Emergent Formal Structures of Factor Oracle-Driven Musical Improvisations

Isaac Schankler[1], Jordan B.L. Smith[1], Alexandre R.J. François[2], Elaine Chew[1]

[1] University of Southern California, Los Angeles, CA, USA,
{schankle,jordans,echew}@usc.edu,
[2] Harvey Mudd College, Claremont, CA, USA
alex@cs.hmc.edu

Abstract. In this article, improvisations created with the factor oracle, a commonly used data structure in machine models of musical improvisation, are shown to exhibit certain formal structures independent of the musical content. We posit that these structures are in fact emergent properties of the behavior of the factor oracle itself. An expert improviser (the first author) performed a series of improvisations with Mimi, a factor oracle-driven multimodal system for human-machine improvisation, and the formal structures of each performance was independently analyzed by the performer and an experienced music structure annotator (the second author). Quantitative assessment of the similarity between the performer's and the listener's analyses was carried out using techniques from the field of automatic structure analysis. Supported by a comparison to baseline analysis approaches, the results suggest a high level of agreement between the two sets of analyses. Drawing upon this foundation of evidence, we discuss these analyses and their relationship to common classical forms, including canon- and rondo-like forms, as well as forms based on the juxtaposition of rhythmic cells.

Keywords: improvisation, factor oracle, musical structure, emergent structure

1 Introduction

When human improvisers perform, they frequently draw on a stock vocabulary of motives, gestures, and phrases. Ideas are recombined, often with very little creation of source material. The factor oracle is an efficient data structure used to model this recombination aspect of musical improvisation using an online construction algorithm and a stochastic traversal algorithm [1]. Assayag and Dubnov were the first to propose and use factor oracles in music improvisation with Omax [3, 4].

While there has been considerable development of improvisation engines that make use of the factor oracle, there has as of yet been little study of the products of these engines, i.e. the improvisations themselves, and of whether they actually resemble improvisations created by human performers. This article examines the large-scale forms created by factor oracle-driven improvisations to determine if any regular or recognizable structural patterns arise. We posit that these structural patterns, if consistently observed, could be emergent properties of the behavior of the factor oracle, or of the particular kind of interaction that exists between the oracle and a performer. Knowing these patterns can lead us to a deeper understanding of the formal structures in musical improvisation and of the nature of human-machine improvisation, and help us advance the development of more sophisticated models of improvisation.

The remainder of this section introduces the Mimi improvisation system and surveys some related research. Section 2 describes how this research was carried out, including the recording and analysis of three improvised performances and the quantitative tools used to compare them. The analyses themselves, the results of their comparison, and the small- and large-scale formal structures oberved in the analyses are reported in Section 3. Section 4, a discussion following the results, concludes the article.

1.1 The Mimi System

For this study, the particular factor oracle-driven improvisation engine used was the Mimi (Multimodal Interaction for Musical Improvisation) system [9, 8] created by François and collaborators using the Software Architecture for Immersipresence framework for interactive software systems. Mimi was designed for human-machine improvisation on a keyboard or some other MIDI instrument.

While Mimi draws inspiration from the growing family of improvisational systems that use the factor oracle, it differs from its siblings in a few significant ways. In the various incarnations of the OMax improvisation system, the machine learns continually from the human improviser, i.e. the factor oracle continues to grow through the duration of the performance. A human operator at the computer acts as a second performer, managing in real-time the recombination rates (i.e., the frequency with which skips are made while traversing the oracle) and the section of the oracle from which the machine draws musical material.

In contrast, the Mimi systems relegate full operational control to the improvising musician. The interaction modality and format are designed so that the human performer is self-sufficient. They also relegate more direct control to Mimi itself, and by extension the factor oracle algorithm. The human performer, for example, has no control over which section of musical material Mimi will choose to perform. (The human performer does, however, have the option at any moment to permanently clear Mimi's memory of the existing musical material.) This makes Mimi an ideal choice for the study of structure, since the structure of the improvisations will more closely reflect the structure of the factor oracle itself.

