Modelling Emergence in an Interactive Music System

Jonathan Impett
Faculty of Music, University of Cambridge
West Road
Cambridge CB3 9DP
UK
jfi21@cam.ac.uk

Abstract

This paper presents work in progress on the generation and cognition of emergent structure in interactive music systems. It is suggested that the affordance and perception of emergence are central to musical experience. We propose an account of the multi-levelled, dynamically parallel nature of musical activity, and describe a system for interacting directly with this aspect of musical production. Various models of computational emergence are discussed in terms of their descriptions and redescriptions of the musical behaviour of a complex adaptive system.

Keywords: complex adaptive systems, dynamical systems, emergence, interactive music, musical behaviours

1 Introduction

The understanding of musical experience as rooted in a process of learning, description and prediction goes back to (Meyer 1956). Meyer's synthesis of ideas from Gestalt psychology and information theory was rooted in an explicit desire to develop a theory that was both dynamic and contextual. The diachronic nonlinearity of this process has since proved one of the most elusive qualities for learning-based models of music to capture. It is not proposed to give a full review of relevant research at this juncture, as all models of music, computer-implemented or not, can be seen as being positioned somewhere on the plane of learning and prediction, even if some are effectively still at learning (0), prediction (1).

From the point of view of learning and prediction, one of the most "organic" qualities of satisfactory musical experience is the way in which elements appear to the listener to emerge from the musical surface, recombine and transform in a dynamic hierarchy of significance. As the listener or composer interacts with the perceived musical materials, an analogously hierarchical, parallel sequence of new understandings emerges.

The aim of the present project is to try to capture something of these qualities in a model for interactive music composition and performance - a system wherein the music is computed in real time as a function of some expression of the processes and materials of the composition, and of information which represents a current human performance. The system has been implemented using the meta-trumpet described in (Impett, 1994), on a Silicon Graphics O2 which also performs the output (processing of live and recorded sounds) using Max/FTS, a sound-processing package from IRCAM, Paris.

2 Musical Activity and Emergence

2.1 Engagement

Engaged musical activity is here understood as encompassing all its many forms - composition, performance, listening and analysis. All these activities involve learning, description and prediction, but all also share an aspect of generative emergence; that is, unpredicted, novel structure arises from the interaction with the materials at hand. From this point of view, the object of study is not just a computational model for interactive music composition/performance , but of the "other" in musical activity in general. This "other" is the musical object in the terms of this paper. The musical object will be seen to have a dual nature. It both constitutes an extension of the musical faculty of the person with whom it interacts, and has a degree of autonomy - a capacity for internal interaction.

The type of emergence that seems to occur in the non-repeatable richness of musical experience has much in common with what Hendriks-Jansen has described as "interactive emergence" in the fields of situated robotics and Artificial Life (Hendriks-Jansen, 1996). Higher levels of "intelligent" behaviour are not directly specified or represented (that would smack of the grail of total serialism) but are rather the aggregate result of the interactions of simpler, local behaviours with each other and with their environment.

2.2 Listening-in-the-World

A technology that would engage with human experience necessarily brings with it a cognitive theory. Here it is proposed that musical listening parses its input in a multi-levelled parallel fashion by identifying and predicting strands of behaviour. It shares this mechanism with "listening-in-the-world", but constrains both the acoustic and intentional spaces. The behaviours of most interest will be neither fully predictable nor completely random, but exhibit dynamic and occasionally emergent behaviour, as is the nature of natural systems. Mutual change in behaviour patterns - interactive emergence - is therefore the object of interest, generally the result of the interaction between behaviours, with their environment, or with themselves. (These must be understood by similar mechanisms to permit the dynamic changing of boundaries proposed here.) As behaviours change or interact, their activity will have to be qualified or redescribed, with consequences for the behaviour associated with the describer. Thus not only behaviours but also descriptions interact. This process of redescription is seen as a vital dynamic of musical activity.

The dual nature of the object means that one is simultaneously ascribing to it some form of self-awareness. Behaviours are thus subject to description and interaction both between object and musician and within the object itself. The implicit ascription of intentionality to these behaviours motivates a search for coherence, and hence in their evolution a cycle of description, elaboration and redescription. We adopt Crutchfield's understanding of computational emergence as self-redescription (Crutchfield, 1994), and posit the same mechanism in both the internal evolution of the behaviours of the object and in the musician's understanding and interaction.

Musical engagement can thus be understood as the search for emergence in perceived musical behaviours. In addition to redescription at moments of phase transition in identified behaviours, a process of prediction (by simulation, in computational terms) confirms, denies or elaborates the present descriptions. With experience, signs of impending phase transition or emergence will engender preparation for redescription, instability and the consideration of alternative understandings, as well as their implications for the future state of the object and of its relationship with the cogniser. The richness of behaviour is considered as critical as its coherence - a potential conflict which largely drives the evolution of the system.

