MusMat

Brazilian Journal of Music and Mathematics

Vol. IV - No. 1 - June 2020

MusMat

Brazilian Journal of Music and Mathematics

Vol. IV - No. 1

MusMat Research Group (Ed.)

Graduate Program in Music
Federal University of Rio de Janeiro (UFRJ)
Rio de Janeiro - June 2020

MusMat • Brazilian Journal of Music and Mathematics

MusMat - Brazilian Journal of Music and Mathematics. MusMat Research Group (Ed.). Rio de Janeiro: Federal University of Rio de Janeiro (UFRJ), School of Music, Graduate Program in Music, 2020. Vol. IV, No. 1.

ISSN: 2526-3757

Contents

Foreword

It is with great satisfaction that the MusMat Group announces the release of the first number of the fourth volume of *MusMat – Brazilian Journal of Music and Mathematics*. Five very interesting articles integrate this number, covering diversified aspects from the rich confluence of musical and mathematical subjects. A study by **Scott Murphy** opens the issue, presenting an original approach of commmon-time meter, based on the properties of the correlate functions of metric weight and onset frequency. **Jean-Pierre Briot** examines deep-learning theory and techniques under the standpoint of autoencoder architectures, used for enhancing the compression of information for musical composition. **Liduino Pitombeira** presents a quite comprehensive survey concerning compositional systems, including the processes related to systemic modeling. **Arthur Kampela** discusses profoundly the processes associated with the Micro-Metric Modulation theory. **Marianthi Bozabalidou** addresses the theory of general scale systems through the prisms of algebraic groups, which involves the ideas of counterpoint groups and counterpoint spaces.

Carlos Almada
June 2020

Common Rhythm as Discrete Derivative of Its Common-Time Meter

Scott Murphy
University of Kansas
smurphy@ku.edu
Orcid: 0000-0001-7766-0777

Abstract: *Musicians recognize two important functions over the sixteen points in time, or beat classes, distributed evenly over a common-time measure: metric weight and onset frequency. Existing scholarship acknowledges that these functions are similar but not identical, and researchers tend toward one or the other as a model for metric entrainment. However, if the discrete metric-weight function is converted into a continuous curve, then the two functions strongly correlate: the ordering of each beat class by its backwards discrete derivative on this curve perfectly matches the ordering of each beat class by its onset frequency in a classical corpus.*

Keywords*: Rhythm. Meter. Calculus. Conducting. Fourier.*

I. Two Functions Over Moments in Common Time

The bar graphs of Figures 1 and 2, representing two functions, offer two perspectives of a musical measure divided into sixteen equal unit spans of time. The aim of this essay is to elucidate, in both mathematically elegant and musically intuitive ways, not only the structures of each function but also their mutual correlation. Before discussing the graphs individually, a few words are needed about how this information is presented generally. The moments beginning each unit span will be referred to as "beat classes" (bcs), because points in time in previous or subsequent measures are deemed as equivalent to those in this measure. Each function maps each beat class to what I will generally call a "level."[1] The measure that these two graphs describes would be typically indicated in musical notation as lasting the duration of a whole note, with a time signature of $\frac{4}{4}$, common time, or cut time, although other time signatures such as $\frac{2}{4}$ readily accommodate a sixteen-fold division. Although it is customary in musical terminology to assign the first element of a series of beats in a measure (or notes in a scale) to the number 1, the horizontal values for these graphs instead use zero-based numbering. This follows both many mathematical conventions in general and music theory's notation of beat class in particular, whereby "beat" in "beat class" describes not only moments of tactus onset—the more standard use of the word "beat"—but also the beginning of any unit duration that evenly divides a longer span as a modular (i.e. quotient) space, such as a measure.[2] I will mix both meanings of "beat" in what follows. In both graphs, my inversion—higher as lesser—of the more typical orientation of a Cartesian vertical axis—higher as greater—is deliberate and will be explained later.

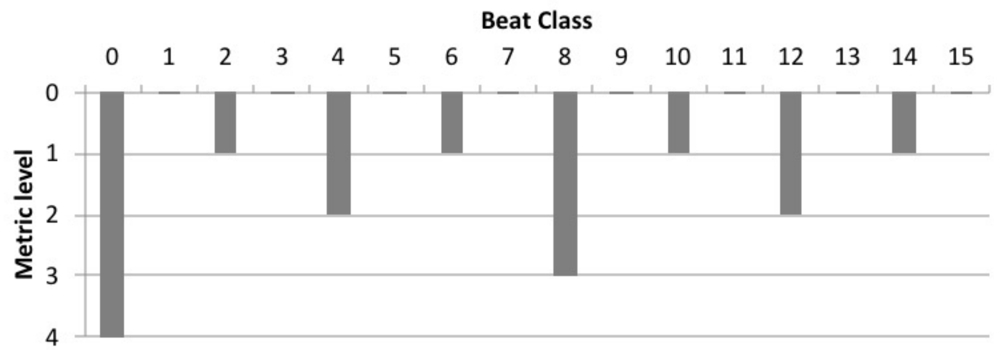


Figure 1: *Metric level of the 16 beat classes (f_1).*

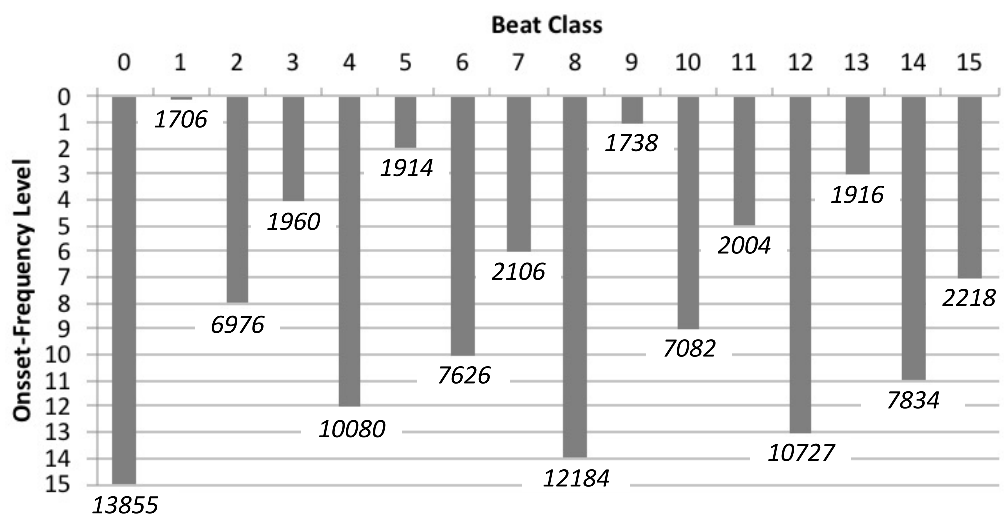


Figure 2: *Number of onsets (in italics) in common-time string-quartet movements by Joseph Haydn assessable by computer, and their ranking into 16 frequency levels (f_2).*

Figure 1 displays a standard interpretation of the relative weights or strengths of metric accent in such a measure.[3] In what follows, the more neutral term "level" encapsulates more subjective expressions like "weight" or "strength" of metric accent. The downbeat (bc 0) is on the highest level, because it carries the most weight of all of the beat classes in this or any other measure. If the lowest level is numbered 0, then this downbeat should be numbered as level 4, as there are five different degrees of metrical accent for this number of beat classes. The third beat (bc 8)—which occurs halfway through the measure—carries the second strongest metrical accent (level 3), and the second and fourth beats (bcs 4 and 12) tie for the third strongest metrical accent (level 2). The four remaining even-numbered beat classes (bcs 2, 6, 10, 14)—the "di" moment in the Takadimi [7] system of metric notation—tie for the fourth strongest metrical accent (level 1), and the eight odd-numbered beat classes—the "ka" and "mi" in Takadimi—tie for the least strong metrical accent (level 0). This mapping of beat-class number (x) to metric-accent level number (f_1) can be generalized for all "pure duple" meters (meters whose adjacent pulses relate by a factor of two) as follows: for a measure of 2^n units ($n \in \mathbb{Z}^+$), $f_1(x,n) = log_2(gcd(2^n, x))$, where gcd stands for "greatest common divisor."[4] For example, in a measure of sixteen units (n = 4), bc 4 has an accent level of $log_2(gcd(2^4, 4)) = log_2(gcd(16, 4)) = log_2(4) = 2$. Another way to define $f_1(x,n)$ is as the number of contiguous rightmost zeros in a binary (radix-2) representation of x. For example, the binary representation of 4 is 100, in which there are two contiguous rightmost zeros; therefore, bc 4 is on level 2.

For a common-time measure in Western classical music, Figure 2 displays a common ranking of onset frequencies among the sixteen equidistant beat classes. In particular, this bar graph shows in italics how many onsets occur in the sixteen beat-class positions for all of the string quartet movements by Joseph Haydn in $\frac{4}{4}$ assessable by computer [4]. These sixteen counts are then ranked into levels. With such a large data set, such that the frequencies of onsets for two different beat classes are highly unlikely to be the same, Figure 2 unsurprisingly has sixteen different levels, numbered 0 to 15, unlike the five of Figure 1. Thus, when described as a function—I will call it $f_2(x,n)$, parallel to f_1's definition—this mapping of beat class to level is bijective, whereas f_1 is surjective. However, despite the higher specificity, the ranking among these sixteen beat classes tends to be the same or quite similar for other Western common-practice corpora of sufficient size. One such instance appears in an article by David Huron and Ann Ommen [8] about syncopation in American popular music: their tally of onsets among common-time monophonic songs in the Essen Folksong Database well matches the rankings in Figure 2.[5]

The distribution of Figure 2 can be related to an understanding of Western common-practice styles. First, an onset chosen at random in a work related to a corpus from which f_2 values are computed is more likely to appear at a beat class with a higher f_2 value, regardless of whether the music is polyphonic (like Haydn's quartet movements) or monophonic (like the folksongs of the Essen Database). Second, as shown in Figure 3, each of the sixteen levels in Figure 2 corresponds to a rhythm within a $\frac{4}{4}$ measure, composed of onsets on the beat classes on that level or higher. For example, the level 13 rhythm is half-quarter-quarter, because these three onsets correspond to beat

[1] See [1] for a relatively early use of this term, although the concept is an older one.

[2] Here, "tactus" refers to a primary pulse near the range of 85 to 120 beats per minute. Justin London [14, pp.30-33] provides a summary of the concept.

[3] Lerdahl and Jackendoff [13, p.19] provide a well-known example.

[4] Cohn [2, p. 194] coins the term "pure duple."

[5] The published bar graph in [8, p. 215] does not contain specific counts. However, in June 2015 Huron kindly shared his data with me, which I will call $f_3(x,4)$ = [19221, 24, 2116, 412, 10226, 25, 5661, 430, 14280, 24, 2380, 348, 13458, 39, 5964, 634]. Of the 120 pairs of 16 beat classes that be ordered, only two of these 120 between $f_2(x,4)$ (my Haydn count) and $f_3(x,4)$ (Huron and Ommen's Essen count) are not the same: $f_2(3,4) < f_2(11,4)$ but $f_3(3,4) > f_3(11,4)$, and $f_2(1,4) < f_2(9,4)$ but $f_3(1,4) = f_3(9,4)$.

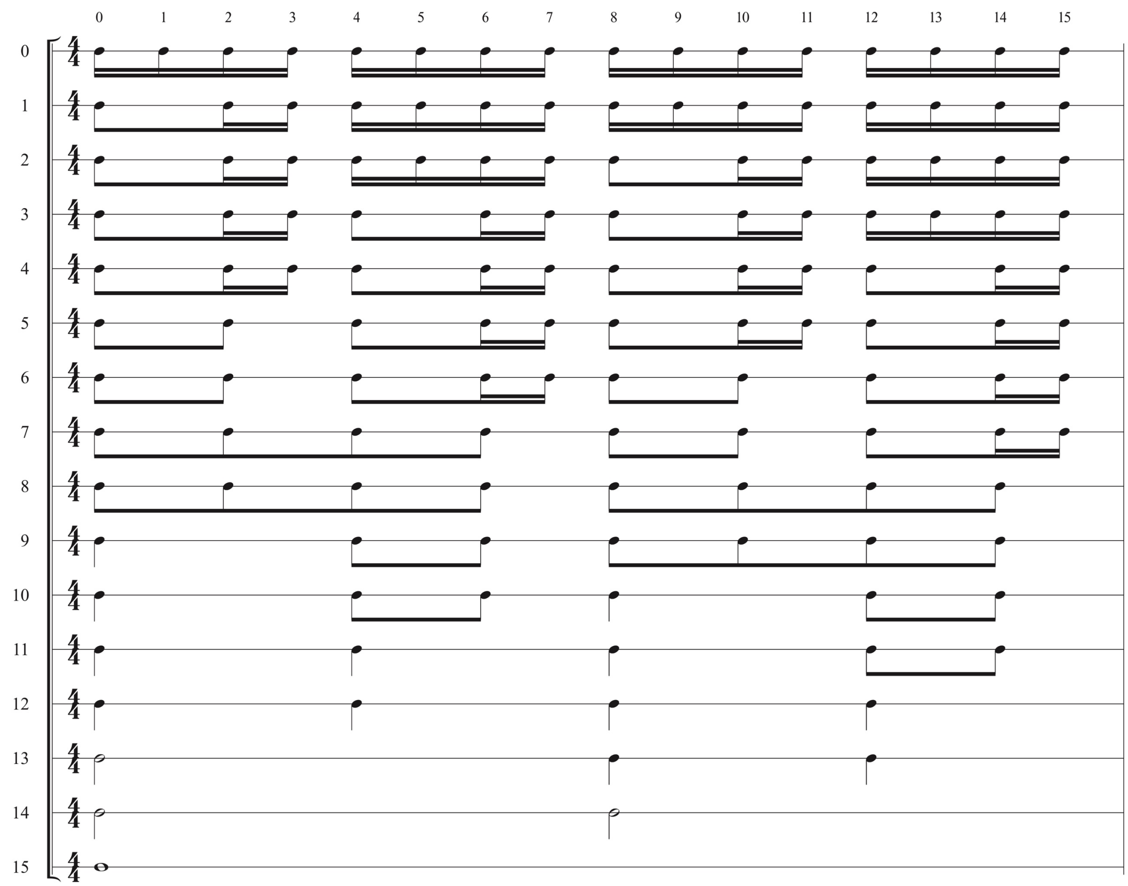


Figure 3: *Sixteen common rhythms in $\frac{4}{4}$, corresponding to the 16 levels and 16 beat classes from Figure 2.*

classes mapped to levels 13, 14, and 15 in Figure 2. Each of these rhythms is relatively idiomatic of, and rather common within, Western classical music. For example, in Johann Joseph Fux's 1725 treatise *Gradus ad Parnassum*, when the student writes a quarter-quarter-half rhythm within a common-time measure, the teacher gently recommends the half-quarter-quarter rhythm as a better solution.[6] These norms may be generalized by the time-honored preference for beginning longer inter-onset intervals on metrically stronger moments, which pushes shorter durations toward the end of metrical spans like measures, half-measures, and beats. To be sure, other rhythms besides those of Figure 3 are also quite probable in Western common-practice music. However, I speculate that, of two rhythms spanning the same kind of measure with the same number of onsets, the one that is more common in Western common-practice music is more likely than not to be the one whose sum of Figure-2 levels is higher.[7]

In the aforementioned article by Huron and Ommen, the authors note how two of their graphs, which correspond to my Figures 1 and 2, "are similar though not identical" [8, p. 214] to each other, acknowledging the imbalance between the frequency of onsets for the second and fourth beats (bcs 4 and 12) that I also cited earlier. They nonetheless continue with the recognition of "a notable correspondence between the hierarchy of event onsets [the counterpart of my Figure

[6] [5, p. 67].

[7] Temperley [17, p. 358] recognizes and addresses the shortcomings of this approach.

2] and the conventional metrical hierarchy [the counterpart of my Figure 1].” However, “[t]he causal relationship here is unknown” to the authors, although they speculate that “[i]t is possible that the metrical hierarchy originates in the distribution of event onsets. . . or the distribution of event onsets might simply reflect a pre-existing metrical hierarchy that influences the composition of music.” The primary purpose of this article is to propose such a causal relationship between Figures 1 and 2. My first step toward this proposal is to appreciate that *both* Figures 1 and 2 exhibit comparable symmetries and formulaic generalizations.

II. Symmetries and Formulas for Each Function

When one looks at both bar graphs, Figure 1’s symmetry is probably discerned more immediately: for example, all of its rankings invert around both bcs 0 and 8; that is, $f_1(x) = f_1(0–x \text{ mod } 16) = f_1(8–x \text{ mod } 16)$. Put in colloquial terms, assuming that the beat-class assignment continues cyclically into adjacent measures, Figure 1 exhibits a vertical mirror symmetry around both an axis at bc 0 and an axis at bc 8. In other words, if two beat classes sum to 0 or 8 mod 16, then they will have the same level in Figure 1. Moreover, Figure 1 admits formulation: earlier, I proposed two formulas for $f_1(x,n)$, one using a logarithm and the other using binary representation.

It is quite reasonable to perceive and understand Figure 2 as a distortion of Figure 1, especially with the latter’s conspicuous symmetry. It is even more reasonable to do so when the data of Figure 2 is displayed on a vertical axis whose units are onset counts and not ranking levels, as Huron and Ommen do in their article. From this vantage point, my Figure 2 may be seen as a distortion not only of Figure 1 but also of the scale of the onset distribution: for example, bcs 7 and 15 are adjacent in the ranking (levels 6 and 7) and differ by 112 onsets, while bcs 15 and 8 are also adjacent in the ranking (levels 7 and 8) but differ by 4758 onsets.

However, this recalibration of the graph’s range helps to reveal Figure 2’s own internal consistency that is at once independent from that of Figure 1, while also obliquely affiliated with it. For example, $f_2(x) = 15 – f_2(1–x \text{ mod } 16)$. In colloquial terms, assuming that beat-class designations continue into adjacent measures, Figure 2 exhibits a vertical mirror symmetry around both the point equidistant from bcs 0 and 1 and the point equidistant from bcs 8 and 9. In other words, if two beat classes sum to 1 mod 16, then they correspond, not with the *same* ranking as they do in the Figure 1 mirror, but rather *opposite* rankings. For one example, bcs 0 and 1 sum to 1 mod 16, and bc 0 hosts the most frequent onsets (level 15), while bc 1 hosts the fewest (level 0). For another example, bcs 12 and 5 sum to 1 mod 16, and bc 12 hosts the third-most frequent onsets (level 13) while bc 5 hosts the third-fewest frequent onsets (level 2).

Furthermore, f_2 may be written formulaically. At the heart of one such formula is a function from applied mathematics that also employs binary representation, a form of representation I used earlier for a formulation of $f_1(x)$. This function is the bit-reversal permutation (rev_n), which maps a set of integers $0, 1, \ldots, 2^n–1$ ($n \in \mathbb{Z}^+$) to itself by mapping each integer to the reversal of the bits in its binary representation.[8] For example, $\text{rev}_4(12) = 3$, because the four-bit binary representation of 12 is 1100, and the retrograde of 1100 is 0011, which is 3 in decimal form. Generalized for any span divided into 2^n equidistant beat classes, $f_2(x,n) = (2^n–1) – \text{rev}_n(\text{-}x \text{ mod } 2^n)$. Table 1 provides the computation of this equation for n = 4 in particular, matching the mapping of beat classes to the rankings of Figure 2.

[8] The bit-reversal permutation has been applied to FFT (Fast Fourier Theorem) algorithms, as demonstrated in [3, p. 918].

Table 1: *Demonstration of $f_2(x,n) = (2^n-1) - rev_n(-x \bmod 2^n)$, for $n = 4$.*

beat class (x)	$f_2(x,4)$	-x mod 16	-x mod 16 in binary	rev_4(-x mod 16) in binary	rev_4(-x mod 16)	$(2^4-1) - rev_4$(-x mod 16)
0	15	0	0000	0000	0	15
1	0	15	1111	1111	15	0
2	8	14	1110	0111	7	8
3	4	13	1101	1011	11	4
4	12	12	1100	0011	4	12
5	2	11	1011	1101	13	2
6	10	10	1010	0101	5	10
7	6	9	1001	1001	9	6
8	14	8	1000	0001	1	14
9	1	7	0111	1110	14	1
10	9	6	0110	0110	6	9
11	5	5	0101	1010	10	5
12	13	4	0100	0010	2	13
13	3	3	0011	1100	12	3
14	11	2	0010	0100	4	11
15	7	1	0001	1000	8	7

III. Relating the Two Functions to Each Other

The resemblances between the manners in which I have described the symmetries and formulations of f_1 and f_2 hint at, but do not themselves furnish, a deep-seated connection between them. The use of binary representation in defining both f_1 and f_2 intimates a utility in a component-wise disassembly and reassembly of both metrical and rhythmic wholes, and the different positions of the f_1 and f_2's axes of symmetry suggests that metrical symmetries hinge on the beat classes themselves, while rhythmic regularities operate in between these beat classes somehow. My investigation into a causal relationship between Figures 1 and 2 continues by revisiting Huron and Ommen's article. At one point, they set two psychological theories of temporal regularity head-to-head:

> Some psychologists have proposed that the metrical hierarchy arises from integrally-related mental oscillators that coordinate auditory attending. However, recent psychological research more strongly suggests that rhythmic perceptions arise from simple exposure to rhythmic stimuli rather than via mental oscillators. This research suggests that patterns of auditory attending arise through the mechanism of statistical learning. Listeners are sensitive to the frequency of occurrence of sound events, and these distributions appear to become internalized as mental "schemas." [8, p. 214].

On the one hand, a series of synchronized mental oscillators operating at different frequencies, all multiples of the downbeat frequency, correlate with metrical weighting schemes such as that in Figure 1. On the other hand, the various schematized rhythms in Figure 3 that follow from the statistical distribution of Figure 2 entrain an acculturated listener to a metrical orientation and flow. The words "rather than" in the quotation above suggest that these two models are mutually incompatible. But could they be mutually reinforcing? Is it possible to replace "rather than" with "in addition to" or perhaps even "derived from"? I believe the answer to both of these questions could be "yes."

The use of a mental oscillator as a model for entrainment to a periodicity expresses not only the regularity of this periodicity and its persistence in the absence of constant support, but also the continuous nature of metric experience, in contrast to the discrete and discontinuous design of Figure 1 and f_1's requirement of integer input. Meter has been characterized as the fluctuation of a listener's attention [6]. The greater the metrical weight, the more a listener is paying attention. This correlation probably stems from evolutionary efficiency. Changes in music, such as shifts to a new harmony, often occur on metrically accented moments. Therefore, it one wishes to discern as much information about the music as possible with the least amount of attentional energy, it makes sense to attend considerably more to downbeats or downbeat-like moments in particular and considerably less to the moments in between. But one cannot change this energy non-continuously. Thus, the model of an oscillator—such as a spinning circle, a swinging pendulum, a vibrating spring, and so forth—captures this continuous change, and has been called upon by scholars to model meter, particularly in the work of Edward Large.[9]

A common visualization of such an oscillating function is as a continuous periodic wave, peaking at the moment of greatest metrical weight and bottoming out at the moments of least metrical weight (although this verticality metaphor has been, and could be, inverted). These waves can be binarily categorized in two ways. First, music scholars have constructed these waves either more causally as visual aids or precisely as mathematical functions. Second, music scholars have constructed these waves so that either its curvature—that is, its second derivative—at any point has the same sign or not. For example, the musicologist and conductor Viktor Zuckerkandl drew inverted cycloid-like waves to depict the periodicity of the downbeat in Chopin's "Military" Polonaise op. 40 no. 1.[10] Although Zuckerkandl's argument is more philosophical than mathematical, his downbeats as the sharp points of an inverted cycloid's maxima suggests the notion that the rate in which one is gaining attentional energy as a downbeat approaches is always increasing. This acceleration and sharpness appears to emulate an idealized metric state, such as the crispness of a conductor's beat pattern, in which the hand moves more quickly as it both approaches and leaves a beat.

This sharpness qualitatively differs from modeling downbeats as the plateaus of a sine wave or an iterated normal distribution, where the rate in which one is gaining attentional energy as a downbeat approaches slows down at some significantly earlier point, which provides a better model for how listeners might "hedge their bets," adjusting to small fluctuations in a periodic stimulus. The use of von Mises distributions [11] allows for a continuum between the two types of curves: a wave with a high κ (its measure of concentration) produces sharp peaks whereas a wave with a low κ creates gently sloped plateaus.

Large and co-author Caroline Palmer [12] also innovated the combination and display of component waves into a composite wave that represents multiple periodicities of a metrical hierarchy. For example, their model of triple meter combines two periodic von Mises distributions of equal amplitude but with one distribution's period three times as long as the other. In an analogous fashion, Figure 4 models the $\frac{4}{4}$-plus-eighths-and-sixteenths meter of Figure 1 by combining—in the manner of Fourier synthesis—five periodic parabolic functions of equal amplitudes but with periods two times as long as the next shorter function. These five periodic parabolic functions signify the five metric levels in a $\frac{4}{4}$ measure subdivided into sixteen equal parts. While other types of waves—such as sine or von Mises—could be used for the components of this composite metric function, the periodic parabolic wave has a constant second derivative, or acceleration of attentional gain. This captures a general intuition that, as a relatively strong beat is approached, the rate in which one is gaining attentional energy is always increasing. This intuition aligns with a conductor-musician synchronization study that concluded that "absolute acceleration along the movement trajectory. . . was the main cue used by participants to synchronize with the conductors' gestures" [15, p. 470]. (Zuckerkandl's cycloid would also provide this continuous acceleration and yield the same results below, but a parabolic function is mathematically easier

[9] [9], [10], and [11] are three early examples.
[10] [18, p. 171].

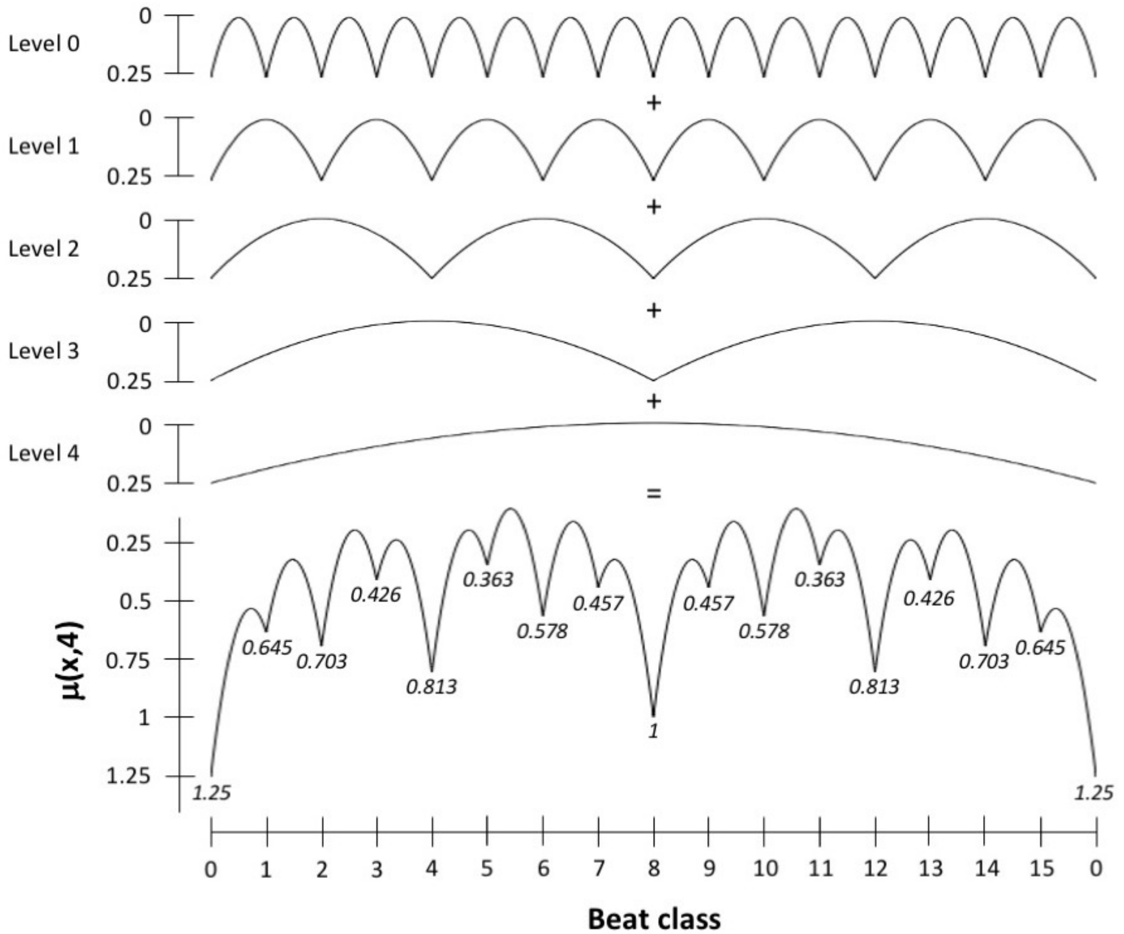


Figure 4: *Demonstration of the synthesis of a composite metric wave composed of five periodic parabolic waves at different frequencies, with values for the 16 beat classes provided in italics, rounded to the nearest thousandth.*

to work with.) This rate then drops non-continuously to a lower value when the strong beat is passed, and then begins to rise once again in its approach to the next. I will generalize this composite function as μ(x,n) for all pure duple meters that are 2^n units long (therefore, Figure 4's composite wave is μ(x,4)) as follows, where $\lfloor\rfloor$ represents the floor function that produces the function's periodicity:

$$\mu(x,n) = \sum_{i=0}^{n} \left(\left\lfloor \frac{-x}{2^i} \right\rfloor + \frac{x}{2^i} + \frac{1}{2} \right)^2$$

This composite wave is attractive to me as a musician and a theorist for multiple reasons, which I will attempt to convey through a listening and kinematic activity that I encourage the reader to try. (It is because of this activity and the ideas that stem from it that the customary orientation of the vertical axis has been consistently reversed in all of my graphs.) Audiate a pure duple meter with five different metric levels such that Level 0's periodicity is somewhere between 200 beats per minute (closer to a $\frac{4}{4}$ measure) and 30 beats per minute (closer to eight measures of $\frac{4}{4}$). Now, in time with the music, move your arm up and down within a range of one or two vertical feet to connect the cusps of $\mu(x,4)$ while flexing your wrist so that your hand (which is perhaps holding a baton) moves along the curves of $\mu(x,4)$. First, despite its limitation to a single spatial dimension, this still feels to me a lot like conducting, a gestural embodiment of the many levels within the meter. Second, the vertical position of your hand at beat classes 0, 4, 8, and 12 corresponds exactly to the metric weight graphed in Figure 1, although the same cannot be said for the other beat classes.

However, this non-correspondence of the other twelve beat classes sets up a third correspondence. In a $\frac{4}{4}$ conducting pattern, the downbeat is preceded by the largest descending motion of the hand. In fact, in standard conducting patterns, the downbeat's significance is experienced somatically by the conductor and indicated visually to the ensemble not by the relatively low position of the hand, but rather by the relatively large *downward* change of position that precedes this low position. It is a *down*beat, not a *low*beat. The same can be said for conducting $\mu(x,4)$: from the immediately preceding beat class, the arm descends the most into beat class 0 than into any other beat class. The values for the 16 beat classes provided in Figure 4 help to see this distinction: the ordered difference between $\mu(0,4)(1.25)$ and the immediately preceding $\mu(15,4)(\approx 0.645)$ is ≈ 0.605, which is larger than the next-largest ordered difference of 0.543 between $\mu(8,4)(1)$ and the immediately preceding $\mu(7,4)(\approx 0.457)$. Described using terms from calculus, beat class 0 has the highest backwards difference, or backwards discrete derivative (∇), of all 16 values in $\mu(x,4)$: $\nabla\mu(0,4) > \nabla\mu(x,4)$ for all $x \in \mathbb{Z}_{16}$, $x \neq 0$, where $\nabla\mu(0,4) = \mu(x,n) - \mu(x-1$ mod16, n).[11]

Beat class 0's superlative position regarding $\nabla\mu(x,4)$ correlates with its highest level in both Figures 1 and 2. While this correlation only extends to the second level in Figure 1, each beat class's ranking by $\nabla\mu(x,4)$ perfectly corresponds to its ranking by onset count shown in Figure 2. Table 2 shows this exact correspondence. In mathematical terms, for all x,y $\in \mathbb{Z}_{16}$, $\nabla\mu(x,4) > \nabla\mu(y,4)$ if and only if f_2(x,4) > f_2(y,4). In practical and embodied terms, the distance one's hand travels downwards to a beat class when conducting $\nabla\mu(x,4)$ matches the degree of likelihood that an onset (or more onsets, if polyphonic) will occur at that beat class in a common-practice work.

This exact correspondence strikes me as more than coincidental. An even division of time focuses a listener's attention more toward some moments equidistant in time and less toward other moments in between. This change of attentional degree is necessarily continuous to some degree. Due to this constant flux, those moments of greater attention can be distinguished not only by the absolute high state of attentional level but also by the *change* of this level from a preceding moment on a lower level to the current moment on a higher level. For relatively strong beats like downbeats, both the state of the moment and the change of state into the moment highly correlate and become interchangeable as indicators. Multiple periodicities, each with their own continuous attentional functions, constitute a composite attentional function. In such music with multiple periodicities, certain points in time that are equivalent in their *state* of composite metrical attention will nonetheless be preceded by different degrees of *change* of state of composite metrical attention, differentiating them. Owing to an overgeneralization inherent in the aforementioned interchangeability of state and change, it follows that a moment preceded by a greater change of state—such as the fourth beat in a $\frac{4}{4}$ measure—could be experienced as more downbeat-like than a moment of equivalent state preceded by a lesser change of state—such as the second beat in a $\frac{4}{4}$ measure.

Earlier, I quoted Huron and Ommen's proposal that "patterns of auditory attending arise through the mechanism of statistical learning" of common rhythmic patterns, such as those in Figure 3 for common time. Nothing I have presented here calls this into question. However, I hope that what I have presented here is a reasonable hypothesis for why at least one meter and its constituent periodicities give rise to certain rhythmic patterns—and not others—affiliated with this meter in the first place.

[11] Reale [16] offers a recent application of calculus's backwards discrete derivative to metric dissonance.

Table 2: *Demonstration of correspondence between $f_2(x,4)$ and $\nabla\mu(x,4)$.*

beat class (x)	$f_2(x,4)$	$\mu(x,4)$	$\nabla\mu(x,4) = \mu(x,4) - \mu(x-1 \bmod 16,4)$	$\nabla\mu(x,4)$ ranking
0	15	1.25	0.60546875	15
1	0	0.64453125	-0.60546875	0
2	8	0.703125	0.05859375	8
3	4	0.42578125	-0.27734375	4
4	12	0.8125	0.38671875	12
5	2	0.36328125	-0.44921875	2
6	10	0.578125	0.21484375	10
7	6	0.45703125	-0.12109375	6
8	14	1	0.54296875	14
9	1	0.45703125	-0.54296875	1
10	9	0.578125	0.12109375	9
11	5	0.36328125	-0.21484375	5
12	13	0.8125	0.44921875	13
13	3	0.42578125	-0.38671875	3
14	11	0.703125	0.27734375	11
15	7	0.64453125	-0.05859375	7

IV. Extensions

I leave it as an exercise to the reader to adapt and extend f_1, f_2, and my proposed manner of relating them to other common-practice meters besides common time, such as what Cohn [2] calls “pure triple” (such as $\frac{9}{8}$) or “mixed meters” (such $\frac{3}{4}$, $\frac{6}{8}$, and $\frac{12}{8}$), each with a potential variety of subdivisions. My preliminary forays into doing so are producing encouraging results. However, to the reader interested in this exercise, I offer a recommendation that I have already built into the present study: use continuous functions like periodic parabolic or cycloid waves. In my initial research into correlating Figures 1 and 2, I built μ from constituent sine waves, and the correspondence between f_2 and $\nabla\mu$ was just as exact. However, neither sine waves nor von Mises curves—regardless of the value of κ—yield nearly as close of a correspondence for mixed meters, at least when the amplitudes for each constituent function are the same, as periodic parabolic or cycloid waves, two continuous functions that are arguably more in line with conducting motions.