The Mimi systems were the first, and currently still the only, factor oracle-based systems to experiment with real-time visualization to aid the performer at the piano keyboard in the act of planning and orchestrating an improvisation. Mimi employs a piano roll-style visual display that gives the user information about the current state of the improvisation engine, a ten-second heads-up on

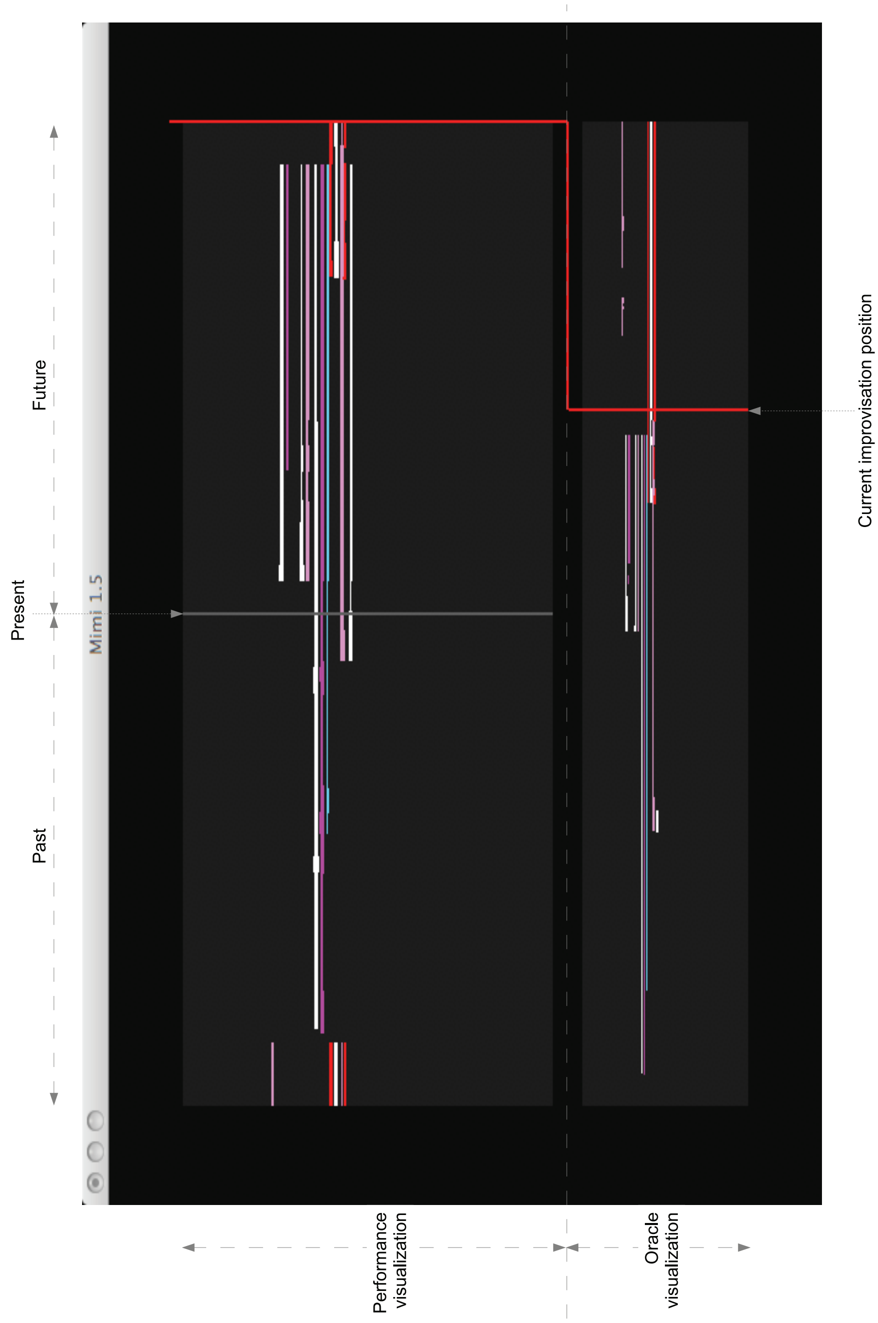

Fig. 1: Annotated screenshot of Mimi's visualization during performance.

the music the machine will soon play, and ten seconds to review the music that the machine and the user recently played together, as shown in Figure 1.