3 The Model

A dynamical systems approach seems most appropriate to a system in which time-evolution and external conditions and environment are critical (Van Gelder & Port, 1995). In a *phase space* consisting of the possible states of the music, with a dimensionality of at least its number of variable parameters plus time, structure of any sort can be represented as an attractor; a trajectory of unique states with some degree of self-similarity. Major changes in the behavioural world of the music correspond to *phase transitions* in the system.

The capacity for self-redescription would imply of a unitary system that it already accounts for its own reformulation, and internal emergence would not strictly be possible. The system at any moment would be an incomplete description of an implicit super-system. Instead, the whole is therefore divided into behaviours and describing agents. This internal division permits something of the dialogue which it was suggested above might be essential to engaged musical activity, and the two elements are here discussed separately.

3.1 Behaviours

By analogy with the above axiom, a theory (whether a cognitive theory or a composition) likewise presupposes embodiment in a technology (which may be a body or an instrument), and situatedness in the environment of its activity. Personal musical activity is the result of the interaction of a specific embodiment of a musical object with a particular environment. The interactive performance nature of this project is a discipline for adhering to this principle. At the lowest level, musical "behavors" are situated, embodied agents (i.e. contingent, constrained and aware), interacting hierarchically and acting in accordance with present circumstances and ideas, not prior plans.

To reflect the hierarchical and parallel nature of music, a Complex Adaptive Systems model has been adopted (Holland, 1995) - a population of micro-behaviours that can multiply, be active, dormant or die, be transformed, and aggregrate dynamically with other behaviours. Their generation, the level and nature of their activity are governed by the environment they inhabit - a changing landscape formed by the other behaviours and the "outside world" (live musicians in this case). Micro-behaviours generated are evaluated and tagged in the light of the behaviour of the musician in the current performance, and in terms of a set of basic cognitive constraints. Each behaviour has a capacity for self-interaction, described below. In the basic model, their activity takes place within a simple time - frequency - intensity space, taking account of rhythm by sampling frequency at small time intervals.

In its simplest form, the landscape consists of a set of sites in n-dimensional space, representing, for example, values of melodic interval, duration and dynamics. In accordance with thresholds set by the composer, agents develop at these sites, born of and fuelled by the incoming musical material. In as far as they have the resources, these agents compete to be allowed to respond to that input, and in the process form hierarchically-structured melodic, harmonic or rhythmic aggregates in accordance with the strength of their bid and their respective tags - state characteristics reflecting their initial birth and current associations. The pitch input is biased in accordance with Schellenberg's rationalisation of Narmour's implication-realization model (Schellenberg 1994; Narmour 1990), to even the playing field until the actions of composer, performer or the describing system begin to shape it. In its current implementation, the output of this system is sent as MIDI data to Max/FTS to control sound processing and synthesis parameters. The description system can change this landscape by controlling the resourcing and thresholds of certain areas, by seeding or inhibiting agents or aggregates, or by filtering their behaviour.

3.2 Description

The mode of description has to be considered in terms of its capacity to generate information which is relevant to the performance function of the system, its predictive power, and descriptive and computational efficiency. The dynamically hierarchical nature of music presents corresponding difficulties for a predictive description of the behaviour of a musical system. Not only must it strive to encapsulate the re-aggregation and scaling discussed above, but it must do so in terms that have both general and specific validity. A single event is a poor indicator of its successor, especially if we consider a uniformly sampled time-series (i.e. wherein most values are not new events). On the other hand, generalised or statistical indicators only relate to general states, and "fuzzy" values stand little chance of being appropriate. In performance terms, music is a succession of *specific* events; the result of a narrative of specific, contextual interactions, possibly widely distributed over time. The need for specific values becomes yet more acute in the machine context.

Considered as a dynamical system, a behaviour can have only one possible subsequent state; that is, the system would be fully deterministic if all the parameters were known (including those of the performer!). The task of the description in reconstructing a dynamical system is to capture the crucial parameters preceding points of bifurcation - moments of selection between forking paths. The problem is not only to find attractors in the phase space of the system, but to identify an appropriate dimensionality and mapping for the phase space itself. The incompleteness of the system's knowledge of the performer (or composer, over a longer timescale) guarantees the need for redescription, affording interaction and the possibility of evolution.

Various techniques have been examined and tested in the course of development, and a brief description of their appropriateness to an arbitrarily complex, improvised musical language gives some idea of the issues that have emerged.

