

Commentary on Ohriner (2016)

NATHANIEL CONDIT-SCHULTZ

McGill University

ABSTRACT: This commentary offers a review of Ohriner’s corpus-based research on rap music, and includes an in-depth comparison of Ohriner’s model with my own. It highlights multiple divergences in our underlying methodologies, and outlines the potential impact of such differences on the research results. Ohriner provides a convincing and widely appealing presentation of how musical corpora may be leveraged to effectively enhance traditional forms of analysis. However, I question his approach to certain processes, such as sampling, transcription, and segmentation. I believe that his analysis would benefit from stronger definitions in these areas.

Submitted 2015 December 15; accepted 2016 March 1.

KEYWORDS: *rap, corpus, rhythm, rhyme, phrasing*

INTRODUCTION

AS an advocate for corpus-based research in popular music, I think it is quite fortuitous that this special issue of the *Empirical Musicology Review* will contain not one but two papers studying rap music. The inclusion of two articles with ostensibly similar methods and goals should offer EMR readers a more complete picture of rap music than either article could alone. Mitch Ohriner’s “corpus-assisted” analysis of OutKast’s “Mainstream” is an excellent piece of work well worthy of inclusion in this very special issue. The greatest strength of Ohriner’s article is his clear presentation of how musical corpora can be used to enhance traditional forms of analysis, in a manner which I believe will appeal to a broad swathe of theorists. Ohriner deftly maneuvers between Meyer’s categories of *critical analysis* and *style analysis*. In contrast, my article (Condit-Schultz, 2016) exclusively focuses on style analysis, and is certainly far less likely to appeal to music theorists with no background in computer-science, empiricism, or statistics.

Other than this broad stylistic difference between our articles, Ohriner and I also have made considerably different choices in our methodologies. It is an interesting case study in itself to compare our approaches in these similar projects. This comparison highlights the importance of the subjective decisions we as researchers must make, even in empirical work. This includes decisions about sampling, transcription, and operationalization, all of which influence the outcome of our final analyses. Typically there are not obvious “right” or “wrong” choices; rather, at every stage we are forced to pick between a set of choices in which there is no clear winner, or are constrained by practical possibilities, forcing us to make semi-arbitrary decisions. As a result, even projects with very similar starting goals can end up looking very different.

Sampling

The first methodological divergence between Ohriner and I is our approaches to sampling. As I mention in my paper, it is often hard to determine what the appropriate sampling plan is in humanistic research. Ohriner, like myself, attempts to balance empirical *representation* with more humanistic interest in *exemplars*. However, Ohriner does not articulate what the population he seeks to represent is, nor why his sampling method is suitable for representing his target of “a broad spectrum of the genre” (p.157). I would presume “the genre” refers to the population of all commercially recorded rap songs. However, Ohriner’s sampling strategy seems to be focused on selecting emcees for the inclusion of the sample, not songs. Songs are selected on a secondary basis, which seems to be related to their popularity in social media. Ohriner then samples a single verse. Overall, it is unclear whether the sampling population is emcees, songs, or verses.

Though Ohriner disavows the aim of representing the “best performances,” he still focuses on identifying “exemplary emcees.” Laudably, he incorporates a variety of characterizations of exemplariness,



including critical acclaim, influence, novelty, and obscurity. The use of these varied criteria for sampling will likely result in a more varied sample than my own, but what exactly this sample “represents” is not clear to me. Overall, I believe there is a stronger logical connection between my explicitly stated target population (“rap listening”) and my sampling source of Billboard. My main concern with Ohriner’s six sources is that the opinions of the professional critics who write these lists are not representative of the view of the majority of rap listeners, whereas Billboard is. I can’t help but think of the discrepancy between Critic and Audience reviews of movies on the website *rottentomatoes*. However, this reliance on “expert” listeners is closer to traditional music theory, and may be appropriate given Ohriner’s greater focus on “critical analysis.”

In a footnote, Ohriner tells us that his sampling is distributed across three time periods and four geographical areas. I believe this material should be reported in the main text. Given the well-known historical shifts in rap between “old-school” and “new-school,” not to mention the trends I’ve discovered in my own work, this breaking into three time periods is appropriate. Given the culture of rap, Ohriner’s inclusion of a geographical criterion in the sampling procedure is also an excellent idea.