An alternative visualization scheme for the factor oracle which draws arcs representing suffix links between different sections of material is provided in WoMax[11], another offshoot of OMax. WoMax offers an overview of the connections created by the factor oracle. In the context of an ongoing improvisation, WoMax's visualization is most useful to a second performer (the one at the computer) who is focused on manipulating the audio generated by the oracle. The goal of Mimi's visualization is to provide timely feedback to a single performer focused on coordinating his or her musical ideas with the oracle's.

In the Mimi setup, the human improviser lays down seed material that the machine recombines to generate new music in the style of the seed material. The user, which could be the original human improviser or another person who chooses to use that particular set of seed material, can then improvise over the music that the machine generates. In the Mimi 1.5 system used in the performances described in this article, notes are color coded by pitch class, pedaling is recorded in the seed material, and volume (attack velocity) is indicated through the thickness of the line segments denoting each sounding tone. A MIDI controller provides an intuitive interface for the performer to have quick access to system commands and parameters before and during performance. Through the MIDI controller, the performer can make on-the-fly adjustments to Mimi's learning state (on, off), improvisation state (start, stop), memory (clear), recombination rate and playback volume.

2 Procedure

This section provides the details for the experimental and analytical procedures employed, including the recording of the human-machine improvisations, the analyses of these performances, and the quantitative analysis of the resulting annotations. The determination of formal structures, such as rondo or canon forms, was done by inspection based on the structure analyses.

2.1 Recording of Human-Machine Improvisations

Over a period of three weeks, an expert keyboard improviser (Isaac Schankler, the first author), produced three separate improvisations, hereafter referred to as Performance 1, 2 and 3, with Mimi. Two recordings of each session were produced: a MIDI recording, with the human improviser's and Mimi's performances saved in separate channels, and a video of Mimi's visual output, recorded using screen-capture software. The two MIDI tracks were used to synthesize a single audio track essentially identical to what the performer heard during the improvisation, and this merged result was the recording consulted during the annotation stage.

It bears mention that the improviser was a practiced Mimi user at the time of the recording of these performances. The performer's experience using Mimi extended several months prior to these recordings, and included several public performances, so that he was already accustomed to working with Mimi's behavior in order to craft coherent improvisations.

2.2 Annotation of Performances

To assess how, and the extent to which, the improvisations with Mimi were structured, the formal structure of each piece was independently analyzed and annotated by the performer and an experienced music structure annotator (Jordan Smith, the second author), hereafter referred to as Annotator 1 and Annotator 2, respectively. Annotator 2 has previously performed and studied analysis of large-scale music structure in his Master's research [13] and as part of the Structural Analysis of Large Amounts of Music Information (SALAMI) project[1].

In order to quantitatively compare the structural descriptions, they had to be encoded in a machine-readable format, preferably aligned with the audio. We therefore constrained the format of the analyses to disallow the overlapping of sections, to require that all portions of a piece be assigned a label, and to require that boundaries at a given hierarchical level were respected at smaller-scale (lower) levels. These requirements are akin to Lerdahl and Jackendoff's well-formedness rules for structural grouping analysis [10].

Each piece was analyzed at two hierarchical levels, the large-scale level was annotated with uppercase letters, and the small-scale level with lowercase letters. In each level, identical labels indicate portions of the piece that were judged to contain similar musical material. These analyses were entered by each annotator into the Variations Audio Timeliner [14], a tool for creating easy-to-read diagrams of formal musical structures, where the section boundaries were precisely aligned with the audio file.

2.3 Comparison of Annotations

To quantitatively estimate the correspondence between Annotator 1's and Annotator 2's analyses, the two sets of annotations were compared using four metrics that are commonly used to evaluate automatic structure analysis algorithms. The measures are: (1) the boundary f-measure; (2) the pairwise f-measure; (3) the K-measure; and, (4) the Rand index. Brief descriptions of each follow; details on their implementation can be found in Lukashevich [12].

The boundary f-measure gives the quality of the match between boundary locations. It is the harmonic mean of *precision*, the percentage of estimated boundaries that are correct, and *recall*, the percentage of annotated boundaries that were retrieved. By convention, either a 1-second or 6-second margin is used to determine whether a matching boundary was found.

