
Software Agents in Music and Sound Art Research/Creative Work: current state and a possible direction

IAN WHALLEY

The University of Waikato, Private Bag 3105, Hamilton, New Zealand
E-mail: musik@waikato.ac.nz

Composers, musicians and computer scientists have begun to use software-based agents to create music and sound art in both linear and non-linear (non-predetermined form and/or content) idioms, with some robust approaches now drawing on various disciplines. This paper surveys recent work: agent technology is first introduced, a theoretical framework for its use in creating music/sound art works put forward, and an overview of common approaches then given. Identifying areas of neglect in recent research, a possible direction for further work is then briefly explored. Finally, a vision for a new hybrid model that integrates non-linear, generative, conversational and affective perspectives on interactivity is proposed.

1. INTRODUCTION

The grounding assumption in this paper is that it is worthwhile to attempt to draw on and integrate technical and human reception perspectives to advance new music research and/or creation, towards implementing robust solutions that can usefully inform both fields of interest. Towards addressing this, a part of the new technology experimentation process is the opportunity to reflect on the balance between technical input and communicative outputs, and to raise issues that might be useful towards developing significant lines of further enquiry.

The emerging use of software agent technology as a tool in music/sound art creation, part of a wider interest in the deployment of interactive music systems such as those applying complementary techniques from evolutionary arts (Brown 2002) and A-Life (Miranda 2003), presents the chance to take a reflective step aside.

This paper is firstly a survey of recent software-based agents in music/sound art in both linear and non-linear idioms. After the technology is introduced, a theoretical framework is put forward towards an understanding of the work, followed by a conceptual overview that characterises and illustrates common approaches. Secondly, areas of neglect in recent research are then identified, particularly in affective (emotion-based) approaches and narrative mapping, and a possible future direction based on affective music briefly explored. Finally, to conclude and

answer the critique, a hybrid model is proposed based on non-linear generative improvisation coupled with a conversational model of human–computer interaction that takes into account music/sound as an affective language.

The intention here is then to broaden debate and stimulate further discussion – and also to signal areas of contention that might be useful to address in future work.

1.1. What are software agents?

An agent is generally understood as somebody or something that acts on behalf of another in a process. A software agent, put simply, is a computer program that works on tasks specified by a user. As with their physical counterparts, software agents can exhibit varying degrees of persistence, independence, communication and collaboration with other software agents or people. In addition, ‘intelligent’ agents might have the capability to make decisions, the capacity to learn within an environment, and the ability to be mobile over computer networks. Increasingly more ‘intelligent’ agents monitor environments, can glean information, make decisions to react or not, and even modify their behaviour according to the results received (see Bigus and Bigus 2001; Consoli, Ichalkaranje, Jarvis, Phillips-Wren, Tweedale and Sioutis 2006).

As the use of software agents has become more widespread, a variety of independent development environments (IDE) have been created for building agent-based systems, including IBM’s *Aglets*, or *Grasshopper*, *Jade* and *Zeus* (see Balogh, Budinska, Dang, Hluchy, Laclavik and Nguyen 2002; Detlor and Serenko 2002). The technology is now used in academia across several disciplines, particularly in simulating multicausal situations, examples being found in the annual conference proceedings of Autonomous Agents and Multi-Agent Systems (AAMAS). It is also increasingly used in real-time industrial and military applications in areas such as scheduling and simulation, fields as diverse as architecture, engineering and construction (Anumba, Ren and Ugwu 2005), and in business (see Aronson, Liang and

Turban 2004). Useful conceptual introductory texts include Bigus and Bigus (2001) and Weiss (2000).

Beyond using a single software agent to do something on your behalf, there are three broad technical software-based deployments of the technology. A Multi-Agent Systems (MAS) is one where a computational system allows agents to cooperate or compete with others to achieve some individual or collective task(s). A Distributed Artificial Intelligence (DAI) approach is a cooperative system where agents act together to solve a problem, and can include agents being used in decision support systems where the outcome is based on interactions between human and agent input. Finally, there are Multi-Agent Based Simulations (MABS) of complex situations, although there is some debate in the field as to the degree to which MABS actually use agent technology (Drogoul, Meurisse and Vanbergue 2002).

2. SCOPE AND THEORETICAL APPROACH

To understand the scope of the field, offer a theoretical framework in which decisions are made, and propose a theoretical test of potential deployment of agent technology in music/sound art works, there are useful perspectives in prior literature that can be drawn on and amalgamated on which to base subsequent understanding.

2.1. Interactivity

Interactivity is integral to agent technology deployment, either in self-contained (closed) or HCI (open)

systems. The scope of music interactivity then gives a sense of the range of the field. Graugaard's (2006: 125) interactivity schema, although not without limitations, integrates various perspectives (see figure 1).