References

[1] Cohn, R. (1992). "Transpositional Combination of Beat-Class Sets in Steve Reich’s Phase-Shifting Music". *Perspectives of New Music*, 30: 146–177.

[2] Cohn, R. (1992). "The Dramatization of Hypermetric Conflicts in the Scherzo of Beethoven’s Ninth Symphony". *19th-Century Music*, 15: 188—206.

[3] Cormen, T., et. al. (2009). *Introduction to Algorithms*. Third Edition. Cambridge: The MIT Press.

[4] Cuthbert, M. *Music21*: A Toolkit for Computer-aided Musicology (http://*web.mit.edu*/*music*21/).

[5] Fux, J. (1725/1965). The Study of Counterpoint from Johann Joseph Fux's Gradus ad Parnassum. Translated and edited by Alfred Mann, with the collaboration of John Edmunds. Revised Edition. W. W. Norton.

[6] Gjerdingen, R. (1989). "Meter as a Mode of Attending: A Network Simulation of Attentional Rhythmicity in Music". *Intégral*, 3: 67—92.

[7] Hoffman, R.; Pelto W.; White J. (1996). "Takadimi: A Beat-Oriented System of Rhythm Pedagogy". *Journal of Music Theory Pedagogy*, 10: 7-–30.

[8] Huron, D.; Ommen A. (2006). "An Empirical Study of Syncopation in American Popular Music, 1890–1939". *Music Theory Spectrum*, 28: 211—231.

[9] Large, E. (1994). Models of Metrical Structure in Music. Annual Conference of the Cognitive Science Society, 16. *Proceedings...*

[10] Large, E.; Kolen, J. (1994). "Resonance and the Perception of Musical Meter". *Connection Science*, 6: 177–208.

[11] Large, E.; Jones M. (1999). "The Dynamics of Attending: How People Track Time-varying Events". *Psychological Review*, 106: 119—159.

[12] Large, E.; Palmer C. (2002). "Perceiving Temporal Regularity in Music". *Cognitive Science*, 26: 1—37.

[13] Lerdahl, F.; Jackendoff R. (1983). *A Generative Theory of Tonal Music*.Cambridge: The MIT Press.

[14] London, J. (2012). *Hearing in Time*: Psychological Aspects of Musical Meter. Second Edition. Oxford: Oxford University Press.

[15] Luck, G; Sloboda, J. (2009). "Spatio-temporal Cues for Visually Mediated Synchronization". *Music Perception*, 26: 465—473.

[16] Reale, S. (2019). "The Calculus of Finite (Metric) Dissonances". *Music Theory Spectrum*, 41: 146—171.

[17] Temperley, D. (2010). "Modeling Common-Practice Rhythm". *Music Perception*, 27: 355—376.

[18] Zuckerkandl, V. (1956). *Sound and Symbol*: Music and the External World. New York: Pantheon Books.

Compress to Create

Jean-Pierre Briot
Sorbonne Université, CNRS, LIP6, F-75005 Paris, France
UNIRIO, Rio de Janeiro, RJ 22290-250, Brazil
Jean-Pierre.Briot@lip6.fr
Orcid: 0000-0003-1621-6335

Abstract: *The current tsunami of deep learning has already conquered new areas, such as the generation of creative content (images, music, text). The motivation is in using the capacity of modern deep learning architectures and associated training and generation techniques to automatically learn styles from arbitrary corpora and then to generate samples from the estimated distribution, with some degree of control over the generation. In this article, we analyze the use of autoencoder architectures and how their ability for compressing information turns out to be an interesting source for generation of music. Autoencoders are good at representation learning, that is at extracting a compressed and abstract representation (a set of latent variables) common to the set of training examples. By choosing various instances of this abstract representation (i.e., by sampling the latent variables), we may efficiently generate various instances within the style which has been learnt. Furthermore, we may use various approaches for controlling the generation, such as interpolation, attribute vector arithmetics, recursion and objective optimization, as will be illustrated by various examples. Before concluding the article, we will discuss some limitations of autoencoders, introduce the concept of variational autoencoders and briefly compare their respective merits and limitations for generating music.*

Keywords*: Deep learning, Autoencoder, Latent variables, Music generation, Control.*

I. Introduction

One of the first documented case of algorithmic composition (i.e., using a formal process, including steps (algorithm) and components, to compose music), long before computers, is the Musikalisches Wurfelspiel (Dice Music), attributed to Mozart. A musical piece is generated by concatenating randomly selected (by throwing dices) predefined music segments composed in a given style (Austrian waltz in a given key).

The first musics generated by computer appeared in the late 1950s, shortly after the invention of the first computers. The Illiac Suite is the first score composed by a computer [12] and was an early example of algorithmic music composition, making use of stochastic models (Markov chains) for generation, as well as rules to filter generated material according to desired properties. Note that, as opposed to the previous case which consists in rearranging predefined material, abstract models (transitions and constraints) are used to guide the generation.

One important limitation is that the specification of such abstract models, being rules, grammar, or automata, etc., is difficult (reserved to experts) and error prone. With the advent of machine learning techniques, it became natural to apply them to learn models from a corpus of existing music. In addition, the method becomes, in principle, independent of a specific musical style (e.g., classical, jazz, blues, serial). More precisely, the style is actually defined *extensively* by (and learnt from) the various examples of music curated as the training examples.

The two main abstract models which can be induced by machine learning techniques are Markov models and artificial neural networks[1]. Since the mid 2010s, deep learning – the 3rd wave of artificial neural networks – has been producing striking successes and is now used routinely for applications such as image recognition, voice recognition and translation. A growing area of application of deep learning techniques is the *generation of content*, notably music but also images. Various types of artificial neural network architectures (feedforward, recurrent, etc.) are being used[2]. In this article, we will focus on a specific type of artificial neural network, *autoencoders*, and how they turn out an interesting approach for generating music.

II. Related Work and Organization

A recent book about deep learning techniques for music generation is [2], with some shorter introduction and survey in [1]. Some general surveys about of AI-based methods for algorithmic music composition are [22] and [6], as well as books [4] and [21]. A complete book about deep learning is [8]. There are various introductory books (and courses) about artificial neural networks, e.g., a very good course named "Machine Learning", created by Andrew Ng at Stanford, and freely available on Coursera.

In this article, in Section III we will introduce the (actually very simple) concept of autoencoder, the way to represent music and to train it on a corpus of Celtic melodies. Section IV will introduce a straightforward way of using an autoencoder to generate new melodies, illustrated by various examples. Section V will introduce and analyze various approaches for controlling the generation of the melodies. Section VI will discuss some limitations and further developments, notably the concept of variational autoencoder, and analyze its merits and limits concerning generation, before concluding this article.

III. Autoencoder

i. Architecture

An *autoencoder* is a feedforward[3] neural network with the following structural constraints:

- the size (number of nodes) of the output layer is *equal* to the size of the input layer;
- there is (exactly) *one* hidden layer; and
- the size of the hidden layer is *smaller* than the size of the input layer.

The output layer actually *mirrors* the input layer, creating its peculiar symmetric diabolo (or sand-timer) shape aspect, as shown in Figure 1, with two components highlighted:

- the *encoder* component, composed of the input layer and the hidden layer (and their connexions); and
- the *decoder* component, composed of the hidden layer and the output layer (and their connexions).

[1] Some tentative comparison, pros and cons, of neural networks and Markov models for learning musical style and for generation may be found in [2, Section 1.2.3].

[2] See a tentative classification and analysis, e.g., in [1].

[3] A *feedforward* neural network, also named *multilayer Perceptron (MLP)*, is the most basic and common type of artificial neural network. It is composed of an arbitrary sequence of *layers* composed of artificial *neurons*, where each successive layer is analog to *multiclass logistic regression*. Computation proceeds by *feedforwarding* data from the input layer into the successive layers, until reaching the output layer.

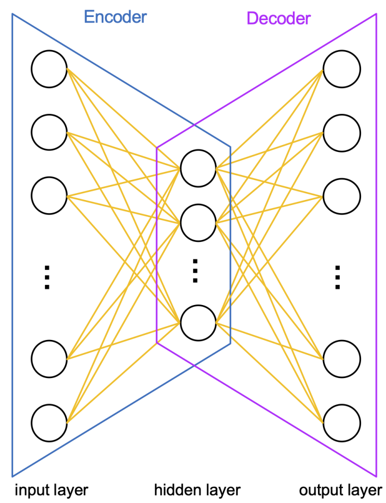


Figure 1: *Autoencoder architecture (with its encoder component in blue and its decoder component in purple)*

ii. Training

An autoencoder is trained with each of the examples both as the input data and as the output target, thus trying to minimize the difference between the reconstructed data and the original input data. As the hidden layer has fewer nodes than the input layer, the *encoder* component must *compress* information[4], while the *decoder* component has to *reconstruct*, as accurately as possible, the initial information. This forces the autoencoder to *discover* significant (discriminating) *features* to encode useful information into the hidden layer nodes, considered as a vector of *latent variables*.

iii. Generation

Indeed the motivation for an autoencoder is neither in just learning the identity function and nor in the direct compressing of data, as opposed to some experiments in using compression for creating art, e.g., the compressed cars by the sculptor César in the 1960s and more recently by the sculptor Ichwan Noor (see Figure 2). The *latent vector* of an autoencoder constitute a compact representation (some kind of label [26]) of the common features of the learnt examples. By *instantiating* this latent vector and *decoding* it (by feedforwarding it into the decoder), we can generate a new musical content corresponding to the values of the latent variables and in the same format as the training examples.

iv. Representation and Encoding

In order to use an autoencoder with music, we need to define a way to represent that music. As in this article we focus on *algorithmic music composition*, we will consider a *symbolic* representation (of notes and durations), as opposed to some *audio* representation (waveform signal or spectrum). We choose a *piano roll* representation, for its simplicity. Piano roll (and its name) is inspired from automated mechanical pianos with a continuous roll of paper with perforations (holes) punched

[4]Compared to traditional dimension reduction algorithms, such as principal component analysis (PCA), feature extraction by an autoencoder is nonlinear, thus more general, but it does not ensure orthogonality of the dimensions, as we will see in Section VI.iii.

Figure 2: *Compressed cars by César (left) and by Ichwan Noor (right)*

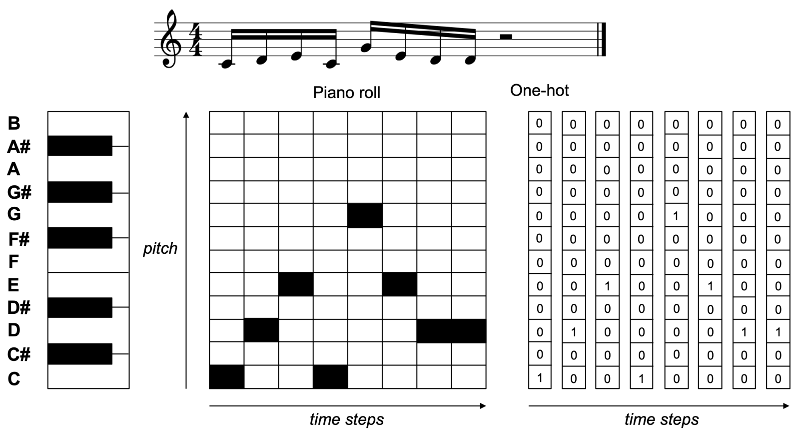


Figure 3: *Example of: (top) score; (middle) piano roll; (right) one-hot encoding – with a time step of a sixteenth note*

into it. In practice, it is a two dimensional table with the x axis representing the successive time steps and the y axis the pitch, as shown in Figure 3.

There is still an additional step from the representation to the artificial network input, this is the *encoding*[5] of a representation (of a musical content). It consists in the *mapping* of the representation (composed of a set of *variables*, e.g., pitch or dynamics) into a set of *inputs* (also named *input nodes* or *input variables*) for the neural network architecture. The most frequent type of encoding is *one-hot-encoding*[6], where a discrete or a categorical variable is encoded as a *categorical variable*, through a vector with the number of all possible elements as its length. Then, to represent a given element, the corresponding element of the *one-hot vector*[7] is set to 1 and all other elements to 0. For instance, the pitch of a note is represented as shown in the right part of Figure 3[8].

[5]Note that this stage of encoding is different and independent of the encoding that will be performed by the Encoder.

[6]The advantage of *one-hot encoding* over *value encoding* (direct encoding of a variable as a scalar) is its robustness against numerical operations approximations (discrete versus analog), at the cost of a high cardinality and therefore a potentially large number of nodes for the architecture.

[7]The name comes from digital circuits, *one-hot* referring to a group of bits among which the only legal (possible) combinations of values are those with a single *high* (hot!) (1) bit, all the others being *low* (0).

[8]The Figure also illustrates that a piano roll could be straightforwardly encoded as a sequence of one-hot vectors to

Figure 4: *"The Green Mountain" (8 first measures)*

Figure 5: *"Willa Fjord" (8 first measures)*

Note that a *global time step* has to be fixed and usually corresponds, as stated by Todd in [27], to the greatest common factor of the durations of all the notes to be learned. In the case of the corpus that we will consider, it is a sixteenth note. Also note that, there is no way to distinguish between a long note and a repeated short note[9]. Therefore, we will use the solution proposed in [10] to consider holding curent note (a *hold*) as a special kind of note (pitch). This solution is simple, but its main limitation is that it only applies to the case of monophonic melodies[10]. We will also consider silence (rest) as a special kind of note[11]. These two special cases will be added to the set of possible note pitches for the one-hot vector.

v. Learning Celtic Melodies

In this article, we will use as corpus a set of Celtic melodies, selected from the folk music repository "The Session" [14]. In practice, we selected 75 melodies, all in 4/4, in D Major key, and tagged as "Reel" (a type of Celtic dance). Two examples, "The Green Mountain" and "Willa Fjord", are shown[12] in Figures 4 and 5, respectively.

The shortest melody in the corpus is 8 measures long and the shortest note duration is a sixteenth note. The lowest note pitch is G_3 and the highest note pitch is B_5. Thus, the number of possible notes within the [G_3, B_5] interval is 29. The size of the final one-hot vector is thus 31 (after adding the hold and rest cases). The size of the the input representation is therefore: 8 (measures) $\times$ 16 (sixteenth notes per measure) $\times$ 31 (size of the one-hot vector) = 8 $\times$ 16 $\times$ 31 = 3,968.

vi. Architecture

Successive melody time slices are encoded into successive one-hot vectors which are concatenated and directly mapped to the input nodes of the neural network autoencoder architecture. In Figure 6, each blackened vector element, as well as each corresponding blackened input node element, illustrate the specific encoding (one-hot vector index) of a specific note time slice, depending on its

construct the input representation of an architecture, as we will see in Figure 6.

[9] Actually, in the original mechanical paper piano roll, the distinction is made: two holes are different from a longer single hole. The end of the hole is the encoding of the end of the note.

[10] Which is the case for Celtic melodies. Polyphonic music would need to be represented as different voices/tracks.

[11] For the reason discussed in [2, Section 4.11.7].

[12] Actually, only their 8 first measures, which is the actual length of the melodies that will be considered, as explained just below.

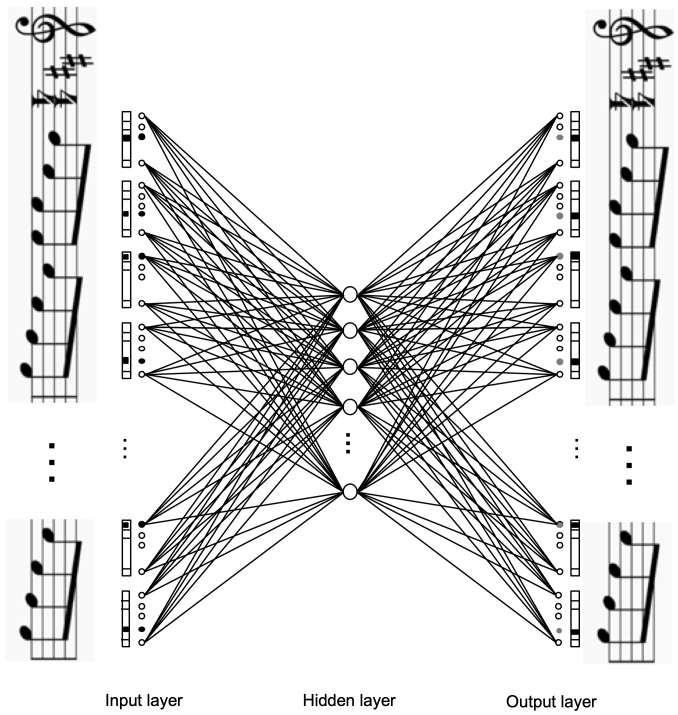


Figure 6: *Autoencoder architecture for learning melodies*

actual pitch (or a note hold in the case of a longer note). The dual process happens at the output. Each grey output node element illustrates the chosen note (the one with the highest probability), leading to a corresponding one-hot index, leading ultimately to a sequence of notes.

The input layer and the ouput layer of the autoencoder architecture have 3,968 nodes. The hidden layer has an arbitrary number of nodes[13], e.g., 1500 nodes. The output layer activation function (as well as the hidden layer activation function) is sigmoid and the loss (reconstruction error) function is binary cross entropy[14]. The optimizer algorithm is ADAM and training hyperparameters are: number of epochs = 100 and minibatch size = 20. We use the Keras framework as front end, Theano platform as the back end and our own made representation library[15]. Purposively, we do not use any additional optimization, as to keep it simple and generic.

Training the architecture proceeds by presenting an example of melody at the input layer and at the output layer (as the target for the reconstruction)[16]. The training procedure will incrementally adjust the connexion weights between neurons in order to minimize the reconstruction error.

IV. Generation

The model having being trained, it may be used for generation. As explained in Section III.iii and shown in Figure 7, we instantiate the latent vector, usually denoted as z, and feedforward it into the

[13]Which will be varied, as we will see in Section IV.i.

[14]See, e.g., [2, Sections 5.5.3 and 5.5.4] for details about the reasons of these choices.

[15]It transforms a musical score into data for the architecture and vice-versa. We designed it for our course at UNIRIO which is available at `http://www-desir.lip6.fr/~briot/cours/unirio3/`. It uses the Music21 symbolic music representation library as pivot and also for reading MIDI and ABC music formats.

[16]Actually, a mini batch of examples, randomly selected from the training set, is used for each epoch.

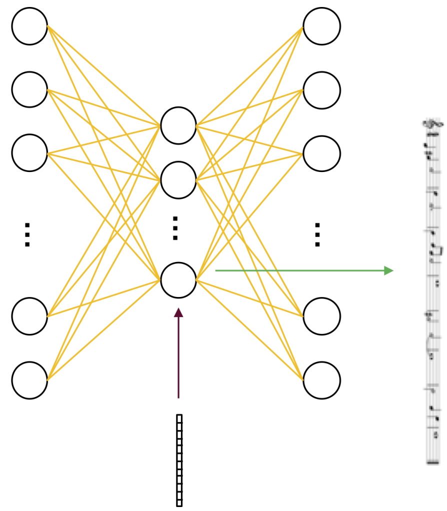


Figure 7: *Generation of a melody by the decoder component of the autoencoder from a latent vector (random or given)*

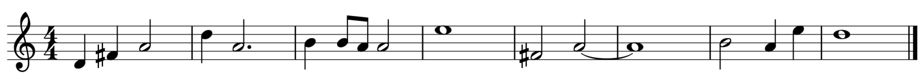


Figure 8: *Example of melody generated from a random latent vector by the decoder component of the autoencoder ($h = 1500$) trained on the Celtic melodies corpus*

decoder component. This will reconstruct some melody corresponding to the compressed version. Figure 8 shows an example of melody generated from a latent vector randomly generated[17].

i. Size of the Hidden Layer

For simplification, we will name h the size of the hidden layer (which is also the size of the latent vector z). Setting the value of h is an important decision. If h is large, close to the input (and output) layer size, reconstruction will be almost perfect or even perfect, but there will be many latent variables to be instantiated in order to generate melodies[18]. If h is small, the control of the generation will be easier to understand, specially in the case of $h = 2$ where the latent vector can be visualized in a 2D-figure[19], as shown for each example of the corpus[20] in Figure 9, but the reconstruction will not be optimal.

Figure 10 shows the loss (the error of the reconstruction) and the accuracy (the precision of the reconstruction), in function of h. Our experiment shows that $h = 1.258$ is the lowest value for which the accuracy is equal to 1, that is, for which the reconstruction of all training examples is perfect.

In Figure 11, we show the reconstruction by the autoencoder of “The Green Mountain” melody, in function of the value of h. We could see that with $h = 1000$, the reconstruction is still perfect[21].

[17] The ranges of the possible values for each latent variable may be determined by computing the lowest and the highest values of latent variables for each training example, see Figure 9.

[18] And moreover, the autoencoder may not be enough forced to extract interesting features.

[19] Let us think of the analogy of approximating a 3D location on earth onto a 2D map and even onto a 1D road on this map.

[20] To compute the value of z corresponding to a given melody, we just need to feedforward the melody data into the encoder component of the autoencoder and retrieve the values of the hidden layer nodes, i.e. the latent variables.

[21] The reconstruction of “The Green Mountain” is perfect but it is not the case for all the training examples, otherwise

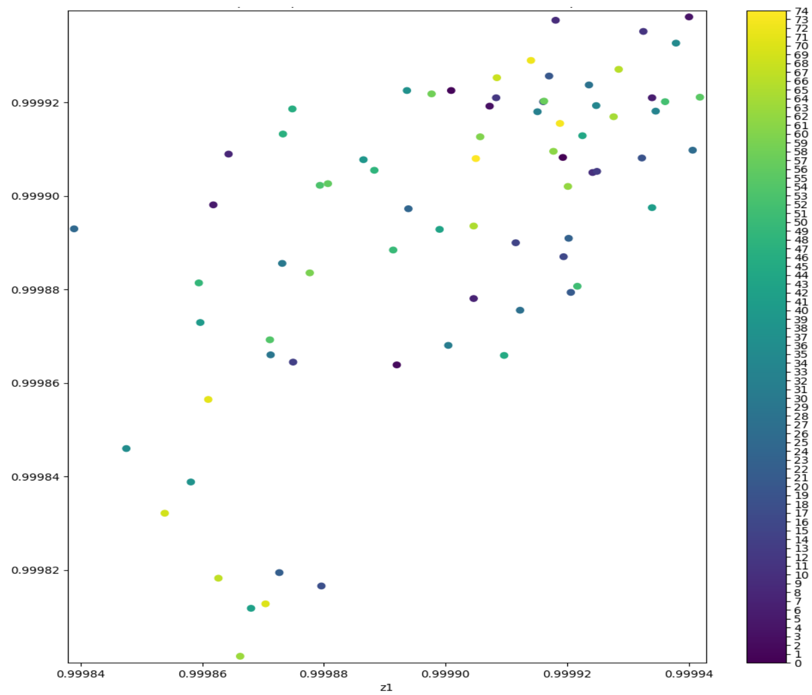


Figure 9: *Values of the latent vector (z) generated by the encoder component of the autoencoder (h = 2) for each of the 75 examples of Celtic melodies. Each dot and its color (from blue to yellow) corresponds to each melody of the corpus (from #0 to #74)*

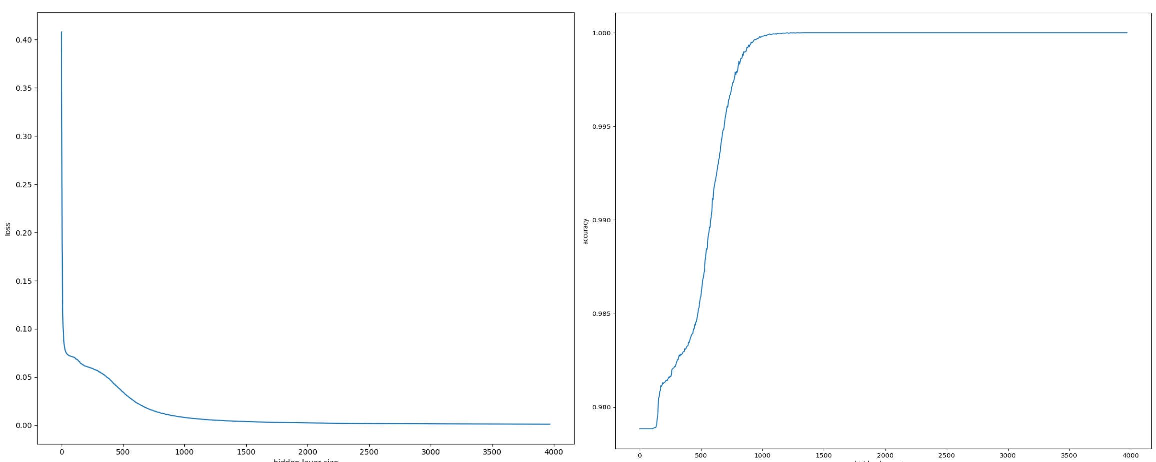


Figure 10: *(left) Loss and (right) Accuracy of the reconstruction by the autoencoder, in function of h*

With $h = 750$, reconstruction has only a minor error: the last note of the 6th measure, a D quarter note, has been substituted by D and B 16ths notes. With $h = 500$, some further errors and

the accuracy (for all examples) would have been equal to 1. Also note that, as opposed to the initial score of "The Green Mountain" in Figure 4 which has a key signature (with two ♯, i.e. D Major) because the actual key was part of the specification, the score of the reconstruction has no key signature but the notes are indeed equivalent. (We used MuseScore to display the scores).

Figure 11: *(from top to bottom) Reconstruction by the autoencoder of "The Green Mountain" for h = 1000, 750, 500, 250, 200, 150, 100, 2*

simplifications appear, although most of melodic motives are still preserved. It is from h = 150 that the melody starts being simplified, with from h = 100 a major simplification trend and quasi stability.

It may seem surprising that, even with a small size of the hidden layer, the autoencoder could reconstruct partially an initial melody. What does the autoencoder is in fact to map the various dimensions corresponding to the various latent variables to some variational characteristics which vary among examples, e.g., main melodic motif, average duration of notes, etc.[22] What is common to all examples is stored into the connexion weights of the network, for it to be able to reconstruct a melody. What is specific to examples is stored into the latent variables. Therefore, changing the values of the latent variables will allow to change the melodies generated.

V. Approaches for Generation

i. Exploration

The latent space (the space of variation for z the latent vector of latent variables) may be explored with various operations to *control/vary* the generation of content, e.g., as summarized in [24]:

- *translation*;
- *interpolation*;
- *averaging*;
- *attribute vector arithmetics*.

ii. Interpolation

We may for instance do *interpolation (morphing)* between two existing melodies, "The Green Mountain" (see Figure 4) and "Willa Fjord" (Figure 5), with, e.g., 5 steps of linear interpolation. In practice, as shown in Figure 12:

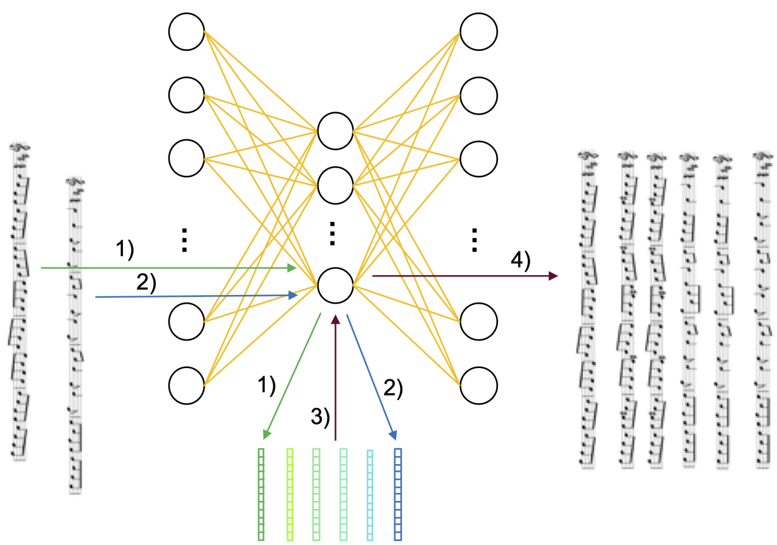


Figure 12: *Generation of interpolation between two melodies by the autoencoder: 1) encode 1st melody into a latent vector; 2) encode 2nd melody into another latent vector; 3) interpolate between them; 4) decode interpolated latent vectors to reconstruct successive melodies*

1. we compute the value of z resulting from feedforwarding "The Green Mountain" into the encoder component of the autoencoder;
2. as well as the value of z resulting from feedforwarding "Willa Fjord";

[22] As will be discussed in Section VI.iii.

3. we compute the various values/steps for interpolating between the two values of z; and

4. we feedforward each value into the decoder component of the autoencoder to reconstruct the corresponding successive melodies.

In the case of $h = 1500$, the resulting melodies are shown in Figure 13. We could see that the interpolation, although correct (it maps the start and the target), is not uniform. Step 1 is equal to start and step 4 is equal to target. This discontinuity limitation will be analyzed and addressed in Section VI.i.

Figure 13: *(from top to bottom) Melodies resulting from the interpolation (5 steps) by the autoencoder ($h = 1500$), from "The Green Mountain" to "Willa Fjord"*

We can also do interpolation between arbitrary values of z. With $h = 2$, we interpolate the

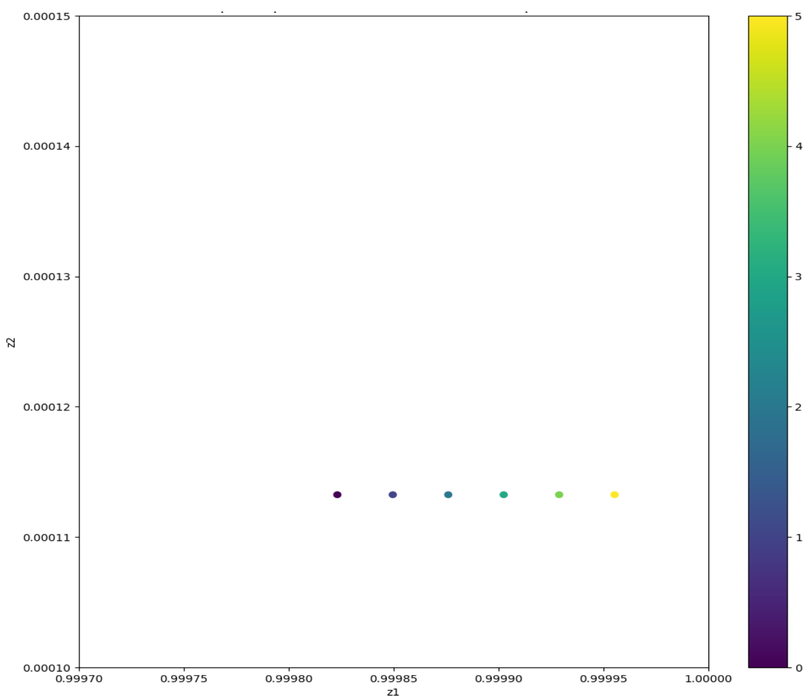


Figure 14: *Values of z for the interpolation (5 steps) of z_1 (from its min value (dark purple blue spot) to its max value (yellow spot)), while z_2 is constantly equal to its mean value*

Figure 15: *Melody resulting from the interpolation (5 steps) by the autoencoder (h = 2) of the value of z_1 (from its min value to its max value), while z_2 is constantly equal to its mean value*

value of z_1 between its minimum value and its maximum value[23], while z_2 stays constant to its mean value, as shown in Figure 14. Unfortunately, the resulting melody, shown in Figure 15, is actually constant for all steps. The same happens, with a different generated melody shown in Figure 16, when interpolating the value of z_2. This is actually another illustration of some limitation of the ways the autoencoder dispatches the melodies in the latent space, as will be analyzed in Section VI.i.

[23]These values, as well as mean values, for all latent variables, are computed from the latent vectors for all training examples.

Figure 16: *Melody resulting from the interpolation (5 steps) by the autoencoder (h = 2) of the value of z_2 (from its min value to its max value), while z_1 is constantly equal to its mean value*

Figure 17: *Bach Chorale BWV 347 soprano voice (transposed into D Major key) (8 first measures)*

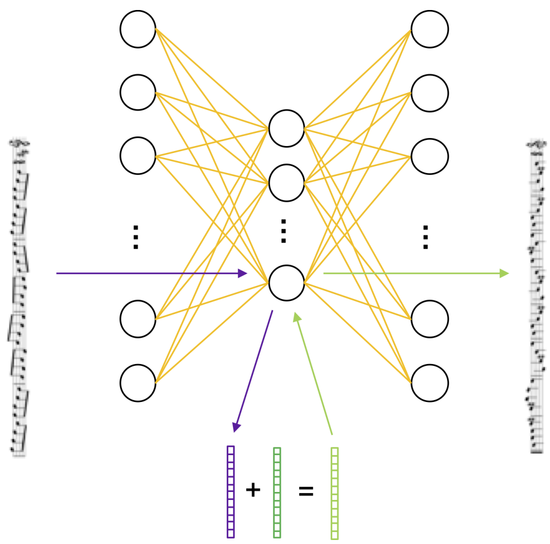


Figure 18: *Generation of the adaptation of a melody by adding an attribute vector to its encoded latent vector and decoding the resulting latent vector*

iii. Attribute Vector Arithmetics

In this approach, the idea is to specify some attribute vector capturing a given *characteristic* (e.g., long notes, high pitch range, etc.) and to apply it to an existing example in order to influence it. An *attribute vector* is computed as the mean of the latent vectors of all examples sharing that characteristic. As an example, let us *augment* the Celtic corpus with a set of (80) soprano voice melodies from Bach chorales – an example, BWV 347, is shown at Figure 17 – and train the autoencoder accordingly[24]. Let us compute the mean of latent vectors of all Bach chorales soprano voices examples. Then, as illustrated in Figure 18:

- let us consider the Celtic melody "The Green Mountain" (shown in Figure 4);
- compute its latent vector;
- add the "Bach chorales" attribute latent vector (computed as the mean of latent vectors of all Bach chorales melodies examples); and
- create the corresponding melody, shown in Figure 19.

We can see that the original melody has been simplified, with longer notes, such as in the Bach chorales melodies corpus, while keeping the basic melodic motif. Figure 20 shows[25] the position of this *chimera* melody regarding the Celtic and Bach corpus. We can see that the original

[24]The training set of examples is the union of Celtic melodies and Bach soprano melodies. We transpose Bach soprano melodies into the D Major key in order to be aligned with the Celtic corpus. Also, as some Bach melodies are outside of the Celtic pitch range, the pitch range must be adjusted and as a consequence the one-hot vector size (and the autoencoder input layer size) is adjusted.

