Title
Implicational Morphologies for Improvised Computer Music

Author
DeFilippo, David

Publication Date
2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Implicational Morphologies for Improvised Computer Music

A Thesis submitted in partial satisfaction of the
requirements for the degree
Master of Arts

in

Music

by

David Morrison DeFilippo

Committee in charge:

Professor Miller Puckette, Chair
Professor Natacha Diels
Professor Shahrokh Yadegari

2018

The Thesis of David Morrison DeFilippo is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California San Diego

2018

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT OF THE THESIS

Implicational Morphologies for Improvised Computer Music

by

David Morrison DeFilippo

Master of Arts in Music

University of California San Diego, 2018

Professor Miller Puckette, Chair

The Thesis describes an approach to generating continuous control data for shaping parameter motion of a synthesizer via a gestalt theory of structural time and of melodic expectation. The set of motions arises from analysis of implicative contents acquired from a musical performer's alteration of a linear potentiometer. The results prove musically relevant, allowing a performer to generate a set of motions from a single slider, each motion operating on different a structural timescale, and allowing a sense of non-linear motion with overall direction to appear in the resultant sound quality.

Chapter 1

Theory

The Thesis entails a modification of gestalt principles to interpret motion by an improviser on a bank of sliders that control a synthesizer. It is my hypothesis that by interpreting motion in this way, the improviser's proprioception engages, creating internal sensations of structure important for relating to changing states of a virtual instrument. The system has a hand in scheduling trajectories to start and stop at moments of valent sensation, either, that of the endings of groups called continuation, closure, or surprise. For scheduling, structural theories relating most closely to James Tenney's theory of Temporal Gestalts and Eugene Narmour's system of analysis of melodic structures are adapted, and posed onto the temporality of motion, in order to find points of structure in the motion.

This instrument generates musical forms when the player moves to the music. In this way it is more like dancing to music than the immediacy of the effect of motion customary to classical instruments. The dancing motion is not primary in terms of the thought way to develop musical forms, but arises from a temporal remoteness to the effect of motion as brought on by algorithmic alteration. An immediacy of interaction is exchanged for a remoteness, and a kinesthetic, dephased motion occurs while engaging the sliders, a motion to the music, that is also the motion from the music in a teleological sense. Ways to consider causation of momentary musical forms include aspects of the overlapping of performer action and computed reaction, and resultant performer entrainment. In more practical terms, this system can be used to synchronize with the musicians playing acoustic instruments, providing morphing lines implied by the psychical expectations of the operator.

In improvisational research, a common thread of experimentation practice involves the use of probes in a musical stream. Once the probe occurs a qualitative account of improvised response is harvested as a set of categories for kinds response. One such study (Cannone, 2013) attempted to identify focal points in collective free improvisation to account for the way in which free improvisors, with no text to reference, are perceived by observers to develop cooperative ways of coordinating musical form. The experimenter designed situations to investigate the behavior of improvisers when presented with a salient or probe moment, in this case and in his terms, presented as an accident. In the experiment, he created a tape that for sixty seconds that would play a kind of natural noisy environment. Then the probe was instantiated as a succession of pitched material for only a few seconds. The experiment analyzed the response of 31 musicians, who were told to listen in headphones and improvise freely to the prepared tape. Five categories were derived. In category 1, improvisers responded with a new musical idea. In category 2, improvisers were drawn to think that the idea they were playing slightly before or during the probe had a higher significance and made it a germinal aspect of the next section. Category 3, the improvisers imitated the probe and then morphed back to the state that they were in before the probe. Category 4, the probe produced a hesitation and then an adjustment to what they were playing. Category 5, there was no observable change.

Of importance to the discussion, the principle underlying the construction of the probe was that of contrast. Contrast functions implicitly at many different moments in the formalism, where it can signify the boundaries of groups, such as the start of the group or the ending of a group. The sensation of contrast in the formation groups operates on a proprioceptive level, lending the mentioned signifiers of structure to the performer as focal points.

To further characterize the instrument discussed, the concept of Embodied Generative Music (Eckel, 2012) makes a distinction among the kinds of modeling used in scientific forecasting and compositional tasks. Rather than forecasting the dynamical effects of supra-causal systems, like the weather, the compositional modeling task "is used to generate the dynamics determining the way a system evolves over time." (Eckel, 2012, p.143) The system is more about poiesis than epistemology. Using a trace of the performer's motion as a source of directing presence, the instrument extrapolates musical figures implied.

1.1 Implication-Realization

Implication-Realization (Narmour, 1990) uses cognitive research to theorize the formation of anticipation of musical structure. In general, moments of time are sequentially situated as having a cognitive value, either as implication or realization. As related to the perception of form, music is a time-based medium, where forms unravel horizontally across time. This is unlike the visual domain where forms are spatially distributed all at once in a visual scene. To create the notion of a form distributed across time, Narmour, gives melodic structure a sequential ordering principle. An event that occurs first in a musical stream is implicative and the next event is the realization of that implication. The sequential ordering occurs continuously as an alternation for the entirety of the musical event. For melody, a bundle of two features characterizes the implication and whether, at the next time step, the realization is a surprise or not. These features are the width and registral direction of an interval.

Narmour contends that interval size and registral direction are independent percepts that feed into a sub-conscious gestalt mechanism; working in terms of bottom-up processing for the formation of groups. A comprehensive report entitled Grouping Mechanisms, "There must, therefore, be a set of mechanisms that enable us to form linkages between some low-level elements and inhibit us from forming linkages between others." (Deutsch, 2013, p. 183) Deutsch's work with perception of musical objects relies on gestalt principles, but also hypothesizes that the features of the object are processed by different specialized neural regions, operating in the effect of a perceptual hierarchy. Through a series of psychoacoustic experiments, Deutsch shows the primacy of certain features of sound over others via ordered soundings that evoke auditory illusions. The notion that a sound object is a set of features as gathered by a set of mechanisms, with differing levels of perceptual primacy, serves as a key aspect to the computer model developed as the subject of this Thesis. A derivation of this concept will be used to construct and reorganize parametric scales of sets of features for the generated response of the computer. The reorganization of elements in the scales relates to the kinds of focus the computer assumes while observing performer input.

Returning to Narmour, the set of mechanisms has been theorized to function in a retrospective sense for grouping a succession of temporally elapsed items (Tenney & Polansky, 1980) and it is Narmour's assertion that, "we replace the retrospective notion of a completed whole instantiated in a pre-existent style-structural grouping with the hypothesis of an implied continuation attached to a parametric style shape." (Narmour, 1990, p. 74) This is to say that partial forms as a musical moment beginning to unravel in the perceptual present, imply a certain consequence, as a function of bottom-up processing, relating to continuations or closures of form based on gestalt principles.

Small intervals (A's) are implicative of continuation, or another small interval in the same registral direction. Rule 1 follows as: if given A, expect A. Large intervals (B's) are implicative of closure, which entails a reversal of both features. Namely, a small interval in the opposite registral direction. Rule 2 follows as: if given B, then expect A. If only one of the implied features is realized, this is considered a partial realization. If neither of the implied features are realized, this is considered a surprise. Of interest, the wider the B interval, the stronger the implication of reversal, so the amount of closure can be calculated by subtracting: B minus A.

