Au(or)a: Exploring Attributes of a Live Algorithm

Dr Michael Young
Goldsmiths, University of London
m.young@gold.ac.uk

Abstract

A live algorithm describes an ideal autonomous performance system able to engage in performance with abilities analogous, if not identical, to a human musician. This paper proposes five attributes of a live algorithm: adaptability, empowerment, intimacy, opacity and unimagined music. These attributes are explored in *aur(or)a*, a performer-machine system for Max/MSP that fosters listening and learning. Live improvisation is encoded statistically to train a feed-forward neural network, mapped to stochastic processes for musical output. Through adaptation, mappings are learnt and covertly assigned, to be revisited by both player and machine as a performance develops.

1.1 Introduction

Advances in our understanding of machine intelligence, in areas such as music informatics, evolutionary computation and neural networks, point the way for a new generation of computer music application. New systems might collaborate in many musical contexts, not merely run pre-loaded scores or depend on stimuli, but engage with performers at a commensurate level. This is the vision of the Live Algorithms for Music network, founded in 2004.[1] A *live algorithm* is the function of an ideal autonomous system able to engage in performance with abilities analogous, if not identical, to a human musician (Blackwell and Young 2005). Such a system differs radically from the established paradigms of 'live electronics'; the *computer-as-instrument*, (a tool that relies on human agency), or the *computer-as-proxy*, (a substitute for the 'composer' that implements pre-designed functions laid out in a musical score or rule set). A true live algorithm would offer a high degree of autonomy; capacities to invent, provoke and respond.

Live algorithms are most germane when there are opportunities for such behaviours, when there is creative group interaction. In this scenario there is no 'top-down' control, no hierarchical human-to-human management analogous to user-to-computer control. In truly 'free' improvised music, structure and character – in so far they are evident – are emergent properties, products of group interaction, not a set of pre-defined rules. Musical languages are formulated pragmatically, self-referentially and on-the-fly. Free improvised music offers a model for aspiring live algorithms and a challenging context in which one could be tested to the full.

The *au(or)a* system described below explores a liminal space between 'live electronics', composing and live algorithms. To make music, the live player must improvise freely while the system acquires its own understanding of the performance. It then responds creatively and

---

[1] A collaboration between the Departments of Music and Computing at Goldsmiths, with support of the Engineering and Physical Sciences Research Council.

quasi-autonomously. Although human agency is necessary for the machine to function, the influence is indirect, just as one member of a duo needs a fellow musican.

The following section discusses attributes of live algorithms that are far-reaching and intended only as a framework for discussion and hopefully future investigation. They are reflected in the *au(or)a* system, as well as sister versions that employ the same basic design (Young 2007). There are five attributes; 1. adaptability, 2. empowerment, 3. intimacy, 4. opacity, and 5. (summatively) the unimagined.

1.2 Adaptability is the ability to acclimatise to a shared environment, demonstrable in changes of behaviour. A musical environment capable of change – and therefore to demand adaptation – is unlikely to be pre-determined by fixed rules, stylistic assumptions or other formal constraints. Adaptation is not necessarily conscious or intentional, even though performers may wish to communicate with a machine. Lewis's term *emotional transduction*, defined as a "bi-directional transfer of intentionality through sound" (Lewis 2000: 36), establishes by implication that adaptations should occur in and of the medium itself, not via controllers, irrelevant gestural information, or control data, etc.. However, Lewis's assertion that a performer's original intention – "emotional and mental" – can be preserved and then co-represented in the machine's response is open to question.

Performers adapt to sound, not to one another. *Stigmergy* avoids the problem of intention and emphasises adaptation. This is the process in which an insect population *self-organises* through the adaptation of individuals to their environment, and initially described termites interacting with their environment (Grassé 1959). Individuals do not commune directly, even though the resultant phenomena – nests – can be extraordinarily complex and seemingly designed. Computer simulations of this and other self-organising behaviours are well established (Bonabeau et al 1999). They are evidenced in the application of evolutionary computing to music (Miranda and Biles 2007). Stigmergy as a model avoids the problem of intentionality and machine cognition; consequently, it avoids the potential pitfalls of anthropomorphism. It proposes a flexible, dynamic and adaptable system capable of novel problem-solving: An effective model of human creativity and more specifically, of improvised computer music (Blackwell and Young 2004).