[25]It uses the T-distributed Stochastic Neighbor Embedding (t-SNE) nonlinear dimensionality reduction machine learning algorithm for visualization [28]. t-SNE models each high-dimensional object by a two (or three) dimensional point in such a way that similar objects are modeled by nearby points and dissimilar objects are modeled by distant points. The

Figure 19: *"The Green Mountain" transformed into a Bach chorales-like melody by the autoencoder (h = 1500)*

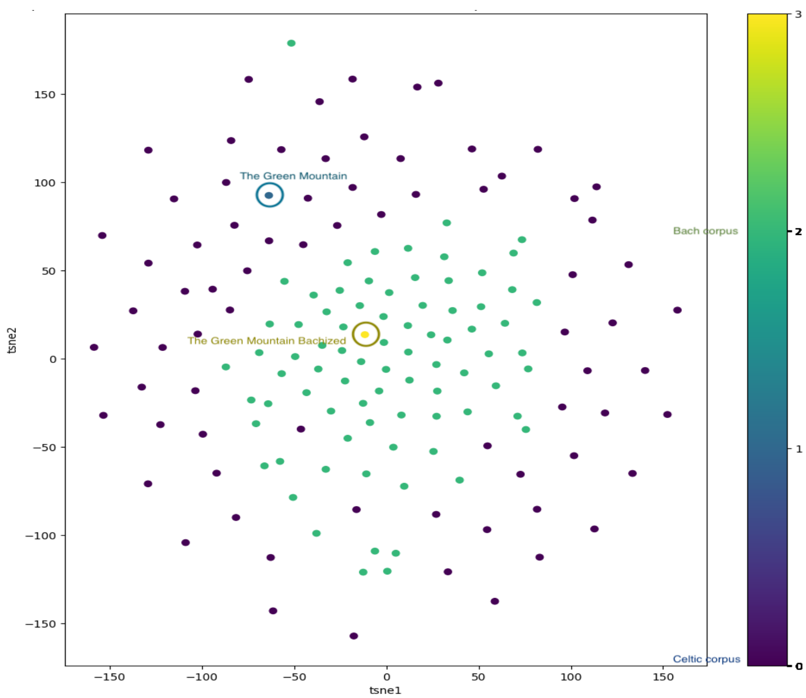


Figure 20: *TSNe representation of the following melodies: Celtic corpus (dark purple blue spots) – including "The Green Mountain" (circled blue spot) –; Bach corpus (green spots); and "The Green Mountain Bach-ized" (circled yellow spot) generated by the autoencoder (h = 1500)*

"The Green Mountain" melody (circled blue spot) is on the left, within the main part of the Celtic corpus (dark purple blue spots), while its "Bach-ized" version (circled yellow spot) is right at the center of the Bach corpus (green spots).

We can also do the other way around, by selecting one of Bach chorales soprano voice, e.g., BWV 347; compute the mean of Celtic melodies latent vectors; add it to BWV 347 latent vector; and obtain a melody, shown in Figure 21. The result of the transformation is much less obvious than for previous case. But, by looking at Figure 22, we can see that the original BWV 347 melody (circled blue spot) is actually already at the center of the Celtic corpus (dark blue spots), thus minimizing the move to its "Celtic-ized" version (circled yellow spot), which results in very little musical change.

difference with an autoencoder is that t-SNE does not try to minimize a reconstruction error but instead tries to preserve the neighborhood distances.

Figure 21: *Bach Chorale BWV 347 soprano voice transformed into a Celtic-like melody by the autoencoder (h = 1500)*

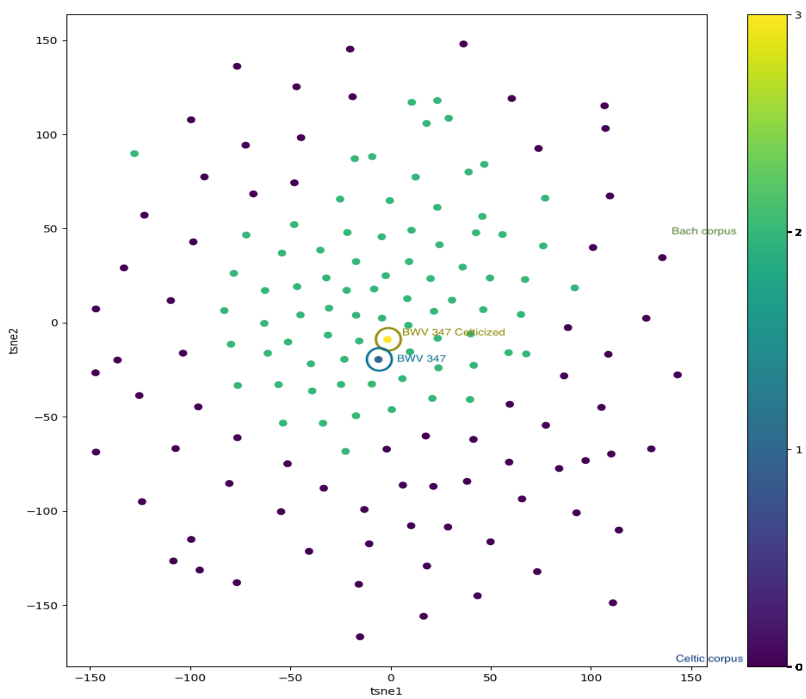


Figure 22: *TSNe representation of the following melodies: Celtic corpus (dark purple blue spots); Bach corpus (green spots) – including "BWV 347" (circled blue spot) –; and "BWV 347 Celtic-ized" (circled yellow spot) generated by the autoencoder (h = 1500)*

iv. Recursion

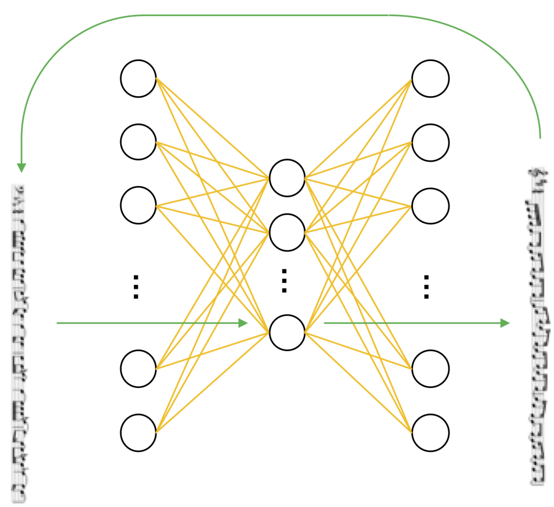


Figure 23: *Generation of a melody by the autoencoder by recursively feedforwarding into the autoencoder an initial melody data (random or given) until reaching a fixed point*

In [13], Kazakçi *et al.* proposed an original way of content generation from autoencoders. The idea is to feedforward a random initial content (random melody representation) into the

Figure 24: *(from top to bottom) Example of successive melodies generated by the autoencoder (h = 1500) for each step of the recursion*

autoencoder and *recursively feedforward* the generated output as the input until the output gets to a *fixed point*. They have experimented with a dataset of handwritten numerical digits (the MNIST dataset for handwritten digits recognition) and generated new types of visual patterns that they name “digits that are not”. We have applied their approach on our autoencoder trained on Celtic melodies, as illustrated in Figure 23, with Figure 24 showing an example of progressive refinement of a melody. Note that each melody generated corresponds to some attractor of the network and their number is finite[26].

[26]This is some kind of *mode collapse*, as for generation by generative adversarial networks (GAN) [16].

v. Objective Optimization

Another approach is by controlling the exploration of the latent space (the input of the decoder) by *optimizing some property*. This idea has been initially proposed in 1988 by Lewis and named by him *creation by refinement* [18]. The idea is to "revert" the standard way of using gradient descent for standard task – adjust the connexion *weights* in order to *minimize* the *classification error* –, into a very different task – adjust the *input* in order to *maximize* an expected *classification result*[27]. This approach can be seen as the precursor of various approaches for controlling the creation of a content by maximizing some target property[28], such as Deep Dream [20] and style transfer [7], see more details in [1].

We apply this approach to optimize the value of the latent vector of the autoencoder to match some objective[29]. In practice, a *vector of random values* is created, as initial values of the latent vector. Then, an *optimization* algorithm[30] is applied on the latent vector to *maximize* the *objective*, as illustrated in Figure 25. We experimented with three different objectives for the melody:

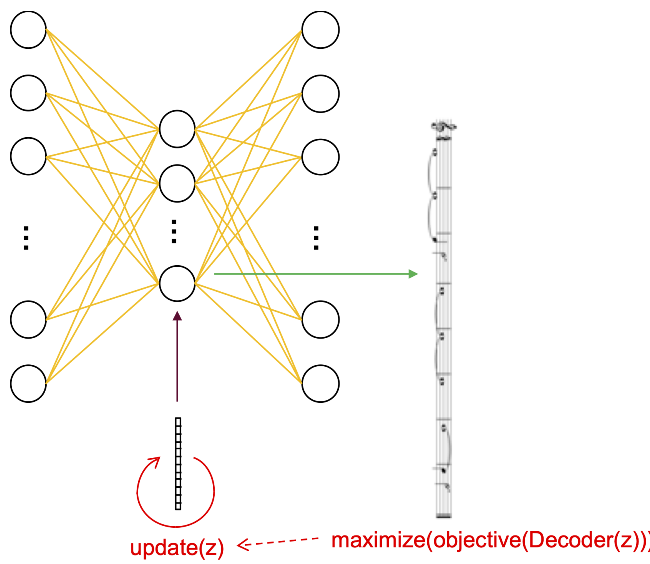


Figure 25: *Generation of a melody by the decoder component of the autoencoder by updating a latent vector (initially random) in order to maximize some objective*

- first note to be a C_4[31], shown in Figure 26;

[27]In Lewis' initial proposal, the neural network which is a feedforward binary classifier is at first trained with positive and negative examples of what he names "well formed" melodies, defined as follows: 1) using only the unison, 3rd and 5th intervals between notes and 2) following some scale degree stepwise motion. Then, a vector of random values is used as the initial input of the network and refined as to obtain a positive classification (i.e. a well formed melody). See more details in [18].

[28]The target property may be of any kind as long as it may be measured and thus optimized.

[29]Sun [26] may have been the first author to propose this approach for autoencoders. In his experiments, the target property is to generate a melody consonant to an existant melody. Note that he used *stacked autoencoders*, i.e. nested autoencoders with a decreasing number of hidden layers.

[30]Gradient-based or even simple random generate-and-test.

[31]Note that the objective is not completely fulfilled. The first note of the generated melody is a D_4. This makes sense because the corpus of melodies is in the key of D Major and many of them start with a D. This is important to remember that we can optimize some objective but within the boundaries of the representation that the autoencoder has learnt. Opening up this structural restriction is possible but with another generation (and architectural) model, see, e.g., structure

- maximize the number of hold (i.e. having notes as long as possible), shown in Figure 27; and
- minimize the number of hold (i.e. having notes as short as possible), shown in Figure 28.

Figure 26: *Example of melody generated by the autoencoder (h = 1500) with the objective of its first note being a* C_4

Figure 27: *Example of melody generated by the autoencoder (h = 1500) with the objective of maximizing the number of hold*

Figure 28: *Example of melody generated by the autoencoder (h = 1500) with the objective of minimizing the number of hold*

VI. Further Developments

i. Variational Autoencoder

As noted in Section V.ii, although producing interesting results, the autoencoder suffers from some discontinuity limitation. We could see that the interpolation, although correct (it maps the start and the target), is not continuous and thus creates discontinuities in the generation when exploring the latent space. The reason, as discussed in [25], is that the autoencoder is solely trained to encode and decode with a minimal loss, no matter how the latent space is organized. The approach is then to regulate the latent space to ensure that the latent space has better continuity properties for the generation.

A *variational autoencoder* (VAE) [15] is a refinement of an autoencoder with the added constraint that the encoded representation, i.e. the latent variables, follows some *prior probability distribution*[32], usually a Gaussian distribution. This regularization ensures two main properties: *continuity* (two close points in the latent space should not give two completely different contents once decoded) and *completeness* (for a chosen distribution, a point sampled from the latent space should provide a "meaningful" content once decoded) [25]. The price to pay is some larger reconstruction error, but the tradeoff between reconstruction and regularity can be adjusted depending on the priorities (as we will see in Section VI. iii).

imposition with a restricted Boltzmann machine (RBM) [17].

[32]This constraint is implemented by adding a specific term to the cost function which computes the cross-entropy between the distribution of latent variables and the prior distribution. The model and implementation is actually more sophisticated, instead of an encoder encoding an input as a single point, a variational autoencoder encodes it as a *distribution* over the latent space, from which the latent variables are sampled, as explained in [25].

	autoencoder	variational autoencoder
interpolation between existing melodies	*Figure 13* good reconstruction but some discontinuity	*Figure 29* better continuity but imperfect reconstruction
interpolation between min and max	*Figures 15 and 16* constant (no interpolation)	*Figures 30 and 31* almost continuous interpolation
vector arithmetics	*Figure 19* convincing result	*Figure 32* ok result
recursion	*Figure 24* convincing result	*Figure 33* convincing result
objective optimization	*Figures 26, 27 and 28* convincing results	*Figures 34, 35 and 36* ok objectives but bad style

Table 1: *Comparison of generation approaches with autoencoder and variational autoencoder*

ii. Generation

Figure 29 shows the melodies generated by the variational autoencoder with $h = 1500$ by interpolation between "The Green Mountain" and "Willa Fjord". By comparing it with the generation by the autoencoder (Figure 13), we could see that the interpolation is more continuous, at the price of some imperfect reconstruction of the two original melodies.

In the case of $h = 2$, we may compare the melodies generated by the variational autoencoder for the interpolation of values of z_1 and of z_2[33] (shown in Figures 30 and 31, respectively), to the melodies generated by the autoencoder (Figures 15 and 16).

Let us now compare in Table 1 the results of the various approaches for generation (interpolation, vector arithmetics, recursion and objective optimization) by a variational autoencoder to the case of an autoencoder. These simple experiments suggest that a variational autoencoder may not necessarily lead to an improvement in the quality of the generation, depending on the generation approach[34].

iii. Interpretation and Disentanglement

Another issue is that the characteristics (meaning, e.g., note duration range, note pitch range, motif, etc.) of the dimensions captured by the latent variables are automatically "chosen" by the autoencoder architecture (variational or not), in function of the training examples and the configuration. Thus, they can only be interpreted *a posteriori*. For instance, in Figures 30 and 31, we can observe that z_1 seems to capture the range as well as the average of note durations, while z_2 captures refinements of the melody motif. This actually heavily depends on the set of training examples and the way they vary[35].

Furthermore, as for the mapping between a genotype and a phenotype, there may be a many-to-many mapping between latent variables and characteristics. In fact, the dimensions captured by the latent variables are not independent (orthogonal), as in the case of Principal component

[33] We use a straightforward linear interpolation of z_1 or of z_2. However, decoding a straight line in the latent space does not necessarily produce melodies whose attributes vary uniformly. See, e.g., [9] for a discussion and a proposed solution.

[34] As pointed out in Section VI.i, there is a tradeoff between continuity and reconstruction. Also, as pointed out in [8, Section 20.10.3], there are still some troubling issues about variational autoencoders.

[35] For instance, when using Bach chorale melodies, the result is different: z_1 captures mostly the range of note durations, while z_2 captures the pitch range.

Figure 29: *(from top to bottom) Melodies resulting from the interpolation (5 steps) by the variational autoencoder (h = 1500), from "The Green Mountain" to "Willa Fjord"*

Figure 30: *(from top to bottom) Melodies resulting from the interpolation (5 steps) by the variational autoencoder (h = 2) of the value of z_1 (from its min value to its max value), while z_2 is constantly equal to its mean value*

Figure 31: *(from top to bottom) Melodies resulting from the interpolation (5 steps) by the variational autoencoder (h = 2) of the value of z_2 (from its min value to its max value), while z_1 is constantly equal to its mean value*

Figure 32: *"The Green Mountain" transformed into a Bach chorales-like melody by the variational autoencoder (h = 1500)*

Figure 33: *Example of melody generated by the variational autoencoder (h = 1500) by recursion*

Figure 34: *Example of melody generated by the variational autoencoder (h = 1500) with the objective of its first note being a* C_4

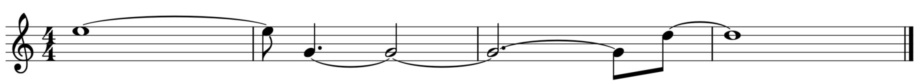


Figure 35: *Example of melody generated by the variational autoencoder (h = 1500) with the objective of maximizing the number of hold*

Figure 36: *Example of melody generated by the variational autoencoder (h = 1500) with the objective of minimizing the number of hold*

analysis (PCA). However, various techniques[36] have being recently proposed to improve the *disentanglement* of the dimensions (see, e.g., [19]). Some recent approaches also propose to "force" the meaning of latent variables, by splitting the decoder into various components and training them onto a specific dimension (e.g., rhythm or pitch melody) [29].

iv. RNN Encoder-Decoder and Variational Recurrent Autoencoder (VRAE)

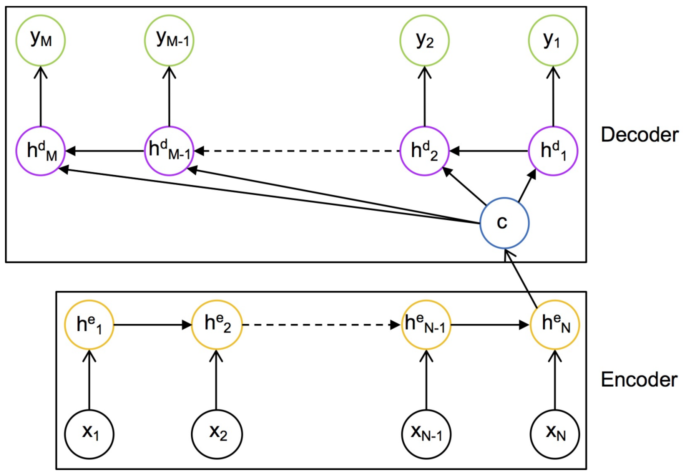


Figure 37: *RNN Encoder-Decoder architecture. Inspired from [3]. The hidden layer h^e_t of the RNN encoder acts as a memory which iteratively accumulates information about the successive elements x_t of an input sequence read by the RNN encoder; resulting in a final state h^e_N; which is passed to the RNN decoder as the summary c of the whole input sequence; which then iteratively generates the output sequence by predicting the next item y_t given its hidden state h^d_t and the summary c as a conditioning additional input*

A practical limitation of an autoencoder is that the size of the input (and output) layer is fixed and as a result also the length of the music generated. The solution is to combine: the *generative* property of the autoencoder with the *variable length* property of a recurrent neural network (RNN) architecture (see [2, Section 6.5]). The idea is to embed a recurrent network (RNN) within the encoder and a similar RNN within the decoder (thus, named an *RNN Encoder-Decoder* [3]), as shown in Figure 37.

A natural further step is to combine this with the variational characteristic of a variational autoencoder, resulting in what is named a *variational recurrent autoencoder* (VRAE) [5]. We will not further detail VRAE architectures here because of space limitation. Please see, e.g., the MusicVAE architecture and details on generation experiments presented in [24] and, e.g., [2] for some comparative analysis of various architectures.

VII. Conclusion

The use of artificial neural networks and deep learning architectures and techniques for the generation of music (as well as other artistic contents) is a very active area of research. In this paper, we have introduced and illustrated the use of autoencoders to generate music. Various

[36] Two examples of approaches are: 1) increasing the weight of the prior distribution conformance (the β-VAE approach) [11]; 2) ensuring that for a given dimension no other dimension will be present by using a classifier to check the equiprobability among the classes along other dimensions (the antagonist approach) [23].

approaches, simple conceptually and to implement, have also been discussed to control the generation. Some lesson from these simple experiments shows that a variational autoencoder, although providing some important improvements on the continuity of the latent space, also suffers from some reconstruction imperfection. Therefore, depending on the generation approach and the priorities, one may consider better using a simple autoencoder. We hope that this article will help at showing the potential of using autoencoders for music generation. MIDI files of examples may be found at: http://www-desir.lip6.fr/~briot/dlt4mg/Papers/compress-to-create-midi/.

Acknowledgements

We would like to thank: Liduino Pitombeira for his invitation to write this article, Matthieu Bizien for his constructive comments, and all participants to our course on deep-learning-based music generation at UNIRIO in 2018–2019 for their interest and feedback.

References

[1] Briot, J.-P. (2020). From artificial neural networks to deep learning for music generation – History, concepts and trends, April 2020. Available in https://arxiv.org/abs/2004.03586.

[2] Briot, J.-P.; Hadjeres, G.; and Pachet, F.-D. (2019). *Deep Learning Techniques for Music Generation*: Computational Synthesis and Creative Systems. Cham, Switzerland : Springer International Publishing.

[3] Cho, K.; van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; and Bengio, Y. (2014). Learning phrase representations using RNN Encoder-Decoder for statistical machine translation, September 2014. Available in https://arxiv.org/abs/1406.1078.

[4] Cope, D. (2000). *The Algorithmic Composer*. Madison, Wis.: A-R Editions, 2000.

[5] Fabius, O. and van Amersfoort, J. R. (2015). Variational Recurrent Auto-Encoders, June 2015. Available in https://arxiv.org/abs/1412.6581.

[6] Fernández, J. D. and Vico, F. (2013). AI methods in algorithmic composition: A comprehensive survey. *Journal of Artificial Intelligence Research (JAIR)*, 48:513–582, 2013.

[7] Gatys, L. A.; Ecker, A. S.; and Bethge, M. (2015). A neural algorithm of artistic style, September 2015. Available in https://arxiv.org/abs/1508.06576.

[8] Goodfellow, I.; Bengio, Y.; and Courville, A. (2016). *Deep Learning:* Adaptive Computation and Machine Learning. Cambridge, Mass.: MIT Press, 2016.

[9] Hadjeres, G.; Nielsen, F.; and Pachet, F. (2017). GLSR-VAE: Geodesic latent space regularization for variational autoencoder architectures, July 2017. Available in https://arxiv.org/abs/1707.04588.

[10] Hadjeres, G.; Pachet, F.; and Nielsen, F. (2017). DeepBach: a steerable model for Bach chorales generation, June 2017. Available in https://arxiv.org/abs/1612.01010.

[11] Higgins, I.; Matthey, L.; Pal, A., Burgess; C., Glorot, X.; Botvinick, X.; Mohamed, S.; and Lerchner, A. (2017). beta-VAE: Learning basic visual concepts with a constrained variational framework. In *Proceedings of 5th International Conference on Learning Representations (ICLR 2017)*, Toulon, France, April 2017.

[12] Hiller, L. A. and Isaacson, L. M. (1959). *Experimental Music:* Composition with an Electronic Computer. New York: McGraw-Hill.

[13] Kazakçi, A.; Cherti, M.; and Kégl, B. (2016). Digits that are not: Generating new types through deep neural nets. In: François Pachet, Amilcar Cardoso, Vincent Corruble, and Fiammetta Ghedini, editors, *Proceedings of the 7th International Conference on Computational Creativity (ICCC 2016)*,188–195. Paris, France, June 2016. Available in https://arxiv.org/abs/1606.04345.

[14] Keith, J. (2016) The Session, Accessed on 21/12/2016. Available in https://thesession.org.

[15] Kingma, D. P. and Welling, M. (2014) Auto-encoding variational Bayes, May 2014. Available in https://arxiv.org/abs/1312.6114.

[16] Kodali, K.; Abernethy, J.; Hays, J.; and Kira, Z. (2017). On convergence and stability of GANs, December 2017. Available in https://arxiv.org/abs/1705.07215.

[17] Lattner, S.; Grachten, M.; and Widmer, G. (2016). Imposing higher-level structure in polyphonic music generation using convolutional restricted Boltzmann machines and constraints, December 2016. Available in https://arxiv.org/abs/1612.04742.

[18] Lewis, J. P. (1988). Creation by refinement: A creativity paradigm for gradient descent learning networks. In: *IEEE International Conference on Neural Networks*, II, 229–233, San Diego, CA, USA, July 1988.

[19] Mathieu, E.; Rainforth, T.; Siddharth, N.; and Teh, Y. W. (2019). Disentangling disentanglement in variational autoencoders, June 2019. Available in https://arxiv.org/abs/1812.02833.

[20] Mordvintsev, A.; Olah, C.; and Tyka, M. (2015). Inceptionism: Going Deeper into Neural Networks, Google AI Blog, June 2015. Available in https://research.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html.

[21] Nierhaus, G. (2009). *Algorithmic Composition:* Paradigms of Automated Music Generation. Wien: Springer.

[22] Papadopoulos, G. and Wiggins, G. (1999). AI methods for algorithmic composition: A survey, a critical view and future prospects. In *AISB 1999 Symposium on Musical Creativity*, April 1999, 110–117.

[23] Robert, T.; Thome, N.; and Cord, M. (2019). DualDis: Dual-branch disentangling with adversarial learning, June 2019. Available in https://arxiv.org/abs/1906.00804.

[24] Roberts, A.; Engel, J.; Raffel, C.; Hawthorne, C.; and Eck, D. (2018). A hierarchical latent vector model for learning long-term structure in music. In: *Proceedings of the 35th International Conference on Machine Learning (ICML 2018)*, 80, 4364–4373. ACM, Montréal, PQ, Canada, July 2018.

[25] Rocca, J. (2019). Understanding variational autoencoders (VAEs) – Building, step by step, the reasoning that leads to VAEs, September 2019, Accessed 23/09/2019. Available in https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73.

[26] Sun, F. (2017). DeepHear – Composing and harmonizing music with neural networks, Accessed on 21/12/2017. Available in https://fephsun.github.io/2015/09/01/neural-music.html.

[27] Todd, P. M.(1989) A connectionist approach to algorithmic composition. *Computer Music Journal (CMJ)*, 13/4, 27–43.

[28] Maaten, L.van der; Hinton, G. H.(2008). Visualizing data using t-SNE. *Journal of Machine Learning Research (JMLR)*, 9, 2579–2605.

[29] Yang, R.; Wang, D.; Wang, Z.; Chen, T.; Jiang, J.; and Xia, G. (2019) Deep music analogy via latent representation disentanglement, October 2019. Available in https://arxiv.org/abs/1906.03626.

Compositional Systems: Overview and Applications

Liduino Pitombeira
Federal University of Rio de Janeiro (UFRJ)
pitombeira@musica.ufrj.br
Orcid: 0000-0002-1141-655X

Abstract: *In this paper the theory of compositional systems is described in detail, taking as a starting point the theoretical framework inherent to systems science. The origins of this science and the definitions of its fundamental concepts are provided in the first part of the article, illustrated with musical examples. The central part of the article contains the definition of the concept of compositional system, its typology, and a series of tools that are useful for implementations. Finally, the design of three types of systems (open, semi-open and feedback) are carried out in order to produce small illustrative musical fragments.*

Keywords*: Compositional Systems. Systems Science. Systemic Modeling. Probability.*

I. Introduction

Systems are ubiquitous in various human activities. However, despite the common-sense familiarity with systems in our everyday life, its concept is far from trivial and the attempts to define it has given rise to a large amount of studies, trends, and quarrels[1]. As pointed out by Robert Rosen [35], "the word system is never used by itself; it is generally accompanied by an adjective or other modifier: physical system; social system", etc. In this paper I am concerned with a special type of system: the compositional system. I start with a brief survey on systems theory, which includes historical aspects and definitions. This will lay the theoretical framework so that the concept of compositional system can be introduced and musical implementations can be performed.

II. A brief survey on systems theory

In this section, I will introduce some historical marks and basic concepts associated with the notion of system, in the general sense, and lay the foundations for the next section, which deals with a particular type of system: the compositional system. For the sake of clarity, I will demonstrate some systems concepts with musical examples, although the literature on systems focus especially on highly complex structures, such as living organisms, society, or even reality. This section will cover the motivations for the development of systems science as well as some of its historical marks (**??**) and definitions of the concept of system (i).

The roots of systems science are mathematics, computer technology, and a group of ideas known as systems thinking [16, p.19]. It emerged from the necessity of dealing with organized

[1] Lars Skyttner [38] has organized a survey on several trends in the field of systems science that includes Klir, Boulding, Laszlo, and many others.

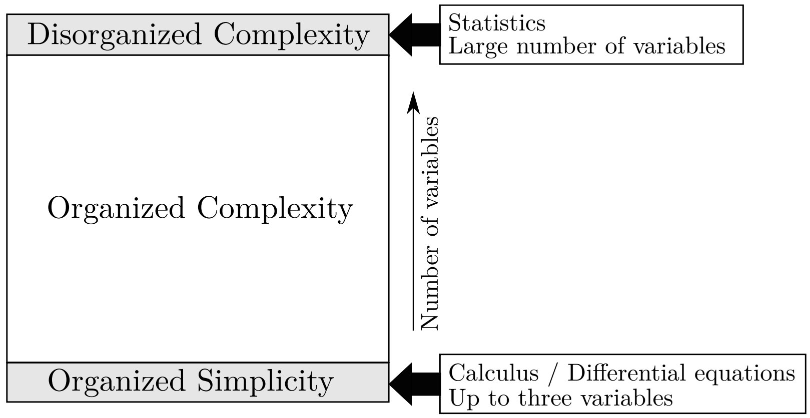


Figure 1: *Organized complexity between its extremes: organized simplicity and disorganized complexity.*

complexity, a category in which is located the vast majority of human problems, including music analysis and composition. In the extremities of this category are, on one side, organized simplicity, which consists of deterministic problems, involving up to four variables, that can be handled, for example, by calculus and differential equations, and, on the other side, disorganized complexity, which involves the use of probability and statistics to deal with an astronomical number of variables (Figure 1). As pointed out by Weaver, in the first half of the 20th century,

> One is tempted to oversimplify, and say that scientific methodology went from one extreme to the other—from two variables to an astronomical number—and left untouched a great middle region. The importance of this middle region, moreover, does not depend primarily on the fact that the number of variables involved is moderate—large compared to two, but small compared to the number of atoms in a pinch of salt. The problems in this middle region, in fact, will often involve a considerable number of variables. The really important characteristic of the problems of this middle region, which science has as yet little explored or conquered, lies in the fact that these problems, as contrasted with the disorganized situations with which statistics can cope, show the essential feature of organization. In fact, one can refer to this group of problems as those of organized complexity [42, p.539].

The limitation of dealing with a great number of variables is a cognitive feature inherently human. According to Halford et al [13, p.70] "a structure defined on four variables is at the limit of human processing capacity".[2] Moreover, structures with five variables are already at the chance level. A computer can make it easier to investigate complex systems involving a large number of variables. Therefore, the computer became a fundamental tool for investigating systems in the realm of organized complexity and its evolution has had a clear impact on systems science.

Besides mathematics and computer technology, a body of ideas connected with systems thinking were crucial to the emergence of systems science. Those ideas include holism, isomorphism, general systems, and cybernetics. Holism, an antithesis of reductionism[3], already known to Greek and Chinese philosophy, reappeared at the beginning of the 20th century in a branch of psychology known as Gestalt theory. It became present also in the field of biology, around the same time, in the *organismic biology* proposed by Paul Weiss and Ludwig von Bertalanffy [3]. Phillips [29, p.6-7]

[2] In music analysis and composition variables may be associated with attributes or parameters of a musical sound: pitch (pitch class and register), rhythm (attack point and duration), dynamics, articulation, and timbre. I have proposed the expansion of parametric repertory by introducing the concept of abstract parameter that includes inversional axis, rhythmic partition, degree of harmonic endogeny [32], melodic contour, and so on [33]. Those abstract parameters (except contour) are not easily detected by perception in the superficial level.

[3] Linked with the analytical method, i.e., a piecemeal approach in which an object is divided into its simple constituent elements.

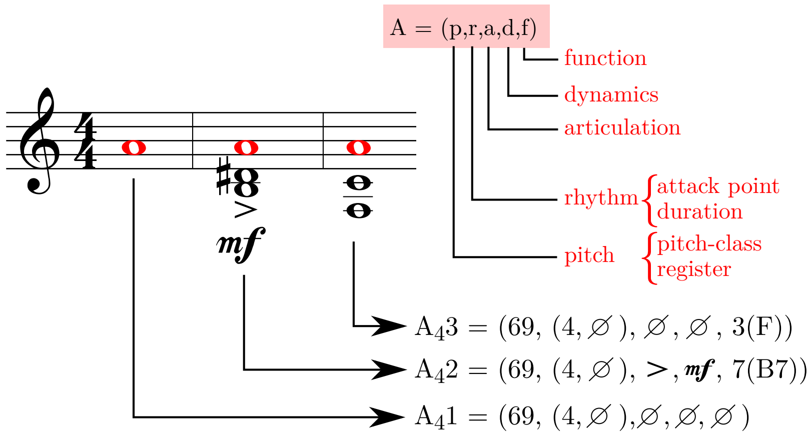


Figure 2: *Three situations in which pitch A4 appears: as an isolated pitch, as part of a B7 chord, and as part of an F7 chord.*

connects holism directly with Hegel's principle of internal relations and, thus, with 19th century organicism. He enumerates its five characteristics[4] :

> 1. The analytic approach as typified by the physicochemical sciences proves inadequate when applied to certain cases—for example, to a biological organism, to society, or even to reality as a whole.
> 2. The whole is more than the sum of its parts.
> 3. The whole determines the nature of its parts.
> 4. The parts cannot be understood if considered in isolation from the whole.
> 5. The parts are dynamically interrelated or interdependent.

The principle of internal relations appears as well in the writing of neo-idealists like Francis Bradley (1846-1924), who enunciates three key points: 1) Relations between entities are possible only inside a whole; 2) Those related entities are altered by the relationships; and 3) Those entities qualify the whole, which, in its turn, qualify them [29, p.8].

The second point is "the heart of the theory of internal relations" [29, p.8] and will be exemplified here by a musical note in three out-of-time contexts[5]. In the first situation, pitch A_4 is isolated from any system. In order to categorize it, let us assign to it the following parameters: pitch (pitch class and register)[6], rhythm (duration and attack point), articulation, dynamics, and function, (p,r,a,d,f). Therefore, this isolated A_4, which will be labeled A_41, can be represented by the expression $A_41 = (69, (4, \varnothing), \varnothing, \varnothing, \varnothing)$, meaning that it has information on pitch (MIDI number 69) and duration (4 quarter-notes), but no information on attack point, articulation and dynamics. Also, its isolated state deprives it of a chordal function. This is shown in Figure 2. In the second situation, the same A_4 plays the role (or has the function) of the seventh of a B7 chord and will be assigned articulation and dynamic values. It is ontologically the same A_4 but it is a different entity in the context of the system[7] formed by the other pitches encapsulated to form a more complex entity: the B7 chord. Its representation, therefore, will be $A_42 = (69, (4, \varnothing), >, mf, \varnothing)$. Figure 2

[4]Throughout his book, Denis Phillips, an adversary of inflexible holism, examines the validity of each one of those characteristics. He classifies Holisms in three types (p.36): I, II, and III. Holism I is the one firmly attached to the five aforementioned characteristics of organicism. Holism II and Holism III gradually accepts some kind of compromise with analytical methods.

[5]Contexts in which the attack (or time) points are not defined and therefore temporal order is disregarded.

[6]Or MIDI number (C_4 = 60)

[7]As it will be seen later in this paper, a system consists of interrelated objects.

shows yet another chord to which the A_4 belongs and functions as its third: F7. For Bradley, A_4 is a part of the B7 chord (which is a system), but when removed from it becomes simply an artifact, something like a dead and functionless entity.

Additionally, Phillips describes connections between systems theory and the philosophy of John Dewey (1859-1952)[8] , whose influence on musicologist Leonard Meyer (1918-2007) is already acknowledged in the literature [34][9]. Phillips also identifies association between systems theory and structuralism, especially in the works of Levy-Strauss (1908-2009)[10] and Jean Piaget (1896-1980)[11]. Structuralism has a methodological impact in the field of music analysis by introducing the synchronic perspective (as a complement to the diachronic one). As Piaget said, "structuralism is chiefly a departure from the diachronic study of isolated linguistic phenomena which prevailed in the nineteenth century and a turn to the investigation of synchronously functioning unified language systems" [30, p.4]. Lévy-Strauss adds:

> Hence the hypothesis: What is patterns showing affinity, instead of being considered in succession, were to be treated as one complex pattern and read as a whole? by getting at what we call harmony, they would then see that an orchestra score, to be meaningful, must be read diachronically along one axis—that is, page after page, and from left to right—and synchronically along the other axis, all the notes written vertically making up one grs constituent unit, that is, one bundle of relations. [20, p.212]

Another issue related to systems thinking is isomorphism (or analogies). In its etymology, isomorphism means simply a similarity of form. It is a concept that permeates several fields of knowledge such as biology, chemistry, sociology, and mathematics, for which it has a particular meaning: bijective correspondence and structure-preserving mappings[12]. In the field of music theory and analysis, for example, isomorphism is one of the formal supports for pitch class set theory, since pitch class space is isomorphic to the Abelian group $(Z, +)$ and, therefore, may inherit the algebraic structure related to groups.