For right now we are only dealing with two intervals, or as to be discussed, two slide segments, producing expectations on the lowest possible level. To obtain higher levels of structure, Narmour introduces a transformational scheme, where structural tones are transformed to higher levels of structural expectation. The intuition of this analysis scheme is that if you have three things, the first and last are structural. A compression occurs by reducing from three to two, where the end two transform to the next level. With two a third is then implied, leading to longer term predictions of the structure unfolding, constrained by the memory capacity of the system. The issue concerning the memory capacity of the system involves the transition from an analysis framework of fixed (or already written) melodic sequences to a real-time system accreting one event at a time, where its capacity is how many events it can hold. Before moving on, one complication exists on the lowest level of implication, which is the source of structural components.

If a component at a realization point of a sequence is contextually weak, then it is not considered a realization, but rather a continuation of implication. A chaining of intervals occurs, by overlapping, until an event of sufficient signification occurs. A common scenario of chaining, is a succession of continuations. Narmour outlines a number of conditions related to classical music form of melody that determine contextual strength. Being that the music intended from this system is more related newer developments, such as electroacoustic improvisation, we now move away from Narmour's theory. At first, to look at James Tenney's preceding theory of Temporal Gestalts and then to redefine I-R to analyze a bank of sliders.

1.2 Temporal Gestalts

Temporal gestalts (Tenney & Polansky, 1980) can be thought of as fundamental stuff of musical structuring, in terms of the perception of boundaries. Considering temporality from the perspective of a historian musical time can be seen as a hierarchically ordered network of "moments, incidents, episodes, periods, epochs, eras." (Tenney, 1980, p. 205) What can be taken from this comparison to the historian's conception of time is the boundaries apprehended exist only on a conceptual level. The demarcation of time spans on different levels, form from within our awareness as an occurrence of sounds that perceptually enclose into units. When a set of occurrences allow an enclosing perpetually, these occurrences are delimited by similarity and proximity in time. If both similarity and proximity in time happen, we have a cohesion of a unit, otherwise we have, in his own terming, segregation.

The structure of the theory can be viewed most readily as a stacking of three levels, each with names. The element is the fundamental first level unit, indivisible into smaller temporal gestalt units always in reference to perception. The clang names the next temporal level constituting of at least two temporal gestalt units. Thirdly, the sequence concatenates the clangs, of two or more. Similar to Narmour, we have a notion of higher level groups implied by the contents of lower level groups. Though the contents differ in substance, Narmour uses edges and Tenney uses averages.

Of note in Tenney's theory applied as an analysis tool is the notion that the two rules for groups take on specific and nuanced formulations, to have knowledge of or predict boundaries. The language of the piece generates precise specifications in perception that garner experiences of cohesion. An example is making a precise measurement of inter-element difference within a clang, where a sharp deviation from this interval optimally lends weight to the distinction of a boundary forming. The kind of deviation functions to enclose a sequence into a kind of group and the kind of group evokes the expectation of a certain kind of event to happen next. In the course of adapting a computer music instrument, the many thresholds as static numbers that allow a kind of representation of possible alternations could be calibrated to reflect the currently involved, temporal or spatial intent of the piece. A piece that intends slow transformations, might require a basis of thresholds that are much longer in duration than the statically imposed 1000 ms. Whether threshold negotiation exists in the form of a setting before the piece or a real-time adjusting location, might be an arena of further investigation. The latter a priori may have advantages, in the sense that larger drifts in the thresholds may signify boundaries, where the dynamics of performer interaction change quite drastically. Meaning that drifts of certain magnitudes allow the system to notice retrospectively that a new state of being in the piece currently, at least as the primacy of dilated time(s) is happening. The explanation of theory now transitions into the adaptation of a real-time technical model for use in an entirely different context, that of computer music improvisation.

1.3 Refiguring Implication-Realization and Temporal Gestalts

To encode slides, we first need to determine what creates a slide segment considered to have a function of implication or realization. To frame slide segments, I propose the use of time-gaps and reversals of direction. Both of these occurrences, observed while moving a slider, are moments of zero-acceleration timing (Wang & Tsai, 2012). Zero-acceleration timing or momentary non-motion events, have been shown in traditions of beat-counting gestures, across cultures, to signify strong and weak beats in rhythmic phrases. The demarcation of timings, in this way, relies on varieties of proprioceptive feedback to reproduce a regular motion that has unique gestural signifiers for weak and strong beats.

A time-gap refers to both equally, the time before the start of a motion or the time after the stop of a motion, except in the case where there was no preceding motion in an absolute sense, it only means the time before a start. The stop to start time, as a time-gap, also relates to the outside time or time static; the time of non-activity spacing active states. The start of a time-gap is recognized as occurring after exceeding a threshold of a quarter second of non-activity. A reversal of direction, is an instantaneous change of direction of sliding with no time-gap.

To create feature bundles from slide segments, I measure the duration, length and direction of each segment. Then applying thresholds to each measured component, a binary triple is obtained. Durations greater than 1000 ms and distances greater than one-half of the slider's path are consider large or a 1, while a direction up or in other terms away from the body are a 1. The triple [1,1,1] is considered [Long duration, Long distance, Away from the body].

Next, a weight relating to the cognitive impact of each feature is assigned to the triple: [44 points, 33 points, 22 points]. In the case the case that the instantiated triple is [1,1,1], a 1 counts for positive points and a 0 for no points. By adding the vector together we create a parametric scale based on a gradient of difference from the initial, shown in table 1.1.

Table 1.1: An example of the parametric scale generated from [1,1,1] as the instantiated feature bundle. The asterisk marks the instantiated implication by the improviser. In this case the instantiated is of type B. The (A) groups all possible A realizations, while the (B) groups all possible B realizations. The weights of the scale are in the order of [44,33,22].

(B)	Weight	Index
111*	99	0
110	77	1
101	66	2
011	55	3
(A)		
100	44	4
010	33	5
001	22	6
000	0	7

The focus of the system can be implemented by changing the order of the weights used to organize the parametric scale. As stated, applying a weight for each feature and adding them, creates a parametric scale in terms of its magnitude of difference from the instantiated input. By changing the order of the weights, the perceptual focus of the system changes. By either considering a common distance or direction to be more similar to each other. Weights in the order of [22,44,33], make the second feature, distance, the most significant feature in the scale. Meaning that if a distance positive feature is instantiated by the performer, the two most adjacent feature bundles are those with ones to signify the kind of the second feature. From viewing six different orders for the weights, what changes in the organization of the parametric is not what is considered a member of the A group and the B group, but rather the order they are in as its similarity to the instantiated. In all weighting arrangements considered, the instantiated retains the same general relation to the what is an A and what is a B. This is important to note, not only to describe effect of changing the weights, it also has implications constraining how the method of selection informs generation of a response. Namely, that if a uniform random number generator is used to select response codes, changing the weights has no discernible effect on the output. To make this effect discernible a nonuniform method needs to be used.

Using six different weighting schemes results in forty-eight parametric scales that all have different orderings of what is an A or B from the instantiated. Moments of closure or surprise are good moments to consider when changes are made to the focus of system. This moment in the course of a musical gesture that in some cases might not be followed by a time-gap on behalf of the performer is interpreted as a kind of group ending. This event of retrospective interpretations happens right before the initial moment of a new group, known as the instantiated for the reason that it will be the result of performer action.

