Simulating the listening of composing

Eliot Handelman
Centre for Interdisciplinary Research in
Music Media and Technology (CIRMMT)
Montreal, Canada
Email: eliot@colba.net

Andie Sigler
McGill University School of Computer Science
and CIRMMT
Montreal, Canada
Email: andiejs@gmail.com

***Abstract*—We address the problem of autonomously generating original music in a dual system, comprising a generative and analytic module—a "listener." The analytic system implements a novel detailed theory of music, involving the recursive application of the concepts of *shape*, *pattern*, and *motion*.**

I. Introduction

The design of an AI music composition system poses nearly metaphysical problems: what is it for a system to produce music *autonomously*, rather than as a stand-in, however disguised, for the implementor? Clearly someone playing the Mozart dice game is not *really* composing: nor is a program, on that account, which merely pulls out its stock and trade. This paper describes work towards a creatively autonomous composition system, in which the central idea is that of a cycle between a generative and a *perceptive* model—a "listener"—that in some sense "understands" music in a human-like way.

The insight regarding understanding involves two factors: first, that pieces of music are structured so as to engage human attention; second, that all music reveals a differential, dynamic flow of *saliences*. By this is meant such things as the shattering patterns of Bruckner symphonies and the like, made of relatively simple repeating elements. These are prepared and offset by *zones* of *non*-occurrence, where particular sorts of patterns *do not* occur. Zones promote difference and structure. A relation between salience and *simplicity* seems evident: it also seems clear that pacing mediates attention. These seem to be matters that *all* composers must think about, regardless of style and culture. At a first approximation, we see a composition as an arrangement of zones with greater or less salience, with attention to the *rhythm* of pacing.

What then are simplicities? They are whatever is apprehended effortlessly, possibly through a specialized neural substrate. Two distinct types of simplicity are easily identified: for pitch, moving in a single orientation (all else equal) is obviously simpler than anything else; in the rhythmic domain, moving in a single pulse. The term "moving," as just used, signals a third type: the idea of transition or change. Berg's variation on one note (from his opera *Wozzeck*) exemplifies the bottoming-out of transition: getting louder, denser, thicker. Clearly change has an orientation—as it does a pattern in the rate of change—suggesting a fairly natural recursion in pattern, shape, and motion. This is in fact the trinity that forms the theoretical core of the Jill analytic system. It is not exhaustive, since a fourth type of simplicity seems to exist in *texture*, posing timbral questions.

It seems startling, on reflection, that there *does* exist a bottoming-out of musical possibilities: perhaps more startling than the reflection that there is a dual in how things "top-in" in noise and unintelligibility. While it is unclear how things become *too* complex, the pulsed beep can be treated as the basic unit of a combinatorial system, offering incremental degrees of complexity in variation. One wonders to what extent composers *necessarily* intuit some such system.

Pulsing is a prototypical *pattern*, a sequence of events involving recurrence. Orientation is is prototypical *shape*, a term intuitively used by musicians connoting the wholeness or gestural integrity of parts of music—motifs, phrases or forms. Musical "shapes" as such are mostly unnamed—an exception is the up/down shape known as arch-form[2] bottoming out in the 'neighboring-note." *Motion*, prototypically, means increase or decrease. Clearly these three terms seem partly to describe each other. Constant orientation is a *pattern* of motion. Constant pattern is a *shape* of orientation, and constant motion is both a kind of shape and pattern.

II. Analytic system

Though partly interchangeable, we find it useful to consider shape, pattern and motion on their own terms. Patterns can be abstracted as sequences of terms, like the familiar ABA, inviting approaches to analysis—i.e., the recovery of simplicities— specific to the structure of this representation. The idea is to *disengage* different dimensions of music as far as possible, treating all components orthogonally. Shape, as pertaining to orientation, is thus decomposed into the three basic orientations, each to be analyzed independently of one another. The problem, later, will be to put these back together again. The method involves *synchronization* of simplicities, where these occur in more than one orientation. The result will be something like a map of simplicities, providing opportunities for segmentation and zoning. The analysis does not stop here: the resultant objects may be again treated as shapes subject to a recursive analysis.

A. Z-shape

The single-orientation view leads to a number of natural recursions. The base object is the so-called oriented "chain" of notes, all moving just one way; we disregard all else. It

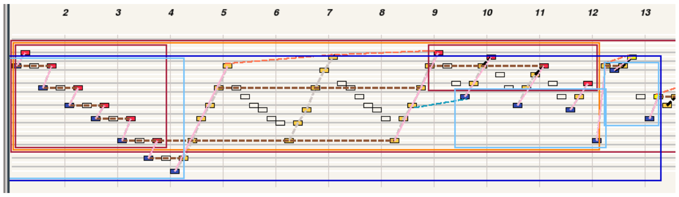

Fig. 1. Two 3rd-order Z-chains in the opening of *Presto* from J. S. Bach, solo violin Sonata I. The red notes (within red boxes) are tops, blue bottoms. The top and bot are synchronized at the beginning and nearly synchronized at the end.