Besides these internal aspects of isomorphism, systems science considers also a larger perspective on isomorphism, that is the connections among different areas. Klir [16, p.32] brings the example of generalized circuit, "a framework within which well-developed methods for analyzing electric circuits were transferred through established isomorphies to less advanced areas of mechanical, acoustic, magnetic, and thermal systems."

Consequently, isomorphism had as a natural result the increasing of interdisciplinarity, which, by its turn, led to the development of the concept of *general systems*, by Ludwig von Bertalanffy (1968), who was aligned with Kenneth Boulding, Ralph Gerard, and Anatol Rapoport. Klir [16, p.16] defines a general system as "a standard and interpretation-free[13] system chosen to represent a particular equivalence class of isomorphic systems". The theory of general systems is extensively discussed in Bertalanffy's book *General Systems Theory* [4].

Besides holism, isomorphism, and general systems, cybernetics was also a key factor for the development of systems science. According to Klir [16, p.37], "cybernetics is a subarea of general systems research that focuses on the study of information processes in systems, particularly

[8] Dewey and Bentley [8, p.509] mention that the world is historically presented to humans in three levels: 1) Self-action: isolated things; Inter-action: things in causal relationships; 3) things and relations forming an unbreakable whole. This last level is clearly related to the idea of system, as understood in a holistic fashion.

[9] Many references to the term system and style-system can be found in Meyer's Emotion and Meaning in Music (1957) [26].

[10] As pointed out by Wilcken [44, p.140], "as its core, structural linguistics worked with a simple, yet revolutionary idea: the notion that language consisted of a formal system of interrelated elements, and that meaning resided not in the elements themselves, but in their relationships to one another."

[11] For Piaget [30, p.5],"a structure is a system of transformations".

[12] An isomorphic relation is equivalent, i.e., it is reflexive, symmetric, and transitive.

[13] Represented by integers or real numbers.

communication and control". It was created by Norbert Wiener, who defines it as "control and communication in the animal and in the machine"[14]. Cybernetics was highly benefited by the theory of information, developed by Claude Shannon and Warren Weaver. As pointed out by Eco [9], communication is an essential factor for cultural phenomena as well as for many scientific fields: Psychology, Genetics, Neurophysiology, etc.

i. Definitions of system

Probably, the earliest (explicit) definition of system comes from the French philosopher Étienne Bonnot de Condillac (1715-1780): "a system is nothing other than the arrangement of the different parts of an art or science in an order in which they all support each other, and where the latter are explained by the former. Those who give reason to others, are called principles; and the system is all the more perfect, as the principles are fewer: it is even to be hoped that they will be reduced to one"[15]. In his definition it is possible to clearly identify two aspects: parts and principles.

For Bertalanffy [4, p.55-56] "a system can be defined as a complex of interacting elements. Interaction means that elements, p, stand in relations, R, so that the behavior of an element p in R is different from its behavior in another relation, R'. If the behaviors in R and R' are not different, there is no interaction, and the elements behave independently with respect to the relations R and R'". Similarly to Condillac's definition, two aspects are also identified here: elements and relations. Bertalanffy went further and specified how should be the interaction of elements within a system. In order to illustrate his definition, consider the set of pitch classes $J = \{B, C\flat, D\sharp, E\flat, F\sharp, G\flat, A\}$ in which two quaternary relations [16] with one element each are defined: $R = \{(B, D\sharp, F\sharp, A)\}$ and $R' = \{(C\flat, E\flat, G\flat, A)\}$.[17] The behavior of pitch class A (the element p in Bertalanffy's definition) in R is different from its behavior in R', since R progresses harmonically to an E chord (set K) and R' progresses harmonically to a $B\flat$ chord (set K'). This is shown in Figure 3.

George Klir [16, p.4-9], inspired, according to himself, by a standard dictionary definition, proposes that a system is "a set or arrangement of **things** so **related** or connected as to form a unity or organic **whole**". I have highlighted in this definition three keywords: **things**, **related**, and **whole**. The latter corresponds to the system itself: it is the whole that emerges from the interaction of things and relations, which, in their turn, constitute the system's components. Klir formalizes this definition in Equation 1, in which S stands for system, T for things, and R for relation.[18] It is noteworthy to verify that for Klir the relational component (R) seems to be the essence of a system, since he associates it with the very property of systemhood. In other words, a set of unrelated things becomes a system when (and only when) these things are connected through some kind of relation. Figure 4A shows a collection of things (T1, T2, and T3). This collection of things is understood as a system when one finds relationships among them. In Figure 4B, the relations are

[14] This definition is on the very title of his 1948 book: Cybernetics or control and communication in the animal and in the machine [43]. In his book he covers topics such as the concept of time in Newtonian and Bergsonian terms, statistical mechanics, Gestalt theory, and information, language and society.

[15] In the original one can read: «un système n'est autre chose que la disposition des différentes parties d'un art ou d'une science dans un ordre où elles se soutiennent toutes mutuellement, et où les dernières s'expliquent par les premières. Celles qui rendent raison des autres, s'appellent principes ; et le système est d'autant plus parfait, que les principes sont en plus petit nombre : il est même à souhaiter qu'on les réduise à un seul » [7, p.1].

[16] A relation may be presented by enumerating its elements in the form of sets, through a matrix, a graph, or even, when the case applies, by its analytical expression.

[17] In the equal temperament tuning system applied to western concert instruments, like the piano, these two chords (represented by relations R and R') sound the same.

[18] T can be any arbitrary set, including the power set, or even other systems. A n-ary relation R on sets $A_1, A_2, ..., A_n$ is formally defined as a subset of a Cartesian product $A_1 \times A_2 \times, ..., \times A_n$.

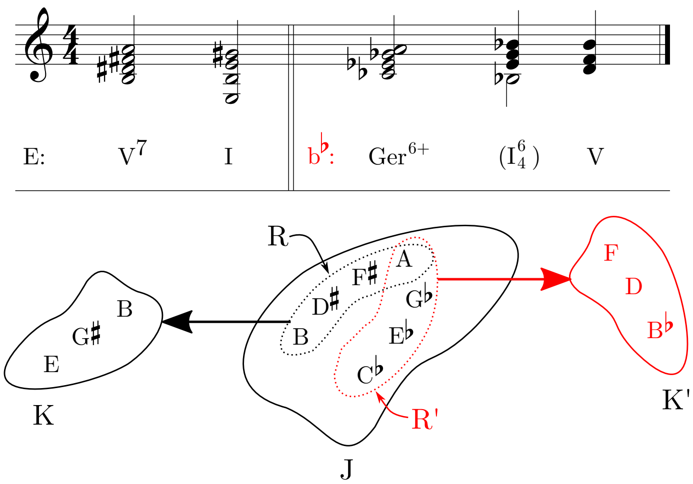


Figure 3: *pitch class A as part of a B7 chord (which resolves to an E chord) and as part of a German sixth chord in b♭ (which resolves to B♭). These chords correspond respectively to relation R and R′*

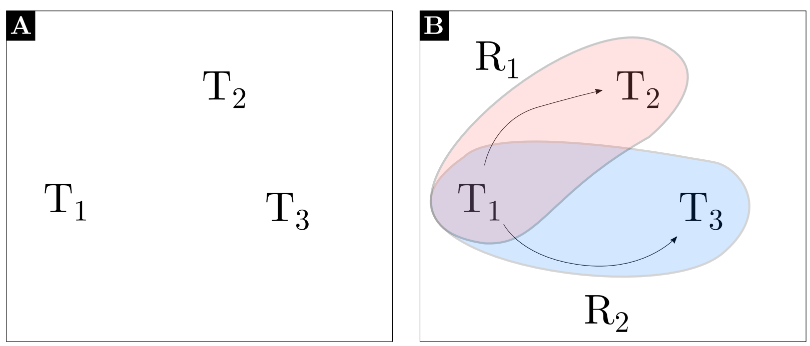


Figure 4: *In A, there is a set of things* $\{T1, T2, T3\}$ *and, in B, a system with three things* $\{T1, T2, T3\}$ *and two relations,* $R1 = \{(T1, T2)\}$ *and* $R2 = \{(T1, T3)\}$.

$R1 = \{(T1, T2)\}$ and $R2 = \{(T1, T3)\}$.

$$S = (T, R) \tag{1}$$

Thus, a collection of pitches in a musical score is not a system *per se*. Only when one identifies some kind of relationship among those pitches a system is cognitively established.[19] If fact, this identification is “the most fundamental act of system theory, the very act of defining the system presently of interest, of distinguishing it from its environment” [12, p.32]. Such identification is an individual task and depends upon the analytical repertory and also the particular choices of the observer. Therefore, different observers may define different systems from the same set of things.

The musical fragment shown in Figure 5 can be understood as different systems. A first analysis can understand the fragment as a melodic line in which the pitch content consists of a

[19] This is a key point in systems science: the musical score exists in the real world independently of our observation but our knowledge about it is only established through the epistemological attitude of making distinctions. This is, as mentioned by Klir [16, p.12], a constructivist perspective and could be traced as early as mid-17th century, in the works of Giambattista Vico (1668-1744). In the 20th century, the works of Jean Piaget (1896-1980), Ernst von Glasersfeld (1917 -2010), Humberto Maturana (1928-), and Francisco Varela (1946-2001) are connected with this epistemological view of the world.

Figure 5: *A musical fragment with intervallic profile (+5,-4,+6,-5,+4,+5,+5,-6,-2,-5,+5).*

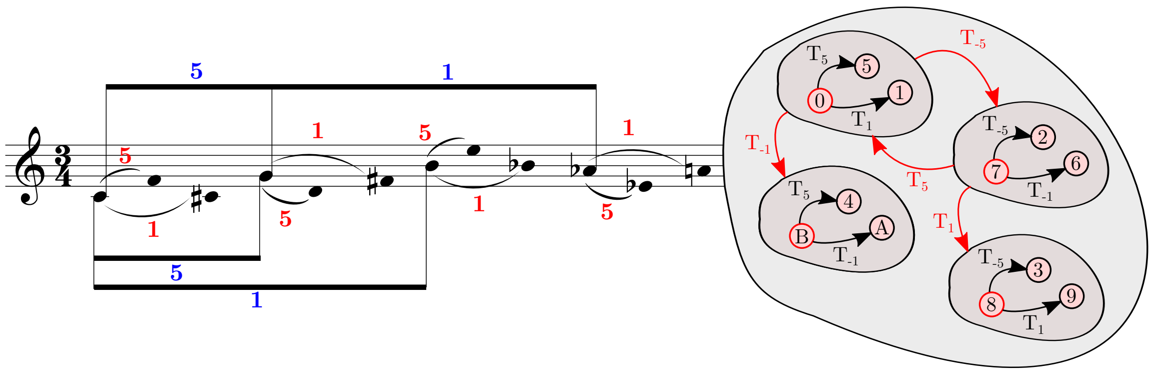


Figure 6: *The same musical fragment of Figure 5 understood now as a self-similar trichordal structure.*

twelve-tone row with intervallic profile $(+5, -4, +6, -5, +4, +5, +5, -6, -2, -5, +5)$.[20] Given the initial pitch, the entire collection can be reconstructed even though information about rhythm (duration and attack points), articulation, dynamics, and tempo is lost.

A second analysis could reveal inner relationships from which one perceives a self-similar structure. This structure is formed by trichords connected to each other through internal transpositional functions[21] that hold their elements together. Figure 6 shows that pitch class set $\{(0), 1, 5\}$[22] , with 0 (inside parenthesis) arbitrarily defined as the main pitch class, and transpositional functions T_5 and T_1 form a system; set $\{(7), 2, 6\}$ and transpositional functions T_{-5} and T_{-1} form another system, and so on. A closer look at the four systems reveals that the interval class[23] from their main pitch class to the other two pitch classes are always 5 and 1. This means that all four systems consist of set class 015. Furthermore, the main pitch classes of the four systems relate to each other through the same transpositional functions used internally. Therefore, they can be grouped together to form a larger system similar to each one of its subsystems.

A third analysis could disregard absolute pitch or pitch class values and consider only the melodic contour of segments. In Figure 7, the melodic line was segmented into four parts and to each part was assigned a contour. It is easy to verify that they can all be related to the initial

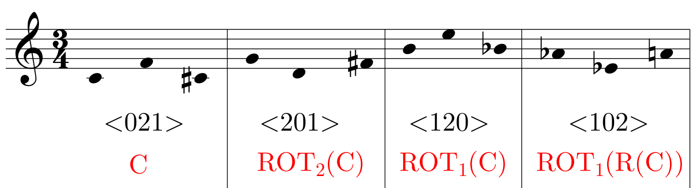
[20] The numbers inside parenthesis indicate the chromatic semitones between two pitches. Ascending and descending intervals are indicated with $+$ and $-$, respectively.

[21] According to Halmos [14, p.30], "if X and Y are sets, a function from (or on) X to (or into) Y is a relation f such that $dom f = X$ and such that for each x in X there is a unique element y in Y, with $(x, y) \in f$."

[22] In this paper, a pitch class set will be represented within braces, its normal form within parenthesis, and its prime form within brackets. A set class is represented unframed.

[23] The interval class (ic) is the smaller distance between two unordered pitch classes (a, b). Formally,

$$ic(a, b) = \begin{cases} 12 - |a - b| & \text{if } (a - b) > 6 \\ |a - b| & \text{if } b \leq 6. \end{cases} \tag{2}$$



C	<012>	R(C)	<210>
ROT_1(C)	<120>	ROT_1(R(C))	<102>
ROT_2(C)	<201>	ROT_2(R(C))	<021>

Figure 7: *The same musical fragment of Figure 3 understood now as chain of inter-related melodic contours.*

contour by rotation (*ROT*) and retrogradation (*R*).[24] Therefore, the second contour, <201> is the second rotation of the original contour, <021>, and so on.

Those three above systems were the result of examinations of a melodic line. From the observations I have proposed three different models.[25] However, a system can also be designed from scratch. In this paper, there is a differentiation between systemic designing and systemic modeling. The latter consists of proposing a model by capturing the architecture of a structure already in place with the purpose of understanding its rules of organization in such a way that it could be replicated or expanded; the former consists of defining a structure of objects and relations from the ground up.

Klir divides systems into five epistemological levels. The most basic level, the *source* system, consists of a set of variables (with no particular values) and potential states.[26] "When the source system is supplemented with data, i.e., with actual states of the basic variables within the defined support set" [17, p.13], the first level, the *data system*, is reached.[27]The data may come from modeling or designing. If information is given on the data generation (deterministic or stochastic), the system is on the *generative* level. So, at this level there are models capable of generating information. Higher levels include the *structure system*, which is a set of generative systems working cooperatively, and *metasystems*, which are systems that describe changes within structure systems (relations of relations).[28]

III. Compositional Systems

Approaches to musical composition through a systems science perspective, although rare, are already known in the scope of music research. Probably, the most comprehensive one is the extensive paper written in the 1990s by Romanian scholars Cosmin and Mario Georgescu [10], which brings strong and innovative associations between music and systems science. They identify several systemic features within a musical work: wholeness, hierarchical order, individualization, and centralization.[29] The last two are mostly useful to explain the appearance of stylistic common practices and the departure from them. They see the musical language as a result of stochastic

[24]Those operation will be formally defined later in this paper.

[25]"Model is a physical, mathematical, or otherwise logical representation of a system, entity, phenomenon or process"[2, p.3]. "Models in ordinary language have their place in systems theory. The system idea retains its value even where it cannot be formulated mathematically, or remains a 'guided idea' rather than being a mathematical construct" [4, p.24].

[26]This is also called a *dataless system* [16, p.49].

[27]It is interesting to find an example of data systems in terms of musical analysis in Klir [17, p.64-67].

[28]Metasystems are particularly important to monitor *morphogenetic systems*, a concept that will be defined below.

[29]A musical work, for them, "is a set of sound objects and processes, organized in a certain way so as to meet an objective–particular overall finalities of a communicational-aesthetic purport" [10, p.17].

procedures and fuzzy indeterminacy and highlight a strong contextual dependency of a musical work with its historical epoch, general trends and individual style. The authors consider that a musical system is, at least through the historical point of view, an open system[30], i.e., a system that interacts with the environment. In fact, they go even further and regard any musical composition as the result of tensions between a structural level and the environment, which produces bifurcation points, zones of uncertainty and fluctuations. Within this context, a key concept is presented: *morphogenetic music*. In contrast with structurally-stable music (the vast majority of Western concert music production), morphogenetic music has a structure that changes drastically over time. It is closely related to catastrophe theory, in which "abrupt changes of state are the result of smooth alteration of the control parameters" in dynamical systems [19, p.359][31]. In terms of Western concert music, the morphogenetic type is present in the transitional periods. If music leaves a steady-state, the bifurcation point leads either to negative entropy or to positive entropy. In the latter, a destructuring process moves towards a white noise state; in the former, a neostructuring process moves towards upper musical states. The operas *Agamenon* and *Oeumenides*, by Romanian composer Aurel Stroe, are respectively examples of both cases, according to the authors. It is noteworthy, however, that, in their article, the term "compositional system" appears only once and no definition is given[32].

A definition for the term *compositional system* was given (maybe for the first time) in Flávio Lima's dissertation [22, p.63], written under my supervision: a compositional system is "a set of guidelines, forming a coherent whole, which coordinates the use and interconnection of musical parameters, with the purpose of producing musical works". Later [31, p.69], I have proposed the inclusion of the word "materials" in the definition in order to considers also the materials as a whole, without breaking them into their various parameters. Therefore, the present definition of system is: a set of guidelines, forming a coherent whole, which coordinates the use and interconnection of musical parameters and materials, with the purpose of producing musical works. The idea of purpose (or function) is inspired by Meadows [24, p.11], for whom "a system must consist of three kinds of things: elements, interconnections, and a function or purpose".

A formal definition of compositional system S is given by the expression $S = (O, R)$, in which O corresponds to objects, i.e., materials or parameters (abstract or concrete) and R to relations (or functions, operations, and transformations). Differently from Klir's definition (Equation 1), relations here may be represented both in the form of subsets of a Cartesian product of the elements of O and by their analytical expressions.[33] Furthermore, the idea of purpose is embedded in the definition and clearly appears during the process of designing a system, as it will be seen next. Why was the word thing (stated in Equation 1 by Klir) translated as object in the initial moments of the foundation of the theory of compositional systems? During the bibliographic research phase for Flávio Lima's dissertation [22], I came across the book *Teoria dos Objetos*, by Abraham Moles [28]. In this book, Moles makes a distinction between the concepts of thing and object. The first is natural, the second is artificial, that is, produced by humans. Thus, as Moles says, a "stone will only become an object when promoted to paperweight, and when equipped with a label: price ..., quality ..., inserting it in the universe of social reference" [28, p.26]. In addition, Moles associates

[30] The concept of open system differs from the concept of open **compositional** system, as it will be seen later in this paper.

[31] Compositional experiments with catastrophe theory have been made by composers Ann Warde [41] and Fani Kosona [19].

[32] The authors frequently use the term "musical system", which appears 22 times.

[33] Relations are used for musical contexts in which an element of a certain domain is mapped onto two or more elements of the range. Functions are used for single parametric elements (a pitch class, a duration, etc.). Operations are functions applied to sets (a pitch class set, for example). Transformations are reserved for operations associated with Lewin's GMIT [21].

the term object with the concept of system when he states that “the whole set of elements or objects linked by functional relations can be considered as a system (...)” [28, p.28]. It is worth mentioning that this author collaborated with Pierre Schaeffer in the definition of the term *sound object*, which is a second use of the term object. Schaeffer divides this term into two categories: 1) musical object, which “is treated as the object of the language established between the composer and the listener (...). It is the spokesman for the musical language” [25, p.59]; 2) sound object, which is related to the sound itself (as much of the so-called musical sounds as of the noises). “The sound object emerges in the functions ’to perceive aurally / to hear (to pay attention)’, while the musical object, inserted in a language, fits in the function ’to understand’"[25, p.65].[34]. Schaeffer relies heavily on Husserl’s phenomenology to arrive at the essence of sound. A third strand for the term object can be found in computer science, more precisely in the programming paradigm known as OOP (object-oriented programming), which deals with the interaction between basic units called objects. Long after the foundation of the theory of compositional systems, there was a happy coincidence in the use of the term object (and not thing) in an article by Goguen [11] that deals with the formalization of musical systems (the second section of this work by Goguen is precisely entitled "Objects and Relations"). The term object in the theory of compositional systems is associated with very specific elements stated in the very fundamental definition of the theory. Objects are parametric structures (concrete, such as pitch, duration, register, harmonic entities, etc., or abstract, such as textural partitions, degree of harmonic endogeny, contours, inversional axes, etc.) or raw materials (fragments extracted from original works). The objects described in the theory of compositional systems encompass Schaeffer’s sound and musical objects and are flexible to the point of encompassing, at least potentially, elements that do not directly involve sound, including elements of a spatial nature.

Compositional-system methodology considers the holistic phenomena (as described earlier in this paper) in a very loose manner—in a more flexible fashion that the third type of holism described in Phillips [29][35]. Therefore, analytical methods are largely employed and even regarded as essential procedures, especially for modeled systems. In terms of design, a compositional system may emerge from a series of formal declarations, diagrams, tables, and computational programs.

With respect to typology, I define three types of compositional systems: open, semi-open, and feedback [33]. Different combinations of these three types yield systems with higher complexity. It is noteworthy to emphasize that our classification is from a different nature when compared with systems science’s typology, that is, it is not associated with the concepts of open and closed systems defined by Bertalanffy [4, p.121], for whom “a system [is] ‘closed’ if no material enters or leaves it; it is called ‘open’ if there is import and export of material”[36].

[34]Schaeffer defines four functions of listening: "écouter", "ouïr", "entendre", and "comprendre" [36]. I am using here the suggestions given by North and Dack in the 2017 English translation of the "Traité des objets musicaux"[37]

[35]See note 4

[36]The dicothomy open/closed systems in the context of social sciences is discussed in length by Luhmann [23]: “Physics has come to the understanding that the universe is a closed system that cannot accept any kind of input from an order that is not contained in itself and that, there, the law of entropy is inexorable. But if this is valid for the physical world, it is not the case for the biological or social order. Hence, the physical lock of the universe was denied as a phenomenon representative of other orders. So it was thought that these different systems would have to be fundamentally open, capable of developing neguentropia. This being open explained the effort of organisms (if you think of biology) to overcome, even partially, the entropic law of the universe. In the original one reads: “La física ha llegado a la comprensión de que el universo es un sistema cerrado, que no puede aceptar ningún tipo de input de un orden que no esté contenido en él mismo y que, allí, la ley de la entropía es inexorable. Pero si esto es válido para el mundo físico, no lo es, sin más, para el orden biológico ni el social. De aquí que la cerradura física del universo se negara como un fenómeno representativo de otros órdenes. Entonces se pensó que estos sistemas distintos tendrían que ser fundamentalmente abiertos, capaces de desarrollar neguentropía. Este ser abiertos explicaba el esfuerzo de los organismos (si se piensa en ia biología) por sobreponerse, aunque fuera parcialmente, a la ley entrópica del universo” [23, p.47].

Basically, in this methodology, open systems have input and output, semi-open systems have only output of data (although it may have some kind of control operators), and feedback systems have the data reinserted into its input. Before proceeding with designs for those three types of systems, it will be necessary to define some operations to be applied to musical parameters and materials.

i. Operations used in the design of compositional systems

In this subsection I will describe several operations that can be applied to parameters and materials in the realm of compositional systems. I will define fourteen operations, some of them original proposals that have already been experimented with compositional students and during my own compositional designs. The list is not exhaustive and their combination are encouraged in order to create compound operations. Some operations have effect only on the surface level (such as transposition, for example), while others (such as retrogradation) have a severe cognitive impact. Some of them are already incorporated in computer applications, such as the Lewin Calculator, developed by Barbosa, Santos, and Pitombeira [1].

1. Transposition (T_n) – rewriting of a segment at another pitch level. Intervallic relationships (and therefore contour) as well as rhythmic structure are preserved. It can be performed in terms of a diatonic set (in this case it is rotation through a modular space).
2. Inversion (I_n) – generates a mirrored outline (in a chromatic context) around the first pitch. The index n is a transpositional factor. It can also be performed diatonically.[37]
3. Prolation (P_t) – temporal expansion/contraction. It consists in rewriting a segment with longer or shorter durational values, according to factor t. Temporal expansion is traditionally known as augmentation and contraction as diminution.
4. Ambitus (A_i) – intervallic expansion/contraction. The line is rewritten with expanded or contracted intervals, according to factor i. Melodic contour and rhythm are preserved.
5. Retrogradation (R) – line (pitches and rhythmic values) rewritten backwards. One can also retrograde just one of the parameters. Although the original material can be visually identified, this transformation drastically changes the profile of the material from an auditory perspective, depending on the number of elements involved.
6. Rotation (ROT_k) – rotation of the elements of a pitch class set. The number of possible rotations depends on set cardinality.[38]
7. Multiplication – I use three types:
   (a) Boulez (Mb) – the intervallic profile of a set is applied to each element of another set.
   (b) Rahn (M_k) – a set (or an entire segment) is multiplied by a constant value (k).
   (c) Rahn expanded (Mr) – the elements of a pitch class set in normal form are concatenated to form an integer (base 12). For example: $J = \{10, 1, 2\} \rightarrow J = A12$; then, this set is multiplied by another set in the same format following the rules of regular arithmetic multiplication of two numbers.[39]

[37] This is different from the T_nI operation of the pitch class set theory, which first inverts around 0 and then applies the transpositional factor.

[38] Rotations and Reflections are the two types of Permutation of a dihedral group, such as a T_n/T_nI group. Reflections here are obtained by applying Rotations to Retrogradations.

[39] This is an original contribution of the present author to the multiplication of pitch class sets.

8. Addition/subtraction of elements (A/S) – insertion of elements (interpolation) or simplification of a line by removing elements.
9. Parametric fixation (Pf) – the pitch structure is maintained and the rhythmic structure is changed or contrariwise (without the obligation to maintain the same contour). Evidently, this concept can be applied to transformations that involve other parameters (density, contour, dynamics, articulation, ...).
10. Octave offset (Of) – consists of a free octave displacement applied to some pitches of a segment.
11. Filtering (Ft) – consists of filtering the notes through a different scale than the one used in the original construction (eliminating those that do not belong to the filter).
12. Conversion (C) – consists of filtering the notes through a scale different from that used in the original construction and converting the notes that do not belong to it (taking this new scale as a reference).
13. Permutation (P) – changes in the order of individual segments or elements.
14. Fragmentation (Fg) – free and repeated use of small portions of the original segment in its original or altered fashions.

I have designed a short example of an open compositional system (with input and output) using only the first six operators mentioned above: Transposition (T_k), Inversion (I_n), Prolation (P_t), Ambitus (A_i), Retrogradation (R), and Rotation (ROT_k). A diagram of this system is shown in Figure 8. The operators, except for R, need to be supplied with control factors (k, n, t, i, k). The first step consists of arbitrarily choosing a small musical fragment (original or taken from another work), in MIDI or *musicxml* format, and insert it into the system. Seven external keys controlled by the composer command the sequence of activation of each operator. They cannot be activated simultaneously. The original fragment as well as the output of each operator are appended to a temporal concatenator that sends the final stream to the system's output. The form of the piece is obtained by the gradual concatenation of the results of each operator. This system, implemented in Python, is a work in progress.[40] One possible musical result is shown at the bottom of Figure 8.

This output [41] may be used as the initial idea for a new composition, after some polishing, or even be reinserted into the system to generate a larger musical segment. I have chosen the first case, made small adjustments in the fragment, assigned it to a musical instrument (clarinet), and added a piano accompaniment. The adjustments are indicated in Figure 9: 1) Operator A/S (addition of elements) – two sixteenth notes (E and D) at the end of measure 2 to connect with the C in measure 3; 2) Rebar of measure 4 to follow the 3/4 metric; and 3) Operator T_{-4} (transposition a major third down) applied to ROT_1. For the piano, I have decided that its pitch classes come entirely from the clarinet's melody (except for measure 2, in which the embellishing note E was added to the original melodic line). Also, I have used the operation called Fragmentation (F) in order to assign the three first figures of measure 4 to the piano's left hand, but adapted to the harmonic constraints. The result is shown in Figure 9, in which the normal forms of the pitch class sets for each measure are indicated below the score.

The way harmony was generated for this excerpt[42] is called *Endogenous Harmony*.[43] It is possible to define other types of harmonic configurations in relation to a given melodic line.

[40] `https://gitlab.com/musmat/open-compositional-system`

[41] A midi rendition of this fragment is available at `https://gitlab.com/musmat/open-compositional-system/-/blob/master/melodiaresultadosistema1.wav`

[42] A midi rendition for this fragment is available at `https://gitlab.com/musmat/open-compositional-system/-/blob/master/melodiaresultadosistema1harmonizada.wav`

[43] This is a pedagogical tool created by the present author for his introductory compositional courses.

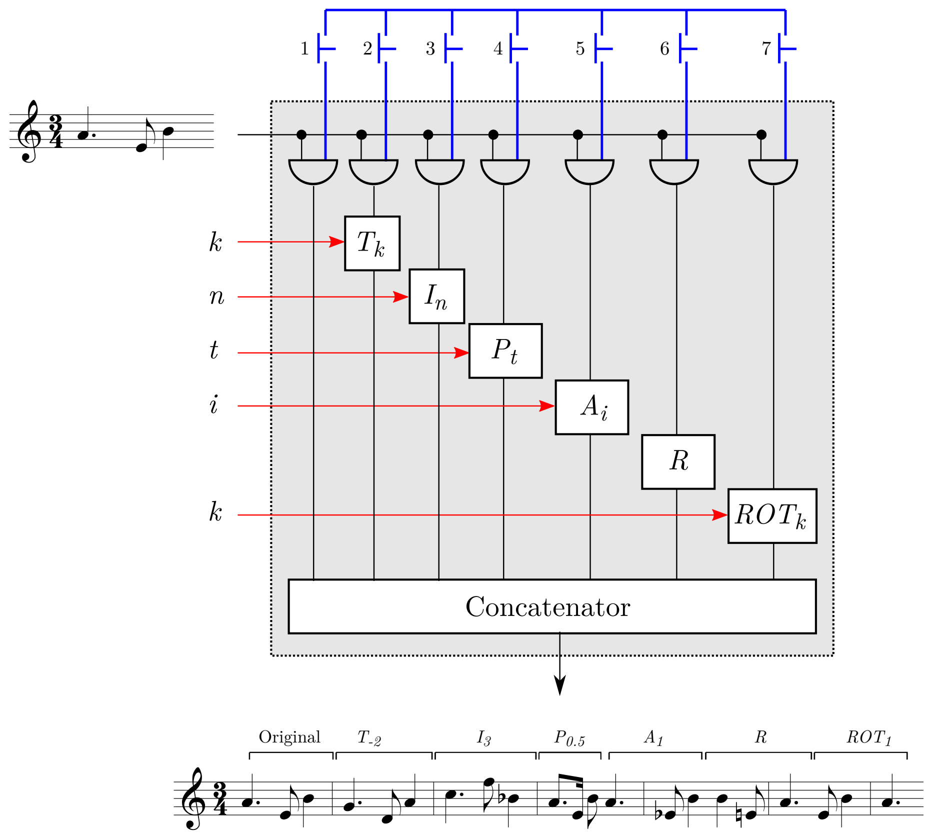


Figure 8: *Diagram of an open compositional system built with six operators and a temporal concatenator.*

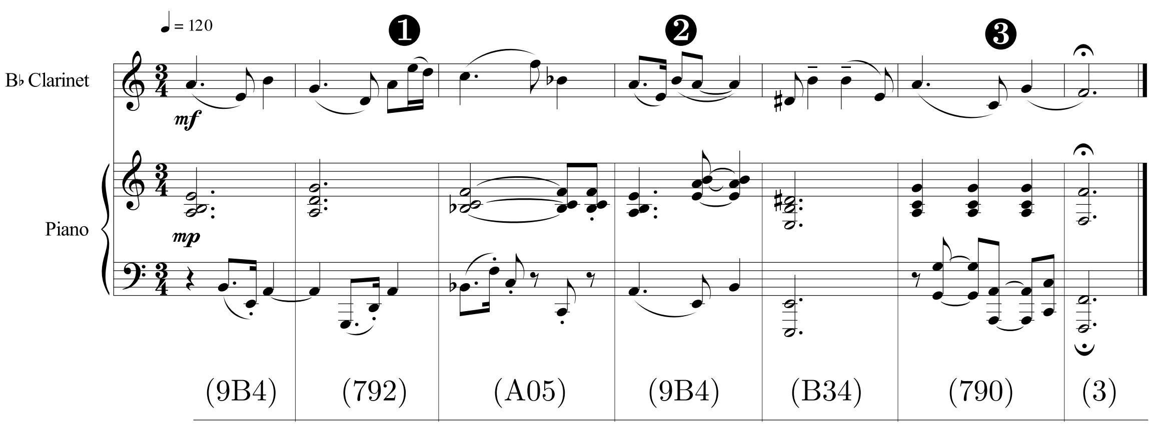


Figure 9: *Use of the fragment produced by the open compositional system in the beginning of a piece for piano and clarinet.*

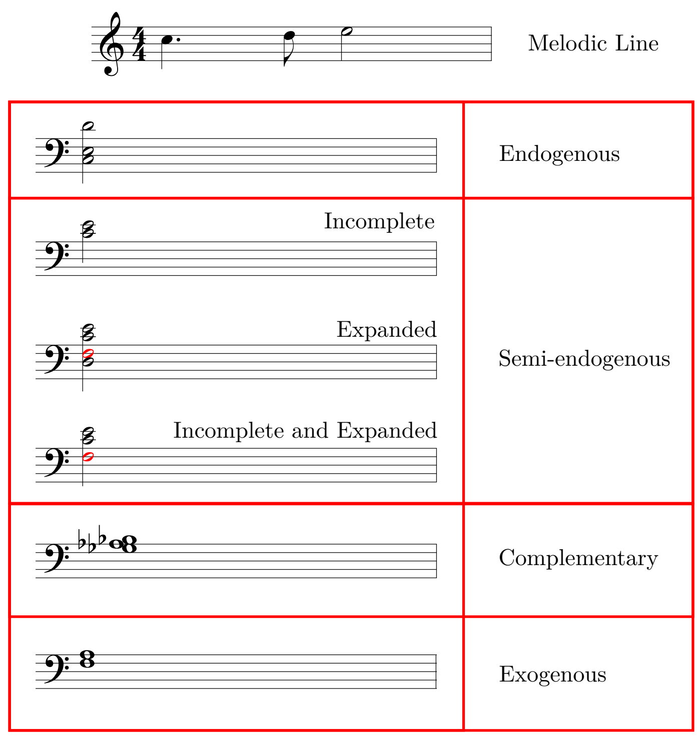


Figure 10: *Types of harmonic configurations derived from a given melodic line.*

Semi-endogenous harmonization takes place when the melody's pitch classes are partially applied to the harmony or when additional pitch classes are included. This type can be divided into three subtypes: a) Incomplete — the melodic pitch classes are partially used (in the example of Figure 10 only the pitch classes C and E are used in the harmony, leaving out the D, which is also part of the melodic line); b) Expanded — inclusion of additional notes (in the example Figure 10 all the pitch classes, but the added F, belong to the melodic line); c) Incomplete and Expanded – a combination of the previous subtypes (in Figure 10 the lower pitch class, F, does not belong to the melodic line, whereas the other two belong; but the D is missing). If the harmony is designed as a complement to the melody's pitch classes, with respect to some larger set, it is said that the harmony is *Complementary*. In the example of Figure 10 the harmony is the complement of the melody with respect to the whole tone scale $C, D, E, G\flat, A\flat, B\flat$. Finally, if the harmony does not have any connection with the melodic line it is called *Exogenous*.