There is a difference regarding what can be instantiated as an A or B and what retains the context of A-ness or B-ness, in the context of a generated parametric scale. Since direction is not an implicative property, only distance and duration apply towards the designation of A or B. Thus, we end up with only two instances of B and the rest A for instantiation of implication as shown in table 1.2. As can be seen in table 1.1, the instantiated is a B (marked with *), moving four steps down the scale to [1,0,0] is an A. The grouping within the generated parametric scale shows which feature bundle would be the correct realization of the instantiated. In the case of figure 1.1, with an implication as [1,1,1], the canonical prediction would be a total reversal to the token [0,0,0], based on the closure rule: if given B, expect A. Other tokens from the A group would be partial realizations and tokens from the A group would be forms of surprise, where [1,1,1] repeated would be a maximal surprise because the basic rules of the formalism disallow: if given B, expect B.

Table 1.2: The classification of feature bundles in the case of instantiation. What can be instantiated as an A or a B is different than what is considered A or B on a generated parametric scale.

(A)	101, 100, 011, 010, 001, 000
(B)	111, 100

Table 1.3: A parametric scale generated from the token [0,1,0]. The asterisk marks the instantiated implication by the improviser. The (A) groups all possible A realizations, while the (B) groups all possible B realizations. The weights of the scale are in the order of [44,33,22]. Meaning that the system considers in importance of features in descending order as duration, distance and direction.

(A)	Weight	Index
010*	99	0
011	77	1
000	66	2
110	55	3
(B)		
001	44	4
111	33	5
100	22	6
101	0	7

Because of the temporal premise of alternating implications and realizations and the fact that we are using a parametric scale based on the gradient difference from the instantiated, implications are always new or ordinal first moments in the succession requiring that a new parametric scale of difference be computed at each instantiation. If the event [0,1,0] is given then the parametric scale is as follows in table 1.3. The weights are now applied relative to [0,1,0]. 0 in the first position is worth 44 points, a 1 in the second position is worth 33 points and a 0 in the third position is worth 22 points. Anything other in the order of the digits described is worth 0 points. This method applies to any instantiated binary triple meaning that there are forty-eight different scales in the formalism, eight of which are shown in table 1.4. This is true when considering the fact that one ordering of weights creates eight scales. As mentioned above there are six ordering of weights in use.

Table 1.4: All 8 generated parametric scales are shown where the weights are used in descending order from time to distance to direction. Weights are organized in the form of [44,33,22], so the system focuses on time as the two most adjacent members, from which it can be said that time is the most primary feature.

111	110	101	011	100	010	001	000
110	111	100	010	101	011	000	011
101	100	111	001	110	000	011	010
011	010	001	111	000	110	101	100
100	101	110	000	111	001	010	011
010	011	000	110	001	111	100	101
001	000	011	101	010	100	111	110
000	001	010	100	011	101	110	111

A concatenation of two successive feature bundles (except in the case of chaining) creates a shape from which to gauge expectations. The scales such as in tables 1.1 and 1.3 are used to gauge the distance of the realization from the implication. We now have 64 different shapes of two concatenated feature bundles, with differing levels of either closure or continuation, exclusively, and surprise, inclusively. Taking chaining into account, there is a combinatorial explosion of possible shapes and kinds of closure, continuation and resultant surprise.

After resolving the existence of shapes on the lowest level, deriving the structural aspects of shapes become an impetus for resolving the perceptual principles that constitute non-linear motion with overall direction. The edges of shapes are starting position and stopping position within the path of a slider. A slider instead of the slider is used here for reasons not yet explained, but relate to the fact that under the right conditions implications produced on one slider can be realized on a different slider. There is this vertical analysis of many sliders occurring at once, as well as a different kind of horizontal analysis that is fragmented across different sliders or 1-dimensional spatial paths.

Returning to the discussion of transforming structural edges of shapes, structural positions when determined transform to the next level, leading to an expectation of the next structural event on that level. Except for the absolute first position (which transforms to all levels immediately), only positions of shapes that share boundaries transform to the next level. To clarify the notion of sharing boundaries: the ending edge of one shape is always the starting edge of the next shape in the sequence, with exception of the first edge of the entire sequence.

In general, the delta of two transformed shape edges or positions, the total time active of the shape, and the direction are computed making a higher-level slide implication. The discussion to follow regarding the implementation introduces a few nuances to the schema for constructing higher level implications. The threshold for the feature of duration (which determines whether a long (1) or short (0) duration has occurred) doubles with each level. Meaning that on the second level the threshold has changed from 1000 ms to 2000ms. In table 1.5, the predicted marker (q) is roughly 30 seconds in the future from (i). The musical system will use these higher level shapes to create paths to reach certain points implied by the past, modifying a layer of parameters of the synthesizer hidden to the improviser.

Table 1.5: The transformation of structure to higher levels. Letters a - i refer to the slider state at either time-gaps or reversals of direction. Letters with an asterisk are the implied destinations or realizations, and the letters above those, not in parenthesis are the actual realizations. The travel from letter to letter is considered a slide segment, as represented by a binary triple.

Lvl 4	a								i								q*
Lvl 3	a				e				i*								
Lvl 2	a		c		e*												
Lvl 1	a	b	c*	d	e	f	g	h	i								

1.4 Generating Paths to Implied Destinations

The computer augments the underlying trends of the source as a future projection gained from the analysis of a framed syntactical unit of the past, but the current model only has a short-term memory. The diagram shown in table 1.5 uses a sequence of only four shapes (as a-c, c-e, e-g, g-i) to generate the prediction of a structural end point four shapes later at the forth level of transformation (q). The first level makes a projection about the current shape after a minimum of two slides. On the second level, the realization of the first shape at c implies the end segment of shape 2 (e). Transformed level 3 finds (i) two shapes later at realization e. The model keeps track only of the vital information of the structural moments at each level updating its projections in a cyclical fashion. The window of memory expands and contracts in time with respect to the temporal density of the motions. The sufficient number of moves in a short duration to allow for a level four projection results in a relatively short window as the sum of the durations, while the opposite results in a relatively long window. Both these scenarios effect the lengths of projections, such as the length of time to reach (q).

It can be seen that with the time window formalism mentioned above the system always reacts in a way that is temporally proportional to user input by following the basic rules of I-R: if A, expect A and if B expect A. Longer duration concatenations of the primary or implicative window, such as table 1.5, window (a-i), imply that a future window (i-q) will be short in duration. Is there a way to break out of this inevitability? The other inevitability stems from the rule: if A, expect A, meaning that a high density window implies a succeeding high density window. Can we only imply high density with every preceding case? Reorganizing the weights of the parametric scales relieves this to some degree. The other way to be discussed in the implementation is that of expanding the temporal response of the system at moments where groups end by closure.

To further elaborate the presentation of table 1.5, the levels of prediction ascending represent increasing amounts of temporal detachment from the improviser's source motion. At higher levels of abstraction, the computer generation increases the improviser's perception of its autonomy by the generation of temporally longer sequences from a start to a destination segment. These sequences gain impetus from sampling a longer span of time as the generation (i-q) relies on the groupings gained from (a-i). In other words, four past shapes should generate four future shapes. The generated shapes should move toward and terminate at (q) as an end position within the specified amount of time.

