in drift results in there being higher directed drift at the ends of pieces as opposed to other positions in a piece. This was statistically confirmed in a supplementary analysis which compared drift in the final 30 seconds of all pieces (mean = .343) with drift from all other sections (mean = .214; t(10,927) = -37, p < 0.001).

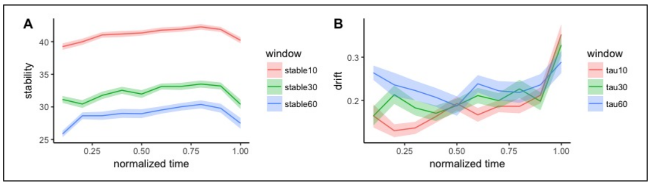

Figure 3. Mean change in stability (A) and drift (B) over the course of improvised pieces. Error ribbon denotes standard error of the mean

Taken together, these results tell us something interesting about how emergent musical structure evolves throughout the course of improvised performance. That stability starts off low, and gradually increases in the first of pieces may reflect a period in which improvisers are "gaining their bearings" and trying to collectively negotiate a coherent musical structure. This doesn't happen immediately, but gradually emerges out of the continued interplay, eventually leading to stable acoustic textures. The sharp increase in directed drift at the later on tells us something interesting about how endings are improvised. As opposed to a sharp, abrupt ending, which might occur in composed musical settings, we observe a systematic tendency for ensembles to gradually "fade out" into silence.

3.3 Relation to with human ratings

Although the current study was not focused on the subjective experience of dynamical regimes in free jazz, we nevertheless collected a small sample of human judgments in order to confirm that our unsupervised methods were, indeed, tracking aspects of free jazz that were subjectively identifiable by listeners. Human listeners (*N* = 5) ranged considerably in musical training, from none whatsoever to extensive training in classical violin, and all listened to free jazz either less than once a month or never. We extracted audio clips that were assigned either low (*n* = 5) or high (*n* = 5) values on each of our three unsupervised measures (i.e, drift, stability, novelty). After a brief description of each dynamic (e.g., stability "means that the entire clip maintains more or less the same acoustic texture"), raters listened to each clip (*N* = 30) and were asked to judge, on a Likert (1-5) scale, how much it exemplified the relevant dynamic (i.e., drift, stability, novelty).

These human ratings confirmed that all three measures captured dynamics that were subjectively apparent to listeners. We analyzed ratings using a linear mixed effects model, thus accounting for variability due to both individual participants and specific audio clips. For clips that were assigned low values by our unsupervised measures, human ratings were also low on the 1-5 scale ($M_{low} = 1.6$); for those assigned high values, human ratings were also high ($M_{high} = 4.3$). Our linear model confirmed that this difference was significant (effect of unsupervised assignment [i.e., high/low] on human ratings: $b = 2.6 \pm 0.3$ SEM, $p < .0001$) and did not differ between the three kinds of dynamics ($ps > .7$; drift: $M_{low} = 1.4$ vs. $M_{high} = 4$; stability: 1.5 vs. 4.3; novelty: 1.9 vs. 4.5). Our unsupervised measures thus succeeded in capturing dynamic aspects of free jazz performance that aligned with untrained human experience.

As a complementary analysis, we used human ratings to establish the 'ground truth' for drift, stability, and transitions. We calculated the mean rating for each clip; clips were considered to exhibit a phenomenon (e.g., drift) if the mean rating was greater than 3 (out of 5). Relative to this measure of ground truth, our unsupervised measures performed well, achieving perfect precision (1.0), excellent recall (0.94), and thus excellent F1 (0.97).

4. DISCUSSION

In the present study we draw on research in cognitive science and complex systems to propose features three features that describe dynamic structure in free jazz music. We devised (borrowed, in the case of novelty), implemented and validated (against listener ratings) unsupervised measures to automatically detect this structure in a corpus of audio recordings of free improvised music. This enabled us to perform a corpus analysis which revealed interesting patterns in how improvisers collectively generate, maintain and evolve musical structure without any a priori template.