Transcription

An important issue in any study of non-notated music, is what to transcribe, and how. When making symbolic transcriptions (as opposed to analyzing actual audio) there is inevitably subjectivity on the part of the transcriber. Ideally, we encode as much “objective” raw information as possible, without interpretation. However, “objectivity” is difficult to achieve and all the raw details of music can be prohibitively time consuming to notate, and ultimately more difficult to fruitfully analyze. By allowing more subjective interpretation in the transcription process we can gather much more immediately useful information, which is in turn easier to analyze. However, we also allow the subjective opinions (and other fallibilities) of the transcribers to color the results. Overall, I think that Ohriner and I include a comparable balance of objectivity and subjectivity in our approach to transcription.

As an example, consider the transcription of rhythm: Musical rhythms are not performed with the precise durations encoded in traditional music notation. In fact systematic micro-timing is clearly evident in most music, including rap. Thus, it would be more objective to encode the exact rhythmic timing of syllables in rap, perhaps in milliseconds, rather than musical durations. However, for a variety of practical and theoretical reasons, Ohriner and I both encode durations in terms of simple musical durations, based on our subjective interpretation of the pieces. In theory, our subjective decisions about the quantized duration or metric position of syllables could affect our conclusions. However, more “objective” measures of rhythm are both far more difficult to encode accurately, and much more difficult to analyze fruitfully. By transcribing a higher-level interpretative conception of rhythm (durations) we are able to quickly make musically relevant analyses. The question is does this quantization lead to any systematic bias in our interpretation of rhythm in rap?

Emcees often lag “behind the beat.” I (and I believe Ohriner as well) tend to encode “lagged” syllables as belonging to the beat they are lagging behind. In my paper I report a graph showing that stressed syllables tend to land on stronger beats than unstressed syllables. If I instead quantized lagged syllables to a later subdivision this graph might look very different. Ultimately, the validity of my graph (as well as some of my other conclusions) is based on the subjective belief in the validity of my quantization method—if you believe that “lagged” syllables are in actuality “early syncopations” then you should disregard my analysis.

One area where Ohriner and I differ significantly in our transcription protocol is the encoding of syllable stress. Ohriner creates an “objective,” deterministic, algorithm for determining stress, whereas I determine stress essentially by ear.[1] Ohriner’s algorithm is intuitively plausible, and from my experience transcribing the stress of tens of thousands of rap syllables, I can attest that his algorithm should do a very good job of identifying stressed syllables. In fact, in personal communication, Ohriner and I have found that our stress annotations show a very high level of agreement. However, I’ve also found that rappers do occasionally stress syllables in ways that violate the rules of Ohriner’s algorithm. In rare cases, I believe that my “by ear” annotations would capture the true performance more accurately than Ohriner’s algorithm.

This raises two important questions: first, who will make more mistakes, me (due to human error) or the algorithm (due to simplicity)? Second, and more importantly, how might my subjective choices potentially affect the outcome of analyses? Regarding the first question, I believe the algorithm will make more mistakes; what's more, it may make mistakes systematically due to some unaccounted for factors whereas pure "human error" should be random. Regarding the second question, I can't think of any examples of how any of my substantive conclusions would be influenced by any slight biases that might creep into my stress transcriptions. What's especially important is that none of my hypotheses, ideas, or conclusions rest very strongly on details of syllable stress. This is in contrast to rhyme.

Both Ohriner and I label rhyme manually. Ohriner laudably admits that his "annotation of rhyme is especially permissive" (p.159), something which is likely true of my annotations as well. Given the nature of our analyses and conclusions, I believe the subjectivity of our rhyme annotations is much more problematic than the subjectivity of either our rhythmic transcriptions, or of my syllable-stress annotations. For instance, if Ohriner had his hypothesis in mind during the transcription of "Mainstream" he may have unconsciously tended to be "permissive" in his annotation of rhymes that "project" triple-meter and "unpermissive" of rhymes which project the duple meter. The same could be said of my annotations. From the beginning of my project, I intended to compare the density of rhymes across artists, so there is a risk that I may have been tempted to be more "permissive" in annotating rhymes in Jay-Z or Eminem, and less "permissive" with the Beastie Boys, in order to maximize the variability between artists. I would like to think that both Ohriner and I are able to put aside our subjective ideas and annotate rhymes with an appropriate level of objectivity, but the possibility of some bias should be acknowledged. Unfortunately, the way around this problem would be to identify rhymes automatically but, as Ohriner points out, this is extremely difficult.