Also there is the irksome problem here, that might lead someone to believe that Narmour's theory does not function from the perspective of a person who is physically producing the sound heard, where the sound is not immediately located, in temporal terms, to the source motion. The model only works from the perspective of an observer. However, this may be a welcome alterity in the system that via the desynchronization of the temporal ordering principle invites unexpectedness to the music, not by introducing a separate randomizer, but by offsetting subconscious sequencing ideation. Here is the scenario: An implicative group has been formed on any level of the hierarchy, and as the system works, the process of generating the realization is in full flight. What if the sufficient conditions have been performed by the player to instantiate another implication, while the realization is still running? The possible response generated will be thrown out on account of the fact that the response from a previous implication is already occupying that level of hierarchy. From the perspective of an observer to the sound, when realization occurs to them, they are not in the act of producing the motion of implication; they are listening to a realization of the previous implication. The system proposed, asks the improviser on many occasions, especially in those of non-chained reversals, to imply the next while listening to the previous next. And if the reversal continues while the previous next is still running, they realize the next while listening to the previous next. This case describes a temporal reduction and a breaking or embellishment of the entrainment of overlapping. The overlapping contrast of especially the case described could be a programmatic signifier, in the variety that there is some sort of intention to change the structuring of the generated response, but I will leave that there for now. A remediation for the situation described will be elaborated in the chapter on implementation, with regards to chaining similar events. The similarity-chainer has a diversity of conditional criteria serving to group many of the long reversal chains into a single implication suspending the realization response until the structural moment. A final thought would be to create separate threads of generation that initiate when a structural moment is reached while a previous thread is running. Branchings of differing, yet concurrent projected expectations would emerge.

1.5 Theoretical Extensions

To conclude this chapter, as we move toward the description of the implementation in the next chapter, the space is used to elaborate theoretical extensions not implemented but useful proponents to add in a future state of the system. If the implied states are closures of form, then this could be a cue for the generator on that level to dropout for a cycle or two. Creating a sense of space or intelligent density increases and reductions based on felt sensation of proprioceptive markers in the improviser. Also, if a log was kept of sequences that caused surprise, the algorithm could then have an offline component that learned under what conditions (possibly an arc of parametric widths) rules if B, expect B and if A, expect B are expected. The use of offline processing could lead to more relevant extrapolations by bringing higher-level rules to the schema that will reduce problems of formalistic inevitability eluded to earlier.

Currently, in the system implementation, the formalistic inevitability is offset by changing the focus system, as its reorganization of the parametric scales, at key structural moments, like that of the closural groups and surprises. Surprises are sometimes considered interchangeable with closures because by way of the formalism, closures can be defined as contrasts in the trace, but not contrasts in subconscious expectation. Closures are contrasts in the trace because the measured quantities of the motion contrast in magnitudes. The surprise as defined as sequence from A to B, is both a contrast in subconscious expectation and a contrast in the trace. The surprise B to B, is a contrast in subconscious expectation, but not in the trace. There are kinds of contrast, where closural forms in Narmour's theory are based on one kind of contrast, however surprises overlap with this kind of contrast, sometimes having both kinds of contrast at once. The B to B marks the orthogonal instance to B to A in terms of contrasts, but is most similar to continuations, A to A, having no contrast in the trace. As seen in table 1.6, new correspondences start to emerge from two part successions when starting to abstract the definition of contrast beyond that which is in expectation to include that which is in the trace. More co-existing forms of contrast exist, one such example is an elaboration of the trace to consider the relation to its environment as contrast or not. But to return to the leading point, surprises contain levels of potential affect useful as structural moments, distinct from continuations and similar to closures, that can be used to refigure the response of the system to focus on different features as primary for the generation of a response. And as a final point of consideration I will start to sketch out how

Table 1.6: Two part successions of slides contain varying forms of contrast. Contrast from expectation and contrast as present among the measurement of two temporally elapsed items shows new relations, especially correspondences relating surprises to closures.

	Trace contrast	Expectation contrast
A to A (continuation)	0	0
B to A (closure)	1	0
A to B (surprise)	1	1
B to B (surprise)	0	1

to start applying the adapted formalism to function on the entire bank of sliders. When using a bank of sliders there are multiple one-dimension motions possible simultaneously, which requires an analysis of aggregates of motion, as well as, a theory of when implications on one slider can be realized on another slider. For the latter, an issue of referentiality occurs that needs breaking down into two categories of possible motions on a slider. In the system there are three types of reference, listed in order from least to most fragmentary: "self-", "intra-", and "inter-," which relates to the basic structure of a synthesizer. With almost any synthesizer multiple parameters connect to and modify a sound source. There are then multiple sound sources that connect to what is abstractly the entirety of the synthesizer, defined here as the concatenation of all outgoing signals from the sources to loudspeaker destinations. With this said, self-referentiality means that the implication refers to the parameter it originates from. Intra-referentiality means that the implication refers to another parameter connected to the same source. Inter-referentiality means that the implication refers to another parameter connected to a different source.

For the two categories of possible motion we have setting and shaping motions. A setting motion is defined as a continuous slide from a stopped position to a time-gap, with no reversal of direction. It refers not only to the self, said in this way because the least fragmentary realization is the most probable, but can refer anywhere on the synthesizer. The shaping motion is distinct from the setting motion in that it is a continuous slide with any number of reversals of direction from the stopped position to the time-gap. Shaping refers only to the self until a time-gap, where it becomes ostensively a setting, and then refers not only to self in the case that it is not at a moment of realization.

Now that the theory has been advanced describing the way in which melodic expectation as groups over time can be leveraged to create a generator system from musical motions on a linear potentiometer, the next step is to describe the process of implementation in Max/MSP. I will interlace considerations about how to synthesize motions from sparse information, that retains the goal in question of how to advance non-linear forms of musical motion that retain interest in the observer of the sound and produce an experience of flow in the performer, while maintaining a sense of direction and excitement in the sound quality itself.

Chapter 2

Implementation

The programmatic implementation entails more complexity than described in the preceding chapter. First the formation of feature bundles is described as a process of segment a Midi Continuous Control (CC) stream for analysis. The analysis is a mode of interpretation that includes grouping as a naming act based the contents encumbered in a temporal succession and of interest includes a mechanism that chains together similar items. Finally the Thesis closes out with a discussion of how the realizations of performer implication are generated in process, leveraging the analysis of group ends as key moments in structural time for the system itself to apply variation to its interpretation and generation scheme. Figure 2.1 appears below as a graph representing the basic flow of the system, encasing the code blocks to be discussed. This diagram will be referenced in the discussion of the code.

2.1 The Formation of Feature Bundles (or Atoms)

What follows is the method used for creating segments out of slides, referred to in figure 2.1 as the parse step. To begin, the slider outputs a stream of continuous control midi messages at a rate of about every ten milliseconds. In order to detect reversals in direction, a stream of the current and the previous value is subtracted them from each other. This yields the momentary