The performance of standard linear techniques with respect to the nonlinearity of a rich musical surface can be understood by analogy with conventional musical analysis. Schenkerian analysis could be regarded as a process of three-dimensional projection of a tonal musical surface (foreground, middle and background) which can capture significant characteristics of the surface, with increasing loss of temporal detail. It relies, however, on a highly-correlated harmonic framework which is not the case in most musical contexts. The dimensionality of a paradigmatic analysis (analysis by contiguous melodic motif) is determined by the number of paradigms identified. In music to which this is naturally applicable, events can be placed within this space with greater short-term predictive power.

Both generalising and surface-specific techniques of this type (correlation and differentiation, for example) are inefficient at capturing emergence or the indicators of nonlinearity in a complex musical behaviour. The *correlation dimension* (Grassberger and Procaccia, 1983) was therefore used as a control parameter for the application and distribution of the results of linear analysis techniques. Recalculated with each new event, it provides an index of the self-similarity of the stream of musical events. Two levels of difficulty appeared at this point. Firstly, such a mechanism introduces a stochastic element into the system, which is then no longer deterministic. Secondly, to track the changing aggregate behaviours of the behavioural system, a corresponding population of describing agents is necessary, none of which characterises the system as a whole and bringing a high computational cost.

Two standard groups of technique exist for the reconstruction of an arbitrarily-dimensioned phase-space: the description of a dynamical system by genetic algorithm (Packard, 1990), and *state-space reconstruction* (Kantz and Schreiber, 1997). The latter uses nested embeddings to find an appropriate space in which to represent any structure the data may express, and in a world of infinite and instant computing power may be very relevant. In this real-time real world context, the specific problem is again that the surface must be constantly sampled as a time series, rather than the mechanism being driven by what we perceive as musical events.

A connectionist approach to issue of behaviour description was ultimately selected. The simple recurrent networks described in (Elman et al, 1996) embody a form of non-representational memory of the evolution of the system, by feeding the current state back with new input; a phase-space filter, in effect. Such a network is fed both the live performance and the output of the "behaving" system described above. By monitoring the changes in the weights of the network, the relevant dimensionality of a particular activity (analogous to musical structural hierarchy, or the aggregate behaviours within a population of autonomous agents) can be judged, and the "meaning" or mapping of the state of the network for the musical behavioural landscape changed accordingly. The results of cluster analysis of the network weights are used calculate control parameters for the landscape.

3.3 Redescription

The assigning of arbitrary numbers of parallel describing agents in earlier versions of this system raises difficult cognitive issues. Presumably the number of *conscious* agents posited must be limited, and yet arbitrary restrictions on the observation of significant inner structure may discourage the interactive emergence sought. In fact, Crutchfeld's computational emergence implicitly takes account of this by relating the moment of redescription to the descriptive limits of the particular computational context. If this implies a cognitive theory (which he does not claim), its implementation is by emulation, not simulation.

In the current version of this system, redescription takes one of three forms:

i) the re-initialising or major reconfiguration of the behavioural landscape.

ii) the resetting of the weights of the network (re-seeding or setting to vectors stored at critical moments of previous runs).

iii) the running of a parallel network which is constantly learning, and to which the weights of the landscape-managing network are set at points of redescription.

The redescription itself can be prompted by two circumstances:

i) observing sufficiently strong indicators of impending emergence in the observed behaviour. This may be a function of an unsatisfactory description, in which case "realisation" is a more appropriate analogy than emergence.

ii) reaching the limit of the system's capacity (real or virtual) to add to the complexity of its current description. The control of artificial limits is a means of managing the degree of redescription of the whole system, an important factor in its perceived energy level. In the case of a fixed-architecture network, the computational complexity clearly does not increase with the "knowledge" represented. Instead, a difference function is maintained, based on prediction error with respect to certain aspects of the performance. Again, the constitution of this function becomes a compositional decision. A threshold can thus be set - dynamically, if required - which triggers the redescription.

This difference is analogous to the "discrepancy signal" of Linear Predictive Coding, and reflects the "noise" unaccounted for by the previous description. It is then used as the initial source of energy and conditions for the newly-expressed behaviour, as it interacts with its own "differend". This "self-environment" thus mirrors the relationship between performer and system/composition.

By understanding each identified behaviour as a dynamical system, emergence can be seen as a process of phase transition. Taking a phase portrait view of each strand or aggregate of musical behaviours, the indicators of emergence, of impending redescription, are the precursors of phase transition - the change to a new attractor or range of possibilities for that behaviour (subsystem). These include an increased criticality, or sensitivity to parameter changes, and can thus be "forced" by interaction of the behaviour is sensitive to some aspect of the musician's performance as an order parameter. It is important to note that the range of possibilities for a behaviour is limited dynamically by external parameters (human interaction and the behaviour of other aggregates), and its own state at the moment of redescription. The initial conditions of a redescribed behaviour determine its activity in the case of a multistable description.