Musical performance, whether between humans or machines, might be regarded as a self-organising system if individuals commune with the shared audio environment rather than directly with one another. This assumption ignores visual/physical cueing and emphasises listening. Musical collaboration in a human/social context involves a continuous process of adaptation based on mutual listening. Goals are identified by group members – who actively assume and cast roles – in order to adapt to the changing audio environment. Goals might be attributable to 'supra-personal' social facts, (such as norms of acceptable behaviour, actions consistent with expectations or requirements). Even so, an entirely new, shared history evolves as the cooperative experience develops. Players become aware of the appropriateness of their response to others' contributions and appraise their own ability to initiate behaviour from others. These processes have been observed in jazz groups (Bastien and Hostager 1992). Arguably, all such social behaviours occur *by proxy* in the shared environment: they are adaptive and essentially indirect.

Whether or not it might ever be possible for computers to have an intentional response is an open question. However, optimisation methods, available in evolutionary computation and machine learning, can present the affect of intention. If interaction occurs only between performer and the environment, and between machine and its environment, the *products* of adaptation might be regarded as equivalent, and equally significant. It is then, arguably, inconsequential whether the interactions depend on machine algorithms or human cognition.

1.3 Empowerment entails control over decisions that impact upon future experience. Decisions have a context; the properties and consequences of options, the strategies that might inform choices and the criteria for their evaluation. In creative practice, decisions do not have easily definable strategies or evaluative criteria, but a framework must be at least implicit. Traditional AI systems that are effective in delivering pre-defined outcomes are not useful. For example, BDI (*belief, desire, intention*) systems must rely on a rule-base, respond only to knowable environmental measures and have clear pre-determined aims, even though they actively respond to input and output while running (Rao and Georgeff 1991).

Algorithms do not cognate! – so they cannot make creative decisions – but they can produce non-arbitrary changes in state. Such changes can be instigated by chaotic, non-linear or complex systems, particularly cellular automata, particle swarms, genetic algorithms and neural networks. The self-organising properties of these algorithms offer potential for novel problem-solving, invention and surprise. They do not necessarily achieve intended goals, but might find new ones. In musical performance, a non-arbitrary change of state is manifest as a 'decision' when it modifies the audio environment. Consequently, it is empowered to demand a response from both human and machine participants alike.

There is mapping problem. What structural and temporal features of music should be determined by a change of state that grant it the status of 'a decision'? It is easy to find examples of generative music where state changes are applied to the very surface of music. For instance in evolutionary computer music, genotype (genetic code) and phenotype (realisation) have been mapped schematically and literally. Various 'biological' models have been hijacked. Should phrases be identified with genes, harmonies with brain waves and notes with swarm particles? This is just the sonification – more accurately, "audification" – of a data space exploration; an experience useful for scientists (Hermann and Ritter 2004) but the artist's narcotic.

Creative decisions and explorations should reference structural properties, language and implicit methodologies of music-making, and offer new possibilities at these levels. Computer 'decision-making' cannot define context, but should engage with it. A common problem is time; algorithms function independently of time, so for music, a real-time clock must be imposed as a function of sonification. It is unavoidable that contexts such as this – however fundamental and transparent – are established by the designer, in order to provide relations for creative behaviour. Eco's term, the *field of relations*, emphasises the finite nature of an open work's discontinuities and its *field of possibilities*. These relations provide a framework for decisions. So, even though a single point of view is absent and there is some devolution of creative responsibility, this does not entail an "amorphous invitation to indiscriminate participation" (Eco 1989: 19). Neither, by extrapolation, does the absence of a point of view (algorithms do not cognate) necessitate a wholesale and literal transfer of state changes to the surface of the music, or to the framework and context for creative acts. Relations are underpinned by the capabilities of the machine system, the technical approaches and aesthetic attitudes of the designer and live player. It is through an interplay of all these relations that empowerment might be perceived.

1.4 Intimacy is experienced – or apparent – if there is a binding understanding shared by performers through informed listening and observation. This is a social process, but can be experienced in and through sound itself. A machine emulation of closeness and intimacy should attend to sonic experience, both in nuance and wider characteristics.

Technological devices that produce control data from a user's actions can only be receptive, not intimate. In music technology, the discourse around intimacy is really about responsiveness, i.e. emulation of a performer's physical interaction with his/her instrument (Wessel and Wright 2002). A truly intimate relationship – as occurs between musicians – is learned, rather than provided, and is an experiential phenomenon within the sound environment. (At least during a performance, before or after is another matter). It is, though, genuinely *interactive*.