While we used a radically different approach, in some sense the present study relates to work in segmentation in pop and classical music, as well as efforts to automatically detect musical highlights in a piece, such as the "drop" in trap music [4, 6, 11]. There is excellent work being done to tackle these problems, but because of the musical conditions under which free jazz operates, those efforts may not port over to this musical domain. In integrating theory from cognitive science allowed us to gain traction on a challenging MIR problem – unsupervised segmentation and structural analysis in a seemingly unstructured musical domain. The methods presented here may be useful in the future for detecting "highlights" in freely improvised music. They may also be of interest to musicologists, as well as cognitive scientists who wish to analyze free jazz audio recordings as a group-level behavioral trace of a distributed cognitive system (i.e. improvising jazz ensembles).

There are a number of extensions and developments we are excited for in the future. Most notably, we would like to increase the size of our corpus so that we can evaluate collective patterns and emergent structure in free improvisation with more statistical confidence. We would also like to obtain listener ratings from a much larger sample of individuals, ideally with diverse listening preferences. It will also be of interest to extend our corpus analysis and investigate some of the questions raised by the current study. For instance, we know that the number of transitions in a piece scales linearly with its duration. But what is the temporal pattern with which transitions occur? Do they occur at a constant rate, or are they clustered in bursts?

5. CONCLUSION

In summary, we proposed three new features specifically relevant to free improvised jazz. The features are inspired by the understanding that improvising music ensembles are a kind of complex, distributed cognitive system that exhibits emergent structure. The proposed features were grounded in implementations of unsupervised metrics, which were verified against listener ratings. Finally, we used said metrics to analyze a corpus of free jazz audio recordings. In so doing we demonstrate ways in which MIR techniques can benefit from integrating theory in cognitive science, and also gain some insight into how improvising music ensembles resemble other distributed cognitive systems.

6. REFERENCES

[1] Daniel Angus, Bernadette Watson, Andrew Smith, Cindy Gallois, and Janet Wiles. Visualising conversation structure across time: Insights into effective doctor-patient consultations. *PloS one*, 7(6):e38014, 2012.

[2] David Borgo. *Sync or swarm: Improvising music in a complex age*. A&C Black, 2005.

[3] Clément Canonne and Jean-Julien Aucouturier. Play together, think alike: Shared mental models in expert music improvisers. *Psychology of Music*, 44(3):544–558, 2016.

[4] David De Roure, Kevin R Page, Benjamin Fields, Tim Crawford, J Stephen Downie, and Ichiro Fujinaga. An e-research approach to web-scale music analysis. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 369(1949):3300–3317, 2011.

[5] Jonathan 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)*, volume 1, pages 452–455. IEEE, 2000.

[6] Jouni Paulus, Meinard Müller, and Anssi Klapuri. State of the art report: Audio-based music structure analysis. In *ISMIR*, pages 625–636. Utrecht, 2010.

[7] Harvey Sacks, Emanuel A Schegloff, and Gail Jefferson. A simplest systematics for the organization of turn taking for conversation. In *Studies in the organization of conversational interaction*, pages 7–55. Elsevier, 1978.

[8] Marten Scheffer, Jordi Bascompte, William A Brock, Victor Brovkin, Stephen R Carpenter, Vasilis Dakos, Hermann Held, Egbert H Van Nes, Max Rietkerk, and George Sugihara. Early-warning signals for critical transitions. *Nature*, 461(7260):53, 2009.

[9] Emanuel A Schegloff. *Sequence organization in interaction: A primer in conversation analysis I*, volume 1. Cambridge University Press, 2007.

[10] Matt Setzler, Tyler Marghetis, and Minje Kim. Creative leaps in musical ecosystems: early warning signals of critical transitions in professional jazz.

[11] Karen Ullrich, Jan Schlüter, and Thomas Grill. Boundary detection in music structure analysis using convolutional neural networks. In *ISMIR*, pages 417–422, 2014.