



Listening as a Creative Act: Meaningful Differences in Structural Annotations of Improvised Performances

Jordan B. L. Smith, Isaac Schankler, and Elaine Chew

NOTE: The examples for the (text-only) PDF version of this item are available online at:

http://www.mtosmt.org/issues/mto.14.20.3/mto.14.20.3.smith_schankler_chew.php

KEYWORDS: form, structure, perception, improvisation

ABSTRACT: Some important theories of music cognition, such as Lerdahl and Jackendoff's (1983) *A Generative Theory of Tonal Music*, posit an archetypal listener with an ideal interpretation of musical structure, and many studies of the perception of structure focus on what different listeners have in common. However, previous experiments have revealed that listeners perceive musical structure differently, depending upon their music background and their familiarity with the piece. It is not known what other factors contribute to differences among listeners' formal analyses, but understanding these factors may be essential to advancing our understanding of music perception.

We present a case study of two listeners, with the goal of identifying the differences between their analyses, and explaining why these differences arose. These two listeners analyzed the structure of three performances, a set of improvised duets. The duets were performed by one of the listeners and Mimi (Multimodal Interaction for Musical Improvisation), a software system for human-machine improvisation. The ambiguous structure of the human-machine improvisations as well as the distinct perspectives of the listeners ensured a rich set of differences for the basis of our study.

We compare the structural analyses and argue that most of the disagreements between them are attributable to the fact that the listeners paid attention to different musical features. Following the chain of causation backwards, we identify three more ultimate sources of disagreement: differences in the commitments made at the outset of a piece regarding what constitutes a fundamental structural unit, differences in the information each listener had about the performances, and differences in the analytical expectations of the listeners.

Received March 2014

I. Differences Among Listeners

[1.1] Listeners perceive grouping structure in music. Since this phenomenon is universal, cognitive scientists seek to understand how these groupings are perceived. Different listeners may disagree about what grouping structure most accurately describes a given piece of music. Such non-uniformity among humans is not surprising; what *is* surprising is that little attention has been devoted to it in the field of music cognition. We suggest that examining the differences between listeners' conclusions about a piece's structure could help explain how these conclusions about structure were reached to begin with, one of the ultimate goals of cognitive science.

[1.2] The problem of modeling grouping structure has usually been approached in a constructive, ground-up manner: a theory seeks to explain how the tiniest sonic units (e.g., notes) are identified by a listener, how these are "chunked" into larger units (e.g., triplets), and how this chunking procedure continues at higher hierarchical levels (to melodic motives, gestures, phrases, and sections). For example, in Tenney and Polansky (1980) and later in Lerdahl and Jackendoff's (1983)

Generative Theory of Tonal Music (GTTM), simple gestalt rules are proposed to describe how listeners perceive and group sounds. The intuitive rules manage to replicate conventional analyses of musical excerpts of moderate size and complexity, and subsequent studies have confirmed the perceptual validity of some of the rules postulated in GTTM. In an experiment by Clarke and Krumhansl (1990), listeners indicated boundaries between chunks while listening to entire pieces and afterward freely explained their choices for each boundary; most of the reasons offered related to the grouping preference rules of GTTM. Frankland and Cohen (2004) showed that quantified versions of the rules of GTTM could be used to predict how listeners segmented short melodic extracts, although not all of the preference rules tested were shown to be equally effective.

[1.3] The most persuasive evidence for GTTM, and for other theories of grouping structure founded on gestalt-based preference rules, such as Cambouropoulos's LBDM (2001) or the Grouper function of Temperley's Melisma Music Analyzer (2001), comes from studies of listening-based segmentation of monophonic melodies of modest length. At a short enough timescale (e.g., the size of a phrase or shorter), where listeners' responses are most consistent, such theories may offer the best explanation of chunking. But chunking at larger scales seems to invoke a complex combination of preference rules based on parallelism, tonal stability/instability, caesuras, and countless other sonic features. Perhaps the reason existing theories have not successfully modeled the interaction of these factors at longer timescales is that listeners genuinely disagree about the structure at these timescales, or perhaps it is that listeners have difficulty processing information at longer timescales after only one or a few hearings.

[1.4] Listener disagreements are noted in all studies on listening-based segmentation. For example, for every song studied by Frankland and Cohen (2004), while the average agreement between listeners' boundary segmentations was high, and in some cases perfect, there were always pairs of negatively correlated analyses. Bruderer, McKinney, and Kohlrausch (2009) had listeners segment full pieces, and found that of all the boundaries indicated by participants, only a few were agreed upon by all listeners. What's more, listeners may even disagree with themselves: in Frankland and Cohen (2004) and Bruderer, McKinney, and Kohlrausch (2009), within-subject agreement was sometimes low, and Margulis (2012) found that after hearing a piece multiple times, listeners indicate different boundaries, in a way that suggests their attention has been drawn to repetitions of greater length. Thus, while differences among listeners can be noted in any study with multiple participants, we see here the benefit of focusing on such differences. Another example of this benefit: Bruderer, McKinney, and Kohlrausch's (2009) found that the number of listeners who detected a boundary was correlated with the average rated salience of that boundary, and this reveals how listener differences are related to perceptual attributes.

[1.5] While such disagreements seem natural and commonplace, constructive music theories such as GTTM and its relatives do not necessarily account for them. Consider the simplifying assumption made in Lerdahl and Jackendoff's GTTM: the first sentence of the book limits its scope to "a listener who is experienced in a musical idiom" (1983, 1). In a world of identically experienced listeners, differences in interpretation would not exist. The authors recognize this but point out that while "the 'experienced listener' is an idealization, . . . there is normally considerable agreement on what are the most natural ways to hear a piece" (3). This assumption aids in their project of deducing a set of gestalt rules for generating analyses of tonal music that resemble the perceptions of humans. Several of the research studies cited so far certainly do corroborate the observation that considerable agreement among listeners exists as far as musical structure is concerned. The evidence may even extend to the neuroscientific literature: Abrams et al. (2013) found brain activity to be synchronized across listeners of the same piece of music. But confidence that agreement among listeners is "considerable" does not explain the origin of the disagreements, which is what we investigate in this article.

[1.6] Outside of the generalist agenda of music perception and music psychology, of course, it is common to study listener disagreements and to share and debate the merits of particular hearings of pieces. In his model of musical perception, Lewin (2006) attributed differences in listeners' interpretations to a deficiency in analytic discourse that fails to account for the fact that listeners are in fact analyzing different phenomena. But Lewin's examples of these "false dichotomies" were hypothetical and did not come directly from listener data. Smith (2006) discussed passages composed by Brahms that suggest two choices for downbeat, or two choices for the tonic, and argued that these are an essential part of the listener's experience. Such views are not uncontested; Smith's article is a response to Agawu (1994), who argued that, given a theoretical context, ambiguities should never be irresolvable and analyses ought not diverge. But the ambiguities discussed by Agawu and Smith are, for the most part, dichotomous ones in which the listener is aware of two alternative but plausible interpretations and must reconcile them. In the present work, as will be discussed in Section II, we are instead concerned with disagreements that arise without the listeners necessarily being aware of the alternative hearing.