Intimacy suggests the psychological process of *optimal flow*; a goal-orientated, mental state that explores the limits of experience and expectation, obtaining pleasure in meeting challenges with appropriate skills (Csikszentmihalyi 1991). It has been conjectured that the effectiveness of group collaborations can be evaluated with this measure (Sawyer 2003). A machine's contribution cannot be evaluated, of course, but a human performer, in his/her musical experience and interaction with the shared sonic environment, might *infer* that flow is occurring for all participants. This is particularly relevant when, for example by using neural networks, a machine can evidence prior learning and experience.

1.5 Opacity is a prerequisite for this flow, an avoidance of naïve processes of cause and effect (and their frequent boredoms for players and audiences alike). Interactivity is a well-discussed term in computer music but its currency is a little devalued. It is often equated with a one-directional transfer of information from user to machine; reaction rather than interaction. A lack of opacity and uncertainty distances the performer from the machine. The relationship is then a familiar 'subject-object', which denies the possibility of intimacy: "…interactivity has gradually become a metonym for information retrieval rather than dialogue, posing the danger of commodifying and . . . reifying the encounter with technology" (Lewis 2000: 36). Lewis offers Voyager's capacity for "variation and difference" as an alternative that avoids transparent and consistent input-output mapping, but still provides against chaos: "…lack of uniformity is not necessarily correlated with 'lack of structure,' as is so often expressed in the vernacular discourse of randomness. Rather, while tendencies over a long period of time exhibit consistency, moment-to-moment choices can shift unpredictably" (Lewis 2000: 36).

A truly interactive system ought to offer an ambiguous and shifting balance between the reactive and proactive, and across the threshold of the apparently chaotic and the readily comprehensible.

1.6 Unimagined. The result of these attributes might be a 'living' computer music, an unimagined music, its unresolved and unknown characteristics offering a genuine *raison d'être* for machine-human collaboration. If computers might extend, not hideously parody, human behaviour, machine music should not emulate established styles or practices, or be measured according to any associated, alleged aesthetic. Why recreate what exists? Machines cannot benefit – as undergraduate students might – in recreating pseudo-Bach or Mozart. In living computer music the contributions of all performers involved – human and machine –have equal significance, but may not necessarily be equivalent. Such music cannot be imagined or reproduced.

Unimagined music, free from *a priori* rules or overt control, moves "… toward a permanent discovery – comparable to a 'permanent revolution'." (Boulez 1960: 32). He refers to compositional method, the exigencies of musical form given a "fluidity of vocabulary", and the consequent need to de-linearise temporal structure. However, a 'living' computer music might be even more apposite, permanently exploring all elements of its emergent language, and in real-time, not just in concept. Successful free explorations would avoid the accusation of "inadvertence", the composer's narcotic (Boulez 1957: 42).

Freedoms, whether open to the player or to computer (e.g. by stochastic methods) might be better described as 'informalities'. Unimagined music is a technological *musique informelle*, emergent and idiosyncratic; its coherence neither derived nor dictated. It "discards all forms which are external or abstract or which confront it in an inflexible way, free of anything irreducibly alien to itself or superimposed on it" (Adorno 1961: 307).

Adorno's term 'informal' is not synonymous with the intuitive, and informal music should "constitute itself in an objectively compelling way". This distinction is apposite to free improvisation and the claims of its practitioners, and has been noted by Ferneyhough in relation to compositional method: "By this term [informal] I do not intend material extruded by some sleight of hand from the inaccessible depths of the 'spirit', but rather musical elements which, however rigorously they may later be employed, enjoy a primal state already imbued with a certain internal differentiation or relational complexity…such elements are in a position to enter into a dialogue with the composing consciousness…" (Ferneyhough 1998). In living computer music, unpredicted acts of a performer, and implicit/virtual acts of the machine should exemplify this "relational complexity", rather than rules or randomness. There is a critical engagement of free association, which might be equated with the "composing consciousness"; associations between intended system behaviours, an appraisal of potential behaviours and response to actual sonic realisations and their unfolding history. The performer-participant can be seen to adopt the role of "subject", in relation to the emerging musical language and the system's behaviour (the "organism"): "the subject must become an integral part of the organism." (Adorno 1961: 306). This relation is intrinsic to a musical language formulated in real-time (Impett 2000). This relationship is far more creatively coherent and profound than the subject-reified object relation criticised by Lewis.