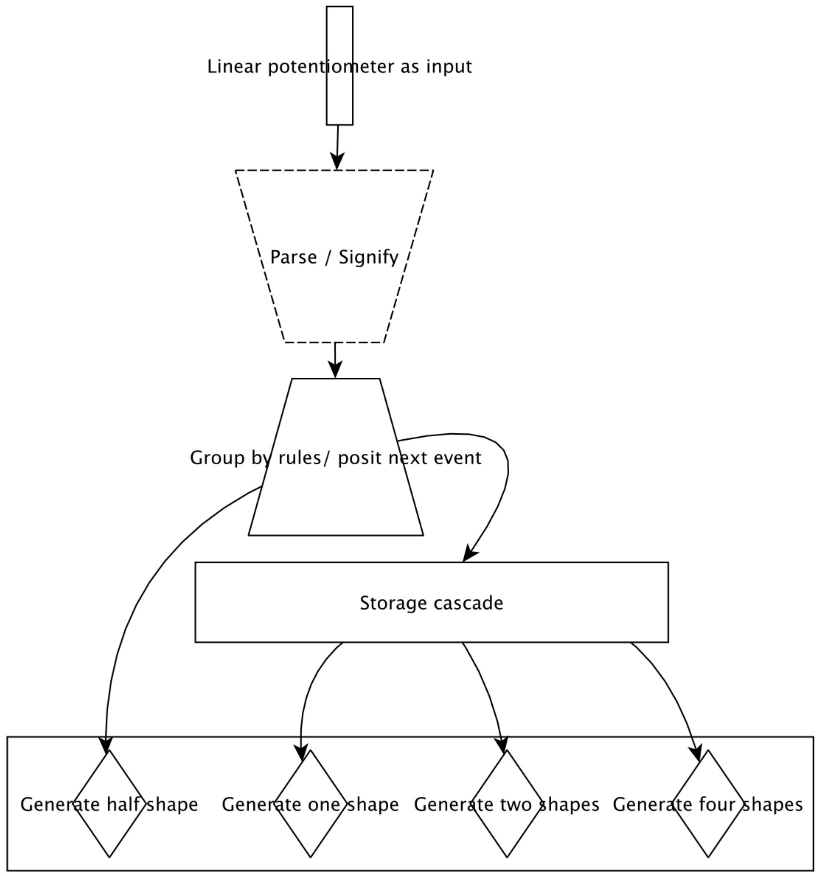

Figure 2.1: The overview of the system implementation. The input is a slider, which is first parsed into segments and signified as atoms. The atoms are grouped, labeled and the next atom is determined. Storages are linked in a cascade to consolidate different chunks of time and fed respectively to the appropriate generator.

interval spacing them. Depending on the sign of interval it can be determined whether the stream, and hence the slider, is moving up or down. Using 0 to signify up and 1 for down, another stream is created of these current and previous significations. If at any point in time, the previous value does not equal the current value, a change in direction has occurred.

Time-gaps are detected with a timeout mechanism. A bang is sent out concurrently with each Midi CC message to the timeout. The timeout duplicates the bang sending the first to the right inlet of a onebang object and the second to the a delay object of 250 ms. Since the delay object reports a delayed bang after the last bang received, but not within a time shorter than the duration of the allocated delay, a bang is reported to onebang's left inlet exactly 250ms after the right inlet does not receive a bang. In this case the slider stopped moving 250 ms ago. When the timer object reports its time of 250 ms a bang triggers, which signifies a segment end by stopping. Down the line in this patch, the 250 milliseconds is subtracted from segments that terminate by stopping, in order to preserve accurate timing.

Storing the start and the stop position within these mentioned demarcations of the edges and taking the difference, yields an interval of a certain magnitude. The direction of the interval is an issue of whether the two stored positions are a minima to a maxima, as up, or a maxima to minima, as down.

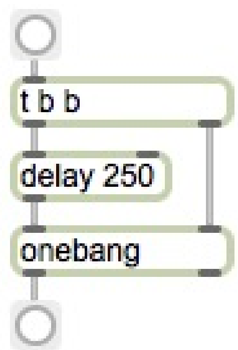

Figure 2.2: A timeout mechanism. A depiction of implementing a timeout mechanism in Max/MSP. Will output a bang signifying that an event has not occurred for exactly 250 milliseconds from the last event.

Separate timers capture the varieties of the combinations of edges, resulting in the use of five timer objects. In terms of the active time when sliders are moving, there are four combinations: the start to reverse, the reverse to reverse, the start to stop and the reverse to stop. In terms of the time when sliders are not moving, there is one combination: the stop to start time. The start can be considered the first moment of activity or the first CC message that occurs after a time-gap. The object onebang is used, only allowing a bang created by any CC message to pass through, if it has received a bang from the timeout in its other inlet. What ensues is an intricate gating system so that the correct timer's duration syncs with the correct variety of motion or non-motion measured and is best shown as in figure 2.3.

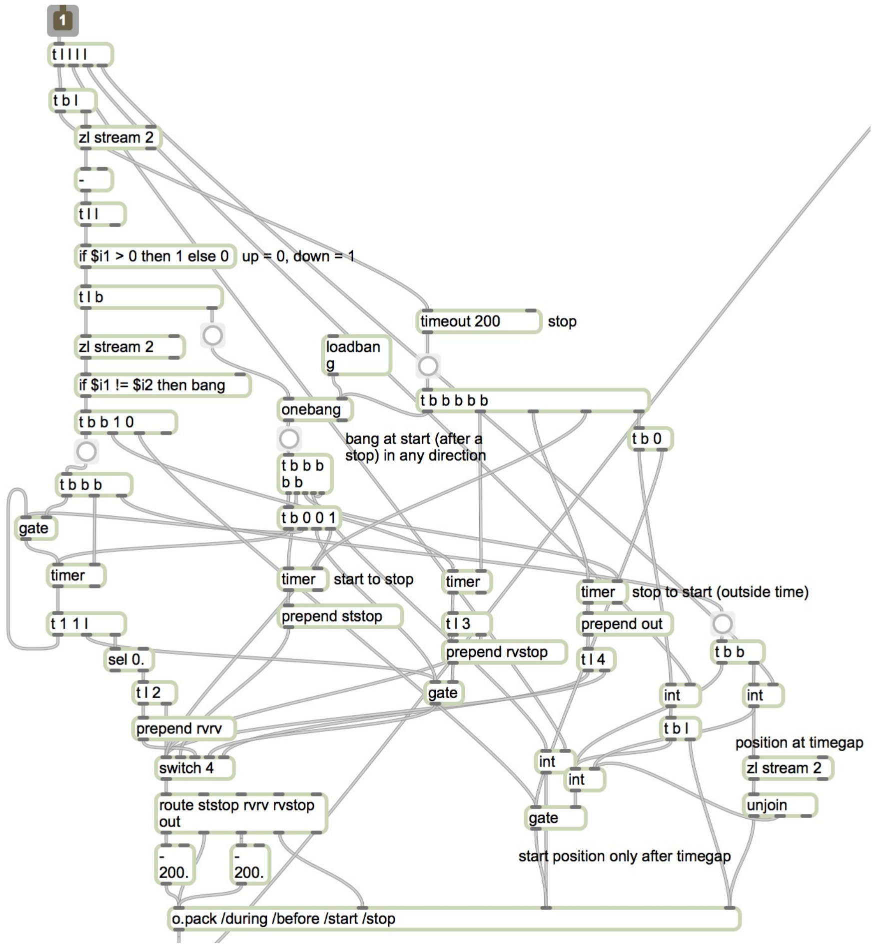

Figure 2.3: The signifier abstraction. The abstraction measures distance, duration and direction.

After obtaining the time before, the time during, the distance and the direction, thresholds are applied to each, except the time before. This step is labelled signify in figure 2.1, occurring directly after Parse as a way to finalize the parsing. To be discussed, the time before is an important condition when determining the chaining of segments. To reiterate, the threshold for duration is 1000 ms, the threshold for distance is half the slider's range which is 64 points in the case of a midi slider and the direction is a function of whether the start and stop points are minima to maxima, as up, or a maxima to minima, as down. The binary triple obtained sends to the next unit, a multi-stage interpreter, which, in general, forms groups of segments.