[1.7] Like Agawu, we are also not interested in musical situations that are "vague," in the sense that the conflicting interpretations are "not sufficiently well-formed to be specifiable," to quote Agawu (1994, 90). Nor are we interested in those passages that are precisely ambiguous, the musical analogues of the visual duck-rabbit illusion (Karpinski 2012). Rather (and *unlike* Agawu), we are interested in how music is heard when a clear music-theoretical context is not imposed. Clarifying how a theoretical context influences a hearing of a piece is central to Hanninen's (2012) recently published theory of analysis. Hanninen describes three fundamental orientations that a listener may take: a focus on discontinuities, a focus on associations between passages, or a focus on how a particular theory of music applies to the piece at hand. In these terms, the present work concentrates on disagreements that arise from discontinuity- and association-oriented hearing, with the

music-theoretical context suppressed as much as possible.

Accounting for Listener Differences

[1.8] We are not the first to observe that listeners can disagree, and other research has suggested refinements to GTTM that could account for such differences. For example, Deliège (1987) suggested that listeners might apply the same gestalt rules as one another, but with slightly varying weights depending on their musical experience, and obtained some evidence for this hypothesis: musicians in her study produced segmentations more often concurrent with GTTM's Grouping Preference Rules than non-musicians. On the other hand, Frankland and Cohen (2004) found that listeners with varying musical backgrounds parsed melodies quite similarly to one another.

[1.9] A second refinement could be to model how listening-based segmentation is affected by a listener's familiarity with a piece. As noted above, Margulis (2012) found that an individual's attention was drawn to longer repetitions after hearing a piece multiple times, and Frankland and Cohen (2004) found that a listener's second and third hearings of a piece agreed more closely with one another than their first hearing did to their second or third, suggesting that listeners were refining their interpretations as the piece became more familiar. Palmer and Krumhansl (1987) found that the more familiar a listener was with a piece, the closer the agreement was between how they segmented two different versions of the piece, each retaining only the rhythm or the melody. The evidence collectively suggests that becoming familiar with a piece involves crystallizing one's interpretation of it.

[1.10] The problem of explaining listener differences is hypothetically sidestepped when a rule-based theory of the cognition of grouping structure is replaced with a probability-based one. For example, Temperley (2006) and Pearce, Müllensiefen, and Wiggins (2008) have both proposed segmentation models whose parameters may be set according to the statistical properties of a corpus of music. Differences among listeners in this framework could be attributed to their having different listening histories. In another (also probabilistic) view, listener differences could arise due to perception being a stochastic process: for example, individuals may perceive boundaries with a probability proportional to the boundary's intrinsic salience. This view aligns with research by Bruderer, McKinney, and Kohlrausch (2009), who found that of all the boundaries indicated by participants in a segmentation task, the few that were agreed upon by all listeners happened to be those with the highest rated salience. All of these probabilistic interpretations of human perception allow one to explain differences among listeners as the variance in the input and output of a perceptual mechanism that is common to everyone.

[1.11] While a probabilistic interpretation of listening is appealing, it might not be a satisfying description of the conscious experience a listener has when they interpret the structure of a piece of music. This brings us to a second simplification admitted by Lerdahl and Jackendoff: they are concerned only with the "final state of the [listener's] understanding," and not the "mental processing" that precedes it (1983, 4). While it is true that the structural descriptions provided by listeners are the most concrete evidence that can be examined, in order to understand how these descriptions deviate from one another, surely we must interrogate the listeners about their mental processing.

[1.12] In this article, we report on a study that seeks to do exactly that. We compare structural analyses of three improvisations as heard by two listeners with very different perspectives on the music: one is the improviser, the other an independent listener. Our goal is to observe a number of disagreements and, with the help of the listeners' introspections about their analyses, to locate the origin of these disagreements.

[1.13] A case study similar to ours was reported by Bamberger (2006). She conducted interviews of three listeners with different musical backgrounds, and compared their hearings of a Beethoven minuet in an effort to understand musical development—specifically, how people learn to have more complex hearings of pieces of music. Although the focus of her study differs from ours, she touched on issues relevant to us here. Most importantly, she discussed how the differences among hearings of a piece could be understood as what she terms "ontological differences" (a musical ontology being a determination of what musical ideas count as genuine abstract entities or units). She also suggested that a listener's musical knowledge can influence which musical features and relationships they deem relevant. We will see in Section IV how these factors—listeners' differing musical knowledge, beliefs, and ways of attending—led, in our case, to diverging musical ontologies and differing interpretations of musical structure.

[1.14] Section II describes the justification for our choice of material and the method for collecting the annotations. Referring both to the annotations and the listeners' written accounts of why they analyzed the music as they did, Section III studies the differences between the analyses of each piece. The results of these comparisons are summarized and discussed in Section IV, and our conclusions are presented in Section V.

II. Procedure

[2.1] Our goal in this study is to develop a better understanding of how and why listener disagreements occur. To do so, we compare the different listeners' analyses of pieces of music. In this section, we describe the compromise we struck between the size of our experiment and the level of detail of the responses gathered, and justify our choice of materials, procedures, and participants.

[2.2] Most significantly, we have opted to limit the “participants” of our study to ourselves: the two main authors. While this precludes the possibility of drawing unbiased or statistically powerful conclusions from our observations, our approach facilitates a deeper examination of the differences between our analyses. As will be explained in this section, our choice of methodology is intended to maximize the number and diversity of listener disagreements observed, while allowing as deep an investigation as possible into the causes of these disagreements.

[2.3] Studies of listeners’ analyses usually tout their large size as an advantage: with increased size comes increased statistical power and greater generalizability. Indeed, with many participants (e.g., [Bruderer, McKinney, and Kohlrausch 2009](#)) or many pieces of music (e.g., [Smith, Chuan, and Chew 2014](#)), it is possible to observe broad patterns in how listeners perform chunking, or in how chunking decisions relate to the music that was heard. However, when studying listener disagreements, increased size can be a liability. Firstly, the information we are most interested in—the listeners’ justifications for their responses—is information that is difficult to quantify or categorize, and hence difficult to interpret in large quantities. Secondly, we would also like to have the participants reflect on each other’s analyses and explain why they did not respond in the same manner, and this information can only be collected after the first part of the analysis takes place. By using only ourselves as participants, we simplify this process.

Choice of Materials

[2.4] The choice of music to study was guided in part by the first author’s experience collecting the dataset for the Structural Analysis of Large Amounts of Music Information (SALAMI) project ([Smith et al. 2011](#)). The SALAMI dataset consists of over 2,400 annotations of nearly 1,400 recordings in a wide variety of musical styles, ranging from Renaissance motets to Dixieland jazz to electronica. It was observed that some styles of music, such as song-form popular music, inspired far fewer disagreements than others, such as avant-garde jazz. Through-composed and improvised works in particular seemed to demand more willful interpretation from the listener.

[2.5] Since we wanted the music in our case study to elicit as many and as diverse a set of listener disagreements as possible, we chose to focus on a human-machine improvisation scenario, described below, that presents unique challenges for grouping and segmentation.