The other three quantities measure the similarity between the labels provided by the two descriptions. The pairwise f-measure estimates the similarity of two sets of labels. Suppose the two descriptions are divided into very short time

[1] http://salami.music.mcgill.ca

windows. Consider the set of all pairs of windows in the annotated description with the same label − as an example, these are indicated by the arcs in Figure 2. We may then calculate the precision and recall with which this set of similarity relationships has been retrieved.

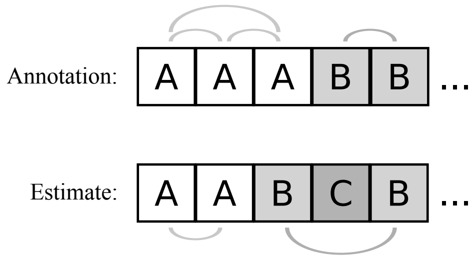

Fig. 2. Example of two sequences of windows being compared. Each arc connects two matching time segments. Precision is 50% because 1 of the 2 estimated pairwise matches were correct, and recall is 25% because 1 of the 4 matches in the annotation were retrieved. Their combined f-measure is thus the harmonic mean of the two, 33%.

The K-measure follows a similar derivation but is harsher when a labeling error matches one section in a description with many in another, rather than just two or three. The K-measure is based on "average cluster purity" and "speaker purity," terms analogous to precision and recall, respectively, and borrowed from a comparable metric used in speech processing research. The Rand index also resembles the pairwise f-measure, except that it additionally measures how well the set of dissimilarity relationships was retrieved.

The values output by each metric lie on a scale from 0 to 1, with 1 indicating the best and 0 the worst possible agreement between descriptions. However, the values alone are not very informative without a baseline against which to compare, since previous evaluations have shown that even random descriptions can be quite similar to a given description [13].

Baseline segmentation approaches include placing boundaries every n seconds (for various values of n), placing ten evenly-spaced boundaries throughout the piece, placing ten boundaries randomly, and placing no boundaries. For each segmentation, two labeling approaches apply: the segments may all be given the same label, or random labels may be drawn from a set of arbitrary size m. To set m fairly, when the baseline was measured against Annotator 1, m was the size of the set of labels in the matching description by Annotator 2, and vice versa. All of the above baseline segmentation and labeling approaches were used to generate descriptions that were compared to each human-generated description.

3 Results

3.1 Annotation of Performances

Figures 3 through 5 show the Variations Audio Timeliner diagrams created by the annotation method described in Section 2.2, in which differently colored "bubbles" represent different sections.

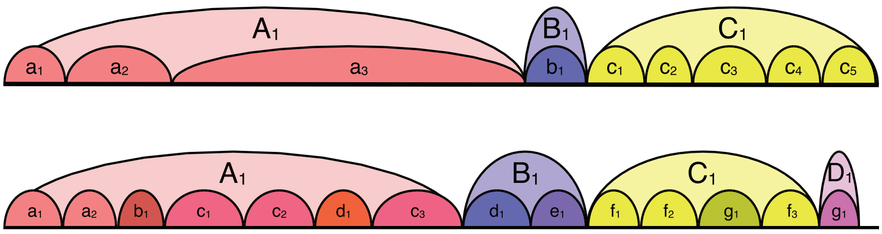

Fig. 3. Analysis of Performance 1 (top: Annotator 1; bottom: Annotator 2).

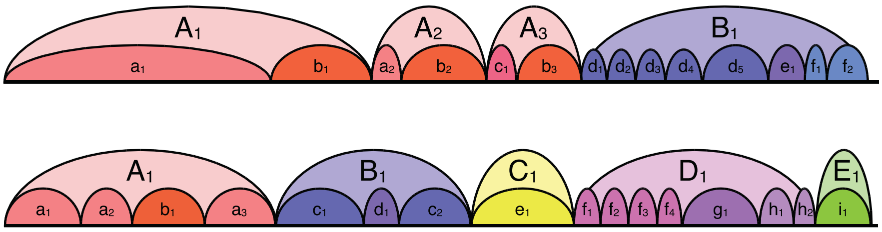

Fig. 4. Analysis of Performance 2 (top: Annotator 1; bottom: Annotator 2).

