



Commentary on Justin London's MTO 0.2 article

Stephen W. Smoliar



REFERENCE: mto.93.0.2.london.php

[1] There is one minor nit I would like to dispense with quickly concerning the matter of style. I felt as if this document had rather a parade of straw men in it. Each paragraph let me to raise my eyebrows and say, "Yes, but what about . . . ?," only to find that the but-what-about was covered in the following paragraph! It would have been nice had Justin not led us down quite so many garden paths in order to make his point, but perhaps I just happen to feel that way because I am in the thick of this stuff right now.

[2] What I *really* want to write about is a but-what-about stone which was left unturned by Justin's discussion. It's a pretty heavy stone, though: But what about the fact that there is already a researcher who has worked out a potentially interesting quantitative model which not only accounts for the dynamic nature of meter but may even provide a viable quantification of just how loud some of those rests are? The research in question is named Peter Desain, and I want to address his work because I have been hard a work reviewing a recent book, *Music, Mind and Machine: Studies in Computer Music, Music Cognition and Artificial Intelligence*, which Desain wrote with his colleague Henkjan Honing.

[3] (As an aside, my original intention was to write this review for *Artificial Intelligence*. However, it began to grow into something more like a paper than a review; so I ended up sending it to *Computer Music Journal*. Until I read Justin's paper, it had not occurred to me that it might be suitable for *Music Theory Spectrum*. For now, however, I just want to summarize one particular aspect of the work reported in this book.)

[4] Desain's model is called an *expectancy space*. It was actually introduced to comparatively evaluate systems concerned with the detection of metric beat in performances of music. For example, given a timetable of MIDI events from a keyboard performance, the system being evaluated should be capable of translating the real-time durations of events into the discrete symbols of music notation. An algorithm to solve this problem was first proposed by Christopher Longuet-Higgins in the Seventies, and Desain wanted to compare the performance of this algorithm with a system of his own design based on a neural network.

[5] The principle behind the expectancy space is similar to that of Meyer's expectations. Given a past history of duration events, the question is whether or not that history predisposed the model in favor of certain durations rather than others. For example, if the last six events have all been interpreted as the duration of an eighth note, the expectancy space gives what amounts to a high probability that the next note will also be an eighth note, somewhat lower probabilities that it will be a sixteenth or quarter note, and so on down to a very low probability that it will be a whole note. Note that the purpose of the algorithm is to reflect how the interpretation system actually performs, but that means that each interpretation system in turn may be viewed as reflecting a particular kind of listening behavior.

[6] Neither of the systems being compared present particularly convincing expectancy spaces. This is because they are both based on the rather trivial goal of trying to establish simple integer ratios between successive durations. Thus, everything is evaluated on a note-by-note basis without any attempt to hypothesize how notes are grouped into measures or any other higher-level construct. However, the expectancy space could be used to evaluate any other system which tries to take this sort of rhythmic dictation. What is important is that it treats such a system as a dynamic function processing data in real time and displays the relationship between specific data and the behavior of that function.

[7] The most important element of this technique is that it is quantitative. One is not dealing with highly subjective measurements which try to capture how strong an expectation it. At any moment in the course of a performance, the system gives a numerical weight of predisposition for the duration of the next event. If that next event does not happen, as would be the case with a rest, it would not be too far fetched to interpret that weight as the “loudness” of the rest.

[8] The only real problem with Desain’s results to date is that you have to have a model implemented before you can evaluate it with an expectancy space. Thus, the main thing we learn from his report is that note-to-note relations do not give us a particularly effective model, particularly when they only take duration into account. If one were to try to develop more realistic expectancy spaces, one would first have to assemble a more comprehensive model, taking account not only the recognition that duration is organized at a higher level than individual notes but also the roles of other parameters of performance, such as the pitches of the notes being performed, their dynamics, and perhaps their articulation. Such a model may still be a ways in the future, but the expectancy space now obliges US to think much more seriously and quantitatively about how it could be implemented.

[9] Let me close with one final nit. In paragraph [14] Justin writes: “Meter is neither a parameter like pitch or timbre, nor is it a part of a nested measuring of durational patterns and/or periodicities. It is something that is heard and felt.” Are not *all* aspects of musical sound elements that are “heard and felt?” Justin’s acknowledgement of phenomenology is all very well and good, but I do not think he gives it sufficient attention. *anything* which is either a musical object or a parameter of a musical object is ultimately a construction of the interpreting mind. That is as true of the sonority of a minor triad in first inversion as it is of a ternary metric pattern. The real question concerns the nature of the operations of construction which are brought into play in the course of listening. Justin is quite right that they are dynamic for meter; but, most likely, they are dynamic for all other aspects as well. The dynamic nature is not the issue. More important will be how well we shall be able to describe that nature in quantitative terms.