[2.6] Mimi (Multi-modal Interaction for Musical Improvisation) is a software system designed for human-machine improvisation. Using a MIDI keyboard, an improviser’s performance is recorded into a buffer (called the “oracle”) and modeled by Mimi (see [François, Chew, and Thurmond 2007](#); [François 2009](#)). Mimi then walks through the oracle, recombining parts of the improviser’s performance into new musical material; in this manner, Mimi and the musician are able to perform concurrently in an overlapping, improvised duet. The performer retains control over some aspects of Mimi’s behavior, including the content of the oracle (which can be added to or deleted altogether), the recombination rate (which controls how likely Mimi is to juxtapose fragments of the oracle), and whether Mimi is generating music or not (naturally, this control must be used in order to end a piece). A visualization accompanying the performance gives the performer information about what Mimi is about to play and has just played, as well as a display of all the musical material currently in Mimi’s memory. In addition to its utility for performance, this visualization also provides useful data for later study.

[2.7] Performances with Mimi provide interesting challenges for the listener seeking to understand musical structure; first of all, there is the improvised nature of the performance, which is, in the words of George Lewis, characterized by a “refreshing absence of the moral imperative concerning structure” (2009). Put simply, improvisation is not necessarily bound to formal structures traditional in popular, classical, or other music.

[2.8] A second and perhaps more intriguing challenge is interpreting the actions of Mimi: Mimi has no knowledge of how, nor the ability, to intentionally create an ending of a phrase, a section, or the entire piece. Any perceived structure could be said to be partly derived from the creativity of the human improviser, whose performance provides the basis for Mimi’s material and whose decisions in response to Mimi may reinforce previous patterns or introduce new material. It may also be partly and serendipitously due to the probabilistic connections Mimi makes between similar note material of disjoint sections. But in the absence of these chance connections or the improviser’s interventions (as when the improviser clears the oracle or tells Mimi to stop generating music), the material Mimi generates tends to be structurally amorphous, especially at larger scales.

[2.9] The third and final challenge is that of integrating the improviser’s musical ideas and Mimi’s concurrent, perhaps not compatible, layers of musical material. For instance, at any given moment, the listener must decide who is in the foreground, Mimi or the improviser. But, as in an Escher drawing, there may be more than one interpretation of the same lines. The focus of the listener’s attention—whether they are concentrating on the improviser, on Mimi, or on both—may thus have a significant impact on the perception of structure. This task is further complicated by the fact that, depending on the instrument patches chosen for the improviser and Mimi, the two voices may not always be distinguishable.

[2.10] Over the course of three weeks, the second author produced three separate improvisations (hereafter referred to as Performance no. 1, no. 2, and no. 3) with Mimi, all on a Yamaha P90 weighted-action keyboard (see **Figure 1**) in a laboratory setting. The three performances were recorded as MIDI files, from which audio tracks equivalent to the original

performances could be made. These were the recordings consulted during the annotation stage and provided here in Section III. Mimi’s visualizations in each performance were also screen captured; these screen-capture videos are provided in Section III with synchronized audio, but they were not consulted for analysis.

Annotation Procedure

[2.11] The annotation procedure was also inspired by previous work with the SALAMI project. The formal structure of each piece was independently annotated by the second author (the improviser, hereafter referred to as Annotator 1) and the first author (an independent listener, hereafter referred to as Annotator 2). Using similar software tools—Annotator 1 used the Variations Audio Timeliner (<http://variations.sourceforge.net/vat>) and Annotator 2 used Sonic Visualiser (<http://www.sonicvisualiser.org>)—the listeners analyzed each piece at two hierarchical levels. In accordance with common practice in formal musical analysis, the large-scale level was annotated with uppercase letters, and the small-scale level with lowercase letters, to indicate which portions of the piece were judged to contain similar musical material. In keeping with Lerdahl and Jackendoff’s well-formedness rules for structural grouping, overlapping sections were disallowed, all portions of a piece were labeled, and boundaries at a given hierarchical level were respected at smaller-scale levels.

[2.12] Each analysis was produced in a single session, each lasting roughly a half hour, although this time was not prescribed beforehand; indeed, aside from producing annotations in the same format, the annotators had total freedom. For example, unlike many psychology experiments, in which analyses are recorded in single passes (with listeners pressing a button when boundaries are perceived, for instance), here the annotators were free to listen to the pieces as often as they liked, and to return to particular spots or repeat short excerpts.

[2.13] In a departure from the procedure used by SALAMI, both listeners also wrote brief notes explaining their choice of boundaries and groupings in a separate session after annotating each piece. The responses were worded freely, but at a minimum the listeners were expected to justify, with reference to the recording, each boundary and the similarity of sections labeled with the same letter.

[2.14] These justifications did not generate explanations from both participants for every moment where the interpretations diverged. Consider the case in which listener no. 1 perceives a boundary where listener no. 2 does not. We may refer to the first listener’s explicit justification for this perception, but listener no. 2’s remarks may not include an explanation for not experiencing this perception. The process of identifying and explaining differences thus required more than just collating responses. Instead, after enumerating all the significant differences between our analyses, we (the two listeners) discussed each one, reflecting on our listening experiences and elaborating on our interpretations of the pieces. The next three sections recount the outcome of these conversations for the three pieces.

III. Differences Between Annotations

[3.1] In this section, we consider the three performances separately. For each, we list the differences between our annotations and offer reasons to account for these differences. We will collect our observations and attempt to generalize from them in Section IV. **Figure 2** provides annotations with synchronized audio for Performance no. 1; the upper part is from Annotator 1 (the improviser) and the lower part is from Annotator 2 (the independent listener). **Audio Example 1** provides the recording on its own, and **Video Example 1** shows how Mimi traversed the oracle and how the performer controlled Mimi’s behavior.

Performance no. 1

[3.2] In Performance no. 1, Annotator 2 roughly agreed with all of Annotator 1’s small-scale boundaries (the smaller bubbles in Figure 2), but Annotator 2’s version has more small-scale boundaries, and it also differentiates subsections within each main section (e.g., *A1* includes *a*, *b*, *c*, and *d* subsections). This leads to two compelling divergences in the large-scale segmentation. Setting aside the small deviations in timing (e.g., the few seconds difference in the boundary between Annotator 1’s *a2/a3* and *a3/b1* transitions, and the disagreement about when the piano stopped ringing at the end of the piece), the differences that require explanation are:

[3.3] (1) *Why is Annotator 2’s A1 section much more segmented than Annotator 1’s?*

Both annotators identified the same initial sequence as a single musical idea *a1*, but they conceptualized this passage differently because they focused on different musical parameters. To Annotator 1, the idea was defined by its mood—an amorphous, ethereal melody with pedal—and the segments *a2* and *a3* were distinguished by the melody moving to a different voice (Mimi) or to a new register. On the other hand, Annotator 2’s hearing was marked by a strong sense of rhythmic phrasing, established when the four-part opening phrase *a1* is answered by Mimi with a similar phrase *a2*. This pace is only followed roughly for the rest of the *A* section, but because the material is very open, containing relatively short gestures with long pauses in between, it is easy to imprint a loose pace of phrases onto the music.