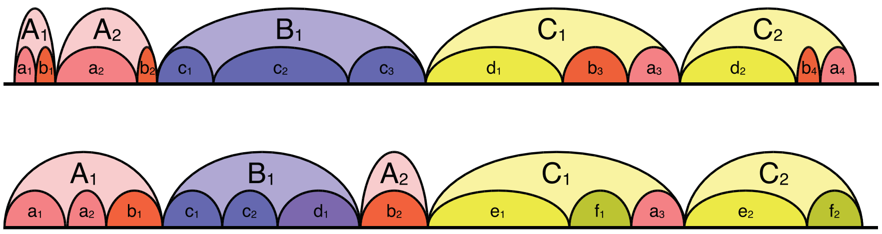

Fig. 5. Analysis of Performance 3 (top: Annotator 1; bottom: Annotator 2).

3.2 Comparison of Annotations

Although they were produced independently, the structural descriptions generated by the two annotators were quite similar one to another for each piece. Inspection of the three pairs of analyses, pictured in Figures 3–5, reveals that both Annotator 1 and Annotator 2 perceived similar large-scale structures. While the small-scale structure labels are less clearly alike, both annotators at least appeared to strongly agree on the placement of the boundaries. This is not a trivial result: given that the pieces were improvisations with no exact repetition (recall that Mimi's and the performer's tracks were merged), and given that the annotators did not consult each other except to establish a common format, one might not have expected the analyses to concur significantly.

These observations are reflected in the quantitative measurements as described below. They indicate that the placement of small-scale boundaries was generally quite similar between listeners, with an average f-measure of 0.737 using a window of plus or minus three seconds. Labeling was also quite similar, especially at the larger scale: across all performances, the average pairwise f-measure, K-measure and Rand index were 0.79, 0.80 and 0.85, respectively.

The detailed results are reported in Tables 1–5. The baseline generation method that matched the human annotations best overall was that in which ten boundaries were placed randomly and the labels were randomized. Because this baseline is random, the values reported in Tables 1–5 are the average over 20 trials. The standard deviation of these values ranged from 0.03 to 0.10, with an average of 0.069 and a median of 0.07.

As can be seen from Tables 1–5, in all but one case the two listeners' descriptions were more similar to each other than either was to the baseline description, and usually by a wide margin. These results suggest that the similarity between the listeners' descriptions is not accidental, and support the claim that Figures 3–5 are a fair representation of the formal structure of the corresponding performances. In the following section, we speculate as to why these structures may have arisen from the interaction between Mimi and the performer.

3.3 Formal Structures

This section describes the formal structures deduced from the analyses shown in Figures 3 through 5, and posits some attributes of Mimi's behavior that may contribute to the creation of these structures.

The Canon. The canon is one of the forms most obviously and fundamentally related to Mimi's performance habits. Consider the beginning of an improvisation session with Mimi. By design, Mimi's actions are delayed by 10 seconds to give the performer a visual heads-up on what Mimi is about to perform. If Mimi chooses not to recombine any of the subsequent material – a distinct possibility, especially if the recombination rate is low – Mimi's performance will be an exact copy of the original performance, delayed by 10 seconds, as shown in Figure 6.

Table 1. Estimated boundary f-measure (± 1s) between the two annotators' descriptions (Ann1-Ann2), and between the baseline and each annotator's description (baseline-Ann1, baseline-Ann2). Results are reported for the small- and large-scale annotations of each performance. Values range between 0 and 1, with 1 being the optimal value. Highest values are boldfaced.

Boundary f-measure ± 1 second	Performance 1		Performance 2		Performance 3	
Scale:	Small	Large	Small	Large	Small	Large
Ann1 - Ann2	**0.27**	**0.29**	**0.28**	**0.40**	**0.39**	**0.53**
baseline - Ann1	0.02	0.01	0.02	0.00	0.05	0.03
baseline - Ann2	0.02	0.01	0.02	0.01	0.04	0.04

Table 2. Estimated boundary f-measure (± 6s) between each pair of descriptions. See Table 1 for details.

Boundary f-measure ± 6 second	Performance 1		Performance 2		Performance 3	
Scale:	Small	Large	Small	Large	Small	Large
Ann1 - Ann2	**0.74**	**0.50**	**0.65**	**0.40**	**0.82**	**0.85**
baseline - Ann1	0.18	0.10	0.14	0.09	0.25	0.17
baseline - Ann2	0.24	0.15	0.13	0.10	0.29	0.17

Table 3. Estimated pairwise f-measure between each pair of descriptions. See Table 1 for details.