2.2 Interpreting Successions of Segments

The interpreter functions to designate segments in there appropriate time-step as either implication or realization, generate the accordant parametric scale when new groups instantiate, describe the group as a continuation, closure or surprise, and finally to chain similar segments together to prolong, if necessitated, the implication step. Contained in this section is a description of figure 2.1's block called group by rules / posit next event. To start off with an account, when instances of encoded feature bundles are passed through to the interpreter, a bang is sent to a counter, whose role is to keep track of the time-step of the feature bundle. If the count is zero, then it is the first implication in a possible chain of implications. If the count is greater than zero, a chaining of similar segments is currently occurring and the motions are still implicative. Once the similarity-chainer detects a certain amount of contrast among segments, based on a variety of conditional criteria, it reports a realization, and sends out a bang directly after. Before describing the variety of conditional criteria, it is necessary to elucidate the procedures that occur before the similarity-chainer.

A stream of the most recent and previous state of the slide segment is created. The bang directly after the realization produces a superseding moment, clearing the previous state of the slide segment. Breaking the chain or burden of causation functions operatively to declare a new instantiated token for a parametric scale generator to project an expected value of realization. Each time a new token is instantiated, even if the interpreter is in the course of chaining these instantiations (as from contents actually measured from the improviser), an expectation is formed from this instantiation in the way of generating its accordant parametric scale.

The para abstraction (for parametric scale creator) takes a Max message with a binary triple of numbers such as [1 0 1]. This triple is first sent to an subpatch that splits the triple's bits across three conditional statements as each statements second variable, the third variable in each statement are the 3 weights: [44, 33, 22]. Then an uzi object iterates through the contents of a coll object holding all eight triples. These triples are split up becoming the first variable of each conditional statement tested against the instantiated. The conditional statement reads: if the feature at 1 equals feature at 2 then output the value of the weight, and if not output 1. So if [1 0 1] is the instantiated, we will get a value of 99, [1 0 0] will get a value of 78, and so forth. In the implementation, features dissimilar to the instantiated are given 1 point instead of 0 for various reasons. This makes the value after 99, 78 not 77. A list is then created out of the triple and the value, at first in the form of 99 1 0 1. A regular expression is then used to remove the white spaces making it a number : 99101. The eight triples are formatted in this way and sorted from high to low. Another regular expression puts back the single white spaces. Then the digits representing the weights are taken off the front of the number, i.e. 9 9 1 0 1 goes to 1 0 1, and sent (now in correct order) to a coll object. Now any of the tokens of any of the 8 parametric scales can be called simply by an ordinal value, as an index from 0 to 7. The final copy of the instantiated triple goes to a subpatch that determines if the instantiated is in fact an A or a B. If it is an A it calls the canonical prediction following the rule: if A, expect A as index 0. If it is a B it calls the canonical prediction following the rule: if B, expect A as index 7.

As mentioned, the algorithm described stores the token from the previous time step, known as the was now. Para takes the was now token and posits the expected value, known as the expected. The third token in consideration is the now, one time step after was now. The first step in determining the marking for each shape as either continuation, closure or surprise, is to create an expected weight calculated from the bit difference of was now and expected. This is done in the form, as described in the para object: if the feature at 1 equals the feature at 2, then output the value of the weight, and if not output 1. The next step is to create a now weight as the weighted bit difference of was now and now. Finally, a difference at now is constructed by taking the absolute difference of the expected weight and the now weight. The difference at now proves crucial to determining the marking for the shape.

When either an A or B is instantiated and the difference at now returns a number less

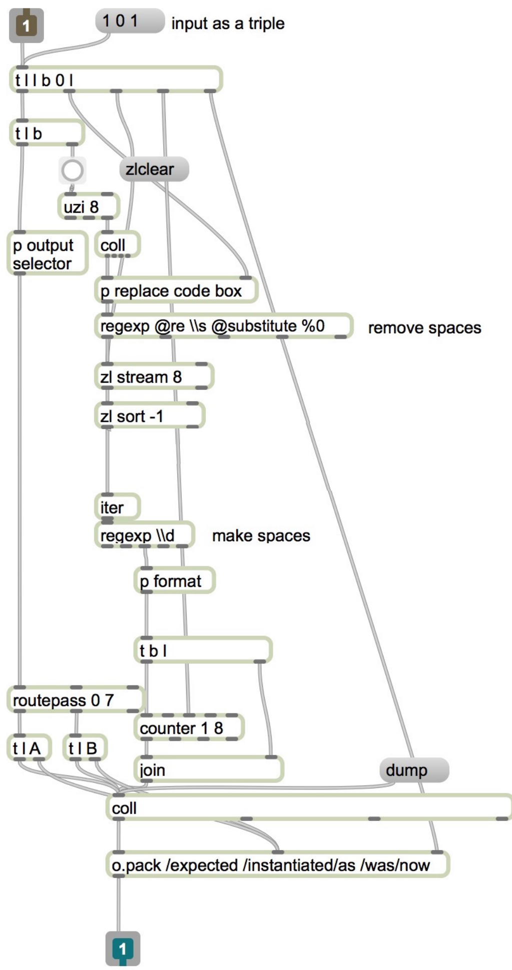

Figure 2.4: The implementation of the para abstraction functioning to create the parametric scale from the instantiated and posit the expected value.

than or equal to 43 weighted points, the resultant marking is not a surprise. If the group is instantiated as an A, and the difference at now proves below the threshold, the group is labeled a continuation. And finally, if the group is instantiated as a B, and the difference at now proves below the threshold, the group is labeled as a closure. We now have groups ostensively terminated as either continuation, closure or surprise. But the terminations of the groups are not finalized until passing through the similarity-chainer for a final assessment.

The similarity-chainer operates on conditional criteria to assess whether two segments in adjacent time steps are sufficiently similar. When any of the criteria prove false, the chain breaks. The first criteria has to do with relative average speed of the two segments. A simple calculation divides the distance traveled by its travel time, for each segment, and then divides the resultant values to get a number representing the relation of the speeds. If this number is greater than 0.75 and less than 1.666, the speeds are considered similar; otherwise the chain breaks. A second criteria is, the time of non-activity shall not exceed 1.5 seconds. Another criteria is, as referencing the parametric scale, the succeeding similar segment cannot be more than 2 indexical steps from what was implied by the former. So in the case of A's the index of the successor need be from 0 to 2 and the in case of B's the successor need be in the neighborhood of 7 and 5. The final criteria is that the group passed to the similarity-chainer cannot be labeled a closure.

The similarity-chainer has three outlets. The first goes directly to the first level generator only containing contents encumbered by time steps labeled implication. The second goes to an abstraction that aggregates the stream of possible features chained, by way of a recursion, in a list. The third emits a bang directly after a realization on the lowest level, which has the function of ending the list mentioned and propelling the structural components to higher levels of generators, to be discussed.

2.3 Generating Realizations from Interpretations

The first level generator gives one time step of generation from one instantiation of analyzed motion as any implicative event. In other words, directly after the implicative moment of the performer, the mechanism rolls out a suitable extrapolation, ideally, before the performer realizes themselves. The extrapolation mentioned projects either a possible chained event when given a token from the A group or a possible realization when given a token from the B group. The discussion to follow relates to the generate blocks in figure 2.1. Each generator receives a tap of memory from the correct location in the storage cascade block depicted in the same figure. In the storage cascade, each level clears after the limit amount of implicative shapes are output, in order to not include repetitions of the same implicational shapes for the next generated group. This means, for example, that after four shapes are output on the fourth level the cache is cleared.