2.1 *au(or)a*

These aspirational qualities of live algorithms are addressed in the performance system designed for *au(or)a*. The title refers to the *aurora borealis*, alluding to harmonic 'illuminations'. It also refers to Walter Benjamin's term *aura*, a "unique phenomenon of a distance however close it may be" (Benjamin 1936). This *aura*, sense of uniqueness and permanence, is lost or extracted from a work by technological reproduction. In *au(or)a*, perceptual distance is experienced between an imagined pianist and clear visual evidence that a computer is responsible. (A recorded performance is another story). There is a presumed, irreconcilable, distance between the human performer and machine itself; the presence and absence of cognition. Most of all, although the live music is infinitely, perpetually, reproducible – at least at a higher level of system behaviour – it is also entirely unique on each occurrence, and cannot be reproduced, or imagined.

What follows is a technical explanation of the *au(or)a* system, which shares its basic design and function with other works under a generalised title, NN Music, presented at ICMC 2007

(Young 2007). Overall, there is a stigmergetic process, in which a neural network learns about its environment, responds and consequently reshapes it; the performer is invited to behave in the same way. There are opportunities for intimacy in the appraisal and reaction to the musical contributions. Responses can be opaque, and at best offer an intimate sense of collaborative music-making.

2.2 Analysing and transforming.

Figure 1 shows the pitch analysis function. Audio to pitch conversion produces a stream of data, accurate to the nearest quarter-tone, filtered by the *attentiveness*; the probability that a pitch will be allowed to update $\mathbf{P}_{\mathbf{chord}}$, a list of most recently admitted pitches {x0, x1,.., x6}. The filter is deployed dynamically, mapped from the mean onset density detected over an adjustable time Δt, so relative inactivity on the performer's part fosters more attentive machine listening.

When $\mathbf{P}_{\mathbf{chord}}$ is updated with a new pitch, a generative function recalculates the eleven hexachords by cross-multiplying each pitch within the primary set of six. This method emulates the post-serial technique of chord multiplication, devised by Boulez (as, for example, identified in the 'L'artisanat Furieux' cycle of "Le Marteau sans Maitre" (Koblyakov 1990)). The obvious difference is that this function continuously updates $\mathbf{P}_{\mathbf{chord\_set}}$ in real time, as new pitches are admitted. $\mathbf{P}_{\mathbf{chord\_set}}$ is a dynamic pitch corpus, deployed as a resource for the synthesis function Q (explained below). So, the system constantly adapts to the pitch content of the player, providing opacity through note selection and harmonic transformation.