OTHER CRITICISMS

Rhyme

I disagree with Ohriners' suggestion that there is a lower-bound for the duration between rhymed syllables. Consider this line from Biggie Smalls' "Hypnotize":

Never loose, never choose to, bruise crews who...

The rhyming syllables "loose," "choose," "bruise," "crews," are each separated from the next by one, one, and zero syllables respectively. I completely agree with Ohriner that listeners probably don't "ascribe the same experience [to] rhyme between...successive syllables" that they do to "the patterning of the ends of lines." However, I still think these successive syllables are rhymes. What we need to distinguish is the basic phenomena of rhyme from the higher-order phenomena of the "rhyme scheme" (what Ohriner is referring to when he says the "patterning of the ends of lines"). One of the difficulties I've encountered in studying rhyme in rap has been distinguishing the usage of rhyme as part of regular, predictable, rhyme schemes, from other non-schematic (unpredictable) usages of rhyme. Ohriner discusses this issue when introducing "rhyme projection," noting that it is "not as simple as the durations between instances of rhyme in a rhyme class." Essentially, it seems that Ohriner is interested in how schematic rhyme schemes "project," but not in non-schematic rhymes which he refers to as "internal rhymes." Despite raising this issue, and demonstrating it clearly using his example from "Game Theory," Ohriner's operationalization of inter-rhyme durations doesn't seem to differentiate between internal rhymes and end rhymes: he measures durations between rhymes. This lack of separation between schematic and non-schematic rhymes is also assumed in the analysis presented in my current paper. In future work, I intend to do more to differentiate them and hope to see Ohriner doing the same as well.

Another issue is the idea of the "upper limit" on the distance between rhymes. On this point, I wholly agree with Ohriner as regards to *hearing* rhymes—syllables that are too far apart just don't sound like a rhyme. However, I have observed examples in rap that test my commitment to an upper bound on inter-rhyme time. Consider the following, hypothetical example: an emcee uses a particular rhyme class (let's call it **A**) once per measure for eight measures in a row, switches to different rhyme class (**B**) for four measures, then returns to his first rhyme class for four more measures, creating a rhyme scheme that looks

something like AAAAAAABBBBAAAA. The question is, do the two As I've underlined rhyme? Normally, I would agree that a four-measure gap is far too much time for syllables rhyme. However, given the many repeats of A, it seems possible that we might still hear some connection between these syllables. What's more, the "descriptive" approach to music theory, focusing on musical listening, is not necessarily the approach favored in music theory: even if the connection between underlined As isn't easily perceived, the connection may still have some music theoretic significance. The same can be said when a single rhyme class is used across multiple verses (for instance in the song "Beautiful" by Snoop Dogg).

To be clear, the rhyme annotations in my corpus also place an upper limit on inter-rhyme time, though I don't stick to one arbitrary limit (Ohriner's six beats), but allow context to play a role—for instance the tempo and the overall density of rhymes. My intent here is only to raise some questions about the issue of an "upper limit" on the time between rhymes.

Probability density

A minor quailm I have with Ohriner's paper is the density plots he presents in Figures 7 and 8. Though rhythmic time is technically a continuous space, our quantized transcriptions discretize the space into categorical durations. What's more, as a theorist and as a psychologist, I believe that our perception of rhythm is to some extent "quantized" into discrete categories, a phenomena known as "categorical perception." Thus, I would argue that inter-rhyme spans can be either sixteen 16th-notes or seventeen 16th-notes, but not 16.2, 16.3, 16.4, or 16.758 16th-notes. With these points in mind, I am not sure it is appropriate to calculate continuous probability density distributions based on these data. It is preferable to present discrete probability masses for each category, as I do in my bar plots. Whether the human perception of musical time is best modeled as a continuous or discrete space is up for debate, but when working with symbolic transcriptions that necessarily discretize continuous space, I think it is more appropriate to use discrete descriptive statistics.