[3.4] (2) *Why does Annotator 2 bear the transition section B as beginning earlier than does Annotator 1?*

Both annotators agreed that the material beginning at 2:54 (Annotator 1's *b1*, Annotator 2's *f1*) was wholly different from the material in section *A1*. Indeed, Mimi is silent during this section, and it is melodically and rhythmically distinct from all of section *A1*. (See **Notation Example 1**.) However, Annotator 2 perceived a "pre-transition function" in segment *e1*, leading him to place the beginning of the section earlier than Annotator 1. While the material in *e1* is similar to the rest of section *A1*, there are a few cues that arguably distinguish it: a new downward theme from the improviser with a repeated rhythm, and a rising, fading motive that follows, both of which feel like ending material and anticipate the change at 2:54. (See **Notation Example 2**.)

[3.5] (3) *Why do Annotator 1 and Annotator 2 disagree about the differentiation of musical ideas in section C1?*

While Annotator 2 differentiated between subsegments throughout the piece, Annotator 1 did not; he posits that this is because that option did not occur to him at the time. It is hard to say whether the labeling differences of these subsections of *C1* (or the subsections of *A1*) are very meaningful, since Annotator 1 and Annotator 2 also initially employed slightly different naming conventions: Annotator 1 used letters and prime notations (e.g. *A*, *A'*, *A''*), and Annotator 2 used a combination of letters, subscript numbers and prime notations (e.g. *A₁*, *A₂*, *A'₂*). The analyses shown in Figure 1 are adaptations of the original analyses, meant to enable comparison; for the later performances, the annotators used the same format as each other. In the diagrams in this paper, indices are only used to indicate repetitions of musical ideas.

[3.6] (4) *Why do Annotator 1 and Annotator 2 disagree about the labeling of the final section (C vs. C and D)?*

The improviser's performance does change dramatically at Annotator 2's *D1*, while Mimi continues in the same vein. Annotator 2 separated *D1* from *C1* because the figure played by the improviser in *D1* was not only musically distinct but, with its descending triads and relatively thin texture, seemed to have a strong sense of ending function, whereas Annotator 1 attended to the continuity in the melodic material in Mimi's voice.

Performance no. 2

[3.7] **Figure 3** provides annotations with synchronized audio for Performance no. 2; the upper part is from Annotator 1 and the lower part is from Annotator 2. Audio Example 2 provides the recording on its own, and Video Example 2 shows how Mimi traversed the oracle and how the performer controlled Mimi's behavior. The most striking differences between the annotations (Figure 3) are in the grouping and labeling of the first five minutes: Annotator 2's *A1* is subdivided further than Annotator 1's *a1*; the placement of the boundaries near 3'11" and 4'17" are disputed; and the larger section that encompasses Annotator 1's *b1* and Annotator 2's *c1* is disputed. There are also some subtle differences in the labeling of the subsections in the last two minutes.

[3.8] (1) *Why does Annotator 2 subdivide Annotator 1's a1 further?*

Annotator 1 attuned to the textural similarities that joined his *a* sections (their atmospheric quality) and their contrast with the *b* sections (louder and rhythmic, i.e., with a strong and regular pulse). Annotator 2 identified the same contrast between the material designated *a1* and *b1* by Annotator 1, but, as with Performance no. 1, he made further subdivisions and associations based on recurring melodic motifs: treating *a1* as the opening theme, both *a2* and *a3* begin with the first part of the opening theme, and *a2* ends with the ending of the opening theme.

[3.9] (2) *Why does Annotator 1 bear the first five minutes as a series of binary groups (A1, A2, A3), whereas Annotator 2 bears a duo of ternary groups (A1, B1) with an additional section (C1)?*

The annotators agree that Annotator 1's *b1* presents a contrast to all that precedes it, but disagree about the larger structural interpretation of *b1*. To Annotator 1, section *a1* was internally continuous and self-similar, and so the change at 2'25" (at the beginning of his *b1*) struck him as the midpoint of a larger grouping. This hearing was reinforced by the subsequent alternation of atmospheric *a1* material and rhythmic *b1* material as repetitions of this binary structure. In contrast, Annotator 2 had already heard a ternary-like structure in the material preceding 2'25" (*aaba*) and so was inclined to hear the material in section *B*, with the entrance of a new quarter-note triplet motive, as beginning a new section, also ternary (*abc*). In hearing things this way, he overlooked the similarity of *d1* to the opening material, instead focusing on the broad textural self-similarity of his *B1*. (See **Notation Example 3**.)

[3.10] (3) *Why does Annotator 2 not identify either of sections d1 or e1 as being a repetition of previous material?*

Annotators 1 and 2 characterized section *a1* differently: Annotator 1 heard a long self-similar span with a particular texture, and hence easily associated the return of this material in his *a2*. To Annotator 2, *a1* was a melody, which recurred in varied form in *a2* and *a3*. With this in mind, he heard the return of the theme in *d1* as a severe truncation of the theme, a kind of quotation in an otherwise distinct passage.

[3.11] (4) *Why do the annotators disagree about the placement of the boundaries near 3'11" and 4'17"?*

In *a2*, Annotator 1 heard a return to the opening material, and hence his section begins at the onset of the restatement of the

theme (see **Notation Example 4**); in *d1*, Annotator 2 heard a brief reprieve between statements of the *c1* material, and hence identified the moment where we deviate from the material of *c1* as the boundary. Both annotators recognized the introduction of new material by the improviser at 4'17", and Annotator 1 placed his boundary (the beginning of *c1*) in line with this. Annotator 2 placed the boundary (*e1*) earlier, at the onset of a stark registral shift at 4'10".

[3.12] (5) *Why do Annotator 1's d4 and Annotator 2's g1 overlap (6'04" to 6'13")?*

Both listeners perceived that this final section (Annotator 1's *B1*) begins with the improviser and Mimi engaging in an approximate canon with a period of about 15 seconds between voices. This pattern breaks down shortly after the 6'00" mark. Here, Annotator 1 heard a prolongation of the last phrase (*d4*), followed by a new section in which the improviser introduces a new musical idea in the lower register while Mimi continues with the canon material. Annotator 2 did not focus on the new theme, and instead heard at *g1* an accelerated continuation of the canon between the voices. This canon has a much shorter period of a few seconds, the improviser and Mimi now trading gestures rather than phrases.

[3.13] (6) *Why is the span from 7'05" to 7'17" (Annotator 1's f1) grouped with the subsequent material (f2) by Annotator 1, and with the preceding material by Annotator 2 (h1/h2)? And why is Annotator 1's f2 given its own large-scale section by Annotator 2?*

From 7'05", the improviser introduces two contrasting ideas: a loud, downward-leading progression (Annotator 1's *ef*), and an ethereal theme played sparsely in a high register (Annotator 1's *ff*). These are repeated by Mimi in canon in Annotator 1's *ff* and *j2*; in the latter of these, the improviser also provides sparse accompaniment. Since he marked *e1* and *ff* as distinct, it can be seen that Annotator 1 focused on the difference between the themes introduced by the improviser. On the other hand, Annotator 2 focused on the repetition of the louder, more prominent musical idea in sections *b1* and *b2*. This meant that he heard a greater degree of discontinuity between *b2* and *i1* than did Annotator 1. This abrupt change to a sparse texture, suggestive of a concluding function, also led Annotator 2 to indicate a higher-level boundary between large-scale sections.