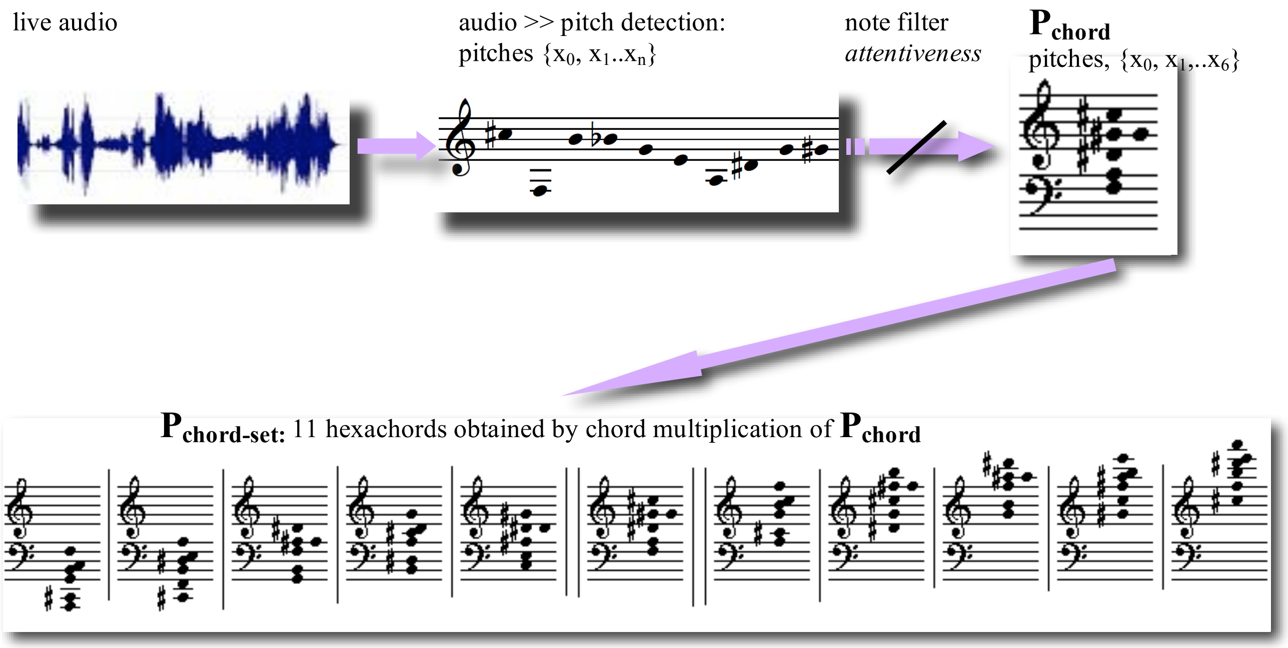

Figure 1: Pitch analysis and generation

The second analysis function, shown in figure 2 below, applies audio descriptors to the live performance (loudness, brightness, duration between events, sustained-ness, frequency etc.) with an analysis window of 50ms. This produces a dynamic performance state $\mathbf{P}_{\mathbf{audio}}$; a statistical representation, measured over time, Δt, comprising the normalised mean and normalised standard deviation of the descriptors, where $5s < \Delta t > 30s$.

The purpose of network **A** is to classify novel performance states, in order to acquire a library of learned states. This can be applied – while the music continues and the network runs – to partially classify new states, if they are not sufficiently novel to be allowed to retrain the network afresh. The aim of the process is to identify musical behaviours that are well defined and contrasting, so the network can be trained to respond effectively to a broad range of subsequent activity. To achieve this the dynamic state $\mathbf{P}_{\mathbf{audio}}$ is considered for retraining only if it satisfies the fitness function f_{fit}, a measure of the similarity of the current $\mathbf{P}_{\mathbf{audio}}$ to all those previously learned. If any similarity value of is less than a predetermined threshold, the current state is allowed to update the network, which is retrained on-the-fly. In the current implementation the threshold is set by the user; to be effective it must adjust to characteristic behaviours of both instrument and performer. The number of output nodes increases every time a new state is classified, $\{O_0, O_1, .., O_n\}$ representing an addition to the network's 'knowledge'.

When the music begins, the network trains several new states, usually within the first few seconds. The time interval between retraining then tends to increase, depending upon the character the improvisation and the consequent variance of $\mathbf{P}_{\mathbf{audio}}$ over time. As the performance develops, new analysis states will approximate one or, more often, several of those previously obtained. The network is continually queried to evaluate how far the current state $\mathbf{P}_{\mathbf{audio}}$ approximates any of those previously learned. Retraining might be thought of as adapting, sensitive to the conditions of the sonic environment.