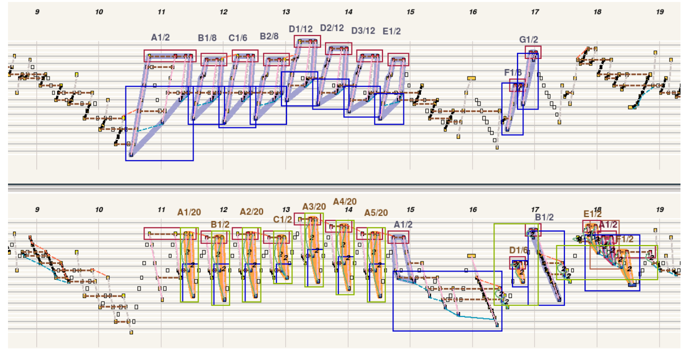

Fig. 2. Clusters of partly synchronized schemas in J. S. Bach: *Gigue*, Solo violin Partita II. The schemas outlined in the top half in lilac are defined by constant pitch top, rising pitch bot. The orange schemas in the lower view are "hotter," comprised of the additional feature of constant chain length, shown with green arcs.

may turn out that the view "up" involves far fewer notes than the view "down": *Happy Birthday* is almost entirely described by the view down. The chain (e.g. "up") may then be taken together with the *next* chain up, and their extrema compared. Each chain has a highest pitch–the "top"—and a lowest, the "bottom"; two tops create a new orientation, as do the two bottoms. Clearly tops and bottoms need not move in the same orientation: in fact they create a counterpoint. It is nevertheless clear that the *synchronization* of these two "outer pitch" levels provides the *simplest* way of creating a structure with two chains. If both take the orientation "same," the result is a trill or alternation. Both going up or down in the same interval results in the familiar figures of "patterned intervals." This illustrates, in a nutshell, the principle of decomposition, object construction, local simplicity analysis, and reconstruction of the whole through synchronizations.

The result of putting two chains together leads to objects with more than one orientation. The chains up may be regarded as going down, relative to their tops; then the chains up are "going down," and may be referred to as "up-down" (placing the lowest level orientation first). The next idea is to recursively "chain" consecutive pairs of chains sharing the same set of orientations, and generating a new set. Note that basic chains occur consecutively but higher-order chains, called "Z-chains," may be separated by gaps. This has several interesting consequences. The first is that in a full recursion, some Z-chains will reliably span entire compositions, including major works as Bach violin suites, Beethoven piano sonatas or concerti by Prokofiev. Through Z-chains, we have an immediate view of a whole composition. Moreover, Z-chains will capture repeating parts of the music, outlining the major components of the form. Further, the *pitch* simplicities of a work are invariably represented as well.

Z-chains require an *ordering* such as pitch but can also be applied to ordered features such as the number of notes in a chain or the intervallic span between the top and the bottom. The results here are not as intuitive as they are for pitch, but they capture simplicities and recognize "additive" procedures common in modern music, where e.g. each chain has one note more than the preceding chain. The result, in the current system, is a quadruple of Z-chains in triple orientation. How might these all be synchronized?

B. *Synchronization*

We see Z-chains as a mathematical model from which a perceptive model can be derived. The derivation involves the location of simplicities in the form of tight synchronizations. To be general, we must allow for different degrees of synchronization, as follows. Consider the features pitch top, pitch bottom, chain length and chain span. Take a *single* chain in a given feature: then which other featural chains are coextensive, if not perfectly synchronous? We regard as a basic object coextensions of *at least* two features. The result is called a "schema," providing a simple description or recipe for construction: e.g., top pitches should go up, the bottom should remain the same, and the length should be constant. Our use of the term does not imply learning, as with e.g. Gjerdingen[1]: it is simply a regularity within the music, a *candidate* for the perceived object.

Schemas provide us with very interesting views of *difference*. In all music analyzed, we encounter the following properties:

- Some schemas occur *only* on one side of a piece, leading to inferences about zoning.
- Some schemas occur *throughout* a piece, possibly reflecting motivic "unities."
- Some schemas present themselves with different *lengths*: schemas are extended or refracted, growing and shrinking, in which we see a trace of musical *development*.

We again have a problem of synchronization, since one set of schemas may potentially overlap others.

Since schemas are only partial descriptions, we may ask *which* instances are literal repetitions, yielding a *pattern* as a sequence of terms where equal terms denote equal objects. Similarly, we may ask which instances of a schema are transpositionally identical, possibly yielding a different pattern. Further, some subsets of schemas are contained within higher-order Z-chains, yielding patterns within Z-chains.