Performance no. 3

[3.14] **Figure 4, Audio Example 3, and Video**

Example 3 provide the analyses, recording, and video for Performance no. 3. In contrast to Performances no. 1 and no. 2, Annotator 1's and Annotator 2's analyses of Performance no. 3 (which were created before the listeners had conferred on Performance no. 2) are largely in agreement, especially with regard to the larger sections (i.e., the uppercase letters). Most of the differences can be understood in terms of attending strategies: Annotator 1 paid the most attention to motivic recurrence, while Annotator 2 paid the most attention to surface qualities (e.g., register and texture). However, thematic segmentation also played a role in differentiating the interpretations: Annotator 1 segmented the opening theme into individual motives, while Annotator 2 did not segment the theme. This had implications for the final section of the performance, when this thematic material returns.

[3.15] (1) *Why does Annotator 1 subdivide Annotator 2's opening section a1 into two subsections? (This also applies to the subdivision of the last section, Annotator 2's f2.) Why does Annotator 1 further subdivide Annotator 2's A1?*

Annotator 1 heard the opening 10 seconds (from 0'05" to 0'16") as *ab* and Annotator 2 heard it as simply *a*. The difference may hinge on a matter of metrical interpretation, and since there is no "ground truth" set of intended note lengths, the preferred interpretation is a creative choice. In **Notation Example 5**, the notes at the boundary between *a1* and *b1* are notated as triplets, suggesting rhythmic continuity. However, if the notes are instead heard as quarter notes, as shown in **Notation Example 6**, Annotator 1's boundary now falls between gestures (instead of in the middle of a triplet), emphasizing the shift in register at the proposed boundary. (In both of these examples, barlines are chosen to emphasize certain patterns and divisions; no particular meter is implied.) This is a vivid example of how a structural analysis can depend on how the listener has made sense of the fundamental units of the piece, an issue discussed by Bamberger (2006) to which we will return in Section IV.

[3.16] This initial discrepancy meant that Annotator 1 was more focused on segmenting the rest of the opening *A* section according to the recurrence of these separate *a* and *b* ideas. For instance, after the end of the initial idea, Annotator 1 placed his next boundary at *b2*, where Mimi repeats the material of his *b1* section and the improviser introduces a new gesture (see **Notation Example 7**). Annotator 2 placed his next boundary earlier, at his *b1*, citing a significant change in register and accompaniment; his *b1* section is united by Mimi's use of *a1* material and the improviser's presentation of contrasting, non-thematic material. Annotator 2's conflation of the two parts of the first idea also led to his fusing the two last segments in Annotator 1's analysis (*b4* and *a4*) into one (*j2*).

[3.17] (2) *Why do the annotators disagree about the grouping of the material from 0'27" to 0'37"?*

Annotator 2 perceived a strong change at 0'27" (*b1*) as the improviser introduced new accompanying material. In contrast, Annotator 1 heard a continuation of the *a1* material in Mimi's voice, and Mimi eventually returns to the *b1* material at 0'37".

[3.18] (3) *Why does Annotator 2 mark boundaries at 0'27" (b1) and 1'13" (d1), when Annotator 1 does not?*

As stated before, Annotator 2 heard a discontinuity at 0'27" in the improviser's material. But Annotator 2 partly attributes both boundaries to the pauses that precede them. In both cases, the thematic continuity of the section led Annotator 1 to forego an additional boundary.

[3.19] (4) *Why does Annotator 2 recognize a return of material from the opening section at his b2 when Annotator 1 does not?*

To Annotator 2, *b1* was characterized by Mimi's playing fragments of the original motive, with the improviser adding novel accompaniment. Thus *b2* represented a return to this configuration. In contrast, Annotator 1 felt that this section continued the chaotic, fragmented feel of section *B1*.

[3.20] (5) *Why does Annotator 2 further subdivide Annotator 1's B1?*

Given that Annotator 1 did not identify *c3* as a return to material from the previous section *A1*, his choice of the large-scale grouping (*AB*) is no surprise. Annotator 2 did identify a return to the previous section at *b2*, and the significance of this return led him to hear a larger-scale ternary grouping, *ABA*.

[3.21] (6) *Why does Annotator 2 not recognize a return of material from A at his f1?*

At Annotator 1's *b3*, Mimi repeats the pattern played at *b1* by the improviser, who then responds with a melodic inversion of the material. The counterpoint is repeated at Annotator 1's *b4*, with the parts swapped: the improviser plays the original ascending *b1* motif, and Mimi repeats the inverted theme from *b3*. As the improviser, Annotator 1 recalls these imitations being deliberate, and hence was aware of their relationship to the earlier material at the time of performance. However, Annotator 2 was not aware of the repetition until it was pointed out to him! This oversight can possibly be explained by Annotator 1's *b* idea having less primacy in Annotator 2's analysis. Since Annotator 2 did not hear it as a "head" of any section, he was less apt to hear just the "tail" of the opening theme return, either at 2'33" or 3'37" (Annotator 1's *b3* and *b4*)—even though he heard these as repetitions of each other.

IV. Discussion

[4.1] The questions we ask in this article are: in what ways may two listeners disagree about the structure of a piece of music, and what factors cause or explain these differences? In the previous section, we presented the analyses produced by two listeners of three improvised pieces, and enumerated the differences between them. We also sought to explain how each difference arose by referring to the listeners' introspective notes on why they made the decisions they did. We are now interested in following the chain of causation backward, first considering the proximate causes of the disagreements—the circumstances that explain the disagreements most immediately—and extrapolating from these possible ultimate causes. In this section, we discuss these causes in a loose progression from most to least proximate. As the causes get deeper, they become more speculative but also, we suggest, more important and illuminating.

Factor 1: Analysis Method

[4.2] First and least importantly, we acknowledge that some of the differences between the annotations arose from non-identical analysis methods. This is most important for Performance no. 1, in which Annotator 2 distinguished the subsections by letter, but Annotator 1 did not, saying it did not occur to him as an option. This was noticed immediately after the first analysis, and the issue was corrected before the next pieces were annotated. Since the question of whether different methodologies can lead to different analyses was not part of our study, we do not consider the issue further.

Factor 2: Musical Features

[4.3] The simplest and most expected explanation for why the two listeners disagreed is that they paid attention to different musical features. For example, in Performance no. 1, the annotators segmented *A1* differently because Annotator 1 paid attention to shifts in register, while Annotator 2 paid attention to the pauses and melodic gestures that supported a regular phrase rhythm. They also gave these subsections different labels because the former focused on the textural similarity between them, and the latter on the slightly different motives in each.

[4.4] Both annotators reported attending to a similar set of musical parameters at various times: melodic themes and their repetition; rhythm, texture, and register; and whether Mimi or the improviser were playing a particular part (recall that these two voices had different timbres). Still, sometimes annotators attributed their decisions to parameters not mentioned by the other; for example, Annotator 2 invoked the perceived function of a section to justify some of his decisions, but Annotator 1 never indicated that this was an important attribute. (This occurs with the concluding sections that Annotator 2 heard at the end of Performance no. 1 and no. 2, and in the preparatory *e1* section that he heard in Performance no. 1.) On the other hand, in Performance no. 3, Annotator 1 identified a melodic inversion at *b3*, which Annotator 2 did not attend to.