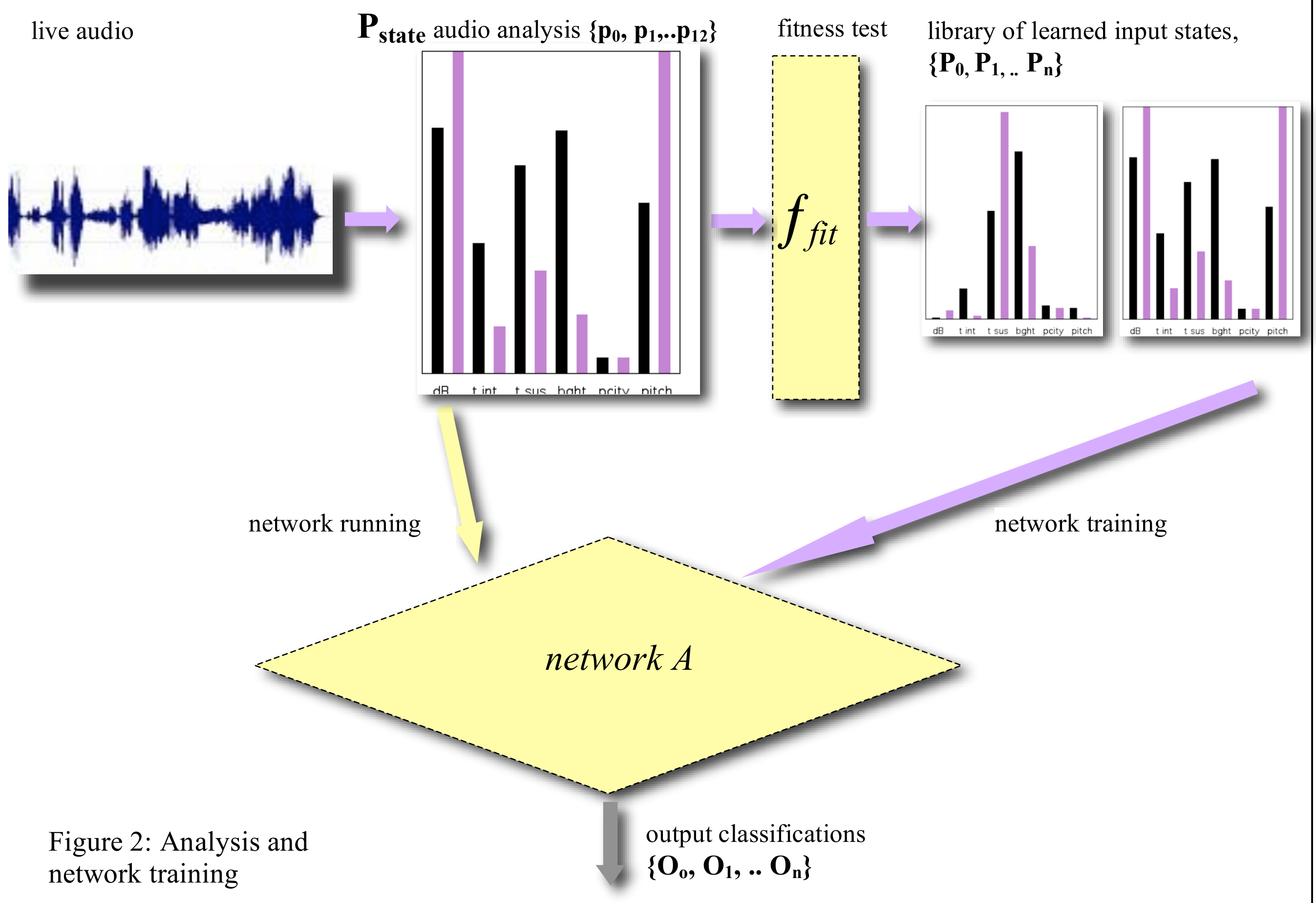

Figure 2: Analysis and network training

2.3 Making sound.

The expanding list of outputs (i.e. classifications) from network **A** $\{O_0, O_1, .., O_n\}$ are not mapped directly to network **B**. A function, *f*map, relays the list to the second network by randomly re-sorting the indices of the data. This jumbling up of output and input nodes provides genuine opacity; it is covert, challenging the player to adapt as the system's behaviour diversifies. Network **B** creates new input nodes as the list $\{M_0, M_1, .., M_n\}$ increases, which in turn allows the network to access more data from its previously learned set of outputs; this library of potential outputs constitutes the 'knowledge-base' of the system. It is decisive in characterising the music; a framework, a *field of relations* for aesthetic judgement.

The behaviour of network **B** is entirely dependent on the classifications made by network **A** as it runs. If a player suggests two previously learned behaviours, this will be reflected in a fusion of two output states. The output of network B is $\mathbf{Q_{state}}$**;** a list parsed into subsets according to parameter type. In figure 3, values for $\mathbf{Q_{state}}$ are shown at a given moment, each subset shown as a separate table. (Normally, due to the varying outputs of network **B** as it runs, $\mathbf{Q_{state}}$ is constantly changing).. $\mathbf{Q_{state}}$ is accessed as a probability distribution; each time a note is to be selected, all subsets are consulted to determine the various aspects and modifications of that note. The values indicated by the y axis in each subset denote the relative likelihood of a particular x axis value to be selected. For example, there are 6 potential note simultaneities (i.e. densities of 1-6 notes); in this scenario it is most likely that densities of 1 (and to a lesser extent, 2) will predominate, and there is a small chance of densities of 6 notes.

The rhythmic behaviour is more complex, and is determined by the first three parameter sets: 1. A geometrically expanding series of 11 values: 53ms to 53 raised to the power i 11 times. (In this example, smaller duration values are preferred, on a rough sliding scale down to the longest possible duration). 2. The likelihood of selecting ay one of three values for i (stretch factor). 3. The likelihood (about 50% here) that the rhythmic pattern will stabilise into a rhythmic pattern. This functions by recording the last 11 selected values; for every new iteration there is a probability that these values – or a selected number of them – will be recalled rather than fresh values generated. These three processes aim to provide a sophistication of rhythmic vocabulary and structural syntax akin to those available to an improviser.

Pitches are determined by two subsets: 1. the hexachords available (in this e.g. mid- to high hexachords are possible) and 2. the notes within each hexachord, determined by their vertical position in the chord (in this e.g. all 6 are equally likely). The outcome of the hexachord/note position is then referred to the current $\mathbf{P_{chord\_set}}$ from which the actual MIDI pitch is obtained.