Pairwise f-measure,	Performance 1		Performance 2		Performance 3	
Scale:	Small	Large	Small	Large	Small	Large
Ann1 - Ann2	0.47	**0.85**	**0.66**	**0.62**	**0.68**	**0.90**
baseline - Ann1	**0.50**	0.51	0.50	0.52	0.41	0.48
baseline - Ann2	0.34	0.50	0.40	0.43	0.36	0.48

Table 4. Estimated K-measure between each pair of descriptions. See Table 1 for details.

K-measure	Performance 1		Performance 2		Performance 3	
Scale:	Small	Large	Small	Large	Small	Large
Ann1 - Ann2	**0.60**	**0.84**	**0.70**	**0.70**	**0.68**	**0.87**
baseline - Ann1	0.56	0.53	0.55	0.56	0.44	0.51
baseline - Ann2	0.43	0.54	0.50	0.50	0.42	0.52

Table 5. Estimated Rand index between each pair of descriptions. See Table 1 for details.

Rand index	Performance 1		Performance 2		Performance 3	
Scale:	Small	Large	Small	Large	Small	Large
Ann1 - Ann2	**0.70**	**0.90**	**0.88**	**0.71**	**0.87**	**0.93**
baseline - Ann1	0.67	0.63	0.76	0.59	0.68	0.58
baseline - Ann2	0.58	0.61	0.72	0.52	0.63	0.58

In other words, Mimi will create a canon at the 10-second level with the original performer, regardless of what the performer chooses to play.

Human:	A	B	C	D	E	F	etc.
Mimi:		A	B	C	D	E	etc.

Fig. 6. Canon between Mimi and human performer.

This pattern was found in Performance 2 (see Figure 4), where a near-exact but delayed repetition contributed to the sense of regular phrasing (segments $d_1, d_2, d_3 \ldots$ in Annotator 1's analysis, $f_1, f_2, f_3 \ldots$ in Annotator 2's).

Even when this exact sequence does not occur, Mimi's habits create an abundance of canon-like, or heavily imitative, forms, such as that shown in Figure 7. Imitative forms can be described as those in which fragments of musical material are passed from one voice to another in a polyphonic texture – in the case of Figure 7, it is passed from the original performer to Mimi. While this and further examples in Figures 8 through 10 are hypothetical examples, they serve to illustrate patterns observed in the structure analyses.

Human:	A	B	C	D	E	F	etc.
Mimi:		A	B	C	A	D	etc.

Fig. 7. Canon-like form created by Mimi and human performer.

The Rondo. As Mimi begins to revisit different sections of musical material, more formal resemblances may emerge. In its conception, Mimi does not privilege any material over any other, but Mimi's design allows it to learn continuously from the performer during the course of an improvisation, if the performer chooses to allow it. In this scenario, material that enters into Mimi's memory earlier will more likely be heard more frequently by virtue of simply having been in memory for a longer period of time. 'A' is more likely to be heard than 'B,' which is more likely to be heard than 'C,' and so on.

As a result, over the course of a typical three- to five-minute improvisation, the very first thing Mimi records is likely to be the most significant piece of musical material. If Mimi revisits this initial idea often enough, it may take on the quality of a refrain, creating a rondo-like form in which the initial idea recurs several times over the course of a performance, interspersed with different, successive episodes.

Performance 3 (see Figure 5) presented an instance of this design, with both annotators identifying a regular recurrence of material from the opening 'A' section. This will happen quite often even if the human performer proceeds in a through-composed fashion (not revisiting ideas), as shown in Figure 8. However,

Human:	A	B	C	D	E	F	G	etc.
Mimi:		A	B	A	B	A	C	etc.

Fig. 8. Rondo-like form with Mimi and human performer (through-composed).

a sensitive performer may also choose to respond to Mimi's behavior in a way that brings out these formal divisions, as shown in Figure 9.

Human:	A	B	C	A	D	A	E	etc.
Mimi:		A	B	A	C	A	D	etc.

Fig. 9. Rondo-like form with Mimi and human performer (rondo-like).