As an example of a semi-open compositional system, I have designed a system in which the internal data is generated through probability: a binomial distribution for the pitch parameter and a uniform distribution for the rhythmic structure. In the uniform distribution all the outcomes have the same probability. The binomial distribution is a discrete distribution that counts the amount of "successes" or "failures" in binary experiments. If n is the number of trials of a probabilistic experiment, p is the probability of success of each outcome, and k is the number of desired success, the probability of k is given by Equation 3.

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \tag{3}$$

For example, if one flips a coin eight times to obtain "heads", a binomial experiment is taking

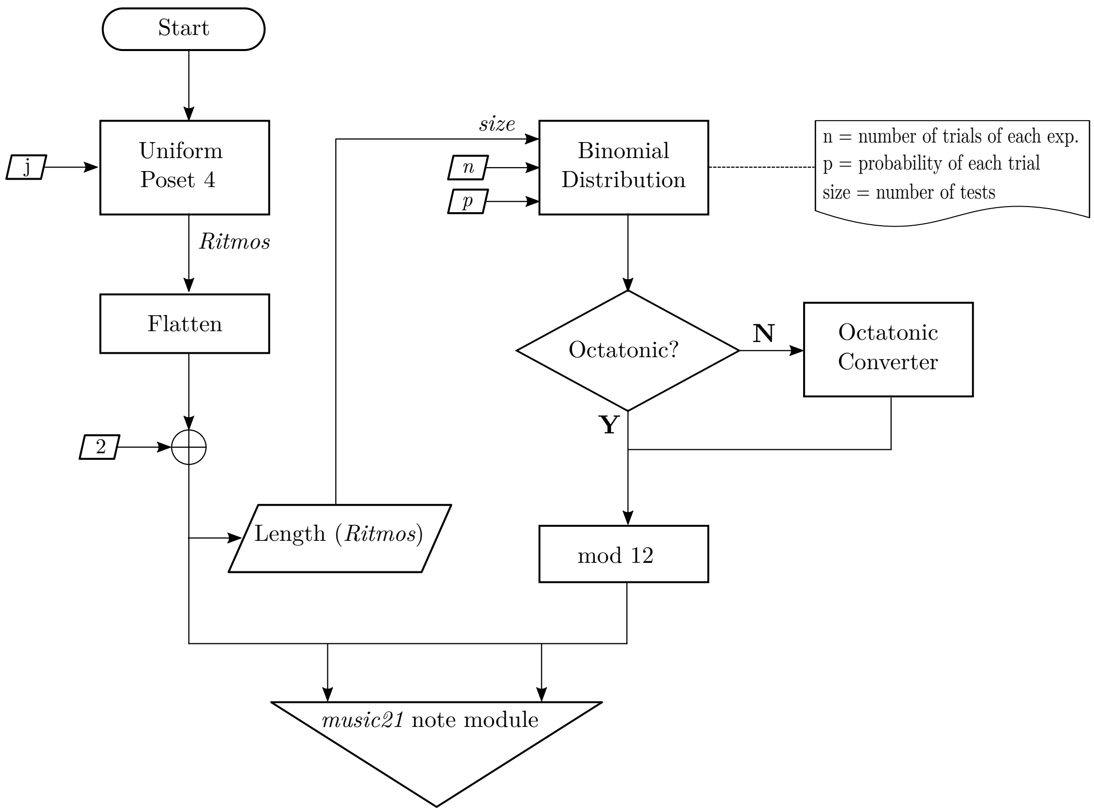


Figure 11: *Flowchart for a semi-open compositional system.*

place. Its parameters are 8 (number of trials) and 0.5 (probability that a head will occur in a single attempt). If one wishes to know the probability of having three successes the result will be given by Equation4:

$$P(X = 3) = \binom{8}{3} 0.5^3 (1 - 0.5)^{8-3} \tag{4}$$

The flowchart for the semi-open compositional system is shown in Figure 11. It starts by uniformly randomizing a *poset* formed by the compositions[44] of integer 4, which are isomorphic to the eight possible integer subdivisions of a quarter note (Figure 12).[45] The composer chooses the number of randomizations (j). The result is transformed into a flat list, i.e., without the separation of rhythmic figures (which are already guaranteed through the ordered randomization). An arbitrary rhythmic figure of half note is appended at the end of the list (called *ritmos*) to promote a rhythmic cadence and the result is sent to the note module of the *music21* Python package.[46]

[44]Posets are partially ordered sets, i.e., sets in which the elements are ordered but not all of them required to be hierarchically comparable. Compositions are partitions in which the order is essential. The number of compositions (C) of an integer n is given by $C_n = 2^{n-1}$.

[45]The operation that connects the nodes from the bottom (1.1.1.1) towards the top composition (4) is the sum of consecutive parts. Therefore for a composition $C_i = p_1, p_2, \ldots, p_n$, with n parts, a composition C_j, with $n-1$ parts is$C_j = p_1, p_2, \ldots (p_k + (p_{k+1})), \ldots, p_n$. So, the composition 1.1.1.1 may progress to 2.1.1, which is the sum of its first and second parts $(1 + 1)$ or may progress to 1.2.1, which is the sum of its second and third parts, and so on.

[46]The *music21* package is a Python library to handle musical objects developed by Michael Cuthbert, at the MIT, and available at https://web.mit.edu/music21/

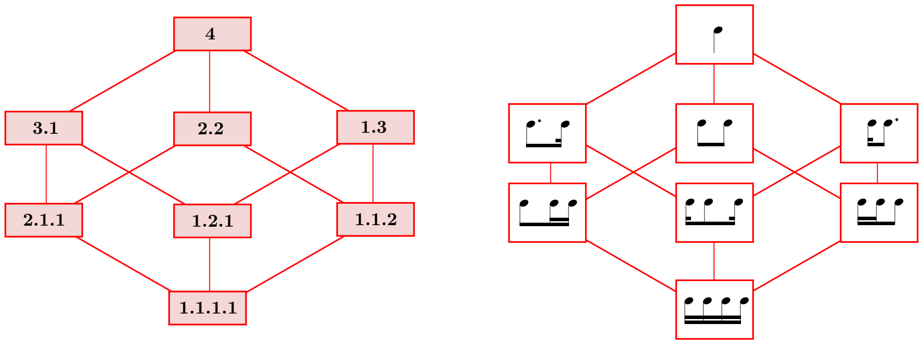


Figure 12: *Two isomorphic Hasse diagrams: the compositions of integer 4 and the eight possible rhythmic integer subdivisions of a quarter note.*

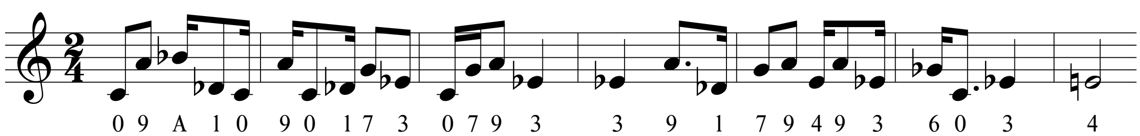


Figure 13: *A melodic fragment generated by the semi-open compositional system for* $j = 12$*,* $n = 100$*, and* $p = 0.5$*. The numbers below each note indicate the corresponded pitch class in integer notation.*

Each value of the list will be assigned to the *note.Note.quarterLength* object.

The length of the list *ritmos* will be passed as a parameter (*size*) to the binomial module of the system, which is responsible for the pitch configurations. The *size* corresponds to the number of tests. Two other parameters are inserted into the binomial (or pitch) module by the composer: the number of trials of each experiment (n) and the probability of each trial (p). The result is filtered through an octatonic converter and through a mod 12 operator. These modules perform respectively the conversion and the octave offset operations previously defined at the beginning of this section[47]. The result will also be inserted into the *note.Note* object of the *music21* package.

In order to generate a melodic fragment, I have chosen $j = 12$, for the uniform distribution module (assigned to rhythm), and $n = 100$ and $p = 0.5$, for the binomial distribution module (assigned to pitch). This is analogous to a probabilistic experiment consisting of flipping a coin 100 times. If this experiment is performed *size* times, one may ask how many times the result will be 1 head, 2 heads, 3 heads, and so on. As I have mentioned, the value of *size* in this system comes from the rhythmic module. The result is show in Figure 13. A histogram of the output reveals a shape typical of a binomial distribution. From this histogram one can easily recognize a centricity in A.

As it was done with the fragment for the open system, I also propose to harmonize this fragment. This time, for clarinet trio (two *B*♭ clarinets and a *B*♭ bass clarinet). The harmony will be extracted from the melodic line using the criteria of complementary harmony taking as a reference the same octatonic scale of the melodic line. The window size will be one measure, which means that every measure will have an octatonic aggregate, i.e., the entire octatonic scale used in the melody. The result is show in Figure 15.[48]

The third type of compositional system, according to our taxonomy, is the feedback system.

[47] Those are respectively operations 12 and 10 in the list of operations given at the beginning of this subsection

[48] A midi rendition is available at https://gitlab.com/musmat/open-compositional-system/-/blob/master/binomiamelodia2_.wav

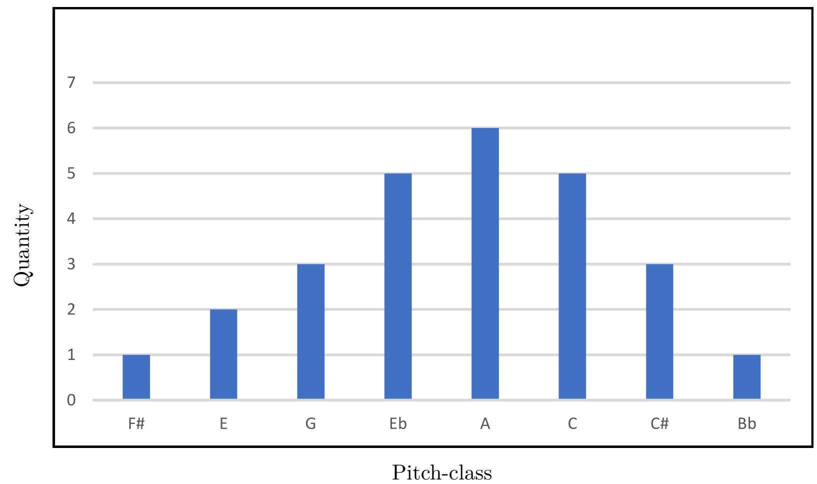


Figure 14: *The histogram of the system output revealing the binomial archetype and indicating a centricity in A.*

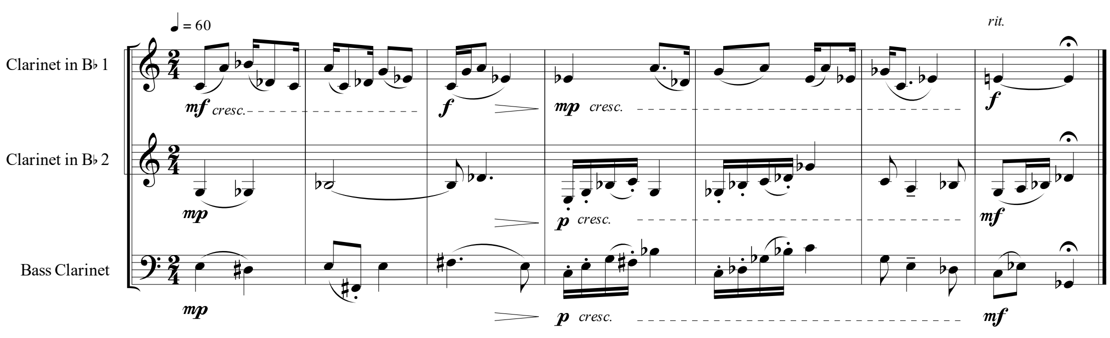


Figure 15: *The fragment generated by the semi-open compositional system harmonized through the criteria of complementarity (with the octatonic scale of the melodic line).*

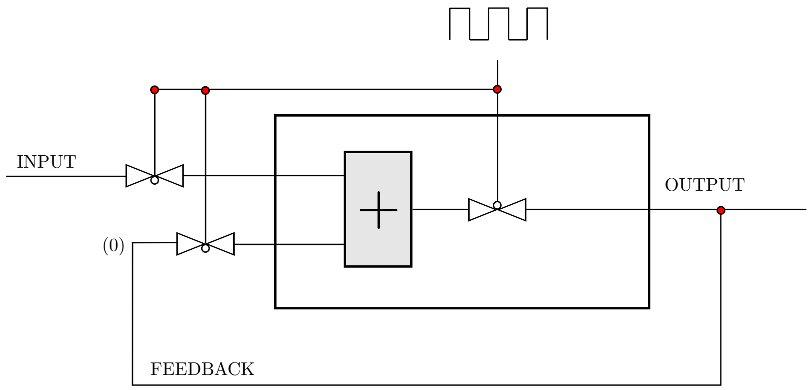


Figure 16: *A feedback compositional system with an addition operator*

This type of system has the output reinserted in the input, establishing what is known as *iterative process*.[49] There are two possible subtypes of input: a stream of information (a MIDI file, for example) or an initial trigger (an initial value employed only to start the process). The compositional system shown in Figure 16 has as its input a MIDI file (the melodic line of the previous example will be used). The output of the system is reinserted into its input. Both the MIDI file and the feedback information are controlled by transmission gates.[50] A square wave controls the cycle input/output of the system, in such a way that when its value is 1, the input and feedback ports are allowed to flow into the system and the output is blocked; when its value is 0 the input is blocked and the output flows. The system has one only operator that adds both the MIDI file and the feedback and apply mod 12. Figure 17 shows a step-by-step cycle of operation for the first five pitch classes inserted into the system. At the start, the MIDI file has pitch class 0 and, as the feedback has no value yet, an initial temporary value is given (0). These two values enter the system and are added, yielding 0. This value flows to the output and is sent back to the input. At this point the MIDI file has the pitch class 9, which is added with 0 (feedback value) yielding 9, which is sent to the input again and added with the next pitch class read from the file (10), yielding 7 (19 mod 12), and so on. Figure 18 shows the fragment produced by this compositional system. The contours of both melodic lines (input and output) are shown in Figure 19. In those graphs, the x-axis corresponds to each event (i.e., first pitch class, second pitch class, and so on) and the y-axis corresponds to the pitch class value.

The other subtype of feedback compositional system falls into a category known as chaotic maps. Those maps may be classified in terms of the number of their space dimension. This is an important factor for musical applications because the system's output may be mapped onto musical parameters. The output of a bidimensional chaotic map can be assigned to pitch and rhythmic parameters, for example. With a four-dimensional map one may have pitch, rhythm, dynamics and preset timbres controlled by its output. There are several known chaotic maps: one-dimension (Gauss, Logistic, Lambic, etc.), two-dimension (Hénon, Mandelbrot, Lozi, etc.), three-dimension (Lorenz, Ueda, Shimizu-Morioka, etc.) and four-dimension (Hyper-Lorenz, Hyper-Rössler).[51]

---

[49] According to Miranda (2004, p.83), "an iterative process is defined as a rule that describes the action that is to be repeatedly applied to an initial value x_0. The outcome of an iterative process constitutes a set, technically referred to as the orbit of the process"[27].

[50] A transmission gate works as an AND logic gate through which regular information (besides binary information) may flow.

[51] It is also possible to find five-dimension implementations[40]

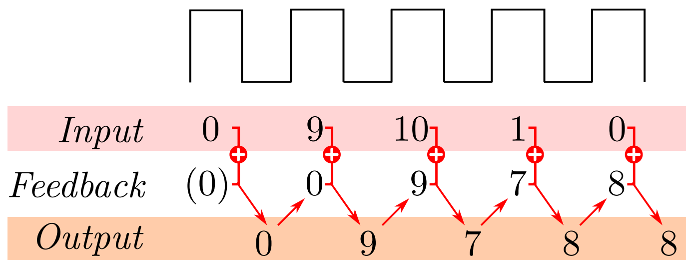


Figure 17: *Tracing the pitch class data within a feedback compositional system.*

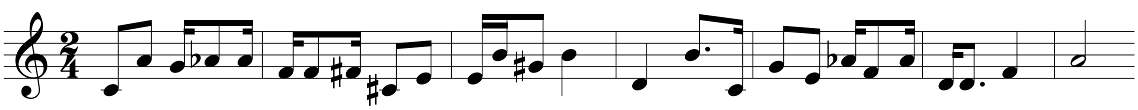


Figure 18: *The output fragment of our feedback compositional system using as input the melodic fragment of the previous compositional system.*

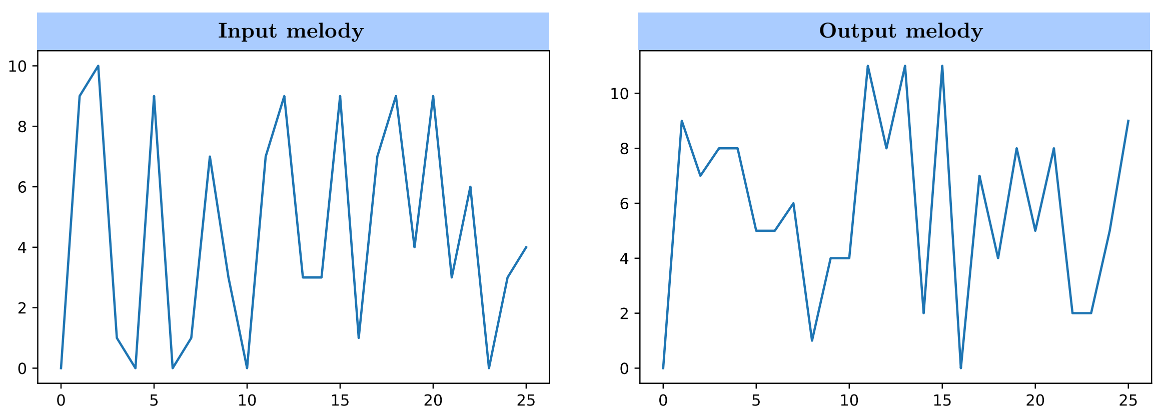


Figure 19: *The contours of the original melodic line inserted into the input of the feedback compositional system (left) and the resulted melody (right)*

Table 1: *The values of $f(z)$ for $c = 1$ and $c = -1$.*

$c = 1$		$c = -1$	
z_n	z_{n+1}	z_n	z_{n+1}
0	1	0	-1
1	2	-1	0
2	5	0	-1
5	26	-1	0
26	677	0	-1
677	458330	-1	0

I will select the Mandelbrot map, which has already been largely applied to the musical domain ([5], [39], [15], [18]). The Mandelbrot map is a representation in the complex plane of a Mandelbrot set M, "which is defined as the set of c $\in \mathbb{C}$ for which the sequence $c, c^2 + c, (c^2 + c)^2 + c, ...$ does not tend to ∞ as n tends to ∞" [6, p.75]. Graphically, this set is located inside the black region of the fractal structure shown in Figure 20, which is built computationally through the iteration of the function $z_{n+1} = z_n^2 + c$. For $z_0 = 0$, in each iteration, if $f(z)$ does not tend to infinity c is a member of the Mandelbrot set.[52] Table 1 shows the values of z_{n+1} for $c = 1 + 0j$ and $c = -1 + 0j$. In the first case the function tends quickly to infinity and, therefore, $1 + 0j$ does not belong to the set; in the second case the function is bounded and, so, $-1 + 0j$ belongs to the Mandelbrot set. Algorithm 1, shown below, receives the real and imaginary components of a complex number, tests if this number makes the function "to explode" under a certain number of iterations (which is also a value sent to this function), and returns the number of iterations, the complex number and its modulus. The number of iterations will be used to fill in an array which will correspond to the color of pixels in a screen. The black pixels correspond to the complex numbers that keep the modulus of z equal or smaller than 2 under iteration. The real and imaginary components of the members of the Mandelbrot set are extracted and assigned to pitch and duration (after a normalization). The result (after metrical adjustments) is shown in Figure 21. One can clearly see the intrinsic symmetry of the melodic line, a characteristic already presented in the fractal (Figure 20).

Algorithm 1 Mandelbrot's algorithm

```
procedure MANDELBROT (R, I, max_iter)
    c ← complex(R, I)
    z ← 0.0j
    counter ← 0
    while abs(z) <= 2 and counter <= max_iter do
        z ← z^2 + c
        counter ← counter + 1
    return count, c, abs(z)
```

[52] It is important to mention that if the modulus of z ever becomes larger than 2, the result will escape to infinity. Therefore, the set is formed by the complex numbers that remain inside a region centered at the origin with radius 2 [6, p.81].

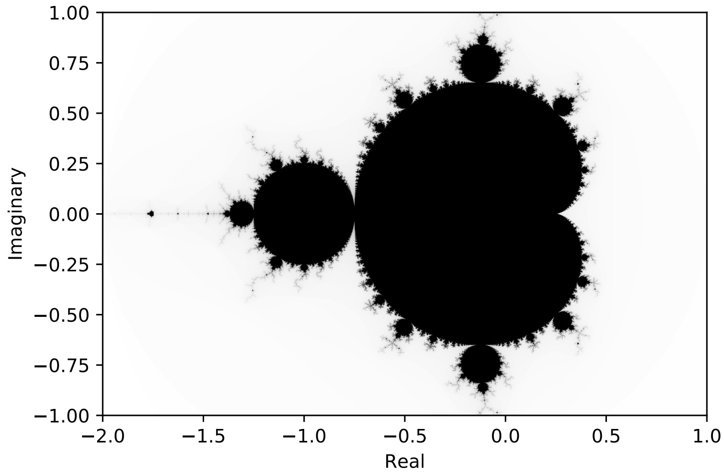


Figure 20: *The Mandelbrot Fractal*

Figure 21: *A possible musical representation of the Mandelbrot Fractal*

IV. Concluding Remarks

I have walked through various topics related to systems science and compositional systems. As it happens to all theories developed by humans, the theory of compositional systems, which appeared around 2009, is a work-in-progress and constantly benefits from the academic and artistic interchange with researchers and composers, as well as from the contributions given by my undergraduate and graduate students, who constantly collaborate with new ideas and pose new questions and challenges that require adjustments and yield expansions. Therefore, I believe that a mature formalization of this theory will be proposed in the near future. At this moment, our role is to support it with experiments, research, and reflections in order to gradually bring it to a more comprehensive, embracing and flexible state. One must remember that arithmetic and probability were formalized as late as 1889 and 1931, respectively by Peano and Kolmogorov.

At this point, the focus is primarily the steady-state compositional systems in open, semi-open and feedback formats. Exceptions are the few experiments with permutation systems (mobile forms) conducted by one my graduate students.[53] However, in those permutation systems, the score, once produced, is fixed, and the various possibilities appear only during a performance. A future goal includes research toward an elaboration of a score that could present changes over time.

Finally, I am very grateful to my friends in mathematics and music research who solved many of my questions related to formalization and notation during the process of writing this paper: Petrucio Viana (UFF), Carlos Almada (UFRJ) and Francisco Aragão (UFC), who read the text and sent detailed suggestions, and mainly Hugo Carvalho (UFRJ) who, in addition to reading the entire text, carefully studied several mathematical aspects and discussed the best strategies to make the formalizations clear and precise.

References

[1] Barbosa, H.; Santos, R.; Pitombeira, L. (2015). Lewin: desenvolvimento de uma calculadora de classes de altura com funções analíticas e composicionais. Congresso da Associação Nacional de Pesquisa e Pós-Graduação em Música, 25. *Proceedings...*, Vitória: UFES.

[2] Banks, C. (2010). Introduction to Modeling and Simulation. In Sokoloswski, John; Banks, Catherine (ed.) *Modeling and Simulation Fundamentals* Theoretical Underpinnings and Practical Domains. New Jersey: Wiley: 1–24.

[3] Bertalanffy, L. (1952). *Problems of Life*: An Evaluation of Modern Biological Thought. London: Watts & Co.

[4] Bertalanffy, L. (1968). *General System Theory*: Foundation, Development, Application. New York: George Braziller.

[5] Bidlack, R. (1992). "Chaotic Systems as Simple (But Complex) Compositional Algorithms". *Computer Music Journal,* 16: 33–47.

[6] Branner, B. (1989). The Mandelbrot Set. In Chaos and Fractals: The Mathematics Behind the Computer Graphics. *Proc. Sympos. Appl. Math.*, 39. *Proceedings...*, Providence, RI: American Mathematical Society: 75–105.

[53] Alexandre Ferreira's Ph.D. thesis, already in its final stage, describes some compositional systems he has developed in the context of Game Theory.

[7] Condillac, É. (1771). *Traité des systèmes, ou l'on démèle les inconvènients et les avantages.* Amsterdan: Arkstée & Merkus.

[8] Dewey, J.; Bentley, A. (1946). "Interaction and Transaction". *The Journal of Philosophy,* 43(19): 505–517.

[9] Eco, U. (2001). *A Estrutura Ausente* São Paulo: Perspectiva.

[10] Georgescu, C.; Georgescu, M. (1990). "A System Approach to Music". *Interface,* 19(1): 15–52.

[11] Goguen, J. (1977). "Complexity of Hierarchically Organized Systems and the Structure of Musical Experiences". *International Journal of General Systems,* 3(4): 233–251.

[12] Goguen, J. A.; Varela, F. J. (1979). "Systems and Distinctions; Duality and Complementarity". *International Journal of General Systems.* 5(1): 31–43.

[13] Halford, G. ; Baker, R.; McCredden, J. ; Bain, J. (1948). "How Many Variables Can Humans Process?" *Psychological Science.* 16: 70–76.

[14] Halmos, P. (1974). *Naive Set Theory.* New York: Springer.

[15] Kerkez, B. (2011). Mandelbrot Sound Map - A Tool for Mapping Fractals into Sounds. In: *Bridges 2011*: Mathematics, Music, Art, Architecture, Culture: 535–538.

[16] Klir, G.(1991). *Facets of Systems Science.* New York: Plenum.

[17] Klir, G.; Elias, D. (2003). *Architecture of Systems Problem Solving.* New York: Springer.

[18] Knapp, C. (2012). *Creating Music Visualizations in a Mandelbrot Set Explorer.* Degree Project (Computer Science). Växjö, Sweden: Linnaeus University.

[19] Kosona, F.; Hadjileontiadis, L. (2011). Catastrophe Theory: An Enhanced Structural and Ontological Space in Music Composition. In Agon C., Andreatta M., Assayag G., Amiot E., Bresson J., Mandereau J. (eds). *Mathematics and Computation in Music - Proceedings of the Third International Conference, MCM 2011, Paris, France, June 2011*. Berlin: Springer: 358–361.

[20] Lévy-Strauss, C. (1963). *Structural Anthropology.* New York: Basic Books.

[21] Lewin, D. (2007). *Generalized Intervals and Transformations.* New York: Oxford University Press.

[22] Lima, F. (2011). *Desenvolvimento de sistemas composicionais a partir da Intertextualidade.* Dissertação de Mestrado. João Pessoa: Universidade Federal da Paraíba.

[23] Luhmann, N. (1996). *Introducción a la teoria de sistemas.* México: Universidad Iberoamericana.

[24] Meadows, D. (2009). *Thinking in Systems:* A Primer. London: Earthscan.

[25] Melo, F. (2011). *De "Introduction a la musique concrète" ao Traité dês objets musicaux: gênese do solfejo dos objetos musicais de Pierre Schaeffer.* Dissertação de Mestrado. Belo Horizonte: Universidade Federal de Minas Gerais.

[26] Meyer, L. (1957). *Emotion and Meaning in Music.* Chicago: University of Chicago Press.

[27] Miranda. E. (2004). *Composing Music with Computers.* London: Elsevier.

[28] Moles, A. (1981). *Teoria dos objetos.* Rio de Janeiro: Tempo Brasileiro.

[29] Phillips, D. (1976). *Holistic Thought in Social Science.* New York: Macmillan.

[30] Piaget, J. (1970). *Structuralism.* Trad. Chaninah Maschler. New York: Basic Books.

[31] Pitombeira, L. (2015). A produção de teoria composicional no Brasil. In Ilza Nogueira e Fausto Borém (org.). *O pensamento musical criativo*: teoria, análise e os desafios interpretativos da atualidade, 1. Salvador: Universidade Federal da Bahia: 61–89.

[32] Pitombeira, L. (2017). Formal Design, Textural Profile, and Degree of Harmonic Endogeny as Modeling Factors. Congresso da Associação Brasileira de Teoria e Análise Musical, 2. *Proceedings...*, Florianópolis: Universidade Estadual de Santa Catarina: 42–51.

[33] Pitombeira, L. (2018). "A Systemic Model for Debussy's Prélude No.1". *MusMat*, 2(2): 37–57

[34] Reimer, B. (1962). "Leonard Meyer's Theory of Value and Greatness in Music". *Journal of Research in Music Education,* 10(2): 87–99.

[35] Rosen, R.(1986). "Some Comments on Systems and Systems Theory". *International Journal of General Systems,* 13(1): 1–3.

[36] Schaeffer, P. (1966). *Traité des objets musicaux.* Paris: Éditions du Seuil.

[37] Schaeffer, P. (2017). *Treatise on Musical Objects.* Translated by Christine North and John Dack. Oakland, California: University of California Press.

[38] Skyttner, L.(2005). *General Systems Theory*: Problems, Perspectives, Practice. New Jersey: World Scientifique.

[39] Sukumaran, S.; Dheppa, G. (2009). "Generation of Fractal Music with Mandelbrot Set". *Global Journal of Computer Science and Technology,* 9(4): 127–130.

[40] Wang, Y.; Kang, S.; Lan, C.; Liang, Y.; Zhu, J.; Gao, H. (2016). A Five-Dimensional Chaotic System with a Large Parameter Range and the Circuit Implementation of a Time-Switched System. In: *International Conference on Reliability, Maintainability and Safety (ICRMS).*, 11. Hangzhou: 1–6.

[41] Ward, A.(1999). "Change Over Time: Responsability and Power in the Midst of Catastrophe". *Leonardo Music Journal,* 9: 95–101.

[42] Warren, W.(1948). "Science and Complexity". *American Scientist,* 36(4): 536–544.

[43] Wiener, N. (1961). *Cybernetics or Control and Communication in the Animal and the Machine.* Cambridge: The MIT Press.

[44] Wilcken, P. (2010)1 *Claude Lévi-Strauss:* The Poet in the Laboratory. London: Bloomsbury.

Rhythm and Entropy: The Exile of the Metric in the Dance of Pulsation*

Arthur Kampela
Federal University of the State of Rio de Janeiro
arthurkampela@earthlink.net
Orcid: 0000-0003-1548-7694

Abstract: *This essay intends to demonstrate that the very system that codified the rhythmic subdivision subsumed to the metronome's beat or pulse might offer, through its own mechanisms, a window to its deconstruction and yet, to a new, intrinsic, development. When metric subdivisions occur that are farther away from the metronomic beat's referential, (as when rhythmic deviations by many sub-ratios accumulate underneath a certain rhythmic figure), the performer experiences a cognitive loss in the sense of immediate metronomic adjacency. New ways to perform a certain rhythmic outcome buried within the grounds of complex subdivisions require mechanisms to momentarily suspend the main, overarching, beat, to impose emergent,* micro-metronomes*. These devices are codifiers of speeds whose regularity opens up terrain for new, rhythmical deviations and sub-ratios. They also allow the performer to negotiate between rhythms that present diverging metric configurations, linking their speeds, through rhythmic bridges. As the performer reaches these bridges, located at a deeper level of rhythmic subdivision, he/she ought to return to the main metronomic surface using the speed managed within these momentary micro-metronomes. Such performative and cognitive inversion, lies at the center of the Micro-Metric Modulation Theory.*

Keywords: *Micro-Metric Modulation. Ratios and Sub-ratios. Complex Rhythms. MicroMetronomes. Diverging Metric Configurations. Commutative and Associative Properties of Rhythm.*

I. The Metric Paradigm and the counting of beats

This paper intends to bring to the fore an unusual aspect of the metric unfolding: the one that is not associated to an immediate correlation to the metronomic pulse. One that suspends the *rhythmic anchor* of the pulsating metronome that regulates durations and rhythms in order to negotiate new rhythms that are farther away from the metronomic reference, but encapsulated in the inner musical fabric. What I am trying to show is that our usual way to deal with rhythms offers implicit mechanisms that can be of great help to achieve a broader understanding of the way the pulse is relative to its context and not an absolute construct that precedes every performative action. When complex rhythmic situations are at play, performers deviate continuously from the beat in order to go deeper within the interstices of a given rhythm. Consequently, a sudden loss of correlation between macro (metronomic) and micro (ratios and sub-ratios) tempi, starts to tear the musical logic of a predominant (above all) pulse, to entertain proximal references regarding the agency of a neighboring subdivision. It is necessary, at farther deviations from the main beat or metronomic pulse, to create emergent, new *sub-metronomes* that will function as an instantaneous metric basis to perform a sound rhythmic correlation with the rhythm that immediately preceded it. [1]

*This article is one of the publications related to the III Congress of the Brazilian Association of Musical Theory and Analysis (TeMA) / IV International Congress on Music and Mathematics, held in Rio de Janeiro, from October 21 to 25, 2019. Full information about the event can be accessed at https://ppgm.musica.ufrj.br/tema-musmat-congress/

[1] The system and concepts presented in this paper are developed from the author's theoretical work on rhythm (i.e., Micro-Metric Modulation), and reflect some of his compositional and poetic practices (see [6] [7] [8] [9]).

i. *Il Tempo de la figura* (after Ferneyhough)

Figures 1 and 2 demonstrate the way a rhythm grid appears to an interpreter firstly anchored in the metronomic pulsation to slowly loose its reference relative to the said pulsation when new subdivisions farther away from the reference beat, start to appear.

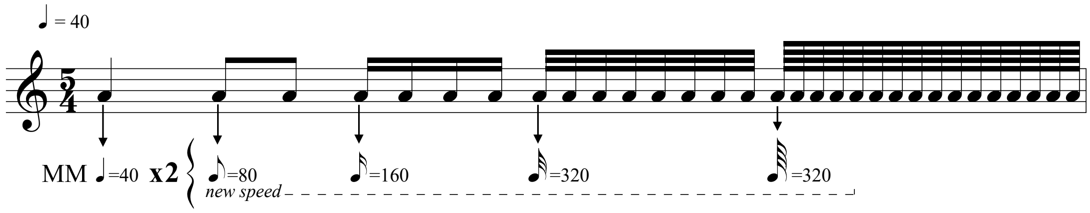


Figure 1: *Stronger to weaker relation to the Metronome given pulse/speed. However, they keep their tight correlation with the initial metronomic pulse as they present a clear* base 2 *subdivision with the initial beat/speed*

Note at Figure 1 that all the subdivisions are tightly correlated with the pulsation or the *feeling* of the main beat, which is the metronome. Even when smaller subdivisions are required they keep a strong reference to the metronome marking because a) they are direct subdivisions (halves and halves of halves) of the metronomic beat and b) because they act as a perceptual *gestalt*, that re-spells the main beat through a regular span (or distance) whose return is clearly expected even when smaller subdivisions are at play. Although rhythmic unfolding within a regular beat might contain irregular pulsations (i. e., a 2/4 bar can also be felt as the conjunction of a 3 + 5 rhythmic accentuation of dislocated sixteenth-notes) they are still *under the spell* of a propulsive metronomic regularity that reaffirms itself at each rhythmic cycle. This is extremely important for performers, since they can coordinate their respective and independent inner rhythms, the micro-fluctuations that their parts might contain, with a projected sum given by the main metronomic beat.