Phrasing

My biggest issue with Ohriner's work is his approach to segmenting rap. His segmentation into *phrases* by the location of breaths is laudable for its simplicity and objectivity. However, I would think it is entirely common for emcees to deliver multiple musical phrases in a single breath. What's more, recording techniques sometimes eliminate audible phrases from recordings—single rap verses can be composed of edits of several performances which are pasted together, eliminating breaths.

My greater concern is with Ohriner's vague definition of *lines*. Ohriner does not explicitly differentiate syntactic closure from prosodic closure—his lines seem to be influenced by both, which limits the possibility of studying effects like elision and enjambment. What concerns me most is his stipulation that "there are as many [lines] as there are measures in the verse" (p.158). It is unquestionably the case that a ratio of one segment to one measure is the most common ratio in rap. However, making this a defining requirement seems like a huge assumption, which will force the interpretation of flow into narrow boundaries. In my corpus there are a significant number of songs in which two-measure or two-beat segmentation is clearly dominant—what's more, occasional passages contain highly irregular segmentation or extended cross-rhythms. Perhaps, Ohriner's smaller corpus[2] does not contain examples like these, but if there are, it seems to me that his definition of lines is untenable. In the case of "Mainstream," are the lines broken up to match the 4/4 measure?

Finally, it is not clear to me what the two different segmentation schemes (lines and phrases) he uses are meant to represent. All criticism aside, overall I doubt that Ohriner's segmentation differs that much from my own—there is probably a high level of agreement between us—but I would like to see a better definition of lines.

Interpreting the "Mainstream" Hook

My final comment is a more musical one. In Figure 3, Ohriner contrasts triple and quadruple interpretations of the hook of "Mainstream." He argues that "listeners familiar with hip-hip" might prefer the 3/4 interpretation, mainly because the meter matches the clear internal repetition. I would suggest that a more

likely hearing is 6/4 (2+2+2), or more accurately a measure of 4/4 alternating with a measure of 2/4. To me, the rhythm of each of the two lines in the hooks sounds very clearly like 4/4—hearing “peaches and cream” as landing on the downbeat of a measure seems unlikely. To me, the oddity is that the two lines are separated by only two beats of rest instead of four, which can be faithfully represented by an alternation of 4/4 with 2/4 (**Fig 1**).^[3]

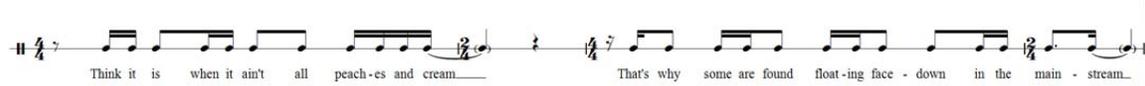


Fig. 1. Another metric interpretation of the hook of “Mainstream.”

CONCLUSION

I am extremely excited to have my own corpus-based paper (Condit-Schultz, 2016) on rap flow published alongside Ohriner’s excellent paper. Having these two papers published side by side offers a more complete picture of rap artistry than either does in isolation, and at the same time presents EMR’s readers an interesting case study in the ways methodological decisions shape research outcomes. I hope that our papers will inspire more music theoretic research into rap artistry and, along with the other fine work in this special issue, inspire more corpus-based music research in general. Most of all, I am inspired by Ohriner’s ability to balance empirical “style analysis” with more humanistic “style analysis”; I think that all of us interested in corpus-based approaches to the study of music can learn from Ohriner’s approach.

REFERENCES

Ohriner, M. (2016). Metric Ambiguity and Flow in Rap Music: A Corpus-Assisted Study of Outkast’s “Mainstream” (1996). *Empirical Musicology Review*, 11(2), 153–179.
<https://doi.org/10.18061/emr.v11i2.4896>

NOTES

[1] I actually use a simpler algorithm to get transcriptions started, then edit them by ear.

[2] My corpus contains about fifty more songs than Ohriner’s. In addition, I transcribe all verses of each song rather than just one, so the total number of measures/syllables in my corpus is likely much larger. Ohriner does not report exact numbers of measures or syllables.

[3] Incidentally one of the things that I find interesting about this hook is that the end-rhyme syllables (“cream” and “stream”) are separated by seven beats, which is neither triple nor duple.