This process can be generalised easily. Versions of *aur(or)a* have deployed audio manipulation of the disklavier sound (including granular synthesis and spectral shifting); these processes can be similarly parmeterised and integrated fully into the **Q** function. The two sample-based/DSP versions of this system, *piano_prosthesis* and *cello_prosthesis*, deploy a much larger set of synthesis functions (Young 2007)[2].

---

[2] For audio examples www.myoungmusic.com

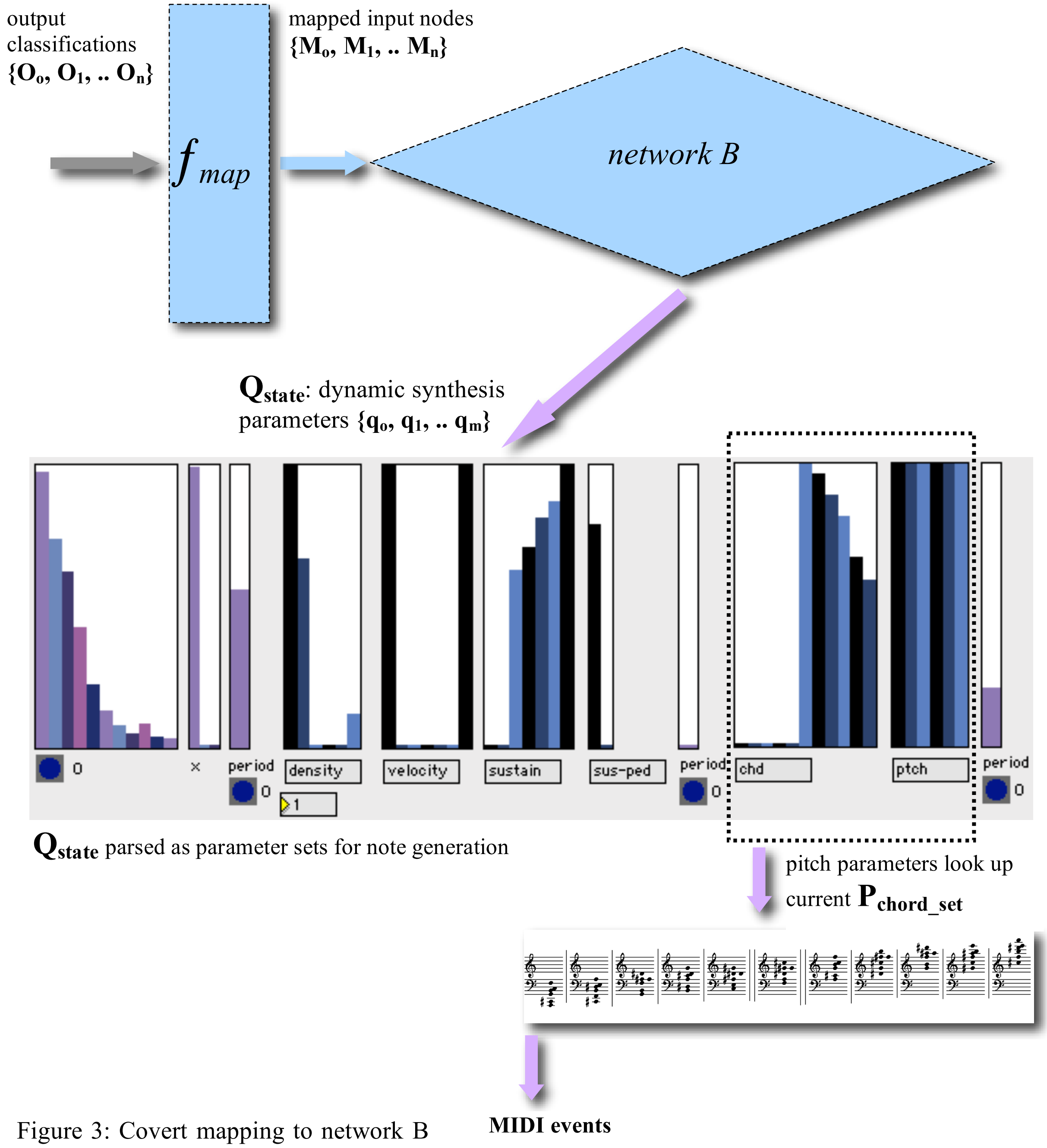

Figure 3: Covert mapping to network B and synthesis

3. Conclusion

The processes of *au(or)a* – pitch-based listening and promulgation, audio analysis and mapping, stochastic sound-event generation – can produce a wide variety of behaviours that challenge and respond to a human musician's contribution. The result is intended to be coherent but, in context, unimaginable. It is hoped that all five properties proposed for a live algorithms are evidenced. This might offer further understanding of the musical possibilities of human-machine musical interactions, and the properties of 'living' computer music. Future developments could explore the accuracy and relevance of audio description, at both the micro-level of timbral description, but also at a higher level of sound event and phraseology. The far-reaching areas of machine learning and machine consciousness are highly apposite. These are fertile and complex areas for research.