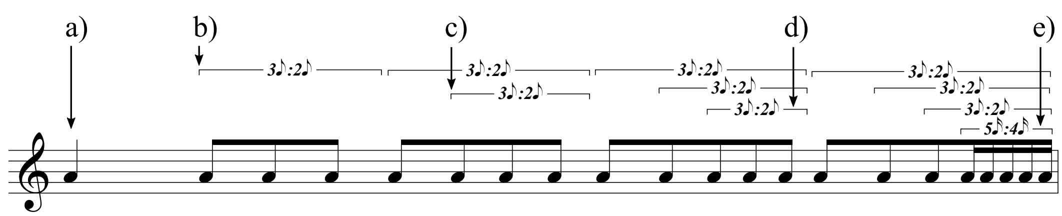


Figure 2: *Levels of subdivision:* a) *Zero or Neutral Level.* b) *First level of subdivision.* c) *Second level of subdivision.* d) *Third level of subdivision. e) Fourth level of subdivision.*

At Figure 2, a similar process is at play initially. Even if irregular subdivisions of the main beat (like 3 eighths in the space of 2 eighths) appear, this first deviation is still strongly related to the speed of the metronomic beat. Only when the process starts to acquire a further redundancy and the deviations are *compressed* towards an irregular part of the figure (like the ones found on the second, third and fourth levels of subdivision) then our notion of a projected beat defaulted by a predictable rhythmic regularity of the figure starts to weaken. As the main metronomic beat suffers an acceleration when novel subdivisions are formed, we are caught between two metric pulls: a suspended beat coming from the metronomic speed, and the inner metric subdivisions of the rhythmic figure. Notice that the metric figure is not a slow accelerando politely written and following therefore an even scale of speed. It is a metric route that demands specific rhythmic contours. To recap: while the metronomic beat acts as a frame of metric reference towards which all rhythms are subsumed, it is, nevertheless, slowly *suspended* when new metric deviations are called forth. Thus, in order for the performer to mediate between the idea of an even out acceleration whose purpose is just to cover the space of the metronomic beat, she/he will have to articulate such acceleration according to specific rhythmic demands written within it. This dichotomy between prospective, overarching beat, and the inner complexity of figural speeds, entails a mechanism that must be accounted for when there is the need of any type of rhythmic *carving* in the path of performing the figure and thus, crossing the span/speed of the metronomic beat.

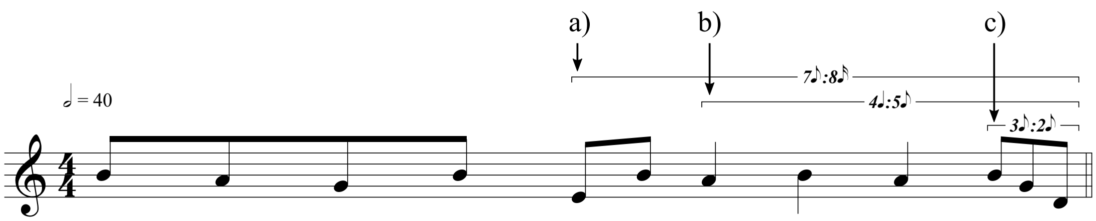


Figure 3: *Differing levels of speed to traverse the second half-note beat of the bar:* a) *First level of subdivion accelerates from metronomic beat's speed.* b) *Second level of subdivion decelerates (slower pulse than 1^{st} level).* c) *Third level of subdivion accelerates. Faster pulse than the second level.*

What exactly does that mean for the mind of the performer? It means, as shown in Figure 3 at second half-note beat, that a *gestaltian* comprehension of the path to be covered can't be solely overcome by the maintenance of the metronomic beat irrespective of the inner mediations of figural demands. Thus, a new concept emerges here where the prospective acceleration must be *halted* by the understanding of which speed exactly is the performer deviating from. As we can see, the performer can't simply accept that all rhythms accelerate towards a prospective beat in a random, or regular, speed. She/He ought to consider which speeds are being required so that every step is compatible with the inner logic, or rhythmic design, of the figure. Therefore, a new strategy to cross a specific metric figure emerges, one that creates *instantaneous* spans similar to the metronomic beat, that again, offers to the performer a new metric hierarchy from which a specific rhythmic speed will be articulated and properly fit. Notice too, another tantamount aspect regarding the flexibility of the rhythmic figure within this process. At the second-level of subdivision (4 notes in the time of 5's ratio) the rhythmic figure acquires a distinct metric configuration and it is shown as a quarter-note value instead of an eighth-note. This might seem to contradict somehow, at the level of figural representation, the first ratio layer, which is supposed to be slower than the second. Here, it is necessary to understand that similarities between figural representation and a hierarchy of speeds is illusory within the very system of metric subdivision. That's a lesson hard to be understood by the performer's intuitive grasp of rhythm, many times. It comes from the simple principle that any metric figure maintains its figural identity until it is twice as fast or twice as slower than its current configuration. Thus, a stream of sixteenth-notes, for instance, starting with 4 sixteenth-notes, won't change its configuration till it reaches a speed twice as fast, becoming finally a thirty-second note rhythmic figure. The same process happens in the slower direction. So, for instance, when a triplet that fits a quarter-note's duration is shown as a rhythmic figure comprising 3 eighth-notes and not 3 sixteenths it is because the quarter-note is first divided in half by two eighth-notes, then in three equal parts by 3 eighth-notes, and only when it has to fit 4 notes it acquires the figural representation of the sixteenth-note since it is twice as fast as the eighth-note. This logic serves for us to understand that when a slower tempo is called forth, as in the second-level of Figure 3, it is a consequence of similar process. It is easy to understand such process just noticing that the 5 eighth-notes (span of the ratio) might only double its figural configuration when reaching a next level of sixteenth-notes. Till then, it maintains the same eighth-note configuration. Only at the point that 10 sixteenth-notes occupy the same space as 5 eighth-notes, a new rhythmic configuration is formed. Thus, 4 quarter-notes belong to a metric hierarchy that precedes the rounding off tempo of 5 eighth-notes, as they come from a slower figural configuration. (Figure 4).

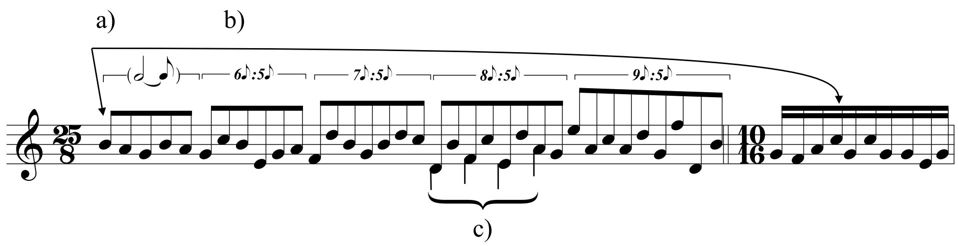


Figure 4: a) *Span of the figure (5 eighth-notes).* b) *It will only change its rhythmical configuration when it reaches a speed twice as fast becoming a sixteenth-note (10/16 bar). Thus, it is possible to see that the 4 quarter-notes placed as sub-ratio (see* Figure 3*), are simply the twice-slower configuration of the 8 eighth-notes of the 8:5 ratio.* c) *4 quarter-notes from a slower rhtyhmic layer forming the ratio 4:5 instead of the 8:5 above*

While in Figure 3 a constant negotiation between *accelerandos and decelerandos* is shown in order to cover the whole span of the beat, at Figure 2 a straight metric acceleration proposes a diverse rhythmic trajectory. Both examples, however, oblige the performer to *halt* the *speeding up* strategy according to an *emergent metronome* found at a deeper level of the rhythmic unfolding.

Thus, as we can observe at the last quarter (beat) of Figure 2, there are four types of metric accelerations being woven by the performer to cross this figure.

a) At the first-level of subdivision the performer is deeply connected with the metronomic beat from which rhythmic deviations are managed;

b) At the second-level of subdivision a new span hierarchy is at place: the performer has to mediate between the span of the metronomic beat, and the span of a part (two-thirds) of the triplet that was just managed at the first level of subdivision. At this point, a new *quarter-note* emerges. It is the result of the addition of the last two eighth-notes of the triplet. Note that this emergent *quarter-note* is not subsumed to the metronomic hierarchy's immediate subdivision, and therefore exhibits a different span/speed, even if its figural embodiment has an identical metric/figural correspondence with the metronomic quarter-note;

c) At the third-level of subdivision, the new triplet is found when the performer carves a new span or a new frame of metric reference out of the last two eighth-notes (or *legs*) of the triplet (located at second level of subdivision) in order to be able to *fit* its new triplet within these last two eighth-notes or *legs* of the figure. Again, this is managed through a similar method used above to acquiring a new quarter-note span out of the last two eighth-notes of the second-level triplet;

d) At the fourth level of subdivision the same method ensues. However, at this time, a new metric configuration is called forth. One that requires a quintuplet-sixteenth as the last rhythmic subdivision before a new prospective beat at the *upper* or neutral-level of the metronome marking is reached.

The propelling energy of the metronomic span encapsulates a regularity that coordinates all types of subdivisions found within a regular beat. Only when under the *spell* of a metronomic *regularity* we manage to pre-calculate the necessary steps we will need to cross it, no matter if these steps are built with regular or irregular subdivisions of the beat. Thus, with the reference of the metronomic beat it is equally possible to envision a strategy, given a certain amount of time or a specific span, for taking for instance, either *seven steps* (septuplet-sixteenths) or *four steps* (four regular sixteenths) to effectively cross the proposed beat or the metronomic span.

To make things transparent, under a certain metronome we have regular subdivisions of eighths, sixteenths, thirty-seconds, sixty-fourths or, for that matter, even slower values like the half and whole, notes, as shown in Figure 1. This simple fact underlies a theoretical construct that foments rhythmic contrast and diversity within the combinatorial vectors of the Western Music's rhythmic practice. They act as well as a great *cognitive glue* to arrange *packets* of diverse, albeit limited, number of figures in repetitive gestural 'condensates' which reinforces mechanical associations with a cemented understanding of rhythm.

Nevertheless, it is possible to combine these regular subdivisions of the main beat to forge a rhythmic discourse of highly contrasted semantic outcomes. More, these rudimentary tools of treating time as a regular multiple or divisor of the main beat, entail a huge combinatorial output. However, as much as the contrasting potential amounts to a great metric flexibility of the rhythm, they reveal, as well, an underlying understanding of meter that becomes detached from the way *Time*, as a general and multidimensional phenomenon, behaves and unfolds. In fact, they promote a deviation from the rhythmic units that, in themselves, are just isolated *syllables* of a yet non-spelled aggregate. In order for these rhythmic units to transpose their mere mechanical and arithmetic relationship they need to create a hierarchy of figures that become associated with gestures that fit the metronomic span. For instance, they can unite to create diverse arrangements of the units used for subdividing time: syncopations, an eighth-note followed by two sixteenth-notes, four sixteenth/thirty-second/sixty-fourth-notes in a row, among innumerable arrangements. They start to point to a state-of-affairs where the very apprehension of time is mediated by the perceptual gestalt of the figure as a unit. As they condense the inner rhythmic flexibility given by isolated subdivisions of the beat (no matter which particular one is being used), a higher order of hierarchy is foregrounded to the performer's perceptual apprehension. Some of these units, when placed together, or adjacent to each other, form *gestures* that suspend momentarily the counting of micro-rhythms derived from the beat's speed. They enhance the view of a larger or higher gestural coherence, creating a quasi-mnemonic association with a specific rhythmic object. Similar to the way we understand words within a phrase. We are not spelling every letter in order to construct the word. We are apprehending a *syntactical rule* that underlies such construction or correlation. In fact, seeing a known arrangement of rhythmic units can be compared to dealing with pieces of legos. These small pockets of information become, by and in themselves, the very underlying fabric of rhythmic construct. As we can see, there is a *slight departure* of the metric subdivision principle, to favor the faster apprehension of information as a whole. One that is build (or cemented) in known monads of gestural and figural detachment. The cognitive apparatus of the interpreter apprehends such new metric hierarchies as a gestalt, and it is brought, by fastidious practicing, to the level of ergonomic memorization. Scales and phrases are formed to guarantee the agreement of a diverse arrangement of these types of second-order metric aggregates. Consequently, an *aesthetic* is built based in the modular rationalization of information into contrasting and yet, repetitive possibilities. Such state-of-affairs is none but the embodiment of a comprehensive, privileged, view, that divides the world into poles of assimilation and irrelevance. There is much more in the story/history of rhythmic unfolding than a mere theoretical construct of base two might address in spite of its arithmetical *neutrality*. And I advance, that no system, no matter how comprehensive, will be able to provide a complete apprehension of time within any graphic or written, representative, model.

ii. Towards a parametric interchangeability of tempo, meter, figure and duration

After the above intro and a rudimentary demonstration of the evolving state of metric subdivision within Western Music, emphasizing the very system's apprehension and appropriation of time/duration, it is possible, at least tentatively, to peruse a more general understanding of some of the implications of metric subdivisions. Regarding new possibilities of the metric unfolding it is important to understand the contrasting and very practical aspects that differentiate the concept or parameter of span from the concept or parameter of rhythm. *Span* means, literally, the full extent of something from end to end; or, in other words, the full amount of space that something covers. In musical terms, span might be seen as accretions of rhythms whose durational trajectory traverses a certain amount of time. For our purposes in this paper we will simply state that the accretion of speeds that differ from each other in terms of non-correspondent numerical/rhythmical scales are boundaries that lie outside the scope of a specific, regular, stream or span. A span must have an underlying common denominator speed that is graspable by a specific rhythmic figure. While we can measure the span that a regular pulsation of sixteenth-notes covers (for example, 3, 4 or 7 sixteenth-notes, at Figure 5a), we can't possibly grasp (within the musical framework) the durational span that starts, say, at the third quintuplet-sixteenth of a quintuplet rhythm and stops at the 5 th septuplet-sixteenth of a septuplet rhythm (see Figure 5b). The latter does not cohere as a regular totality from which we can calculate a homogeneous quantity having a common, underlying rhythm.

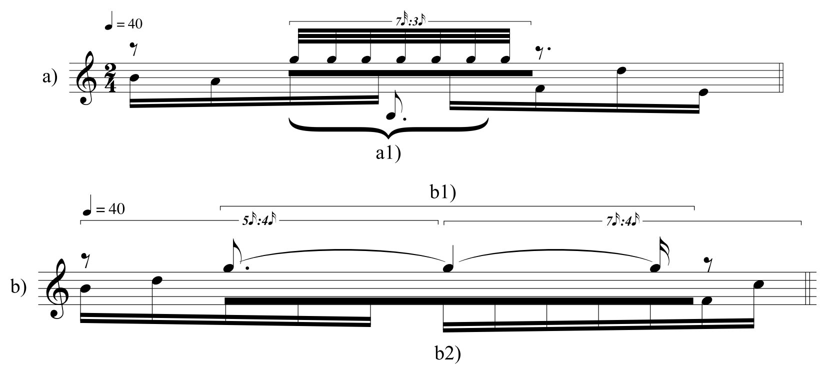


Figure 5: a) *Representation of a regular span from which a new subdivision might occur.* a1) *Metric span coherent to be notated as one isolated, whole, rhythmic figure, a dotted-eighth in this case. We can add any amount of notes from this regular pulsation of sixteenth-notes to create a regular stream/span from which new rhythms can be inferred from.* b) *Undefined metric span not coherent to be notated as one isolated rhythmic figure.* b1) *What's the size and figural representation of this span? With which coesive and regular rhythmic figure you can cross this span? Not possible since it accelerates and looses common ground from which you could deviate from.* b2) *Undefined metric span not coherent to be notated as one isolated, whole, rhythmic figure. While the total of the span can be thought of as an addition of two distinct rhythmic figures they can't cohere into a specific rhythmic unity since they belong to two differing metric hierarchies or two unrelated scales of speed.*

The second parametric aspect of time is rhythm. Rhythm cannot be understood as a diverse pull of quantifiable elements forming disconnected aggregates. It is not a mere game of accretions and eliminations whose parametric independence might be measured as an isolated layer of materials. As said before, all elements of duration are interconnected. (We also know that we can define rhythm as the speed of frequencies that, if above a certain acoustic threshold, enables the perceptual foregrounding of pitch, another parametric strata of time.) At a very basic level, rhythm is but the intervallic obviation of flowing time, not its concrete representation. It is possible to demonstrate even rudimentary that rhythm can be understood, at a primary level, as a metric concomitance with the metronomic speed. At that point, the metric parallelism between metronome and rhythmic pulse is leveled. Both are heard as one. The maintenance of musical duration is given by a periodicity that needs to reenact itself to reacquire *musical* pertinence. In order to re-potentialize the audibility of a repetitive *chunk of information* that otherwise would fade in the acoustic background, it is necessary the introduction of small disturbances. Therefore, whenever a rhythmic deviation from the beat is introduced we start to hear back the very beat. The idea of rhythm as an intervallic default of the metronomic is more akin of my understanding of the concept of rhythm. And it makes it less petrified and more fluid to the idea of parametric interchangeability. Between the *vertical* prospectiveness of the metronomic and the *horizontal* filling of micro-units of time, lies no boundaries. The musical figure might be re-contextualized and detach from the grip of the regular metronomic subdivision when the metric figure itself contributes to create an independent, asymmetric level of perceptual spans that neutralizes, at least temporarily, the regularity of the overarching, metronomic, beat/speed.

Figures 6 and 7a illustrate two instances of parametric interchangeability: one of metric subordination (Figure 5, where rhythm is presented as a polyphonic strand), the other of detachment (between metronomic prospectiveness and rhythm-as-new-metronome), where the eruption of asymmetric figuration starts to impose itself as perceptual irregularities. At Figure 6 two scales of metric units are at play: one (sixteenth-notes) belongs to a regular, repetitive unit, that divides the metronomic beat into small equal parts. Running in parallel and below it (bass line), we find rhythmic figures of differing spans. These figures present either regular or irregular *sizes* or durations. However, under the performer's perspective, these accretions of notes are not immediately perceived as a metric discrepancy weakening the main beat's coordinates. They are polyphonic strands still subsumed to the underlying stream of sixteenth-notes and therefore seen (or read) as a part of the beat's metric unfolding and regularity.

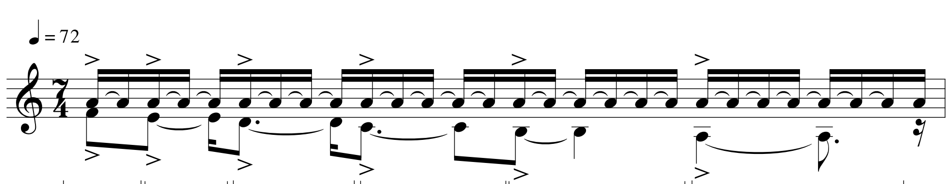


Figure 6: *Polyphonic strand of linked rhythms forming the bass line. They are just regular and irregular spans subsumed to the beat's subdivisions. Because of their strong connection with the Metronome's pulse they are not seen as detached, independent, isolated, rhythms yet.*

Notice, at Figure 7a, that when the bass line is written in such a way as to tentatively forge independent, more autonomous rhythmic figures, a new layer of information erupts, one that departs from the immediate *tutelage* of the counting beat. Their sudden graphic autonomy forming a metric scale that begins at an eighth-note and reaches the doubled-dotted quarter-note shows the totality of a rhythmic figure not obviously subscribed to a part of the beat's metronomic pulsation.

This clearly points out to a privileged stance of parametric interconnectedness: the rhythmic figure carved out of the metronomic pulse might acquire an implicit role as potential *new metronomic span*. Here, a new paradigm is at play that although not contradicting the pulsation of the main beat or metronome marking, weakens nonetheless, the feeling of regularity and prospectiveness. Thus, they can be suddenly understood as irregular accretions, becoming somehow anti-intuitive for the performer and consequently, not necessarily being felt as part of the metric narrative of the metronomic beat.

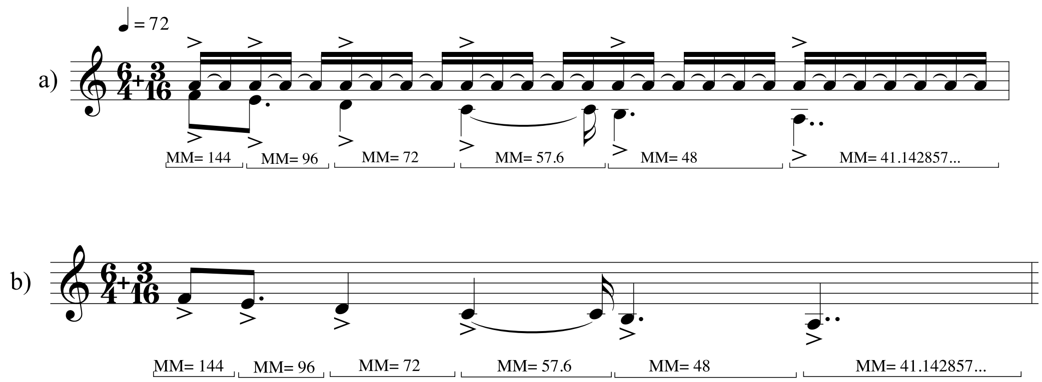


Figure 7: a) *The rhythmic figure within a specif metronomic marking (bass line) might be seen as a potential new Metronome Marking (MM).* b) *Below the bass line detached from the sixteenth-notes' context shows a clear independence and can acquire the dimension of durational span itself.*

These new durations might be considered as a quasi second-level of rhythmic deviation and are, in themselves, spans that might function as the base or reference for further deviation. In fact, the perception of *fertile ground* for deviation indicates that an independent terrain is available to insert novel rhythmic variations. They are the result of a cognitive inversion where the inner time of rhythmic units starts to be perceived as *nodes* of metric coordination, or small metronomic fields. What was initially seen as the very codification of purely gestural affairs, acquires now a privileged, metric, autonomy where new materials will reinstate with devious, clever, features, a subliminar metronomic pulse, reenergizing the underlying pertinence of the metronomic beat and the coherence of *the musical*.

Again, at Figures 7a and 7b the metronome marking placed below each rhythmic figure indicates, precisely, the amount of time necessary to cross a certain sum of sixteenth-notes under a specific, overarching, metronome marking. Thus, each MM can be understood within two immediate levels:

a) They reflect the speed of the sixteenth-note (taken here as a temporary figural unit) under a certain metronome marking (in this case quarter = 72);

b) They can also become in themselves, a sudden metronome, whose figural representation might be equally seen as a new quarter-note if there is a metric modulation to the span shown by them. For instance, a dotted eighth-note under the metronome marking of quarter = 72, shows a (faster) sum of MM = 96. Its figural representation is, therefore, of a dotted eighth-note. However, if I want to metric modulate to quarter = 96, the figural representation of these three sixteenth-notes' aggregate that fill the dotted-eighth under MM=72, ought to suffer a new figural representation when addressed as quarter = 96 and, under such metronomic beat, it will be seen as an eighth-note triplet (Figure 8).

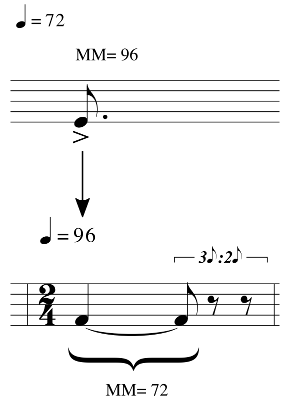


Figure 8: *Metric equivalency between distinct rhythmic figures: a dotted-eighth MM = 72 is rhythmically identical to a quarter = 96. Thus, the sixteenth-note and the triplet share a similar rhythmic speed*

There are three variables that work together to coordinate rhythmic hierarchies. At the primary level of duration, we will encounter the metronome marking variable, setting the overall speed of the (*regular*) beat, under which all other rhythmic elements will be subsumed. The second element is the meter (time signature) where a certain amount of rhythmic information will be *squeezed into* such that it creates perceptible (or countable) frames of reference to the fluency of musical materials. It functions as a type of *sieve* (filter) where groups of notes can comfortably fit and accommodate a certain number of beats, or part of it. It promotes the understanding of phrasings as well, as it helps to *pack* rhythm in diverse monads of energy, justifying the underlying waves of the text's *semantic fluency* and rationalizes the distribution of information. The third aspect is the musical figure (i.e., rhythm) as the residual consistency of periodic quantities whose fast parametric coordination exhibits the most condensed way to join spans of divergent speeds. It works by *conjoining or atomizing* a diverse array of durations derived from subdivisions of the main beat, forming small groupings of regular or irregular assemblages. These variables, seen initially as independent parameters, are co-dependent and work as the underlying fabric of our apprehension of time and durations. Figure 9 shows the parametric exchange between rhythmic figures.

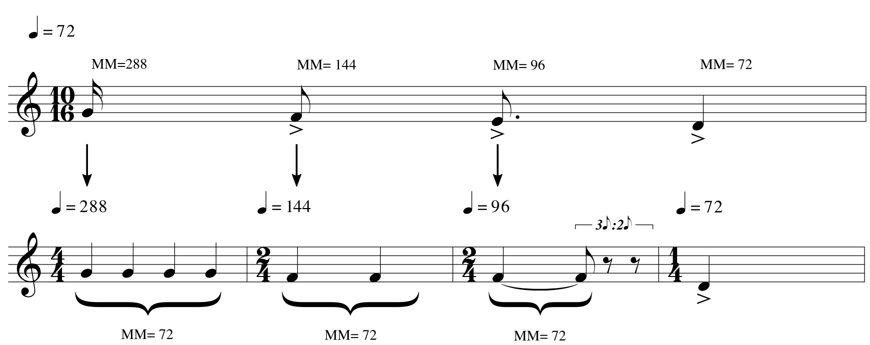


Figure 9: *Parametric exchange between rhythmic figures.*

If every span derived from Figure 9 is understood as a new metronome marking, and not uniquely as a rhythmic figure merely subsumed to the metronomic time, whose intrinsic duration covers a certain amount of sixteenth-notes, then we are forced to reconfigure the two other variables of meter and rhythm in order to create a new environment and presentation for that specific durational *quantity*. Thus, while the equivalency of durational quantities is maintained across two different rhythmic configurations (or metronome markings), their figural and metric representations are now re-contextualized under a new beat or metronomic speed.

Notice, at Figure 9, how we depart from the very tiny figure of the sixteenth-note under the metronome marking of MM=72 to a first equivalency that transforms the speed of the note (that runs at a speed of 288 beats per minute, which is 4 x 72) into a new metric coordinate redressed as metronome marking. Not only that, we decided arbitrarily and for the sake of clarity, to suppose that this new metronome is relative to a quarter-note figure. (We certainly could have decided that it belongs to any other figural equivalency: either an eighth-note, a dotted-eighth-note, a whole-note, etc. And this would imply a distinct reconfiguration of meter and rhythm, obviously.) The new metronomic quarter-note that beats at the speed of 288 beats per minute, has to manifest itself under a new metric and figural representation. Having a metronome of 288 implies that the quarter-note is fitted under a specific metric frame that is subsumed to a clear metronomic correlation. Thus, I could choose any time-signature that has the capacity to encapsulate such quarter-note. I can fit it in a 1/4, 2/4, 3/4, 4/4, etc., bar, under MM=288. I choose a 4/4 bar to keep a strong equivalency to the metronome marking's span of 72 where 4 sixteenth-notes suffice to be encapsulated in the total metronomic span. The next figure which transforms the speed of the note into a new metronome, has similar characteristics, as it is strongly related to a regular part of the beat. The same operation ensues (2 x 72 or 288/2 = MM=144). Here, the span acquired is equivalent to the exact span of 2 sixteenth-notes under MM=72. Again, I create a new quarter-note metronome and consequently a new metric frame of reference in order to fulfill an entire beat. My choice of bar is relative to context but is still strongly related to the beat hierarchy as it maintains a regular coordination with it. Again, under MM=144, I choose a 2/4 bar, since a quarter-note in MM=144 fills the equivalent span of an eighth-note under MM=72. But since I want to cover the MM=72 span I add to the bar one extra quarter-note/beat. Next, I have a dotted eighth-note, an irregular quantity formed by the sum of 3 sixteenth-notes under MM=72. The span of such figure is MM=96 (MM 72 x 4/3). This span can also be converted into a quarter-note durational/metronomic span. Such choice is deliberate in order to exemplify a new exchange that happens this time at the level of the figure itself. When choosing to represent a dotted, irregular figure, into an equivalent one however subsumed to a regular representation (a quarter-note), I am obliged, as well, to reconfigure the rhythmic figure and the time signature in order to fit it within this new metronomic marking. So, if my new quarter-note metronome indicates a speed of MM = 96 beats per minute, I have to fit 3 notes under this beat. When my metronome marking was MM=72, I needed 3 sixteenths-notes to fill that figure. As my new metronome marking is quarter equals MM=96, the three sixteenth-notes have to change their figural representation in order to acquire a metric correlation with the new metronome. Thus, if three equally spaced rhythms have to fit this new quarter, I can only place an eighth-note triplet as a valid figural equivalency with the speed of three sixteenth-notes under MM=72.

The precedent demonstration brings us home towards the understanding of the equivalencies and discrepancies between rhythms and metric figures. It implies the parametric interchangeability as a temporal given under which the hierarchization of figures and speeds are subsumed to a flexible coordination of variables. When one aspect of the durational hierarchy is changed all others follow suit and exhibit a novel configuration either at the metronomic, metric or rhythmic levels.

Figure 10 illustrates a typical metric modulation between MM=60 and MM=75. To pass from

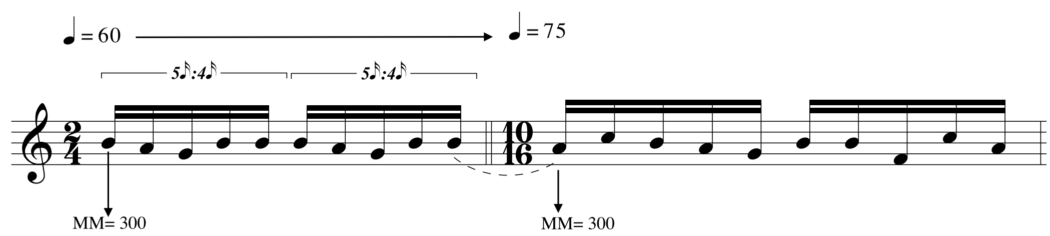


Figure 10: *Realization of sharing possibilities between meter, rhythms and metronomes and calculating the parametric exchange between the components of duration.*

the first MM to the next, it is necessary to understand how these two metronomes connect through a common denominator's speed. For that, it is necessary to factor each metronomic number in order to find how they are formed and their numerical similarities. Below, the resulting factorization of each MM: $\frac{60}{75} = \frac{2235}{355}$

Since we are going from MM=60 toward MM=75, we are trying to subdivide 60 into 75. Or to be more transparent, we are trying to figure what type of operation will enable us to match their speeds. When numbers are factorized they exhibit the building blocks of their total span/speed. For example, 60 is first divided by 2, resulting in = 30; then it is again divided by 2, because 30 is still a multiple of 2 and therefore can be divided by it which results in = 15. At this point the smallest prime factor of 15 is 3, (not 2) which results in = 5; finally, 5 is only divided by 5, which brings the division to an end, as it results in = 1. If you do the same operation for the denominator which is 75, you will obtain the three lower numbers in the fraction above (3, 3, 5). At this point we can see how these numbers match into each other. Thus, if a fraction contains 3 at the top and 3 at the bottom these numbers can be scratched off as they cancel each other. The same happens with the 5 located at both, the top and bottom of the fraction. What is left is the most condensed form of the fractional expression between both, numerator and denominator, which is $\frac{4}{5}$ because we can't eliminate the 2's in the numerator as there are no 2's in the denominator neither the 5 in the denominator for similar reason. The 2's in the numerator are then multiplied and become a 4. At the denominator, the only number left is the 5 as the other one was previously cancelled by its counterpart at the numerator. Now, we can clearly see through this fractional expression, that in order for one to go from MM=60 to MM=75 and even out their speeds, or, in other words, to match them, it is necessary to multiply (or accelerate) MM=60 by the $\frac{4}{5}$ fraction. That amount will suffice to find common rhythmic ground between both metronomic speeds. I want to use the example just discussed to call attention for the final figural representation of both bars shown. If I may use a bit of *poetic license* to illustrate the operation, I could describe it by saying that an *invisible knob* was turned so that MM=60 is now suddenly seen as MM=75. What happens next with the other parameters when basic speed information is changed? The *metric machine* is now obliged to convert meter and rhythm as well in order to convey the same relationships we found previously when our metronomic beat's pulse was at a slower speed. Note that at MM=60 we have a meter/bar of 2/4 (that could be either 8/16 or 4/8), and a rhythmic stream of quintuplet-sixteenths. On the side of MM=75 we will encounter a completely different configuration for meter and rhythm. This points already to the fact that the rhythmic figures that exhibit dissimilarities in their presentation or configuration, are not necessarily implying discrepancy of speeds. An eighth-note under a certain metronomic speed or under a sub-ratio (which can be seen as another *buried manifestation of a metronome*) might be faster than a sixteenth note under a slower metronomic tempo or a comparative rhythmical (sub-ratio) speed.

There are many strategies for sharing rhythms and we could therefore program them in endless ways (Figures 11, 12, 13, and 14).