Formal Divisions. The performer may also choose to clear Mimi's memory at some point, creating a blank slate to be populated with new material. In this way, the performer is able to create large-scale formal divisions between sections that do not share the same pool of material. Figure 10 gives an example of a large-scale binary form with nested rondo-like forms.

Human:	A	B	C	A	D	E	F	G	E	H	etc.
Mimi:		A	B	A	C		E	F	E	G	etc.

Fig. 10. Large-scale binary form with nested rondo-like forms.

This kind of formal division is the most salient aspect of Performance 1 (see Figure 3), which both annotators interpreted as consisting of two contrasting sections ('A' and 'C') with a transitionary middle section ('B'). This structure is also apparent in Performance 2 (Figure 4), in which the canon section ('B' in Annotator 1's analysis, 'D' in Annotator 2's analysis) contrasts strongly with the first half of the piece. This binary structure is apparent despite the annotators' disagreement over the labeling of the first half.

Up until now, much attention has been paid to the large-scale structures that Mimi tends to create, but not much has been said about the small-scale structures they are comprised of. The large-scale section boundaries indicated in the annotations only capture the most broad structures found in the improvisations. A close examination of a short excerpt of Mimi's performance shows that, on a smaller scale, Mimi operates quite differently than one might expect from looking at the large-scale structure.

Rhythmic Cells. In Figure 11, Mimi's opening phrases from the MIDI file of Performance 3 are transcribed and quantized. Repetitions of the same motive

Fig. 11. Opening phrases of Mimi's portion of Performance 3, with annotation of rhythmic cells. Note values were quantized to the nearest 16th note, while allowing for simple triplet divisions.

are bracketed and indicated with the same letter – this helps to indicate motivic connections that are obscured by the quantization, e.g., motive 'A' in m. 1 and motive 'A' in m. 6, which are identical in performance despite the slightly different notation. Here we see at work what Messiaen called *personnages rythmiques*, or rhythmic cells, individuated by particular sonorities, that can be expanded or contracted [5].

Mimi performs four distinct cells ('A' through 'D') before recombining them in various ways. Contractions of a cell are more common (e.g. the abbreviated form of 'A' at the end of m. 5) but expansions (such as the "stutter" in the middle of the version of 'C' that appears in m. 10) are also possible. Mimi's overwhelming preference for its initial idea can also be observed here. Motive 'A' recurs a total of 7 times, versus 4 iterations of 'B,' 2 of 'C' and 3 of 'D' before moving on to a new idea ('E').

Unlike the canon and the rondo, this particular kind of juxtaposition of rhythmic cells is regarded as a 20th century innovation, and Mimi's juggling of these particular cells is not wholly unlike the jumpy, erratic rhythms found in the pivotal "Danse sacrale" from Stravinsky's *Le sacre du printemps*, which informed Messiaen's conception of *personnes rhythmiques* as well as successive generations of composers who sought a new working theory of rhythm [5, 6]. However, since Mimi itself has no theory of rhythm, its improvisations lack the rigorous rhythmic structure and symmetries found in "Danse sacrale" and other composed music.

4 Discussion and Conclusion

The fact that large-scale forms can be identified in improvisations with both Mimi and a human performer gives rise to a question: do these forms arise from the human performer's conscious or unconscious intent, Mimi's behavior, or some combination of the above? In the present case, the human performer had improvised extensively with Mimi over several months before performing the improvisations analyzed here. During this time, the performer adapted his usual improvisational technique in order to create successful performances with Mimi (since Mimi is incapable of certain kinds of adaptation, like rhythmic or harmonic coordination with the human performer). In some cases, these adaptations even ran counter to the performer's usual technique (for example, certain kinds of harmonic movement would be restricted by Mimi's inability to follow along). In other words, while the performer had significant control over the improvisation's moment-to-moment *content*, Mimi had much more influence over the *structure* of that content.

We have discussed some of the formal structures observed in the recorded improvisations with the assumption that they emerge, at least in part, from Mimi's inherent behavior. However, whether this is a truly intrinsic aspect of Mimi's behavior or a complex interaction between Mimi and the personal tendencies of the performer is beyond the scope of this paper, and is a definite area for further study involving other performers and more performances.