C. Pattern

The analysis of pattern, understood as sequences of terms like $AABB$, does not necessarily benefit from Z-chain analysis. The imperative here, to detect the obvious, requires specialized treatment in three independent, possibly synchronous, approaches. The simplest is, of course, the detection of recurrent subsequences, for which well-known methods apply. A second concerns "metrical" segmentation, constructing a set of *places*—e.g., the 2nd beat of each measure—each of which can be taken orthogonally, yielding a new set of objects for analysis. We will confine ourselves here to an original concept, the *concentric* pattern.

In this, the insight is provided by the classical minuet, whose Trio, deploying new material, occurs in the middle of a larger piece, a "piece within a piece." Sectional patterns, typically, resemble the following:

$$(ABCABCABCABC(DEFDEF)ABCABC)$$

The parentheses are to be read as follows. Within a set of parentheses, nothing on the inside occurs on the outside: DEF is strictly contained by the outer pattern. The central group is providing something that is entirely new, at least from the perspective that this pattern represents.

The prototypical concentric pattern is a mirror:

$$(A(B(C(D)C)B)A)$$

reflecting a "bottoming out" of the idea, suggesting a combinatorial matrix of variation. An interesting property of the concentric pattern algorithm is a natural *segmentation*, with more than one set of top level parentheses:

$$(ABCABCABCABC)(DEFDEF)$$

in which no terms in any top level group occur in any other top level group. We may need to infer whether we are looking at one piece, or two.

A randomly-generated pattern as long as the minuet above and with as many terms might come out looking like this:

$$(AB(CDCBDD(EBDCCFFAABEAE)DF)B)$$

Here the parentheses mean: *some* terms, not necessarily all, occur within and never without: E, bounding the innermost group, does not occur on the outside. Moreover, in order to get a grouping some terms must be taken together, as a unit: the second level of parentheses binds CDF as terms surrounding the next level. Patterns like these can be hard to interpret: we must fall back on the repetition structure in order to make claims about simplicity.

But as observed, schemas generate as many patterns as we wish, depending on the recurrence function used, and as many sets of patterns as there are schemas. Among these are sure to be some easily identified as "canonical," fully interpretable as structures of repetition within concentricities. The segmentation of a Bach work composed of two repeating sections proves to be a trivial consequence of this analysis.

D. Motion

Motion naturally leads us to inquire into *rhythm*, but here we restrict ourselves to the motion of schemas. We mentioned the concept of schemas *growing* and *shrinking* as a hallmark of musical development: one way to develop music is to allow something to "go on", "completing its course," extending an already established pattern. A simple case of this is found in the well-known Menuet from the Anna Magdalena collection[1], where the simple shape preceding each cadence is lengthened in the parallel passage. We regard "lengthening" of this kind as a sort of "formal motion", contributing to the sense that the music "has gone somewhere," "has done something." In a structure of his kind, we can speak of an orientation—*up*, with regards to length, generating a new set of objects. The "motion" therefore has a "shape."

Schemas allow for the construction of "motion shapes" in the following way. Schemas are *partial* descriptions: nothing precludes a description involving 3 features partially covering a description with just 2 features of the 3. Since schemas reflect underlying constant chains, we may presume that the schema with more features is more regular than that with fewer: in McLuhan's (jazz-influenced) terminology, the more regular schema is the "hotter" pattern, where a "hot pattern" is said to "drive" perception and a "cool" pattern is said to invite perceptual completion. The partial covering of a "cool" pattern—involving few features–by a "hotter" pattern—a superset of the cool pattern—can be thought of as a *heat* transition, or *heat schema*, in case the pattern of transition is detectable elsewhere: we find very many exemplars of this in the Bach works. Figure 3 shows an example.

A further kind of motion bears mention. In a study tracking the longest and hottest shapes in several corpora, we found that these are not uniformly distributed within a piece; rather their disposition tends towards asymmetry. This, we propose, is a key hallmark of a piece sounding as if filled with *agency*: living things do not behave mechanically, and an aim, certainly, of music is to compel a sense of being alive.

This very approximately describes the theoretical underpinnings of the Jill analytic system.

III. Generative System

The Jack & Jill system is based on the idea of a double interaction between a generative and an analytic component. However, as the basic philosophy of design here involves "unyoking" things, we felt compelled to try a version of the system that composed with no analytic help at all, but which simply generated the sorts of things that comprised the descriptive machinery of Jill. Further, it was decided to afford as much randomness as possible, just in case it turned out that the accidental interaction of independent components proved sufficient for our purposes. These were:

- The music generated should reflect subjectively interesting formal organizations which were *not* coded for, directly or indirectly – nothing should be canned.