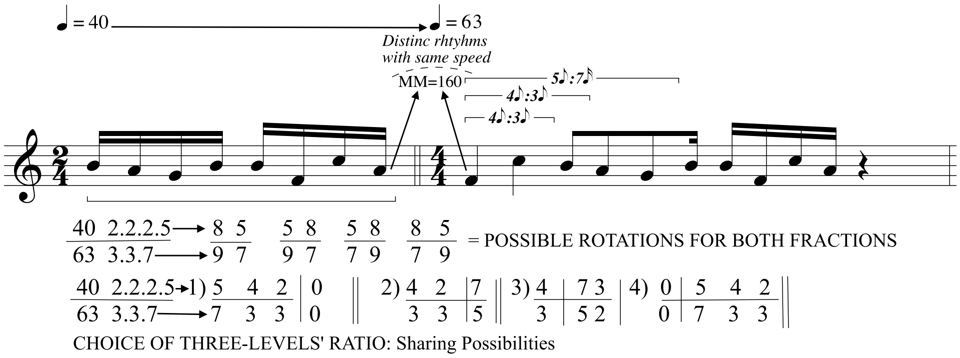


Figure 11: Micro-Metric Modulation*: The exile of the metric in the dance of pulsation*

The *sharing strategy* is a fundamental feature to understand the unfolding of the rhythmic parameter within the Western Music's practice. At Figure 11 we are showing the distributive aspect of factorization. This means that I can express the metronomic speed using any number's product within the smallest prime numbers found within a metronome's factorization. Without changing the total of the factored number it is possible to show any arrangement of numbers that presents the same result, if multiplied. This implies a very important feature inscribed in the factorization: that the amount of numbers found within it reveals the maximum levels of subdivision within which such number can be dismembered. Thus, at Figure 10, the factorization of both metronomes (MM= 40 and MM=63) shows the smallest prime numbers for each of these speeds: $\frac{40}{63} = \frac{2225}{337}$. Initially, the factorization shows that there are at least four layers of possibilities to account for in the distribution of the metronomic fraction. (These layers by the way, map the numerator into the denominator, as I will soon show.) No matter how we shuffle their order of appearance, we will have a maximum of fractional configurations conditioned by the metronome which presents a larger amount of numbers. To exemplify:

a) *Layers of configuration and mapping*: The fractional configuration that maps 40 into 63 can be expressed into a maximum of these four levels: $\frac{40}{63} = \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{2}{7} \cdot \frac{5}{1}$. Shuffling imparts no alteration of results for mapping 40 into 63. Thus, the above fraction could also have an alternate order: $\frac{40}{63} = \frac{2}{7} \cdot \frac{2}{3} \cdot \frac{5}{1} \cdot \frac{2}{3}$. Obviously, that this is a mathematical redundancy. But in musical terms some aspects of this *mathematical nonsense* can be extremely fertile as it will be soon shown, and express a new layer of information buried within a new metric shift, or configuration.

b) *The distributive factor*: Taking these numbers into account means that every arrangement of them is a valid configuration of the total of the metronome's product. At the first numeric row at the Figure 11, we can clearly see that we had distributed the factors of both metronomes into two fractions, each being a part of the total product found within their respective numbers. Since the order of the factor does not alter the final product, we are able to exchange their positions at will. Again, what is a mathematical redundancy becomes a fertile strategy to deal with rhythmic outcomes. Figure 12 illustrates the 4 possibilities of distribution found within the arrangement of the fractions. Note that there are 4 rotations for the fraction:

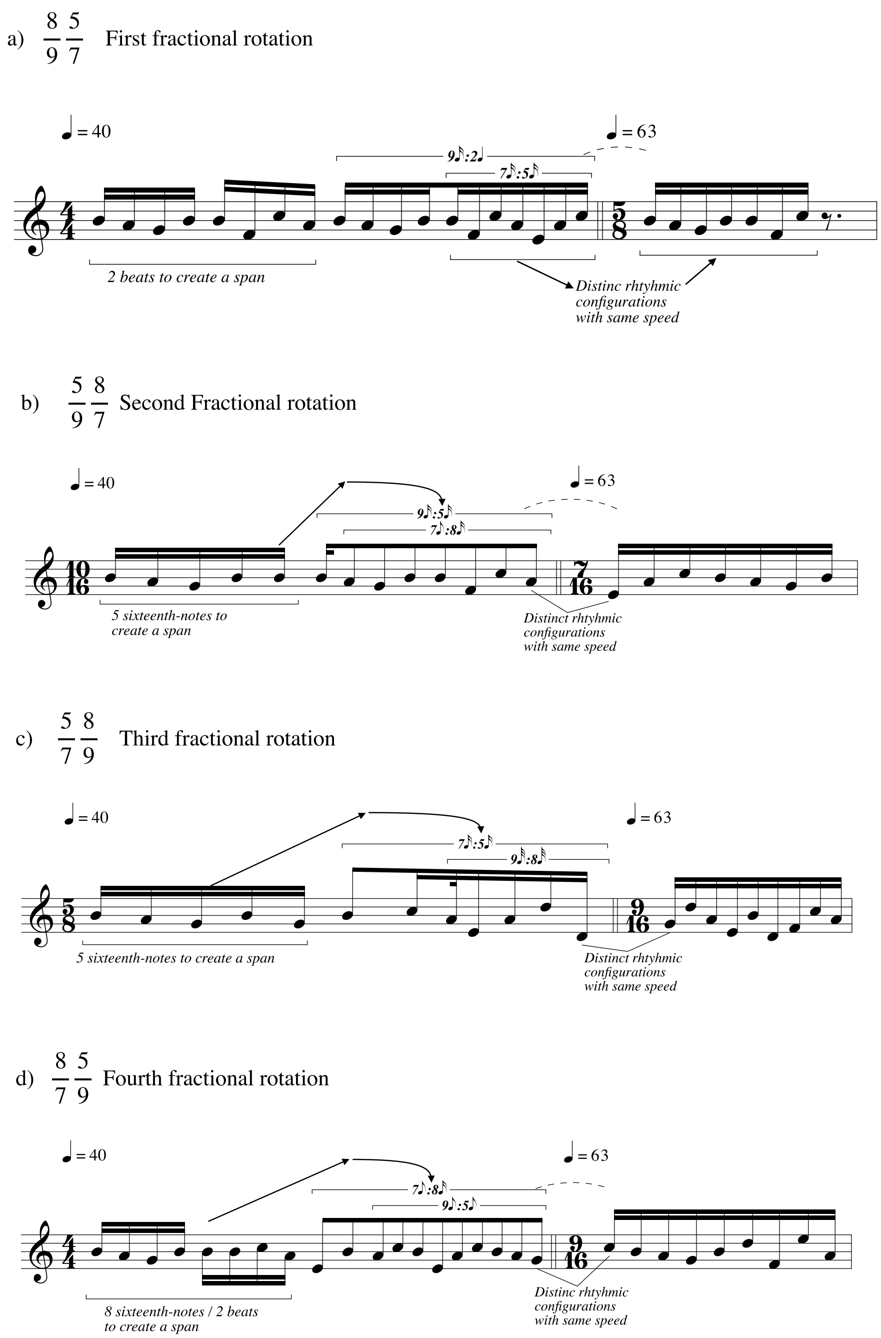
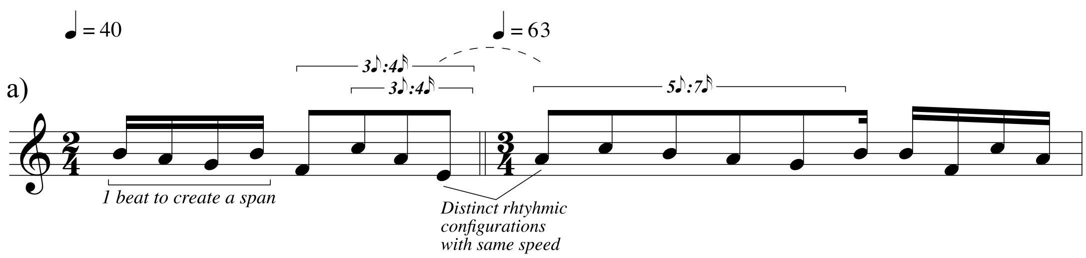


Figure 12: *Fractional Rotations and novel rhythmic outcomes.*

At the examples of Figures 11 and 12, all the operations are done in order to alter the speed of MM= 40 such that a certain rhythmic output is generated that maps this metronome's speed onto the other. However, such mechanism can be easily distributed between both metronomic speeds such that operations are equally shared between both sides of the metronomic chain. To recap, from MM= 40 to MM= 63 we kept the numerator as the departure metronome and the denominator's speed as the arrival point. The arrival metronome didn't go through any type of metric alteration to match the departure metronome. In order to distribute the rhythmic layers between both we need also to create the conditions to map the MM= 63 speed onto the MM= 40 metronomic speed. As we invert the metronomic positions (numerator goes to denominator and vice-versa) we are able to forge a rhythmical bridge towards the slower metronome. $\frac{63}{40} = \frac{337}{2225}$. It is important to see that mapping one number onto another implies cancelling each other's till no number is left. Figure 13 shows some examples where both metronomes are rhythmically *twisted* such that the resulting speeds of each separate rhythmical configuration maps onto one another. This is the principle of *Micro-Metric Modulation* as we have more than one layer to map at the sub-ratio level and we are not necessarily addressing just metronomic changes.



$$\frac{40}{63} = \frac{2}{3}\cdot\frac{2}{3}\cdot\frac{2}{7}\cdot\frac{5}{1} \qquad \frac{63}{40} = \frac{3}{2}\cdot\frac{3}{2}\cdot\frac{7}{2}\cdot\frac{1}{5}$$

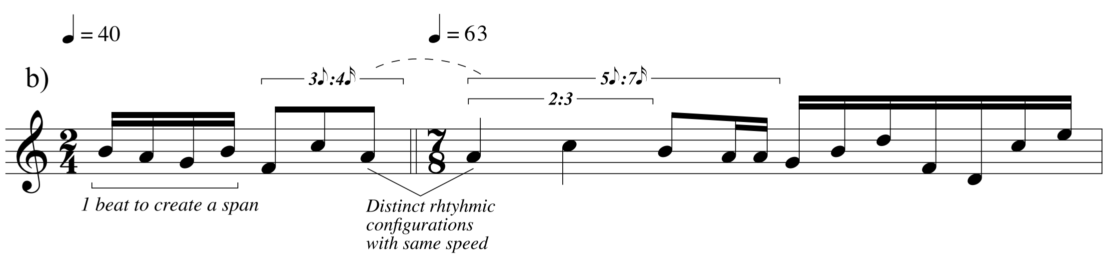


Figure 13: *Mapping strategies to bridge speeds between both sides of the rhythmic chain.* a) *The two-level ratio at the MM= 40 maps onto the one level ratio at MM= 63.Note that instead of simply matching the Metronome's speed at the other side/bar, the two-level figure at he first bar addressed the one-level, irregular, rhythmic figure, at the other side under a distinct metronomic speed.* b) *The one-level ratio under MM= 40 maps onto the two-level ratio at MM= 63. Note that the rhtyhm is distributed between both metronomic speeds. Beyond that, the metric figure at the first bar does not match a similar rhtyhmic figure at the other. Its eighth-note configuration does not match the quarter-note figure on the other side, even if their speeds are the same.*

Whichever fraction is used in one side it is immediately cancelled or eliminated on the other. Above, the first two fractions at the $\frac{40}{63}$ side are used to form both ratios (3 eighth-notes in the space of 2 eighth-notes); thus, they are scratched at the $\frac{63}{40}$ side. At the second bar we used the $\frac{7}{5}$ fraction cancelling out these numbers at the left side ($\frac{40}{63}$ side). Proceeding that way, all the numbers that map one metronome onto the other were used in a distributive fashion. Such operation opens up huge metric possibilities using an elegant formula to bridge rhythms that present distinct metric configurations. And this led us to the *Micro-Metric Modulation* perspective/theory.

The method for finding similar speeds or a numeric equivalency between rhythms is clearly demonstrated by a simple mathematical device: the commutative and associative properties. Any number can be dismembered in smaller multiples. As seen in the previous examples, when fractional representation of larger sums were shown as the product of smaller quantities it became clear that in order to bridge and match rhythmic speeds, it was necessary to coordinate variables found within the metronomic pulse, the meter, and the metric figure context in order to bridge speeds and irrespective figural configurations. Mathematically, the commutative and associative property guarantees that the order of factors will not alter the final product. Therefore, [4 x 5/6] = [4 x 6/5] = 30 for both. The final product of both equations can be easily seen represented in Figure 14.

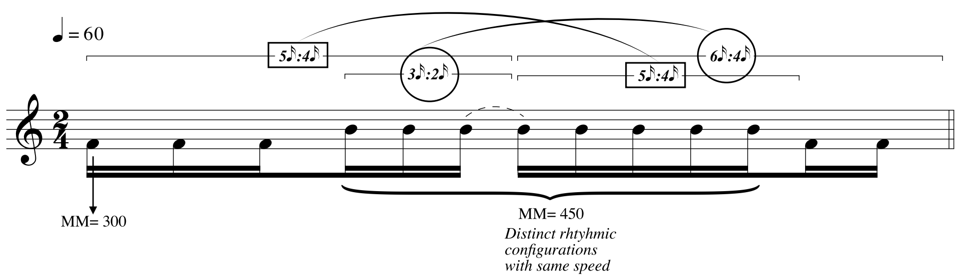


Figure 14: *Commutative and associative properties applied to rhythm: equivalency of speeds on both sides of the ratio chain.*

At the first quintuplet figure the following result was produced: MM60 x 5 quintuplet-sixteenths = MM=300 for each "leg" of the quintuplet sixteenth. Then, we need to acquire the span of 2 quintuplet-sixteenths to fit 3 sixteenth-note triplets within the same space. The span will be twice as slower 300/2 = 150. After that, we multiply the span of 150 x 3 sixteenth-note triplets = 450. That's the maximum speed reached by the first quintuplet figure. The same is done with the next 6:4 ratio. MM60 x 6 = 360/4 (to calculate the span of 4 sextuplet-sixteenths = 90. Within this span we need to fit 5 quintuplet-sixteenths. 90 x 5 = 45 which again confirms the matching of the sub-ratios' speed at both sides of the ratio chain. The operation illustrated above, a typical example of Micro-Metric Modulation, exemplifies some of the most important attributes of rhythmic/metric unfolding:

a) The common route or rhythmic bridge between apparently differing configurations is able to link rhythmically these configurations, offering enough ground for further subdivisions to occur since all the rhythms conjoined by the total span located between both rhythmic configurations can be subsumed to rhythmic deviation, or new metric subdivisions.

b) That in order to enter the next ratio (6:4) a new perspective is given to the performer. She/He has to be able to lift the overall metronomic pulse to blindly enter the next sub-ratio and consequently, the terrain of the next rhythm, attaching him/herself to the speed generated by the sixteenth-note triplets located under the first quintuplet figure. That way the performer can be sure that rhythmic precision linking both configurations is attained. As she/he enters the next rhythm through a sub-ratio the performer will be forced to reconfigure the span to be crossed. In the above case, when she/he enters the 6:4 ratio at the right through a sub-ratio of quintuplet sixteenths, it will be necessary to figure the "emergent/temporary metronome" of a quarter-note since she/he is crossing exactly 4 sixteenth-notes belonging to the 6:4 ratio above. At this point, there is a strategic counting inversion taking place. In order to reach back to the rhythmic surface the performer has to understand that the emergent quarter-note just crossed indicates the rhythmic equivalent of 4 sixteenth-notes under a 6:4 ratio. But being currently subsumed to a subdivision that fits the surface's ratio rhythmic grid, it becomes easier to calculate the speed of the last two sixteenth-notes at the end of the 6:4 ratio - as they belong to the surface level of the very ratio (first level of subdivision).

The theory of Micro-Metric Modulation has a huge wealth of resources to bridge dicothomic figures of regular or irregular configurations as it calculates common rhythms with corresponding speeds. The most important feature opened up by such technique, one that brings immense possibilities to the development of Western Music's rhythmic canon, relates to the passage of sub-ratios between differing ratios' configurations. That is a first in Western Music showing the flexibility not yet thoroughly exploited within the musical metric system. I would like to end this essay with an example of this very last, unique rhythmic feature, I used in several of my pieces. [2]

In the piece *"...B..."* for 10 instrumental soloists, video and electronics premiered in 2012 in Darmstadt by the Linea Ensemble of Strasbourg, it is possible to notice the passing of sub-ratios within sub-ratios as the common rhythm between these configurations was known (Figure 15a and 15b). First an isolated case to make the feature clear and subsequently the first metric page of the score where these novel rhythmic techniques are used simultaneously by the instrumental forces (Figure 16).

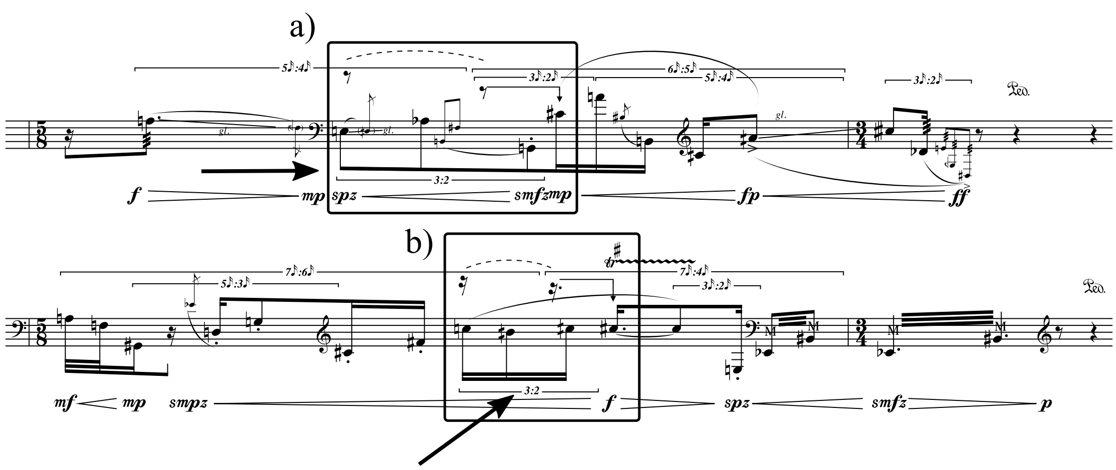


Figure 15: *Fragment of "...B..." for 10 instrumental soloists, video and electronics (2012) from the horn and the trumpet parts. Premiered by Linea Ensemble of Strasbourg - Darmstadt (2012). Differently from the crossing of the sub-ratio at the top staff, the bottom staff places its sub-ratio of three sixteenth-notes triplet crossing two distinct metric figures: under the 7:6 ratio (where the triplet starts) we see a sixteenth-note rhythm (shown by the little pause above). On the other side at the top ratio configuration of 7:4, we see a dotted sixteenth-note (shown as well by a dotted sixteenth-note little pause). After factoring both top ratios, mapping one onto the other (as it was shown on previous examples), an important feature of MMM is foregrounded: one that clearly demonstrates the rhythmic flexibility of the figure, proving its rhythm is relative to context and not a cemented, given, rhythmic speed: both notes while exhibiting a diverse rhythmic configuration, have, nonetheless, the same speed.*

[2] For detailed assessment of similar resources in other pieces, see [1], [2], [3], [4], [5], and [10],

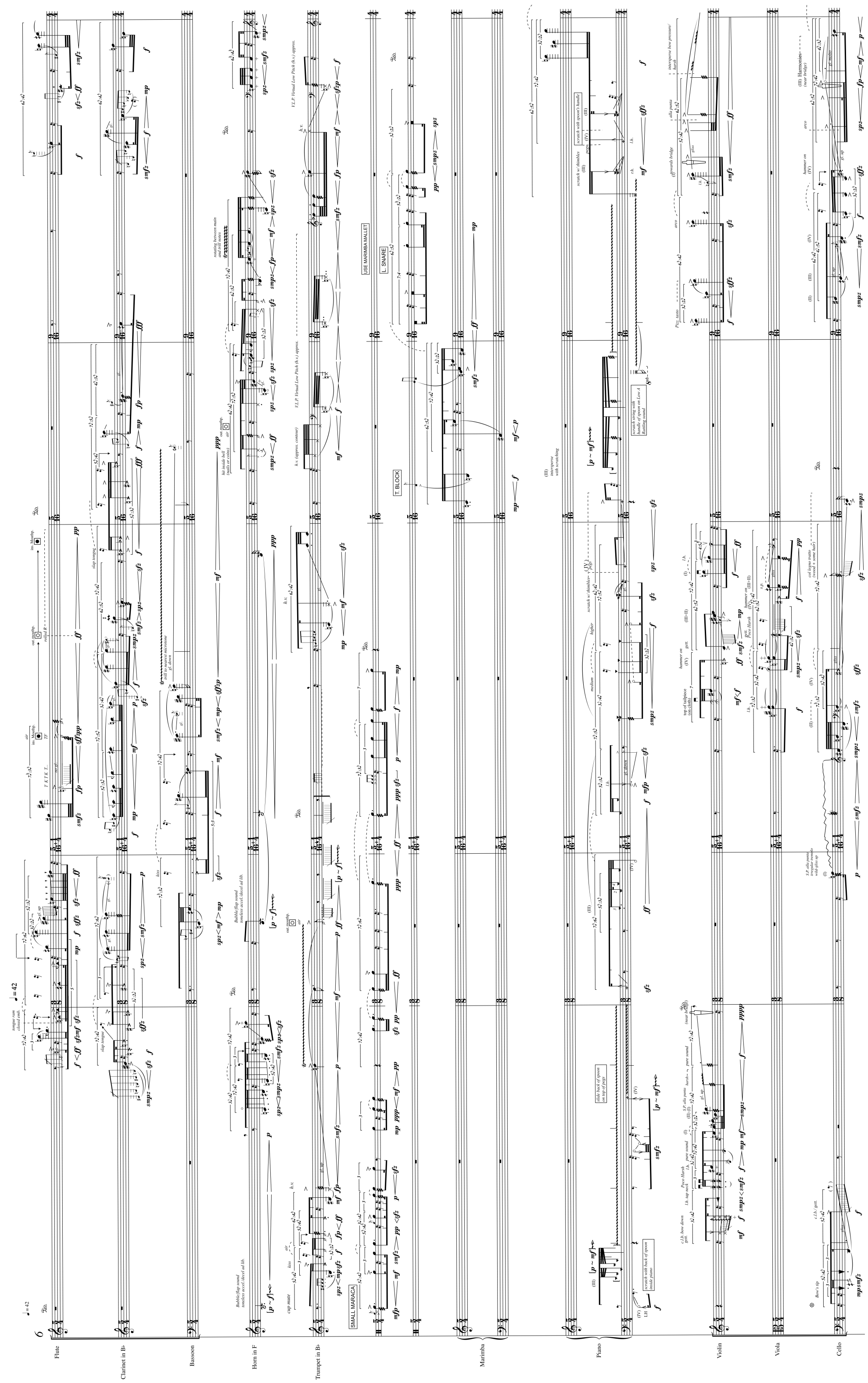


Figure 16: *First metric page of "...B..." for 10 instrumental soloists, video and electronics (2012). Premiered by Linea Ensemble of Strasbourg - Darmstadt*

References

[1] Bortz, G. (2003). *Rhythm in the Music of Brian Ferneyhough, Michael Finnisy, and Arthur Kampela*: A Guide For Performers. D.M.A Dissertation. City University of New York.

[2] Cury, F. (2006). *Percussion Study I, by Arthur Kampela*: A Guide for Interpreters. Masters Dissertation, Universidade Federal do Rio de Janeiro.

[3] Cury, F. (2013). "Análise dos elementos composicionais da obra violonística de Arthur Kampela". *Revista do Fórum Internacional de Estudos em Música e Dança*, 2(2): 78–87. https://proa.ua.pt/index.php/postip/issue/view/41.

[4] Cury, F. (2017). *A obra para violão solo de Arthur Kampela e suas implicações na performance.* D.M.A Dissertation, Universidade de Aveiro.

[5] Fernandes-Weiss, L. (2017). Le geste en lien avec le son: lecture cinétographique du jeu du guitariste dans *Percussion Study I* d'Arthur Kampela. In La Pensée Sonore du Corps: Corps, son et technologies entre théories et pratiques. *Proceedings...*, Nice: Université Côte d'Azur:1–18.

[6] Kampela, A. (1998). *Micro-Metric Modulation: New Directions in the Theory of complex Rhythms*. D.M.A Dissertation, Columbia University.

[7] Kampela, A. (2002). "A Knife All Blade: Deciding the Side Not to Take". *Current Musicology*, 67/68: 167–193.

[8] Kampela, A. (2012). *"...B..." for 10 instrumental soloists video and electronics*. Youtube Video. Ensemble Linea of Strasbourg. Jean-Philippe Wurtz, conductor. https://youtu.be/xh_xRnODctk.

[9] Kampela, A. (2013). Micro-Metric Rhythms and Noises: Emanations from the Stochastic Cloud. In Sharon Kanak (ed.) *Xenakis Matter*. New York: Pendragon Press: 341–362.

[10] Özçelebi, B. (2015). *An Analysis of Arthur Kampela's* Probe*, for 3 Female Voices and 2 Contrabass Clarinets (2015).* Master Dissertation, Hochschule für Musik und Theater Felix Mendelssohn Bartholdy.

Scales, Counterpoint Triples and their Groups

Marianthi Bozabalidou
bozapalidou@gmail.com
Orcid: 0000-0001-9796-5965.

Abstract: *General scale systems are defined to be linearly ordered finite sets of musical objects. Apart from the common pitch scales we may also speak of duration and interval scales, major and minor scale schemes, ancient greek trope scale schemes. The fundamental groups of a scale (clock group and group of rows) are discussed. The principal counterpoint triple of a scale Σ consists of the operators R_Σ (Σ-retrograde), T_Σ (Σ-transposition) and I_Σ (Σ-inversion). The group they generate will be referred to as a counterpoint group of Σ. A wide class of counterpoint triples is presented extending the composition material of n-tone music. Variations of twelve tone pieces may be derived by applying these triples. Counterpoint spaces (CP-spaces) are reachable left actions of counterpoint groups. Such an action is actually simply transitive. Major and minor chords are defined with respect to a pair (p, q) of natural numbers playing the role of major and minor thirds respectively. It is shown that (p, q)-consonant chords in a CP-space constitute a CP-space as well.*

***Keywords**: Scale, Clock and Row Groups of a Scale, Counterpoint Groups, Counterpoint Spaces.*

I. Introduction

Scale is a generic notion in music theory ([4], [5], [6], [8]). Mathematically speaking, a scale Σ is a linearly ordered finite set of musical objects called degrees of the scale

$$\Sigma : \sigma_0 < \sigma_1 < \cdots < \sigma_{n-1}$$

This general consideration allows us to unify various musical scale type situations: pitch scales (chromatic, diatonic, pentatonic, whole-tone, octatonic etc.), scales of durations and intervals, major and minor scale schemes, ancient greek trope scale schemes (section 2). The set of degrees $\Sigma_n = \{\sigma_0, \sigma_1, \ldots, \sigma_{n-1}\}$ with the clock addition

$$\sigma_\kappa \oplus \sigma_\lambda = \begin{cases} \sigma_{\kappa+\lambda} & \text{if } \kappa + \lambda < n, \\ \sigma_{\kappa+\lambda-n} & \text{if } \kappa + \lambda \geq n \end{cases}$$

form the first fundamental group $G_1(\Sigma)$ of Σ. Transposition T_Σ (one step shift upwards) and inversion I_Σ (reflection with respect to a fixed center) are exclusively defined in terms of $G_1(\Sigma)$:

$$T_\Sigma(\sigma_\kappa) = \sigma_\kappa \oplus \sigma_1, \quad I_\Sigma(\sigma_\kappa) = -\sigma_\kappa = \sigma_{n-\kappa}, \quad 0 \leq \kappa \leq n-1$$

A Σ-row is a permutation of the set of degrees of a scale Σ ([7], [9]). The set $S(\Sigma)$ of all Σ-rows is closed under composition and constitutes the second fundamental group of Σ, $G_2(\Sigma) = (S(\Sigma), \circ)$. $(R_\Sigma, T_\Sigma, I_\Sigma)$ is the principal counterpoint triple of Σ, where R_Σ is the mirror image operator on

Σ-rows. Scales with the same height have isomorphic the respective fundamental groups. This leads to the notion of the scale type $\mathbb{Z}_n : 0 < 1 < \cdots < n-1$ (section 3).

Groups generated by retrograde, transposition and inversion operators are the subject of section 4. The counterpoint group $(r/t/i)$ is the free group generated by three letters r, t, i subjected to the axioms.

$$r^2 = 1 = i^2, \;\; rt = tr, \;\; ri = ir, \;\; it = t^{-1}i.$$

$(r/t/i)$ has three remarkable subgroups:

- (r/i), copy of the Klein four group $\mathbb{Z}_2 \times \mathbb{Z}_2$ ([9], [7]),
- (r/t), copy of the commutative group $\mathbb{Z}_2 \times \mathbb{Z}_n$, provided t has order n,
- (t/i), copy of the dihedral group D_{2n}, provided t has order n ([2], [3]).

Section 5 is devoted to exhibit new counterpoint triples which enrich the musical material permitting to compose extensive twelve tone structures. By construction, the chromatic scale

$$C = \{c, c\sharp, d, d\sharp, e, e, f, f\sharp, g, g, g\sharp, a, a\sharp, b\}$$

is the disjoint union of the diatonic scale

$$D = \{c, d, e, f, g, a, b\}$$

and the pentatonic scale

$$P = \{c\sharp, d\sharp, f\sharp, g\sharp, a\sharp\},$$

so that apart from the counterpoint triple (R_C, T_C, I_C) mainly used in twelve tone music, we obtain two partial counterpoint triples

$$(R_1 = R_D \vee id_P, T_1 = T_D \vee id_P, I_1 = I_D \vee id_P)$$

$$(R_2 = id_D \vee R_P, T_2 = id_D \vee T_P, I_2 = id_D \vee I_P)$$

where $T_D \vee id_P$, $I_D \vee id_P$ are the operators on C_{12} given by

$$(T_D \vee id_P)(x) = \begin{cases} T_D(x) & \text{if } x \in D_7 \\ x & \text{if } x \in P_5 \end{cases}$$

$$(I_D \vee id_P)(x) = \begin{cases} I_D(x) & \text{if } x \in D_7 \\ x & \text{if } x \in P_5 \end{cases}$$

Moreover, $R_D \vee id_P$ reverses the longest substring of D_7 inside a string of C_{12}

$$(R_D \vee id_P)(w_0 s_1 w_1 \cdots w_{k-1} s_k w_k) = w_0 s_k w_1 \cdots w_{k-1} s_1 w_k.$$

Composing termwise the previous triples we get the triple

$$(R_D \vee R_P, T_D \vee T_P, I_D \vee I_P)$$

whose group has 140 elements instead of 48 elements of the group $(R_C/T_C/I_C)$ used in twelve tone composition. On the other hand the group generated by

$$(R_1, R_2/T_1, T_2/I_1, I_2) = \{R_1^{\kappa_1} \circ R_2^{\kappa_2} \circ T_1^{\lambda_1} \circ T_2^{\lambda_2} \circ I_1^{\mu_1} \circ I_2^{\mu_2} \mid 0 \leq \kappa_1, \kappa_2, \mu_1, \mu_2 < 2, \; 0 \leq \lambda_1 < 7, \; 0 \leq \lambda_2 < 5\}$$

and so the cardinality of this group is $2^2 \cdot 7 \cdot 5 \cdot 2^2 = 560$. Extension to arbitrary scales is provided.

In section 6 we introduce the notion of counterpoint space in order to describe various musical structures in an abstract setting. A *rti*-space is a triple $\mathcal{A} = ((r/t/i), A, a_0)$ consisting of a left group action and an element a_0 from which all elements of A are reachable, that is

$$A = \left\{ t^k \cdot a_0, rt^k \cdot a_0, rt^k i \cdot a_0, t^k i \cdot a_0 \mid k \in \mathbb{Z} \right\}.$$

Actually, the above action is simply transitive. Major and minor chords are defined with respect to a pair (p, q) of "intervals". We show that (p, q)-consonant chords in a counterpoint space constitute a counterpoint space, as well.

II. Scale Systems

i. Definition and Examples

A *scale of height n* is a linearly ordered set of musical objects

$$\Sigma : \sigma_0 < \sigma_1 < \cdots < \sigma_{n-1}.$$

The objects $\sigma_0, \sigma_1, \ldots, \sigma_{n-1}$ are the *degrees* of Σ.

Common pitch scales:

- the chromatic scale

$$C : c < c\sharp < d < d\sharp < e < f < f\sharp < g < g\sharp < a < a\sharp < b,$$

- the diatonic scale

$$D : c < d < e < f < g < a < b \text{ (white keybords)},$$

- the pentatonic scale

$$P : c\sharp < d\sharp < f\sharp < g\sharp < a\sharp \text{ (black keybords)},$$

- the whole-tone scale

$$H : c < d < e < f\sharp < g\sharp < a\sharp,$$

- the octatonic scale

$$O : c < d < d\sharp < f < f\sharp < g\sharp < a < b.$$

Durations and intervals may be organized into scales

- $DUR(n) : \frac{1}{2^{n-1}} < \frac{1}{2^{n-2}} < \cdots < \frac{1}{2} < 1$
- $INT : 1p < 2m < 2M < 3m < 3M < 4p < 4^+ < 5p < 6m < 6M < 7m < 7M < 8p$ (p=perfect, m=minor, M=Major, 4^+=augmented)

Any increasing sequence of indexes $0 \leq i_0 < i_1 < \cdots < i_\kappa < n$ induces the subscale of Σ

$$\Sigma[i_0, \ldots, i_\kappa] : \sigma_{i_0} < \sigma_{i_1} < \cdots < \sigma_{i_\kappa}.$$

For instance D, P, H, O are subscales of C:

$$D = C[0, 2, 4, 5, 7, 9, 11],\ P = C[1, 3, 6, 8, 10],\ H = C[0, 2, 4, 7, 9, 11],\ O = C[0, 2, 3, 5, 6, 8, 9, 11]$$

ii. Scale Schemes

Let s, t be two symbols connected with the ordering $s < t$ meaning "s smaller than t". Any string of $\{s,t\}^*$

$$w = w_1 w_2 \cdots w_\kappa, \quad w_i \in \{s,t\}$$

is called *scale scheme*; it generates by "prefix ranking" the scale

$$1 < w_1 < w_1 w_2 < \cdots < w_1 w_2 \cdots w_\kappa.$$

For instance, the major and minor scale schemes are

$$M = ttsttts \text{ and } m = tsttstt$$

respectively.

An *interpretation* of schemes into a scale

$$\Sigma : \sigma_0 < \sigma_1 < \cdots < \sigma_{n-1}$$

is a right action

$$\Sigma_n \times \{s,t\}^* \to \Sigma_n$$

fully determined by the assignments

$$(\sigma, s) \mapsto \sigma s, \quad (\sigma, t) \mapsto \sigma t \quad (\sigma \in \Sigma_n).$$

Then the *scale of root* $\sigma \in \Sigma_n$ with respect to the scheme w above, is

$$\sigma < \sigma w_1 < \sigma w_1 w_2 < \cdots < \sigma w_1 w_2 \cdots w_\kappa.$$

Implementing s and t in C as "semitone" and "tone" respectively, we get the ordinary major and minor scales with root $\sigma \in C$.

$$M(\sigma) : \sigma < \sigma t < \sigma tt < \sigma tts < \sigma ttst < \sigma ttstt < \sigma ttsttt < \sigma ttsttts,$$

$$m(\sigma) : \sigma < \sigma t < \sigma ts < \sigma tst < \sigma tstt < \sigma tstts < \sigma tsttst < \sigma tsttstt.$$

The ancient greek trope schemes (GTS) are lexicographically ordered from top to bottom with respect to the relation $s < t$ meaning that s is smaller than t:

s	t	t	s	t	t	t	:	Locrian scheme	(LO)
s	t	t	t	s	t	t	:	Phrygian scheme	(F)
t	s	t	t	s	t	t	:	Aeolian scheme	(A)
t	s	t	t	t	s	t	:	Dorian scheme	(D)
t	t	s	t	t	s	t	:	Mixolydian scheme	(M)
t	t	s	t	t	t	s	:	Ionian scheme	(I)
t	t	t	s	t	t	s	:	Lydian scheme	(LY)

That is

$$GTS : LO < F < A < D < M < I < LY$$

([10]).

III. The Fundamental Groups of a Scale

i. The Clock – Group

Consider the scale

$$\Sigma : \sigma_0 < \sigma_1 < \cdots < \sigma_{n-1}.$$

The clock group of Σ is $G_1(\Sigma) = (\Sigma_n, \oplus)$, with Σ_n denoting the set of degrees, $\Sigma_n = \{\sigma_0, \ldots, \sigma_{n-1}\}$ and the binary operation

$$\oplus : \Sigma_n \times \Sigma_n \to \Sigma_n, \ (\sigma_\kappa, \sigma_\lambda) \mapsto \sigma_\kappa \oplus \sigma_\lambda$$

is the clock addition

$$\sigma_\kappa \oplus \sigma_\lambda = \begin{cases} \sigma_{\kappa+\lambda} & \text{if } \kappa + \lambda < n, \\ \sigma_{\kappa+\lambda-n} & \text{if } \kappa + \lambda \geq n. \end{cases}$$

The neutral element is σ_0 and the opposite of σ_κ is $\sigma_{n-\kappa}$, $-\sigma_\kappa = \sigma_{n-\kappa}$, $\kappa = 0, \ldots, n-1$.

Transposition and Inversion with respect to Σ are exclusively defined in terms of $G_1(\Sigma)$. Precisely

$$T_\Sigma, I_\Sigma : \Sigma_n \to \Sigma_n, \ \ T_\Sigma(\sigma_k) = \sigma_\kappa \oplus \sigma_1, \ \ I_\Sigma(\sigma_\kappa) = -\sigma_\kappa, \ \ 0 \leq \kappa \leq n-1.$$

We observe that $T_C(f) = f\sharp$ and $T_D(f) = g$ and $T_P(c\sharp) = d\sharp$ and $T_C(c\sharp) = d$, etc. Likewise, $I_C(d) = a$ and $I_D(d) = g$, $I_P(a\sharp) = d\sharp$ and $I_C(a\sharp) = d$, etc

Proposition 1. *Consider two scales of the same height*

$$\Sigma : \sigma_0 < \sigma_1 < \cdots < \sigma_{n-1}, \ \ \Gamma : \gamma_0 < \gamma_1 < \cdots < \gamma_{n-1}.$$

Then the function

$$\phi : \Sigma_n \to \Gamma_n, \ \ \phi(\sigma_\kappa) = \gamma_\kappa \ \ (\kappa = 0, 1, \ldots, n-1)$$

is an isomorphism of $G_1(\Sigma)$ onto $G_1(\Gamma)$ preserving transposition and inversion

$$\phi \circ T_\Sigma = T_\Gamma \circ \phi, \ \ \phi \circ I_\Sigma = I_\Gamma \circ \phi$$

where "$\circ$" designates function composition performed from right to left.