Interactive technology demands its own critique in music. To parody: The overwhelming multitude of interactive systems can't be counterbalanced any longer by comfortable returns to fixed media. Does it seem adequate to force computer music into the Procrustean bed of tapes and DVDs without acknowledging the doubts and contingencies that pertain to it?

4. References

Adorno, T. 1961. Vers une Musique Informelle. In T. Adorno (ed.) (1963/1994) *Quasi une Fantasia*, trans. Rodney Livingstone. London: Verso.

Bastien, B. T. and Hostager, T. 1992. Cooperative as communicative accomplishment: a symbolic interaction analysis of an improvised jazz concert. *Communication Studies* **43**: 92-104.

Benjamin, W. 1936. The Work of Art in the Age of Mechanical Reproduction. In Arendt H (Ed.) (1999). *Illuminations*. Pimlico, London.

Blackwell, T. and Young, M. 2005. Live Algorithms. *Artificial Intelligence and Simulation of Behaviour Quarterly*. **122**: 7-9.

Blackwell, T. and Young, M. 2004. Self-Organised Music. *Organised Sound*. **9**:2: 123-136. Cambridge University Press.

Bonabeau, E., Dorigo M. and Theraulaz T. 1999. *From Natural to Artificial Swarm Intelligence*. Oxford University Press, New York

Boulez, P. 1957. Alea. Reprinted in P. Boulez (1991*) Stocktakings from an Apprenticeship*, trans. S. Walsh. Oxford: Clarendon Press.

Boulez, P. 1960, Sonate, que me veux-tu*?* Reprinted in P. Boulez (1986) *Orientations: Collected Writings* (1986) London: Faber and Faber.

Eco. U. 1989. *The Open Work.* Trans. Anna Cancogni. Harvard University Press.

Csikszentmihalyi, M. 1991. Flow: The Psychology of Optimal Experience. Harper Collins.

Ferneyhough, B. 1998. Talk at Goldsmiths College. Published as La 'musique informelle'. Brian Ferneyhough, (1999). Trans. into French Peter Szendy L'Harmattan-Ircam, www.entretemps.asso.fr/Adorno/Informel/Ferneyhough.htm#_ftnref1. Accessed 01/08/2007.

Impett. J. 2000. Situating the invention in interactive music. *Organised Sound*. **5**:1 27-34.

Grassé, P. 1959. La reconstruction du nid et les coordinations inter-individuelles chez Bellicosi-termes natalensis et Cubitermes sp. La theorie de la stigmergie: essai d'interpretation des termites constructeurs. *Insect Societies* **6**: 41-83

Hermann T. and Ritter H. 2004. Sound and Meaning in Auditory Data Display. *Proc. of the IEEE (Special Issue on Engineering and Music - Supervisory Control and Auditory Communication).* **92**:4 730-741.

Koblyakov, L. 1990. Pierre Boulez: *A World of Harmony*. Harwood Academic.

Lewis, G. E. "Too Many Notes: Computers, Complexity and Culture in Voyager". Leonardo Music Journal. Vol. 10, pp.33–39, MIT Press, 2000.

Miranda E.R and Biles J.A. 2007. *Evolutionary Computer Music*. Springer-Verlag: London

Lewis, G. E. 2000. Too Many Notes: Computers, Complexity and Culture in Voyager. *Leonardo Music Journal*. **10**:33–39, MIT Press,.

Rao, A. S. and Georgeff, M. P. 1991. Modeling rational agents within a BDI-architecture. In Fikes, R. and Sandewall, (Eds.), *Proc. of Knowledge Representation and Reasoning*. 473-484. Morgan Kaufmann, CA.

Sawyer, R. K. 2003. *Group creativity: Music, Theater, Collaboration*. Lawrence Erlbaum Associates.

Wessel, D. and Wright, M. 2002. Problems and Prospects for Intimate Musical Control of Computers. *Computer Music Journal*.**26**:3.

Young, M. 2007. NN Music: Improvising with a 'Living' Computer. *Proc.of the International Computer Music Conference.* Copenhagen.