[4.5] The annotators did not seem to consistently prefer one musical attribute over another: in the disagreement over the labeling of the final three subsections of Performance no. 2 (*eff* vs. *bbi*), it was Annotator 2 who found the overall texture salient, whereas Annotator 1 paid attention to the different themes being played by the performer. But in their analysis of

section A1 in Performance no. 1, the annotators focused on the opposite features.

[4.6] The instances where the function of a part was cited as a reason to segment or differentiate it recalls the observation of Peeters and Deruty (2009) that music structure is multidimensional, consisting of attributes that can be independent, such as musical function, similarity, and instrumentation. In this view, some disagreements could be attributed to listeners focusing on different dimensions of structure, although it remains to be explained why some people focus on different dimensions to begin with. Peeters and Deruty thus proposed an annotation format that would separate these dimensions, a scheme that was adopted for SALAMI. The notions that musical similarity could be similarly decomposed, and that attention to different musical features could explain disagreements between listeners, are explored in Smith and Chew (2013). In that work, repetitive patterns in different audio features and the analyses of listeners were compared to deduce what listeners may have focused on in their analyses, and the results were found in some cases to provide good explanations for the disagreements.

Factor 3: Opening Moments

[4.7] While most of the differences seem well explained by referring to the listeners' attention to different musical features, it is more concise to attribute later differences between two annotations to earlier differences. That is, how the listener happens to perceive the opening moments of a piece—what they initially perceive as the basic units in their chunking, or what they initially call *A* and *B* appears to greatly determine how the rest of the analysis will proceed.

[4.8] For example, in Performance no. 2, Annotator 2 heard the opening section *A1* as having a basically ternary structure; this may have encouraged him to perceive the following material (*B1*) as a ternary grouping as well. Similarly, Annotator 1 heard a binary contrast within the opening section (*A1*), which would reinforce the binary interpretation of the next two sections (*A2*, *A3*).

[4.9] It makes sense that the opening moments would lay the framework for the rest of the piece, since they would strongly affect one's expectations. In Performance no. 1, Annotator 2 identified a regular four-phrase structure in the first section *a1*; this seemed to lead him to expect a similar phrase rhythm in subsequent material, resulting in more regular section lengths. The opening moments establish for the listener what design principles the composer or improviser is using; what contrasts are relevant and what units can be repeated.

[4.10] The opening moments were clearly crucial in Performance no. 3. Here, the opening 10 seconds crystallized in the mind of Annotator 1 as two distinct themes (*a1* and *b1*), but as a single theme to Annotator 2 (*a1*). The fact that the material Annotator 1 calls *b1* did not strike Annotator 2 as a distinct theme likely explains why Annotator 2 did not recognize the return of this *b* material later on as *b2*, *b3*, and *b4*—even though he did recognize that *b3* and *b4* were similar to each other.

Factor 4: Difference in Information

[4.11] Assuming that the perception of the opening moments is crucial in forming an analysis, how is it that listeners differ in how they perceive these opening moments? A deeper explanation ought to include differences between the listeners that are present before the analysis is begun. Generally speaking, differences in information are anticipated as an important factor in psychological studies; for example, participants are classified as musicians and non-musicians (i.e., people with and without specialized musical knowledge). Here, we consider a more specific difference: a difference in the type and thoroughness of the knowledge each listener has about the piece.

[4.12] Annotator 1, as the improviser in the performances, had a more intimate understanding of how the piece was constructed than Annotator 2 before each later listened to and analyzed the performances. This difference had an impact on the slightly different procedure used by the listeners: Annotator 1 tended to analyze pieces section by section, nearly finalizing his analysis of the first half before listening to the second half, for example. The ability of Annotator 1 to work through the large sections in series suggests that the large-scale analysis (or at least the large-scale segmentation) may have already been decided at the beginning of the annotation process. In contrast, Annotator 2 tended to work in parallel: he annotated boundaries in real time while listening through the whole piece several times, and in between auditions he re-listened to specific parts to adjust his annotations. This discrepancy between the listeners suggests an important difference in the initial information each had about the performance. To Annotator 1, the lay of the land was already well known; Annotator 2 had to do more scattered scouting before he could finalize his understanding of the large-scale patterns. While this observation may seem particular to the scenario at hand, comparable situations arise frequently among listeners, with some analyzing a piece only after becoming familiar with it as a performer or in casual listening, and others analyzing as new listeners.

[4.13] Elizabeth Margulis (2012) has found that listeners who are less familiar with a piece of music are more likely to focus on shorter repetitions, while those who are more familiar are likely to focus on longer repetitions. Extrapolating from repetition (which never occurs exactly in the three performances studied here) to similarity, we see the same pattern reflected in the differences between our annotations: in Performance no. 1, Annotator 2, the newer listener, subdivides *A1* more than Annotator 1 on the basis of a perceived phrase rhythm and on local changes in texture, whereas Annotator 1 focuses on the self-similarity of the entire passage. Similarly, in Performance no. 2, Annotator 1 points out what unites sections *A1*, *A2*, and

A3 at a large timescale, whereas Annotator 2 does not recognize these similarities. Finally, in Performance no. 3, although Annotator 1's conception of the opening moments at first appears more fine-grained than Annotator 2's, it leads to an analysis that recognizes more repetitions and returns globally, requiring only four distinct section types (*a* to *d*) compared to Annotator 2's six types (*a* to *f*).

[4.14] The different levels of familiarity with the pieces also seemed to influence the musical features to which the listeners paid attention. Annotator 2 (whose annotations are generally more segmented than Annotator 1's) attributed more of his boundaries to surface features, such as long silences, sudden note clusters, and changes in register, than did Annotator 1. For example, Annotator 2's large-scale boundary in Performance no. 3 between *B1* and *A2* is attributed to a long pause. In the same performance, Annotator 2 starts his section *C1* where Mimi plays some disruptive clusters, whereas Annotator 1 begins *C1* a few moments later, when the improviser takes up the new theme.

[4.15] One final difference in information is quite specific to the present circumstances but nonetheless bears mentioning: the fact that Annotator 1 was the improvising performer and hence had memories of creating the music. Annotator 1, being thus more aware of details such as what part of the oracle Mimi had access to and when Mimi was active and inactive, may have been less willing to give an analysis that did not reflect these events. For example, in Performance no. 1, his section *B1* exactly aligns with when Mimi was turned off; Annotator 2, however, heard parts of the previous section as being a part of this transition section. In Performance no. 2, the oracle is cleared and reset only at the boundary between Annotator 1's *A3* and *B1*; perhaps Annotator 1, knowing this, was less inclined to differentiate the large-scale subsections of each half with different letters, as Annotator 2 did. Memories of the performance may also have helped ensure that intentional but subtle repetitions, such as the return and inversion of an earlier motif in Performance no. 3 (at *b3*), were reflected in Annotator 1's analysis. While the difference in information between the listeners in our case was extreme by design, listeners certainly differ along similar lines: access to the score may radically affect how a listener perceives the structure of a piece, and listeners may differ in their insight into the relevant instrumental practice (e.g., pianists and non-pianists analyzing a piano sonata) or prior knowledge of the specific piece being performed.