Proof. We are going to show that ϕ preserves the clock addition, i.e. that

$$\phi(\sigma_\kappa \oplus \sigma_\lambda) = \phi(\sigma_\kappa) \oplus \phi(\sigma_\lambda) \ \text{ for all } \kappa, \lambda.$$

Indeed, if $\kappa + \lambda < n$, then

$$\phi(\sigma_\kappa \oplus \sigma_\lambda) = \phi(\sigma_{\kappa+\lambda}) = \gamma_{\kappa+\lambda} = \gamma_\kappa \oplus \gamma_\lambda = \phi(\sigma_\kappa) \oplus \phi(\sigma_\lambda).$$

In the case $\kappa + \lambda \geq n$, then

$$\phi(\sigma_\kappa \oplus \sigma_\lambda) = \phi(\sigma_{\kappa+\lambda-n}) = \gamma_{\kappa+\lambda-n} = \gamma_\kappa \oplus \gamma_\lambda = \phi(\sigma_\kappa) \oplus \phi(\sigma_\lambda).$$

The rest of the proof is straightforward. □

According to the previous result, scales with the same height behave alike from the transposition/inversion point of view.

ii. The Group of Σ-rows

Rows in an arbitrary scale will be discussed below. A pivotal axis of twelve-tone music is the restriction in the repetition of each of the twelve pitch classes. In a twelve-tone sequence, a pitch class cannot reappear before the remaining eleven pitch classes are heard, thus creating a permutation of twelve different pitch classes ([7], [9]). An *n-row* or *n-aggregate* is a rearrangement of the numbers $0, 1, \ldots, n-1$, that is a bijective function p of $\mathbb{Z}_n$ onto itself that can be represented by a matrix of the form

$$p = \begin{pmatrix} 0 & 1 & 2 & \cdots & n-1 \\ p(0) & p(1) & p(2) & \cdots & p(n-1) \end{pmatrix}$$

or shortly

$$p = (p(0), p(1), p(2), \cdots, p(n-1)) .$$

Transposition and inversion can be pointwise extended on rows:

$$\begin{aligned} T\,(p(0), p(1), \ldots, p(n-1)) &= (T(p_0), T(p_1), \ldots, T(p_{n-1})) \\ &= (p(0) \oplus 1, p(1) \oplus 1, \ldots, p(n-1) \oplus 1) , \end{aligned}$$

$$\begin{aligned} I\,(p(0), p(1), \ldots, p(n-1)) &= (I(p_0), I(p_1), \ldots, I(p_{n-1})) \\ &= (-p(0), -p(1), \ldots, -p(n-1)) . \end{aligned}$$

Retrograde is the mirror image operator

$$R\,(p(0), p(1), \ldots, p(n-1)) = (p(n-1), \ldots, p(1), p(0)) .$$

(R, T, I) will be referred to as a *counterpoint triple*. The set S_n of all n-rows is closed under composition

$$(p(0), p(1), \ldots, p(n-1)) \circ (q(0), q(1), \ldots, q(n-1)) = (p(q(0)), p(q(1)), \ldots, p(q(n-1))$$

and constitutes a group $(S_n, \circ)$.

Given a scale of height n

$$\Sigma : \sigma_0 < \sigma_1 < \cdots < \sigma_{n-1}$$

a Σ-row is a rearrangement of the elements $\sigma_0, \sigma_1, \ldots, \sigma_{n-1}$ represented as

$$\pi = \begin{pmatrix} \sigma_0 & \sigma_1 & \cdots & \sigma_{n-1} \\ \sigma_{p(0)} & \sigma_{p(1)} & \cdots & \sigma_{p(n-1)} \end{pmatrix}$$

or shortly

$$\pi = (\sigma_{p(0)}, \sigma_{p(1)}, \cdots, \sigma_{p(n-1)}),$$

where $(p(0), p(1), \ldots, p(n-1))$ is in S_n.

The set $S(\Sigma)$ of all Σ-rows is closed under composition

$$(\sigma_{p(0)}, \sigma_{p(1)}, \ldots, \sigma_{p(n-1)}) \circ (\sigma_{q(0)}, \sigma_{q(1)}, \ldots, \sigma_{q(n-1)}) = (\sigma_{p(q(0))}, \sigma_{p(q(1))}, \ldots, \sigma_{p(q(n-1))})$$

and constitutes the second fundamental group of Σ, $G_2(\Sigma) = (S(\Sigma), \circ)$.

Proposition 2. *If* Σ, Γ *are scales as in the statement of Proposition 1, then the assignment*

$$(\sigma_{p(0)}, \sigma_{p(1)}, \ldots, \sigma_{p(n-1)}) \mapsto (\gamma_{p(0)}, \gamma_{p(1)}, \ldots, \gamma_{p(n-1)})$$

is an isomorphism of $G_2(\Sigma)$ *onto* $G_2(\Gamma)$.

These data lead to introduce the notion of scale type, a means to classify scales. We call *type of height n* the scale

$$\mathbb{Z}_n : 0 < 1 < 2 < \cdots < n-1.$$

Its associated group $(\mathbb{Z}_n, \oplus)$ is the well known group of modulo n integers. For instance the types of the scales C, D, P, H, O are

$$\begin{aligned}
\mathbb{Z}_{12} &: 0 < 1 < 2 < \cdots < 11, \\
\mathbb{Z}_{7} &: 0 < 1 < 2 < \cdots < 6, \\
\mathbb{Z}_{5} &: 0 < 1 < 2 < \cdots < 4, \\
\mathbb{Z}_{6} &: 0 < 1 < 2 < \cdots < 5, \\
\mathbb{Z}_{8} &: 0 < 1 < 2 < \cdots < 7,
\end{aligned}$$

respectively. If two musical scales have the same height, then the isomorphisms of their corresponding fundamental groups describe equivalent musical mathematical structures besides the nature of the objects they act upon.

The triple $(R_\Sigma, T_\Sigma, I_\Sigma)$ defined below is the principal counterpoint triple of Σ:

$$\begin{aligned}
R_\Sigma(\sigma_{p(0)}, \sigma_{p(1)}, \ldots, \sigma_{p(n-1)}) &= (\sigma_{p(n-1)}, \ldots, \sigma_{p(1)}, \sigma_{p(0)}), \\
T_\Sigma(\sigma_{p(0)}, \sigma_{p(1)}, \ldots, \sigma_{p(n-1)}) &= (\sigma_{p(0)} \oplus 1, \sigma_{p(1)} \oplus 1, \ldots, \sigma_{p(n-1)} \oplus 1), \\
I_\Sigma(\sigma_{p(0)}, \sigma_{p(1)}, \ldots, \sigma_{p(n-1)}) &= (\sigma_{-p(0)}, \sigma_{-p(1)}, \ldots, \sigma_{-p(n-1)}),
\end{aligned}$$

where $\oplus$ is the $\bmod n$ addition and $-p(k)$ is the opposite of $p(k)$ with respect to this addition.

Proposition 3. *In a scale Σ of height n the following equalities hold*

$$R_\Sigma^2 = id = I_\Sigma^2, \quad T_\Sigma^n = id, \quad R_\Sigma \circ I_\Sigma = I_\Sigma \circ R_\Sigma,$$

$$R_\Sigma \circ T_\Sigma = T_\Sigma \circ R_\Sigma, \quad I_\Sigma \circ T_\Sigma = T_\Sigma^{n-1} \circ I_\Sigma.$$

Proof. We only establish the last equality. First we show that for every $\sigma_\kappa \in \Sigma_n$ we have

$$(I_\Sigma \circ T_\Sigma)(\sigma_\kappa) = \left(T_\Sigma^{n-1} \circ I_\Sigma\right)(\sigma_\kappa).$$

Indeed

$$\begin{aligned}
(I_\Sigma \circ T_\Sigma)(\sigma_\kappa) = I_\Sigma(T_\Sigma(\sigma_\kappa)) = I_\Sigma(\sigma_\kappa \oplus \sigma_1) = -(\sigma_\kappa \oplus \sigma_1) = (-\sigma_\kappa) \oplus (-\sigma_1) = \\
= (-\sigma_\kappa) \oplus \sigma_{n-1} = T_\Sigma^{n-1}(-\sigma_\kappa) = T_\Sigma^{n-1}(I_\Sigma(\sigma_\kappa)) = \left(T_\Sigma^{n-1} \circ I_\Sigma\right)(\sigma_\kappa).
\end{aligned}$$

Furthermore, for every $(\sigma_{p(0)}, \sigma_{p(1)}, \ldots, \sigma_{p(n-1)}) \in S(\Sigma)$ we have

$$\begin{aligned}
(I_\Sigma \circ T_\Sigma)(\sigma_{p(0)}, \sigma_{p(1)}, \ldots, \sigma_{p(n-1)}) = \left((I_\Sigma \circ T_\Sigma)(\sigma_{p(0)}), \ldots, (I_\Sigma \circ T_\Sigma)(\sigma_{p(n-1)})\right) = \\
= \left(\left(T_\Sigma^{n-1} \circ I_\Sigma\right)(\sigma_{p(0)}), \ldots, \left(T_\Sigma^{n-1} \circ I_\Sigma\right)(\sigma_{p(n-1)})\right) = \left(T_\Sigma^{n-1} \circ I_\Sigma\right)(\sigma_{p(0)}, \sigma_{p(1)}, \ldots, \sigma_{p(n-1)})
\end{aligned}$$

hence the desired result. □

Remark. Given that the diatonic scale may be constructed with fifths or fourths, the pentatonic with stacked fifths, etc. the scales listed in II.i can all be considered as symmetrical generated constructions. As Andreatta points out, group is the dominating algebraic structure utilized to describe symmetry in music ([1]).

IV. Counterpoint Groups

They are groups generated by retrograde, transposition and inversion operators playing a significant role in serial composition. More precisely, the $(r/t/i)$–group is the free group generated by three letters r, t, i subjected to the axioms

$$r^2 = 1 = i^2, \;\; ri = ir, \;\; rt = tr, \;\; it = t^{-1}i. \tag{1}$$

Its elements are of the form

$$t^\kappa, \;\; rt^\kappa, \;\; rt^\kappa i, \;\; t^\kappa i \;\; (\kappa \in \mathbb{Z}).$$

Three subgroups of the above group are of interest: (r/i), (r/t) and (t/i). The first one $(r/i) = \{1, r, iri\}$ is isomorphic to the Klein four group $\mathbb{Z}_2 \times \mathbb{Z}_2$ ([7], [9]). The other two groups are $(r/t) = \{t^\kappa, rt^\kappa \mid \kappa \in \mathbb{Z}\}$ and $(t/i) = \{t^\kappa, t^\kappa i \mid \kappa \in \mathbb{Z}\}$.

Notice that the groups $(r/t/i)$, (r/t) and (t/i) are infinite unless t has finite order, say n, in which case (r/t) and (t/i) coincide with the commutative group $\mathbb{Z}_2 \times \mathbb{Z}_n$ and the dihedral group D_{2n} respectively ([2], [3]). In addition $\text{card}(r/t/i) = 4n$, $\text{card}(r/t) = 2n = \text{card}(t/i)$. In the sequel we deal with left group actions of the form

$$(r/t/i) \times A \to A, \;\; (u, a) \mapsto u.a$$

Due to the free character of the group $(r/t/i)$ the above action is completely determined by the values $r.a$, $t.a$, $i.a$ $(a \in A)$ compatible with equalities (1), which means that

$$r^2.a = a = i^2.a, \;\; ri.a = ir.a, \;\; rt.a = tr.a, \;\; it.a = t^{-1}i.a, \tag{2}$$

for all $(a \in A)$.

Examples. The operation of row composition defines a left action of $(r/t/i)$ on S_n in the following manner:

$$r.p = R \circ p, \;\; t.p = T \circ p, \;\; i.p = I \circ p,$$

for all $p = (p(0), p(1), \ldots, p(n-1)) \in S_n$. For $n = 12$, choosing the Schoenberg op.36 aggregate $p = (0, 1, 6, 2, 7, 9, 3, 4, 10, 11, 5, 8)$ we get

$$\begin{aligned} r.p &= (8, 5, 11, 10, 4, 3, 9, 7, 2, 6, 1, 0) \\ i.p &= (0, 11, 6, 10, 5, 3, 9, 8, 2, 1, 7, 4) \\ t^3 i.p &= (3, 2, 9, 1, 8, 6, 0, 11, 5, 4, 10, 7). \end{aligned}$$

More generally, for any scale Σ the relations

$$r.\tau = R_\Sigma \circ \tau, \;\; t.\tau = T_\Sigma \circ \tau, \;\; i.\tau = I_\Sigma \circ \tau \;\; (\tau \in S(\Sigma))$$

actually define an action of $(r/t/i)$ on $S(\Sigma)$.

Further actions:

- $(r/t/i) \times \mathbb{Z}_{12}^* \to \mathbb{Z}_{12}^*$ with

  $$r.s = s_\kappa \cdots s_2 s_1, \;\; t.s = (s_1 \oplus 1)(s_2 \oplus 1) \cdots (s_\kappa \oplus 1), \;\; i.s = (12 - s_1)(12 - s_2) \cdots (12 - s_\kappa),$$

  for all $s = s_1 s_2 \cdots s_\kappa$ in $\mathbb{Z}_{12}^*$

Likewise

- $(r/t/i) \times \mathbb{Z}^* \to \mathbb{Z}^*$, with

  $$r.s = s_\kappa \cdots s_2 s_1, \;\; t.s = (s_1 + 1)(s_2 + 1) \cdots (s_\kappa + 1), \;\; i.s = (-s_1)(-s_2) \cdots (-s_\kappa),$$

  for all $s = s_1 s_2 \cdots s_\kappa$ in $\mathbb{Z}^*$.

In all these cases the conditions (2) are verified.

V. Partial Counterpoint Operators

The aim of this section is to present new counterpoint triples, which enrich the musical material permitting to compose extensive twelve tone structures. By construction, the chromatic scale is the disjoint union of the diatonic scale and the pentatonic scale, $C = D \cup P$.

Therefore, apart from (R_C, T_C, I_C), we obtain two partial counterpoint triples

$$(R_D \vee id_P, T_D \vee id_P, I_D \vee id_P) \text{ and } (id_D \vee R_P, id_D \vee T_P, id_D \vee I_P), \tag{3}$$

where $T_D \vee id_P$ (resp. $id_D \vee I_P$) is the operator on C_{12} coinciding with T_D (resp. T_P) on the set D_7 (resp. P_5) and leaving unchanged the elements of $C_{12} - D_7$ (resp. $C_{12} - P_5$). Likewise, $(I_D \vee id_P)(x) = I_D(x)$ if $x \in D_7$, $= x$, else. Finally, $R_D \vee id_P$ reverses only the longest string of D_7^* occurring inside any string of C_{12}^*

$$(R_D \vee id_P)(w_0 s_1 w_1 \cdots w_{\kappa-1} s_\kappa w_\kappa) = w_0 s_\kappa w_1 \cdots w_{\kappa-1} s_1 w_\kappa$$

for all $s_i \in D_7$, $w_i \in (C_{12} - D_7)^*$. Similar definitions for $id_D \vee I_P$ and $id_D \vee R_P$ can be stated. In the context of mod 12 integers the above data are encoded as follows:

$$\begin{aligned} &\underline{0}\ 1\ \underline{2}\ 3\ \underline{4}\ \underline{5}\ 6\ \underline{7}\ 8\ \underline{9}\ 10\ \underline{11} \\ R_D \vee id_P &= (11, 1, 9, 3, 7, 5, 6, 4, 8, 2, 10, 0) \\ T_D \vee id_P &= (11, 1, 0, 3, 2, 4, 6, 5, 8, 7, 10, 9) \\ I_D \vee id_P &= (0, 1, 11, 3, 9, 7, 6, 5, 8, 4, 10, 2) \end{aligned}$$

$$\begin{aligned} &0\ \underline{1}\ 2\ \underline{3}\ 4\ 5\ \underline{6}\ 7\ \underline{8}\ 9\ \underline{10}\ 11 \\ id_D \vee R_P &= (0, 10, 2, 8, 4, 5, 6, 7, 3, 9, 1, 11) \\ id_D \vee T_P &= (0, 3, 2, 6, 4, 5, 8, 7, 10, 9, 1, 11) \\ id_D \vee I_P &= (0, 1, 2, 10, 4, 5, 8, 7, 6, 9, 3, 11) \end{aligned}$$

Furthermore, composing termwise the triples (3), we get the triple

$$(R_D \vee R_P, T_D \vee T_P, I_D \vee I_P)$$

where

$$(T_D \vee T_P)(x) = \begin{cases} T_D(x), & \text{if } x \in D_7 \\ T_P(x), & \text{if } x \in P_5 \end{cases}$$

and so on.

In terms of pitch classes

$$\begin{aligned} R_D \vee R_P &= (11, 10, 9, 8, 7, 5, 6, 4, 3, 2, 1, 0) \\ T_D \vee T_P &= (11, 3, 0, 6, 2, 4, 8, 5, 10, 7, 1, 9) \\ I_D \vee I_P &= (0, 1, 11, 10, 9, 7, 8, 5, 6, 4, 3, 2) \end{aligned}$$

Obviously $T_D \vee T_P \neq T_C$, $I_D \vee I_P \neq I_C$, $R_D \vee R_P \neq R_C$. The order of $T_D \vee T_P$ into the group $S(C)$ is $7 \cdot 5 = 35$ and so the group $(R_D \vee R_P / T_D \vee T_P / I_D \vee I_P)$ has $4 \cdot 35 = 140$ elements instead of 48 elements of the group $(R_C / T_C / I_C)$ used in twelve tone music. Consequently, we are able to speak of a multiple enrichment of the organisation of the pitch material, beyond the already known ways of managing it in twelve-tone music and without the encroachment of its main principles.

Our next task will be to determine the group generated by the six partial operators

$$\begin{aligned} R_1 = R_D \vee id_P, \quad R_2 = id_D \vee R_P, \\ T_1 = T_D \vee id_P, \quad T_2 = id_D \vee T_P, \\ I_1 = I_D \vee id_P, \quad I_2 = id_D \vee I_P. \end{aligned} \tag{4}$$

By taking into account that generators (4) commute to each other except that

$$I_1 \circ T_1 = T_1^{-1} \circ I_1 = T_1^6 \circ I_1 \text{ and } I_2 \circ T_2 = T_2^{-1} \circ I_2 = T_2^4 \circ I_2$$

we get

$$\begin{aligned} &(R_1, R_2/T_1, T_2/I_1, I_2) = \\ &\quad = \left\{ R_1^{\kappa_1} \circ R_2^{\kappa_2} \circ T_1^{\lambda_1} \circ T_2^{\lambda_2} \circ I_1^{\mu_1} \circ I_2^{\mu_2} \mid 0 \le \kappa_1, \kappa_2, \mu_1, \mu_2 < 2,\ 0 \le \lambda_1 < 7,\ 0 \le \lambda_2 < 5 \right\} \end{aligned}$$

and thus

$$\operatorname{card}(R_1, R_2/T_1, T_2/I_1, I_2) = 2^2 \cdot 7 \cdot 5 \cdot 2^2 = 560.$$

Let us discuss more complex situations. Suppose that $(A_1, \ldots, A_\kappa)$ is a partition of a scale Σ

$$\Sigma_n = A_1 \cup \cdots \cup A_\kappa,\ A_i \cap A_j = \emptyset \text{ for } i \neq j$$

and denote by $\bar{A}_i$ the set complement of A_i in Σ_n, $\bar{A}_i = \Sigma_n - A_i$, $1 \le i \le \kappa$. Since subsets of Σ are scales with the induced ordering, we may define the counterpoint triples

$$R_i = R_{A_i} \vee id_{\bar{A}_i},\ T_i = T_{A_i} \vee id_{\bar{A}_i},\ I_i = I_{A_i} \vee id_{\bar{A}_i},\ 1 \le i \le \kappa.$$

Theorem 4. *The cardinality of the group generated by the above operators is*

$$\operatorname{card}((R_i)_i/(T_i)_i/(I_i)_i) = 2^\kappa \cdot \operatorname{ord}(T_1) \cdots \operatorname{ord}(T_\kappa) \cdot 2^\kappa$$

where $\operatorname{ord}(T_i)$ *is the order of* T_i *in the group* $S(\Sigma)$*.*

In another direction we may replace R_i*,* T_i*,* I_i *by a triple of permutations* $(\rho_i, \tau_i, \iota_i)$ *on* A_i*,* $1 \le i \le \kappa$*.*

VI. Counterpoint Spaces

The algebraic structure of counterpoint space (CP-space) is proposed, in order to found basic musical notions in a formal framework.

An *rti-space* is a triple $\mathcal{A} = ((r/t/i), A, a_0)$ formed by a left group action

$$(r/t/i) \times A \to A$$

and an initial element $a_0 \in A$ (playing the role of the pitch c).

These data must comply with the following axiom:

S) the elements of A are accessible from a_0, that is

$$A = \left\{ t^k \cdot a_0, rt^k \cdot a_0, rt^k i \cdot a_0, t^k i \cdot a_0 \mid k \in \mathbb{Z} \right\}.$$

The *ti*-spaces are formulated as before except we replace $(r/t/i)$ by the group (t/i). Axiom S) takes the form

$$A = \left\{ t^k \cdot a_0, t^{ki} \cdot a_0 \mid k \in \mathbb{Z} \right\}.$$

Examples. • $\mathcal{A} = ((r/t/i), A_n, (0,1,\ldots,n-1))$, where

$$A_n = \{T^\kappa, R \circ T^\kappa, R \circ T^\kappa \circ I, T^\kappa \circ I \mid \kappa = 0,1,\ldots,n-1\}$$

Special case $\mathcal{A}_{12} = ((r/t/i), A_{12}, (0,1,\ldots,11))$.

- Given an arbitrary scale Σ of height n, we have the following CP-space

$$\mathcal{A}_\Sigma = ((r/t/i), A_\Sigma, (\sigma_0, \sigma_1, \ldots, \sigma_{n-1}))\,,$$

where

$$A_\Sigma = \{T_\Sigma^\kappa, R_\Sigma \circ T_\Sigma^\kappa, R_\Sigma \circ T_\Sigma^\kappa \circ I_\Sigma, T_\Sigma^\kappa \circ I_\Sigma \mid \kappa = 0,1,\ldots,n-1\}$$

Standard CP-spaces:

- $\mathcal{Z}_n = ((t/i), \mathbb{Z}_n, 0)$ instances $n = 12, 7, 5$
- $\mathcal{Z} = ((t/i), \mathbb{Z}, 0)$

Remark. It should be noticed that the last two examples of actions in section 4 can not be organized into CP-spaces since they are deprived initial elements.

Simply transitive actions are often encountered in mathematical music theory ([3]). Recall that a left group action

$$G \times A \to A, \ (g, a) \mapsto ga,$$

is *simply transitive* whenever for any pair $(a_1, a_2) \in A^2$ there exists a unique $g \in G$ so that $a_2 = ga_1$.

Proposition 5. *If $\mathcal{A}$ fulfils the additional axiom*

S') the function $g \mapsto ga_0$ is injective,

then the action $(r/t/i) \times A \to A$ is simply transitive.

Proof. We follow the guideline proof arguments of the corresponding result in [3]

(Existence). According to the axiom S) any pair $(a_1, a_2) \in A^2$ is written $a_1 = g_1 a_0$, $a_2 = g_2 a_0$ and so $a_0 = g_1^{-1} a_1$ and $a_2 = g_2 g_1^{-1} a_1$.

(Uniqueness). Assume, now, that $a_2 = u_1 a_1 = u_2 a_1$ then $u_1 g_1 a_0 = u_2 g_1 a_0$ and so by axiom S') $u_1 g_1 = u_2 g_1$ and by right cancellation $u_1 = u_2$. □

Corollary 6. *The counterpoint spaces $\mathcal{A}_n$, $\mathcal{A}_\Sigma$, $\mathcal{Z}_n$, $\mathcal{Z}$ fulfill the axiom S') and thus the corresponding actions are simply transitive.*

We are going to indicate how basic musical notation can be defined in the setup of counterpoint spaces. In the traditional context consonant chords are built by overposing thirds. A chord is a triple of simultaneously played pitches. A major (resp. minor) chord consists of a root pitch, a second pitch four (resp. three) semitones above the root and a third pitch seven semitones above the root. Major (resp. minor) chords are successive transpositions of the C-major chord $(0,4,7)$ (resp. f-minor chord $(5,8,0)$). Moreover $(5,8,0)$ is the inversion of $(0,4,7)$:

$$\begin{array}{ccc}
C = (0,4,7) & \xrightarrow{\ I\ } & (5,8,0) = f \\
\downarrow T & & \downarrow T \\
C\sharp = (1,5,8) & & (6,9,1) = f\sharp \\
\downarrow T & & \downarrow T \\
D = (2,6,9) & & (7,10,2) = g \\
\downarrow & & \downarrow \\
\vdots & & \vdots \\
\downarrow T & & \downarrow T \\
B = (11,3,7) & & (4,7,11) = e
\end{array}$$

The foundation of chord theory can be realized in a *ti*-space $\mathcal{A} = ((t/i), A, a_0)$. First we need a building pair (p,q) of natural numbers abstracting the pair $(4,3)$ for thirds. Now, the (p,q) major chords in $\mathcal{A}$ are all triples of the form $(t^\kappa \cdot a_0, t^{\kappa+p} \cdot a_0, t^{\kappa+p+q} \cdot a_0)$, κ running over the set $\mathbb{Z}$ of integers. The (p,q)-minor chords in $\mathcal{A}$ are all triples of the form $(t^\kappa i \cdot a_0, t^{\kappa+q} i \cdot a_0, t^{\kappa+p+q} i \cdot a_0)$, $\kappa \in \mathbb{Z}$.

All together (p,q)-major chords and (p,q)-minor chords form the set $(p,q) - CC(\mathcal{A})$ of (p,q) *consonant chords* in $\mathcal{A}$

$$(p,q) - CC(\mathcal{A}) = (p,q) - MC(\mathcal{A}) \cup (p,q) - mC(\mathcal{A}),$$

where $(p,q) - MC(\mathcal{A})$, $(p,q) - mC(\mathcal{A})$ stand for the sets of major and minor chords in $\mathcal{A}$ respectively.

Remark. Since an ordinary chord consists of simultaneously sounding pitches it is not necessary to insist on a particular ordering of the pitches inside a chord. We adopt this mathematical abuse also in our general setting, i.e. any chord (a_1, a_2, a_3) will be identified with the chord (a_3, a_2, a_1)

The following result confirms that the group (t/i) acts on the left on $(p,q) - CC(\mathcal{A})$.

Proposition 7. *i) Left multiplication by i changes the arity of the chords, that is converts major to minor chords and vice versa.*
ii) Left multiplication by t preserves the arity of chords.

Proof. Indeed, if $(t^\kappa \cdot a_0, t^{\kappa+p} \cdot a_0, t^{\kappa+p+q} \cdot a_0)$ is in $(p,q) - MC(\mathcal{A})$, then

$$\begin{aligned} i(t^\kappa \cdot a_0, t^{\kappa+p} \cdot a_0, t^{\kappa+p+q} \cdot a_0) &= (it^\kappa \cdot a_0, it^{\kappa+p} \cdot a_0, it^{\kappa+p+q} \cdot a_0) \\ &\overset{(+)}{=} (t^{-\kappa} i \cdot a_0, t^{-\kappa-p} i \cdot a_0, t^{-\kappa-p-q} i \cdot a_0) \\ &\overset{\text{Remark}}{=} (t^{-\kappa-p-q} i \cdot a_0, t^{-\kappa-p} i \cdot a_0, t^{-\kappa} i \cdot a_0) \in (p,q) - mC(\mathcal{A}) \end{aligned}$$

where equality (+) comes from the axiom $it = t^{-1}i$. The other assertions can be proved by similar arguments. $\square$

Proposition 8. *If $\mathcal{A} = ((t/i), A, a_0)$ is a ti-space, then*

$$(p,q) - chord(\mathcal{A}) = \left((t/i), (p,q) - CC(\mathcal{A}), (a_0, t^p \cdot a_0, t^{p+q} \cdot a_0)\right)$$

is also a ti-space. Moreover, if $\mathcal{A}$ satisfies the axiom S'), then $(p,q) - chord(\mathcal{A})$ also satisfies S').

Proof. First we show that all major chords are accessible from the initial major chord $(a_0, t^p \cdot a_0, t^{p+q} \cdot a_0)$:

$$(t^\kappa \cdot a_0, t^{\kappa+p} \cdot a_0, t^{\kappa+p+q} \cdot a_0) = t^\kappa (a_0, t^p \cdot a_0, t^{p+q} \cdot a_0).$$

On the other hand, minor chords are accessible from the generic minor chord $(t^{-p-q} i \cdot a_0, t^{-p} i \cdot a_0, i \cdot a_0)$:

$$(t^{-\kappa-p-q} i \cdot a_0, t^{-\kappa-p} i \cdot a_0, t^{-\kappa} i \cdot a_0) = t^{-\kappa} (t^{-p-q} i \cdot a_0, t^{-p} i \cdot a_0, i \cdot a_0).$$

Notice that when κ runs over $\mathbb{Z}$, the minor chord $(t^{-\kappa-p-q} i \cdot a_0, t^{-\kappa-p} i \cdot a_0, t^{-\kappa} i \cdot a_0)$ runs over the whole set $(p,q) - mC(\mathcal{A})$.

Finally,

$$(t^{-p-q} i \cdot a_0, t^{-p} i \cdot a_0, i \cdot a_0) = i(a_0, t^p \cdot a_0, t^{p+q} \cdot a_0).$$

Hence axiom S) is valid in $(p,q) - \mathrm{chord}(\mathcal{A})$. To establish S') we have to show the injectivity of the function $u \mapsto u(a_0, t^p \cdot a_0, t^{p+q} \cdot a_0)$. We have

$$\begin{gathered} u_1(a_0, t^p \cdot a_0, t^{p+q} \cdot a_0) = u_2(a_0, t^p \cdot a_0, t^{p+q} \cdot a_0) \text{ implies} \\ (u_1 a_0, u_1 t^p \cdot a_0, u_1 t^{p+q} \cdot a_0) = (u_2 \cdot a_0, u_2 t^p \cdot a_0, u_2 t^{p+q} \cdot a_0) \text{ implies} \\ u_1 \cdot a_0 = u_2 \cdot a_0 \end{gathered}$$

which, by axiom S') for $\mathcal{A}$, gives $u_1 = u_2$ and the proof is achieved. □

Corollary 9. *The action*

$$(t/i) \times (p,q) - CC(\mathcal{A}) \to (p,q) - CC(\mathcal{A})$$

is simply transitive, provided $\mathcal{A}$ satisfies S').

Proof. It is a consequence of proposition 3. □

VII. Conclusion-Future Work

A generic notion in music theory is that of a scale. It refers to a finite linear ordered set of musical objects called degrees of the scale. This definition covers various musical scale situations. Basic tools in this setup are the fundamental groups of a scale: clock group and group of rows. Scales are classified according to their cardinality: two equivalent scales have isomorphic their corresponding fundamental groups.

Major and minor scale schemes as well as ancient greek musical trope schemes are discussed. Groups generated by retrograde/transposition/inversion operators are used to enrich twelve tone composition techniques. Essentially, a novel view of partiality of the twelve-tone aggregate is proposed, which is not based on the division of the chromatic pitch set into trichords, tetrachords or hexachords, but on the partition $C = D \cup P$. Counterpoint spaces are domains suitable to develop an abstract music theory. The main result is that consonant chords in such a space form also a counterpoint space. A future task will be the study of neo-Riemannian theory in the present general context, which will provide a novel view of the relationship between voice-leading and counterpoint. Infinite scales could have also interest to be investigated.

Acknowledgement

The author is indebted to the anonymous referees for their fruitful comments.

References

[1] Andreatta, M. (2004). On Group-Theoretical Methods Applied to Music: Some Compositional and Implementational Aspects. In Mazzola, G., Noll, T. and Luis-Puebla, E. (ed.) *Perspective in Mathematical and Computational Music Theory*: 169–193.

[2] Andreatta, M.; Agon, C.; Noll, T.; Amiot, E. (2006). "Towards Pedagogability of Mathematical Music Theory". *Bridges. Mathematical Connections in Art, Music and Science*: 277–284.

[3] Crans, A.S.; Fiore, T.; Satyendra, R. (2009). "Musical Actions of Dihedral Groups". *American Mathematical Monthly*: 116(6): 479–495.

[4] Barbour, J. (1949). "Musical Scales and their Classification". *The Journal of the Acoustical Society of America*, 21(4): 461–461.

[5] Dowling, W. (1978). "Scale and Contour: Two Components of a Theory of Memory for Melodies". *Psychological review*, 85(4): 341—354.

[6] Fiore, T. (2009). Scale Theory and Algebraic Combinatorics on Words: Extension of Christoffel Duality to Christoffel Conjugates, and the Characterization of the Ionian Mode via Divider Incidence and Sturmian Standardicity. Available at http://www-personal.umd.umich.edu/ tmfiore/1/music.html.

[7] Morris, R. (2007). "Mathematics and the Twelve-Tone System: Past, Present, and Future". *Perspectives of New Music*, 45(2): 76–107.

[8] Noll, T. (2008). "Sturmian Sequences and Morphisms: A Music-Theoretical Application". Mathématique et Musique. Journée Annuelle de la Société Mathématique de France *Proceedings...*: 79–102.

[9] Peck, R. (2013). A Hypercube-Graph Model for *n*-Tone Rows and Relations. International Conference on Mathematics and Computation in Music *Proceedings...*: 177–188.

[10] Winnington-Ingram, R.(1936). *Mode in Ancient Greek Music*. London: Cambridge University Press.

FEDERAL UNIVERSITY OF RIO DE JANEIRO

Denise Pires de Carvalho
Rector

Carlos Frederico Leão Rocha
Vice Rector

Denise Maria Guimarães Freire
Pro-Rector of Graduate and Research

CENTER OF LITERATURE AND ARTS

Cristina Grafanassi Tranjan
Dean

SCHOOL OF MUSIC

Ronal Silveira
Director

Marcelo Jardim
Vice-director

Maria José Bernardes Di Cavalcanti
Associate Director of Undergraduate Education

Marcelo Jardim
Director of the Cultural Artistic Sector

Andrea Adour
Director of Extension Course

João Vidal
Coordinator of the Graduate Program in Music

Aloysio Fagerlande
Coordinator of the Graduate Program Professional in Music

MusMat Research Group

Members:

- Adriel Viturino
- Alexandre Avellar
- Alexandre Ferreira
- Alexandre Schubert
- Ana Miccolis
- André Codeço
- André Pinto
- Ariane Petri
- Bernardo Ramos
- Carlos Almada (leader / coordinator)
- Carlos Mathias
- Claudia Usai
- Daniel Moreira (coordinator)
- Daniel Mussatto
- Desirée Mayr
- Eduardo Cabral
- Erico Bonfim
- Fabio Adour
- Fabio Monteiro
- Filipe de Matos Rocha
- Gabriel Mesquita
- Helder Oliveira
- Hugo Carvalho
- Igor Chagas
- João Penchel
- Jorge Santos
- Leandro Chrispim
- Liduino Pitombeira (coordinator)
- Max Kühn
- Marcel Castro-Lima
- Nei Rocha
- Pauxy Gentil-Nunes (coordinator)
- Pedro Miguel de Moraes
- Pedro Proença
- Pedro Zizels Ramos
- Petrucio Viana
- Rafael Fortes

- Rafael Soares Bezerra
- Roberto Macedo
- Rodrigo Pascale
- Rodrigo Furman
- Sérgio Ribeiro
- Stefanella Boato
- Vinicius Braga