The algorithm at this stage depends on having the previous time step with the current time step in order to make assessments about whether implications are prolonged in chains or a termination qualifies, necessitating a gestalt pointer. With this in mind, the level one generator functions primarily as a linear interpolation of a slide segment generated from the implication derived from the instantiated token by the performer and the actual measured features from the performer that are referred to, in terms of the time step, by the other segment. It functions as a retrospective interpolation of a imaginary segment and a real segment both related by the time step intended.

The creation of the imaginary segment involves a modified version of the para abstraction described above called paragen. The modification alters the selection of the response to the input. Instead of always selecting the canonical response, as either a duplication for A's or a total opposition for B's, the selection is made by tracking the number of bits that changed from the previous feature bundle to the current. If the expected value is an A, the response is selected from the parametric scale as an index from 0 and 3. If the expected value is a B, the response is selected from the generated parametric scales as an index from the difference of the index maxima of 7 minus a number from 0 and 3. In this scheme, the response selected always conforms to the low-level sub-conscious rule of Narmour: if A, then A and if B, then A. Also of note, the expected token was earlier generated by a para abstraction in the interpreter. So as input to paragen, it is already an imaginary value.

The expected token is passed to an abstraction called focal that has the function of outputting a linear succession of values conforming to the input it receives. The token splits into three places delegating how the time, the distance and the direction unfold. Time is the only real valued feature fed forward and used throughout the system as an input proportional way to construct motion of the generators.

The time from the real event informs the timescale of the imaginary generator. This means that the real time can be greater than the threshold of 1000 ms and the expected token can signify the need for a duration of less than 1000 ms, which occurs frequently due to the closure rule and the fact that the performer may be subconsciously planning a surprise, in the form of B followed by B. Of course, the B could be followed by a long duration A, in the case of [1,0,0]. Much less frequently, the real time might be a duration less than the threshold and the expected token might signify the need for a duration greater than the threshold. This is true with expected tokens [1,0,0] and [1,0,1], only when the improviser performed those exact tokens and paragen randomly selects the canonical prediction as its output.

With the first bit from the expected token and the time from the real segment, the time of imaginary segment is formed in the contime abstraction. Due to the possible conflicts of these two items described in the preceding paragraph, contime has four possible ways to construct the time. A short duration from a long duration, a long duration from a short duration, a short duration from a short duration and a long duration from a long duration.

With the time determined in the contime abstraction, a distance is generated randomly above or below the threshold depending on which bit it is passed. The reason for such a lack of acuity in regards to the distance is, perception of distance in timbre space as opposed to pitch space is less sensitive to the nuance of differing distances. All that can be perceived reliably is the difference of a small change and a large change when operating on timbres. Particularity in the case of timbres with dense spectrums there maybe not even be a real discernible difference distinguishing a small change and a large change. A vast motion by a performer and a limited perceivable change in the sound quality obfuscates the relationship of the sound and the gesture in terms of embodiment and intention on the part of the performer.

In terms of performance, it may be useful to evoke a traditional form of electronic music to characterize relation of gestures on timbres, which is the acousmatic. Sounds produced by diverse means in the studio, at a time before the sounds are heard in concert are not observed in respect to the gestures forming them, they are shrouded by barriers and time gaps. Sounded over loudspeaker, causative gestures do not appear, heightening in the observer a sense of inquiry about the identity of the gesture's activation of a material to account for the sound heard. Exactly what does the event heard reference, a kind of idea in terms of semiotic association as a mental sign or a bodily interaction with a musical material?

Two things emerge from the comparison with acousmatics: invisible reference to the body and the material, and a temporal detachment from when the gesture occurred on the material and when the sound is heard as an electronic reproduction. But what does the invisibility of reference have to do with an expectation of an immediacy of relational action amongst the gesture and material? This is to say, isn't it expected that if the origin of sounding event was visible, it is only visible because there was a distinct relation of similarity relating the course of the performer's motion with the changing qualities of the sounding material. With a simplified example of a long gliding motion producing a negligible change in the timbral quality of the sound, we have acousmatic qualities, even though the body is visible. Taking these cues, the observer could consider the performance of instruments with algorithmic alteration as a kind of live acousmatics that complicate expectations regarding live sound organization.

Returning to the discussion of the implementation of the level one generator, the third bit of the triple stating the direction is used to multiply the generated duration by either a negative one if it down or one if it is up, and is added to the value fed back of the previous destination. The time and the value become the destination point and the duration of travel for a line object. Before they are passed to the line object, a pipe object is used to delay the onset of the line output, by an interval created from the most recent time of inactivity. Immediately after this, a gate is closed, only to open again to accept new time/destination pairs directly after the line object stops. Because the distance implied by the newly created distance can exceed the boundaries of the parameter space, the output of the line is fed through a subpatch that wraps its output within 0 to 127. In terms of wrapping, it may be that time, or a magnitude of displacement as either a setting position to result or as an intensification to the instantaneous experience of shaping change is the primary goal on a subconscious level for what the performer wants to accomplish with a given action. And given the boundary enforced by the limit of a slider, the only thing to do is reverse direction upon approaching the border. Many of the short durations, short time and directional changes measured result from the physical limitation of the boundary imposed, having an overall, observable effect of limiting motion. In terms of the output of this generation, the wrapping technique also tends to slightly randomize the output away from the pole of absolute predictability, making long duration flights have moments of reflection while maintaining constant speed and further fragmenting short durations in the same way.

Even with this slight foray into randomization, the generation especially on this first level was alarmingly predicable. It's next move was frequently highly to related to the new move I was preparing. A solution for this is to simply delay the stream of destinations one time step in relation to the stream of durations. This action makes a large difference in terms of the playability and also supports the notion that the syntax of the formalism described is relevant in terms of combining single features into an atom and furthermore basing the decision about what is the next atom on the relation of the features from the previous atom. A slight temporal offset or disturbance in the formation of the atom makes the next less predictable and more engaging, while still maintaining a thread of coherence to follow on the basis of the progression of individuated feature flows.

For the second level generator, the analysis of one shape yields a generation of one shape, which needs to include both an implication and a realization, possibly in the form of a chain and a resolution. To begin to construct a higher level response, a new code must be constructed for the kind of resolution that is to occur. To do this, the start and stop position are taken from the implicative shape produced on the slider and the total distance is computed. Because of the fact that reversals of direction are common to occur on the same slider, the starting position and the end position can be very near in position. A common scenario is that there was a large number of chained segments that compose the shape or group. So in determination of the distance as either 0 or 1, a condition about the total distance of the group is added, that being, the total distance as a sum of the distances travelled within the group must be greater than 127, otherwise the displacement must be greater than 64. In regards to the time, the threshold is doubled from 1000 ms to 2000 ms for the sum of the total time of implication, the time from the end of an implication and to the start of a realization and the time of realization. Finally direction is computed in the normal way.