The hierarchy and scaling of complex structures can seen as analogous to those qualities of music identified elsewhere as the keys to a cognitive-based analysis. The self-redefinition of interactive emergence parallels the way in which emotional response itself reshapes the mechanisms of response. The musical object might then be an active extension of working memory, to some degree common to different participants, the response of the whole being reshaped by responses not only present but past. These responses are themselves the result of interactions which may be distributed through space, time and artefacts.

4 Further developments

At present, the system described here requires the "hand-tuning" of parameters, to avoid a lengthy learning process which would soon be overtaken by the evolution of other works or techniques. Indeed, this tuning process could be considered an essential part of the narrative of composition itself. In general, the techniques and materials of a composition develop together in some sort of dialogue. The pre-teaching of certain material to the system would be simple to imagine. However, a more interesting development might be observe the evolution of the weights of the describing network as it moves towards critical moments during the composition phase, in order to be able to induce certain types of phase shift in the piece as it is performed.

An implementation of such a reshaping mechanism would permit the landscape inhabited by the behaviours to be shaped by predictions made by the current hierarchy of descriptions as to their stability or the likely nature of any change, and the perceived implications of these for the behaviour of the rest of the system. Memory, by this understanding, is not a process of storage, but more akin to the evolution of resonances in the tuning-in-time of the system; sequences of changes to its parameters, or perhaps a high-order "differential filter" with a time dimension.

5 Conclusion

The hybrid nature of the model proposed here allows for both internal and external interaction - a cycle of description, prediction and action - to be dealt with in the same terms. There can be no "universal" implementation; the construction of a particular matrix of such relationships is essentially an act of composition. Only where knowledge changes state or crosses some inter-modular boundary can it be said to afford anticipation or interpretation. The process of emergence-perception at the heart of the model presented here could be seen as analogous to the "representational redescription" proposed by Karmiloff-Smith (1992) as being a vital mechanism of cognitive development.

The interactive situatedness, contextuality and time-dependency of the phenomena under consideration here are such that whilst their observation in extant music is both instructive and ever-changing, their *simulation* in a dislocated context can only be of limited value. The "voicing" of a model is a quality often neglected in its definition. That "it" does this or shows that is generally only half the story; its developer and user are indispensable to its functionality. Their interaction with the model forms part of a dynamic process of learning and prediction, such that the most static technique or inert tool can be seen as representing one stage of such a process. The approach presented here attempts to take account of this aspect by explicitly situating simple techniques in the activity of music.

References

Crutchfield, James (1994). The calculi of emergence: computation, dynamics and induction. *Physica D* 75, 11-54.

Elman, Jeffrey L., E.A.Bates, M.H.Johnson, A.Karmiloff-Smith, D.Parisi, K.Plunkett (1996). *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge MA: MIT Press.

Grassberger, P. and Procaccia, I. (1983). Measuring the strangeness of strange attractors. *Physica D* 9, 189.

Hendriks-Jansen, Horst (1996). *Catching Ourselves in the Act.* Cambridge, Mass.: MIT Press.

Holland, John (1995). *Hidden Order - How Adaptation Builds Complexity.* Reading, Mass.: Addison Wesley.

Impett, Jonathan (1994). A Meta-Trumpet(-er). In *Proceedings: 1994 International Computer Music Conference.* San Francisco: International Computer Music Association.

Kantz, Holger and Schreiber, Thomàs (1997). *Nonlinear time series analysis.* Cambridge: Cambridge University Press.

Karmiloff-Smith, Annette (1992). *Beyond Modularity.* Cambridge, Mass.: MIT Press.
Meyer, Leonard B. (1956). *Emotion and Meaning in Music.* Chicago: University of Chicago Press.
Narmour, Eugene (1990). *The analysis and cognition of basic melodic structures: the implication-realization model.* Chicago: University of Chicago Press.
Packard, Norman (1990). A genetic learning algorithm for the analysis of complex data. *Complex Systems* 4, no.5: 543-572.
Schellenberg, E. Glenn (1994). Expectancy in melody: tests of the implication-realization model. *Cognition* 58: 75-125.
Van Gelder, Timothy and Port, Robert F. (1995). It's About Time: An Overview of the Dynamical Approach to Cognition. In R.Port and T. van Gelder (eds.), *Mind as Motion.* Cambridge Mass.: Bradford Books/MIT Press.