Addressing rhythmic rigor in Mimi is a compelling avenue for future research. One approach to incorporating rhythmic sensitivity into a factor oracle-based system involves suppressing factor links between contexts with grossly different average rhythmic densities [13]; examination of exactly how this or other approaches might operate in the context of the Mimi system is warranted, as well as further study of the mechanisms by which the observed large-scale structures emerge from these smaller units.

Other recent upgrades to OMax suggest interesting possibilities. Newer versions of OMax traverse the factor oracle using a suffix link tree, improving the contextual continuity during traversal [2]. In addition, the use of an anticipatory model can improve the quality of the suffix links [7]. In this approach, several competing factor oracle models are iteratively improved using reinforcement learning, and the most successful model drives the output. Incorporating these improvements into Mimi may have interesting ramifications for its behavior.

In conclusion, the performances with the Mimi improvisational system appear to be rich in structure, both small-scale and large-scale. Following an investigation of two annotators' analyses of these performances – one being the performer himself and the other an experienced structure analyst – the emergence of several structuring behaviors was observed; these were attributed to Mimi's design and to basic interactions between Mimi and the performer. However, it remains uncertain whether these principles are truly intrinsic to Mimi's programming, or whether the unconscious structuring instincts of the performer were alone sufficient to produce these results. Further investigations, including the analysis of a greater number of performances with other musicians, should address this question.

5 Acknowledgments

This work was supported in part by a United States National Science Foundation (NSF) Grant No. 0347988, and by a University of Southern California (USC) Provost's Ph.D. Fellowship. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors, and do not necessarily reflect those of the NSF or USC.

References

1. Allauzen, C., Crochemore, M., Raffinot, M.: Factor oracle: A new structure for pattern matching. In: Pavelka, J., Tel, G., Bartošek, M. (eds.) SOFSEM'99: Theory and Practice of Informatics, LNCS, vol. 1725, pp. 295–310. Springer Berlin / Heidelberg, Milovy, Czech Republic (1999)
2. Assayag, G., Bloch, G.: Navigating the oracle: A heuristic approach. In: Proceedings of the International Conference on Music Computing. pp. 405–412 (2007)
3. Assayag, G., Bloch, G., Chemillier, M., Cont, A., Dubnov, S.: OMax brothers: A dynamic topology of agents for improvization learning. In: Proceedings of the ACM Workshop on Music and Audio Computing. pp. 125–132 (2006)
4. Assayag, G., Dubnov, S.: Using factor oracles for machine improvisation. Soft Computing 8(9), 604–610 (2004)
5. Boivin, J.: Musical analysis according to Messiaen: A critical view of a most original approach. In: Dingle, C., Simeone, N. (eds.) Olivier Messiaen: Music, Art and Literature, pp. 137–157. Ashgate Publishing, Ltd., Burlington, VT, USA (2007)
6. Boulez, P.: Stravinsky remains. In: Stocktakings from an Apprenticeship. Trans. Stephen Walsh, pp. 55–110. Clarendon Press, Oxford (1991)
7. Cont, A., Dubnov, S., Assayag, G.: A framework for anticipatory machine improvisation and style imitation. In: Proceedings of Anticipatory Behavior in Adaptive Learning Systems (ABiALS) (2006)
8. François, A.R.: Time and perception in music and computation. In: Assayag, G., Gerzso, A. (eds.) New Computational Paradigms for Computer Music, pp. 125–146. Éditions Delatour France / IRCAM (2009)
9. François, A.R., Chew, E., Thurmond, D.: Visual feedback in performer-machine interaction for musical improvisation. In: Proceedings of the International Conference on New Interfaces for Musical Expression. pp. 277–280 (2007)
10. Lerdahl, F., Jackendoff, R.: A Generative Theory of Tonal Music. MIT Press, Cambridge, MA (1983)
11. Lévy, B.: Visualising OMax. Tech. rep., Ircam (2009)
12. Lukashevich, H.: Towards quantitative measures of evaluating song segmentationu. In: Proceedings of the International Conference on Music Information Retrieval. pp. 375–380 (2008)

13. Smith, J.B.: A Comparison and Evaluation of Approaches to the Automatic Formal Analysis of Musical Audio. Master's thesis, McGill University (2010)
14. Yorgason, B., Halliday, J., Colvard, C.: Variations Timeliner [computer software]. Available online at: `http://variations.sourceforge.net/vat` (2008)