For the construction of a second level implication, the abstractions paragen and focal are used again. However, paragen is used in a feedback loop with itself after receiving the last implication in the chain, the number of implications in the chain (which could be one), and the total duration of the implication chain. The number of implications (aI) are used to subdivide the total duration by an random integer from 1 to aI + 2. This creates a uniform duration to be combined with all ensuing codes for interpretation and generation by focal. The random aI sets the limit of a counter, that when reached, triggers the generation of a realization. But before the limit is reached, the implication code of the last in the chain by the performer starts off paragen in a feedback loop. It interprets this code, sends out a similar code to focal, and stores one in a zl reg object. After focal outputs a segment, it triggers the code stored in zl reg as input to paragen, which again sends out a similar code. This process shown in code, as in figure 2.5, continues until the limit of the counter triggers the waiting realization code gained in the way that was described in the preceding paragraph. At higher levels, all threshold numbers for conditional statements classifying higher level ontology of form double. The programmatic process of constructing sequences also doubles with each level. Repeating, categorically similar, forms follow from this approach based on overlapping but offset samplings of shape successions, and resolving with the accentuated motion implied. An example of the traces from four levels of output are shown below show in figure 2.6.

Instead of leaving each level independent of each other, each level has a possible connec-

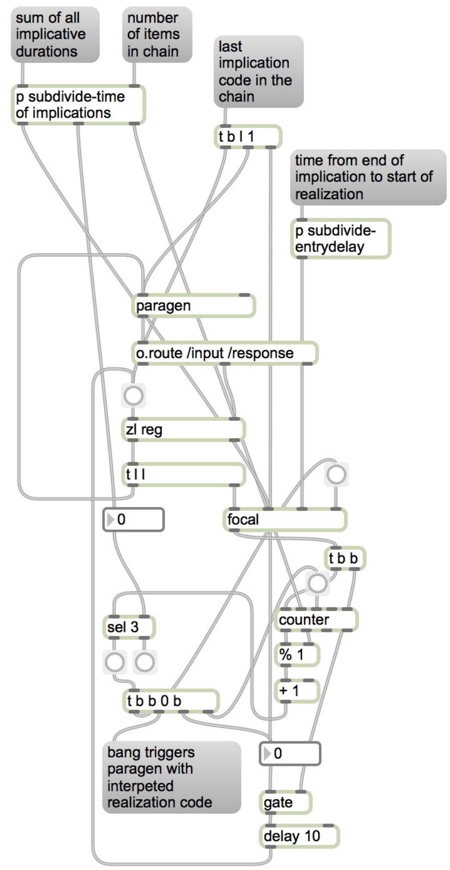

Figure 2.5: A code block used for higher level generators. The depicted functions to determine if the algorithm will generate a chained sequence or a non-chained sequence before triggering the waiting destination value gained from interpreting the implications of the performer's actions. This block of code can be chained to together in a sequence to function on higher levels.

tion with the temporal domain of an alternate level. This describes in part why the traces in figure 2.5 are not always comprised of segments that are linear or continuous. For performance of the system, there are connections made from the first to second level, and the third to the forth. Many different topologies could be devised. In general fewer connections lead to more independent sounding layers, while at the farthest end of the spectrum, connecting everything to everything produces layers that are sonically indistinguishable.

The connections are made by linear interpolation with a weight value derived from the time interval separating the end of the previous segment and the beginning of the next segment, as

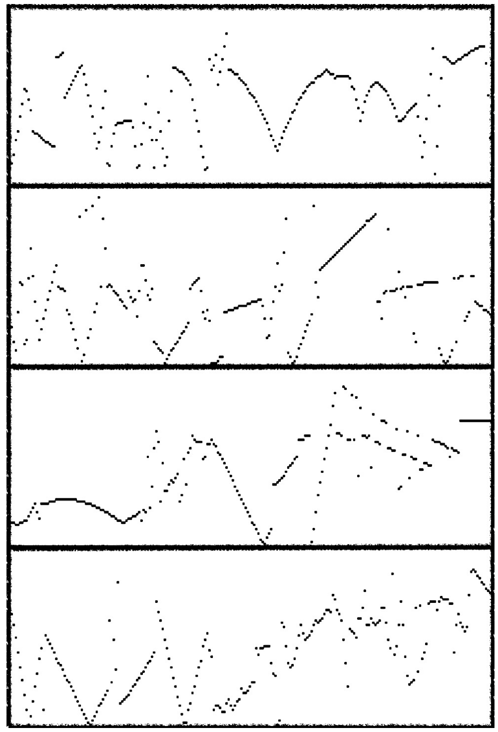

Figure 2.6: Traces of motion generated at four levels of structural time. The topmost container is level one and at the bottom level four. The individual segments of motion generated aren't always continuous or linear. Weights of the scales are in descending order, or in other words, the focus of the system is related to durations.

captured most recently from the impinged level. The derived time value informs the rate at which the weight vacillates between the impinging level and the impinged level, and also the extend to which the weight travels toward, in favor of the impinging level. The last consideration when combining levels is that if the impinging level is inactive, a timeout is used to detect inactivity and a gate is flipped to output only the primary level (what was being impinged upon) at that time. This kind of mechanism to could be extended so that coupling is dependent on the state of the impinging level. So instead of activity being the only condition for coupling, something like a range of accelerations values opens the gate. The visual effect of these relations is again illustrated in figure 2.6, it can be seen that curvilinear motion and discontinuities result from connections in this fashion, interspersed with linear motion when the impinging motion ceases.

The arrivals at generated non-motion appear to have a group ending significance as the result of instituting chaining and terminating by a formalism that predicts the last event before a non-motion. It is of importance to note that zero-acceleration refers to an absence of change in the model's parameters and does not mean silence. To create more pronounced, musically significant stoppages, chaining rules for the realization could be adapted as related to the content of the preceding implicative chain.

Non-motion moments by the performer, like that of closures, are considered significant stoppages for the system to respond. At these moments, it can change its focus to a different feature of the performer's input and also change the relative amount of time it expands the set of durations it schedules. When focus changes suddenly and durations expand greatly, to say, four times relative to the performer. A moment of lacking control over the system arrives, and all that is left for the performer to do is wait for the implied trajectory to take its course and plan the next set of implications for the music to take on. This time off can be important in the reflection and planning of future states in an improvisation.

Bibliography

Canonne, C. (2013). Focal Points in Collective Free Improvisation. *Perspectives of New Music, 51*(1), 40-55.

Deutsch, D. Grouping mechanisms in music. (2013). In D. Deutsch (Ed.). The psychology of music. (3rd ed., pp. 183-248). San Diego: Elsevier.

Eckel, G. (2012). Embodied Generative Music. In D., Peters, G., Eckel, & A. Dorschel (Ed.). Bodily expression in electronic music: perspectives on reclaiming performativity. (1st ed., pp. 143 - 151). New York, NY: Routledge.

Leman, M. (1995). Music and schema theory: cognitive foundations of systematic musicology. Heidelberg, Germany: Springer.

Narmour, E. (1990). The analysis and cognition of basic melodic structures: The implication-realization model. Chicago, IL: University of Chicago Press.

Pearce, M. T., Ruiz, M. H., Kapasi, S., Wiggins, G. A., & Bhattacharya, J. (2010). Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. *NeuroImage, 50*(1), 302-313. doi:10.1016/j.neuroimage.2009.12.019

Tenney, J., & Polansky, L.(1980). Temporal gestalt perception in music. *Journal of Music Theory*, 24, 205-41.

Wang, L. C., & Tsai, C. G. (2012) Embodiment of Metrical Structure: Motor Patterns Associated with Taiwanese Music. Paper presented at 12th International Conference of Music Perception and Cognition, Thessaloniki, Greece. doi: http://icmpc-escom2012.web.auth.gr/files/papers/1120_Proc.pdf