[10] Joel Lester raises some interesting points in his response to the London article. I think it is particular important to recognize the score as a set of instructions for performance whose information content should not be confused with that of sounding music. My guess is that one could augment his list of cues through which the sounding music can guide how one taps one’s foot; but enumerating those cues is not as important as acknowledging that such cues are there to be “picked up” from the audible signal.

[11] However, no matter how rich our supply of cues may be, it is rarely foolproof. Ultimately, there really is no good answer to the question: How do we know when to begin counting? The only absolute answer is: We don’t; we *hypothesize* a count. If we then discover that our count really does not “fit,” we update our hypothesis. It is this updating of a running hypothesis which makes the model “dynamic,” in London’s sense of the word. Unfortunately, his paper only began to scratch the surface of those dynamics (as did Desain’s work, coming from a different direction). The biggest rub, however, has to do with the question of “goodness of fit.” How do we determine whether or not our current hypothesis should be abandoned. That, I think, is where the sorts of cues Joel enumerated enter the picture. If too many of those cues offer too much evidence against where the hypothesis says the downbeat is, then it is time to change hypotheses.

[12] As a final point I think it is probably important that most of the cues which tend to be invoked to assess the running hypothesis are *surface* features. When one analyzes a score, one can find no end of “deep” structural features which offer evidence as to where the downbeat *really* is. However, I content that those features are another part of the landscape of instructions for performance. Listeners tend not to read scores, just as listeners to natural language tend not to diagram the sentences they hear. Rather, they pick up on surface features and respond to them. Perhaps, then, the real *art* of performance concerns how the deep features which are the result of careful analysis may be made available as surface features to the listening ear.

[13] I am getting a bit worried about the way in which we are all jumping on Example 3 in Justin’s paper. What worries me the most is a methodological danger which I shall call “selective denial of context.” It seems as if each interpretation chooses to bar certain experiential elements from the context in order to make its point, and I am not sure this is a terribly healthy way to go.

[14] For example, I have now read several accounts which basically have tried to abstract away from the way in which Example 3 is actually notated, as if any intelligent ear should be able to infer the notation from the listening experience. This strikes me as being akin to looking down the wrong end of a telescope. I prefer Lester’s view of the score as a set of instructions for performance. Thus, in this case the “game” is not one of inferring where, and how hard, to tap your foot. The score tells you that already; and it is the responsibility of the performer to make sure you “get the message.” Rather, the “game” is determining when you bring your foot down which particular emphasis on a rest; where, to some extent, the energy of your stomp may then be taken as a rough measure of the loudness of the rest. This is not a question of the listener resolving any ambiguities which are latent in the score. That’s the performer’s job. Rather, the question is how the performer endows the listener with a mental state based on expectancies which set his foot tapping in the first place. The reason I trotted out my Desain hobby horse at the beginning of this discussion was because his expectancy space provides a means by

which such a mental state may be inferred from strings of perceived durations.

[15] Having said all that, let me now stir up the pot which a bit more context which has received little attention. Having now sung Example 3 to myself so many times that it is beginning to invade my dreams, I have discovered that it is beginning to co-mingle with some more concrete musical memories. For example, while the resemblance is not note perfect, it begins with a gesture which we all know and love from the last movement of Beethoven's first symphony. I feel that such a "family resemblance" is particularly important when considering the "responsibility" of the performer. What I mean is that, because this particular passage is so similar in both pitch and rhythm to a passage which is in so many listeners' memories, the performer really does not have to do very much to communicate this particular set of score instructions. Indeed, the memory may well be triggered before even that first rest has been reached, thus making it all the easier for a mind with a rich memory to control the tapping foot.

Stephen W. Smoliar
Institute of Systems Science
National University of Singapore
Heng Mui Keng Terrace
Kent Ridge, SINGAPORE 0511
smoliar@iss.nus.sg

Copyright Statement

Copyright © 1993 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 Natalie Boisvert, Cynthia Gonzales, and Rebecca Flore, Editorial Assistants