Factor 5: Difference in Analytical Expectations

[4.16] Beyond the information the listeners had about this specific piece, we consider the role of information about music in general, involved here as analytical expectations. Some of our results suggest that the listeners may have had different a priori expectations about what the analyses would look like. Since the two listeners have different backgrounds and experience in music theory, analysis, and musical taste, it is difficult to speculate as to where these expectations would arise. However, the two sets of annotations differ strikingly in one property: the small-scale segments perceived by Annotator 2 tend to have more equal size than those of Annotator 1. For example, in Performance no. 2, Annotator 2's *A1* has 4 subsegments, each roughly one quarter the duration of *A1*. Annotator 1's *A1*, on the other hand, is subdivided highly asymmetrically. The trend appears to be somewhat consistent across the three performances, although a larger study would be needed to confirm this difference. If it were found to be a consistent trend, it may reflect a strong expectation on the part of Annotator 2 that subsegments will be of equal size. This is not an unreasonable expectation, given that composed music often includes repeating or contrasting sections with similar lengths, and may be shared by many listeners. It would be interesting to determine whether this expectation affects how music is analyzed.

V. Conclusion

[5.1] We examined two listeners' analyses of three improvised performances and found the differences between these analyses to reveal several insights.

[5.2] *Attention.* First, we observed that these differences were often due to the fact that the listeners paid attention to different musical features. Attention itself is already widely studied, but usually only as a global concept: researchers are interested in *how much* the listener is paying attention to the music, not *what* the listener attending to in the music. For example, Abdallah and Plumbley (2009) showed that a model of attention and surprise can bear a striking resemblance, in practice, to a theory of musical structure, and attention is a key concept in existing theories of music such as Farbood's model of musical tension (2012). Jones and Boltz (1989) have shown that paying attention to short and long timescales can affect one's perception of time, but it remains to be studied how this can in turn affect one's interpretation of musical structure. Since listeners are able to focus their attention on (or have their attention unwittingly drawn to) particular aspects of a piece of music—patterns in a vocal line, recurrences of a motif, shorter or longer timescales—we recommend following up this research in a way that treats music, the object of attention, as multi-dimensional. The way the attention of the listener wanders between these aspects could become the subject of a new theory of analysis.

[5.3] *Opening moments.* We next noted that differences in how two listeners heard the opening moments prefigured most of the remaining differences. It appears to be in these opening moments that listeners decide what will comprise their basic units of analysis and what types of abstraction—melodic, textural, rhythmic—will serve them best. The mental representation formed here serves as a template, allowing the listener to form expectations about how the material will develop in the rest of the piece. One conclusion from this—that knowing how a listener understands the beginning of a piece allows one to predict how the rest will be understood—is a readily testable hypothesis that would form the basis of

exciting future work.

[5.4] *Information.* On a deeper level, we speculated that access to information could affect the area of attention or focus. In our case, Annotator 1 had more information than Annotator 2 in several ways: he had created and performed the piece, his memory helped him to better disentangle his own and Mimi's contributions, and he had simply heard the piece much earlier than Annotator 2. Several differences in the annotations seemed well explained by these differences in information. While these differences in information may appear to be circumstantial, comparable differences arise between listeners who have access to a score and those who do not, or among listeners who have listened to a piece different numbers of times and whose familiarity with it varies—a factor whose importance has already been demonstrated in previous studies (e.g., [Bamberger 2006](#), [Margulis 2012](#)).

[5.5] *Analytical expectations.* Finally, we found that listeners may bring a priori global expectations to the analysis. In the analyses studied here, this was suggested by the conspicuously regular phrase length indications of Annotator 2, which contrasted with the more asymmetric groupings in Annotator 1's analyses. These global assumptions are formed over the entirety of a person's listening history; they are based on familiarity with the style of music at hand, instrumental experience, and exposure, if any, to music theory, or to the piece in question. These analytical expectations may also influence how a listener initially understands the opening moments of a piece.

[5.6] These four insights resonate with Bamberger's (2006) argument that perceptual disagreements among listeners can be ascribed to differences in ontology, which are in turn affected by a listener's values and belief system (which are shaped by the information they have, and reflected in their expectations) and preferences (which influence the features and relations to which they pay attention). Of course, while this system of beliefs appears to be the source of listener disagreements, a listener's analysis of a piece is still predicated on external cues present in the music: for example, prosodic cues (stresses, pauses, and shaping of tempo and loudness as communicated by the performer) or repetitions that guide a listener's attention or expectations.

Future work

[5.7] As noted in Section II, our present study considered few participants and few pieces to facilitate a very close investigation of our structural analyses. The outcome was a rich set of questions that we hope to address in future work using more participants and simpler pieces. Some of these questions would be straightforward extensions of previous work: for example, to better understand how people's backgrounds affect their perception of structure, experiments comparable to Margulis's (2012) could be undertaken, cataloging listeners' musical background in greater detail and measuring their perception of more complex grouping and associative structures.

[5.8] Other questions for future research are less straightforward: for example, what musical features do listeners pay attention to, and does this directly impact their perception of a piece's structure? It seems that answering this question directly would require an auditory attention-tracking system, some analogue to the eye-tracking systems used to study visual attention. Since none exists, a carefully constructed set of artificial stimuli will be necessary to study this question.

[5.9] We would also like to know how quickly listeners decide on a set of basic musical ideas when they begin to listen to a piece of music, and how definitively this guides their interpretation of the piece. Supposing a listener devises a running hypothesis of the piece's structure while listening to it, how easily or how frequently is this hypothesis revised? What kinds of musical events are capable of causing this? If listeners are permanently beholden to any aspect of their first impressions, this has wide-reaching implications for those who make music.

Jordan B. L. Smith
Queen Mary University of London
Centre for Digital Music
Mile End Road
London, E1 4NS, United Kingdom
j.smith@qmul.ac.uk

Isaac Schankler
Process Pool Music
3527 Isabel Drive
Los Angeles, CA 90065
eyesack@gmail.com

Elaine Chew
Queen Mary University of London
Centre for Digital Music
Mile End Road