[1]By Christian Petzold, misattributed to J. S. Bach

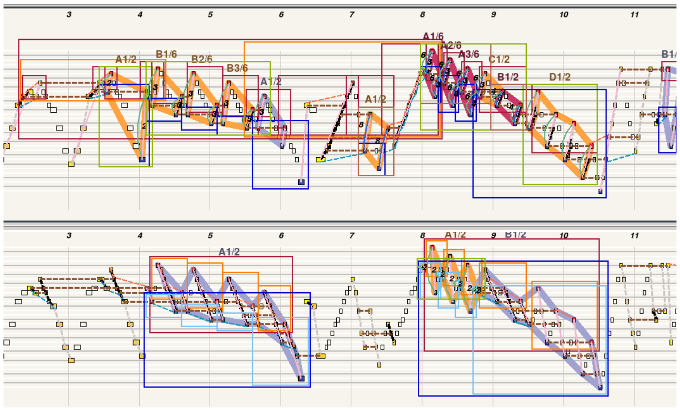

Fig. 3. *Gigue*, near the opening. Differential heat in the same schema, comprising a *heat schema*. Red is hottest, i.e. most regular, followed by orange and lilac. The boxes show the Z-chains synchronized by the schemas.

- Over a span of 1–3 mins., the music should subjectively compel a sense of *purpose*.
- The program should succeed in generating a very wide range of "styles," "idioms," and not be locked into a basic "sound" which it is powerless to overcome.

We refer to the system as Jack I. The low level idea is that in music, all things are "controlled" by patterns, which may vary in complexity but which should, routinely, reflect *simple* organizations. An algorithm was invented that generated *zones*, similar to the concentric patterns described above; these, ultimately, were responsible for differences occurring in the piece. Patterns were independently generated; their synchronization, if any, was random. Patterns operated a recursive machinery that generated shapes that could be extended, and more patterns decided how and where these would be placed. Shapes were independently generated and transposed to a skeletal "supershape," produced with the same machinery. The system ran a "metapiece" which was not descriptive of the results, but which did stipulate sequences of operations. It possessed also a rhythm system, which we cannot describe here. We find it to be in poor taste to comment on the adequacy of one's own music, however produced, but we felt that the methods deployed surpassed expectations; some of the pieces generated seem eerily filled with purpose, agency, and an unanticipated general atmosphere. The deployment of "motion asymmetries" in the growing and shrinking of shapes seemed to be a very effective strategy.

Jack I was of course limited in significant ways. It could not control dynamics, since these require (we think) a view of synchronization mediated by analysis, which was not yet available. It lacked a position on harmony, since this too involves synchronization, and was restricted to diatonic production. It did not explicitly compose polyphonically, although this partly emerged as a random side-effect. Finally, the system did, in the long run, produce music roughly of the same cloth, even though it did manage at times to achieve difference and uniqueness. This is among the main failings that are to be redressed with the analysis system.

IV. Memory and difference

We cast composition as a problem: composing always takes place within a context in which it is expected to produce differences. A composer *must* "make it new," not stumbling over the formulaic, not repeating oneself, not repeating others. The success of (human) composers therefore partly must be attributed to a capacity to recall an enormous amount of music, both at the note-to-note level and in very general terms.

The technical machinery described above seems appropriate for high-level schematic memory. Rather than fixing Jack I, we can simply reject whatever Jill finds to be too similar to what has already been done. We do not say what a piece *should* be, only what it should *not* be. At the same time, some sort of repetition between works is *necessary*: the basic material of music—simple shapes—*is* exhaustible and *must* be reused. This *could* be randomly constrained, but in the interests of design clarity, it seems preferable to create a system that will acknowledge high-level imperatives.

V. Conclusions: creativity and recursion

The pursuit of musical autonomy, though perhaps esthetically (ethically?) problematic, is valid as a means to creating discipline: it forces us to encounter hard problems, does not look for easy/deceptive solutions, and leads to an increase in the abstraction required of the whole system, and of the underlying theory of music.

Genuine autonomy might be acheived as follows. The objects we have described are not yet fully recursive, but will be through Jack and Jill's high-level language. Thus expressions like "the pattern of shapes of motion of patterns" will be composable[2] and intelligible, suggesting a combinatorial explosion.[3] Our solution is stochastic querying, in which recursive expressions are randomly generated. In this, there is a plausible fitness metric in the simplicities that arise, suggesting a genetically-driven theory generator that mutates on the strength of the kinds of patterns it is able to produce. Should such a system succeed, the result might be singular.

Acknowledgment

The authors thank Doina Precup and the School of Computer Science at McGill University for their support.

References

[1] R.O. Gjerdingen. *Music in the galant style*. Oxford University Press, 2007.

[2] David Huron. The melodic arch in western folksongs. *Computing in Musicology*, 10:3–23, 1996.

[2] In both a computational and musical sense.

[3] Z-chains, too, are open to a recursive explosion: we can meaningfully speak of the "bot(toms)s of the tops of the bots of the bots," etc.