Works Cited

- Abdallah, Samer, and Mark Plumbley. 2009. "Information Dynamics: Patterns of Expectation and Surprise in the Perception of Music." *Connection Science* 21, nos. 2–3: 89–117.
- Agawu, Kofi. 1994. "Ambiguity In Tonal Music: A Preliminary Study." In *Theory, Analysis and Meaning in Music*, ed. Anthony Pople, 90–107. Cambridge: Cambridge University Press.
- Abrams, Daniel A., Srikanth Ryali, Tianwen Chen, Parag Chordia, Amirah Khouzam, Daniel J. Levitin, and Vinod Menon. 2013. "Inter-subject Synchronization of Brain Responses During Natural Music Listening." *European Journal of Neuroscience* 37: 1458–69.
- Bamberger, Jeanne. 2006. "What Develops in Musical Development? A View of Development as Learning." In *The Child as Musician: Musical Development from Conception to Adolescence*, ed. Gary McPherson, 69–92. Oxford: Oxford University Press.
- Bruderer, Michael, Martin McKinney, and Armin Kohlrausch. 2009. "The Perception of Structural Boundaries in Melody Lines of Western Popular Music." *Musicae Scientiae* 13, no. 2: 273–313.
- Cambouropoulos, Emilios. 2001. "The Local Boundary Detection Model (LBDM) and its Application in the Study of Expressive Timing." In *Proceedings of the International Computer Music Conference*.
- Clarke, Eric F., and Carol L. Krumhansl. 1990. "Perceiving Musical Time." *Music Perception* 7, no. 3: 213–51.
- Deliège, Irène. 1987. "Grouping Conditions in Listening to Music: An Approach to Lerdahl & Jackendoff's Grouping Preference Rules." *Music Perception* 4, no. 4: 325–59.
- Farbood, Morwaread M. 2012. "A Parametric, Temporal Model of Musical Tension." *Music Perception* 29, no. 4: 387–428.
- François, Alexandre R. J. 2009. "Time and Perception in Music and Computation." In *New Computational Paradigms for Computer Music*, ed. Gérard Assayag and Andrew Gerzso, 125–46. Editions Delatour France / IRCAM.
- François, Alexandre R. J., Elaine Chew, and Dennis Thurmond. 2007. "Visual Feedback in Performer-Machine Interaction for Musical Improvisation." In *Proceedings of the International Conference on New Interfaces for Musical Expression*, 277–80.
- Frankland, Bradley, and Annabel Cohen. 2004. "Parsing of Melody: Quantification and Testing of the Local Grouping Rules of Lerdahl and Jackendoff's A Generative Theory of Tonal Music." *Music Perception* 21, no. 4: 499–543.
- Hanninen, Dora A. 2012. *A Theory of Music Analysis: On Segmentation and Associative Organization*. Rochester, NY: University of Rochester Press.
- Jones, Mari Riess, and Marilyn Boltz. 1989. "Dynamic Attending and Responses to Time." *Psychological Review* 96, no. 3: 459–91.
- Karpinski, Gary S. 2012. "Ambiguity: Another Listen." *Music Theory Online* 18, no. 3.
- Lerdahl, Fred, and Ray Jackendoff. 1983. *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.
- Lewin, David. 2006. "Music Theory, Phenomenology, and Modes of Perception." In *Studies in Music with Text*, 53–108. Oxford: Oxford University Press.
- Lewis, George. 2009. "The Condition for Improvisation." Keynote address at the International Society for Improvised Music. Santa Cruz, CA.
- Margulis, Elizabeth. 2012. "Musical Repetition Detection Across Multiple Exposures." *Music Perception* 29, no. 4: 377–85.
- Palmer, Caroline and Carol L. Krumhansl. 1987. "Pitch and Temporal Contributions to Musical Phrase Perception: Effects of Harmony, Performance Timing, and Familiarity." *Perception & Psychophysics* 41, no. 6: 505–18.
- Pearce, Marcus T., Daniel Müllensiefen, and Geraint A. Wiggins. 2008. "A Comparison of Statistical and Rule-Based Models of Melodic Segmentation." In *Proceedings of the Ninth International Conference on Music Information Retrieval*, 88–94.
- Peeters, Geoffroy, and Emmanuel Deruty. 2009. "Is Music Structure Annotation Multidimensional? A Proposal for Robust Local Music Annotation." In *Proceedings of 3rd Workshop on Learning Semantics of Audio Signals*, 75–90.

- Schankler, Isaac, Elaine Chew, and Alexandre R. J. François. 2014. "Improvising with Digital Auto-scaffolding: how Mimi Changes and Enhances the Creative Process." In *Digital Da Vinci: Computers in Music*, ed. Newton Lee, 99–125. Springer.
- Schankler, Isaac, Jordan B. L. Smith, Alexandre R. J. François and Elaine Chew. 2011. "Emergent Formal Structures of Factor Oracle-Driven Musical Improvisations." In *Mathematics and Computation in Music*, eds. Carlos Agon, Moreno Andreatta, Gérard Assayag, Emmanuel Amiot, Jean Bresson and John Mandereau. Lecture Notes in Artificial Intelligence 6726, 241–54.
- Smith, Jordan B. L., J. Ashley Burgoyne, Ichiro Fujinaga, David De Roure and J. Stephen Downie. 2011. "Design and Creation of a Large-Scale Database of Structural Annotations." In *Proceedings of the International Society for Music Information Retrieval Conference*, 555–60.
- Smith, Jordan B. L., C.-H. Chuan, and Elaine Chew. 2014. "Audio Properties of Perceived Boundaries in Music." *IEEE Transactions on Multimedia* 16, no. 5: 1219–28.
- Smith, Jordan B. L., and Elaine Chew. 2013. "Using Quadratic Programming to Estimate Feature Relevance in Structural Analyses of Music." In *Proceedings of the ACM International Conference on Multimedia*, 113–22.
- Smith, Peter H. 2006. "You Reap What You Sow: Some Instances of Rhythmic and Harmonic Ambiguities in Brahms." *Music Theory Spectrum* 28, no. 1: 57–97.
- Temperley, David. 2001. *The Cognition of Basic Musical Structures*. Cambridge, MA: MIT Press.
- . 2006. *Music and Probability*. Cambridge, MA: MIT Press.
- Tenney, James, and Larry Polansky. 1980. "Temporal Gestalt Perception in Music." *Journal of Music Theory* 24, no. 2: 205–41.

Footnotes

1. This research was supported by the Social Sciences and Humanities Research Council, as well as a PhD studentship from Queen Mary University of London. This material is also based in part on work supported by the National Science Foundation under Grant No. 0347988. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Social Sciences and Humanities Research Council, Queen Mary University of London, or the National Science Foundation.

[Return to text](#)

2. And yet, whether through the latent tendencies of the performer or by the constraints of Mimi's programming, traditional formal structures may still emerge from performances with Mimi. We discuss this in other articles related to the same and other performances; see Schankler, Chew, and François (2014); and Schankler, Smith, and Chew (2011).

[Return to text](#)

3. The notation examples are interpretive transcriptions created by the improviser (the second author) by adding pitch spellings, rhythmic values, and barlines to the MIDI recording. They serve as plausible canvases on which to show the divergent perspectives on grouping in the music being analyzed. They were not used as artifacts for analysis, and do not derive from the recorded performance in an automatic way. Whether they reflect the state of the mind of the improviser cannot be said definitively, although we discuss this issue in paragraph 3.15.

[Return to text](#)

Copyright Statement

Copyright © 2014 by the Society for Music Theory. All rights reserved.

[1] Copyrights for individual items published in *Music Theory Online (MTO)* are held by their authors. Items appearing in *MTO* may be saved and stored in electronic or paper form, and may be shared among individuals for purposes of scholarly research or discussion, but may *not* be republished in any form, electronic or print, without prior, written permission from the author(s), and advance notification of the editors of *MTO*.

[2] Any redistributed form of items published in *MTO* must include the following information in a form appropriate to the medium in which the items are to appear:

This item appeared in *Music Theory Online* in [VOLUME #, ISSUE #] on [DAY/MONTH/YEAR]. It was authored by [FULL NAME, EMAIL ADDRESS], with whose written permission it is reprinted here.

[3] Libraries may archive issues of *MTO* in electronic or paper form for public access so long as each issue is stored in its

entirety, and no access fee is charged. Exceptions to these requirements must be approved in writing by the editors of *MTO*, who will act in accordance with the decisions of the Society for Music Theory.

This document and all portions thereof are protected by U.S. and international copyright laws. Material contained herein may be copied and/or distributed for research purposes only.

Prepared by Rebecca Flore, Editorial Assistant