The focus in this paper is largely on the dialectic between computer science, 'musical', narrative, and emotional parameters relating to agent technology given in figure 1.

2.2. Functional decision space

In making new works, creators do not begin in a vacuum. Gimenes, Miranda and Johnson (2006), for example, note that each artwork one creates is embedded in a cultural inheritance, and pieces of music reflect both composer choices and the history of influences on a work. A computer-based composer/sound artist or programmer in making new works, either consciously or unconsciously, is likely to draw on their view of music/sound art, the purposes for which the output is intended, and the knowledge, techniques and conventions developed through artistic and technical training. A theoretical approach to the multidimensional decision space that integrates software-based agent technology is proposed in figure 2, accepting that this neglects engineering aspects of Graugaard's model (figure 1) that are beyond the focus of this paper.

Elaborating on figure 2, the GOALS a creator has influence the type of system built technically and artistically. Weinberg (2005: 31–2) characterises the two possible approaches as being musical structure/composer controlled (based in the late European

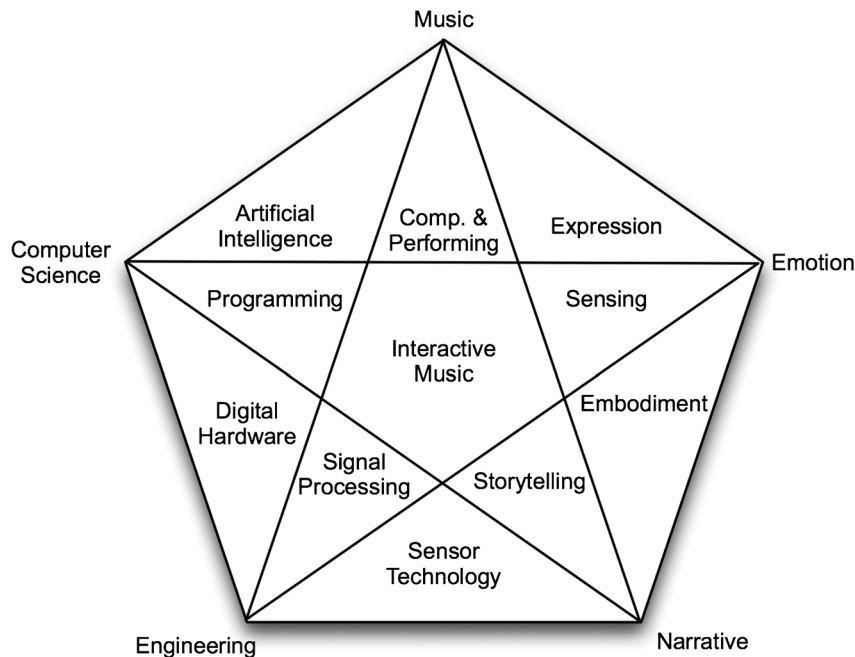


Figure 1. Disciplinary context of interactive music.

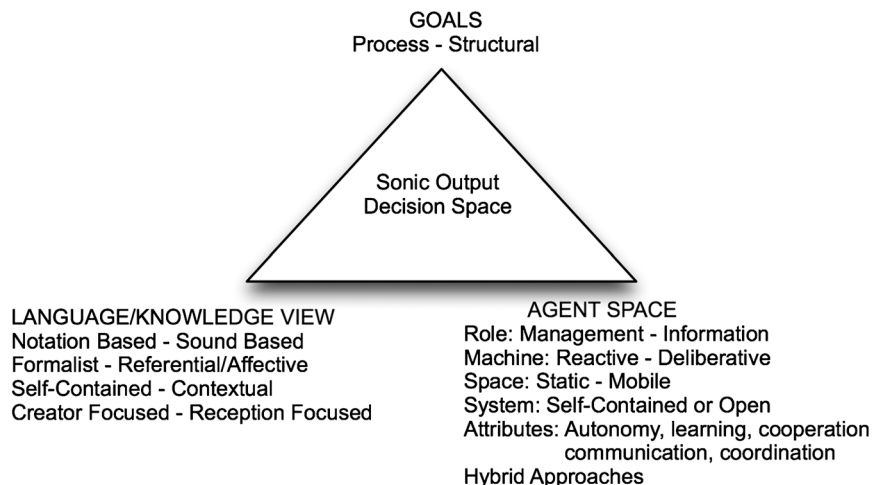


Figure 2. Decision space.

avant-garde position), and a processes-based approach (explored by American composers such as John Cage and Steve Reich). Technically, a structure-centred approach allows participants (agents and/or humans) to fulfil prescribed musical or performance outcomes. A process-centred approach concentrates on exploring possibilities or fulfilling goals through collaboration or competition, and the experience may differ with each session. Expanding the notion of making linear approaches to notation or acousmatic/sound art works beyond one possible 'frozen' form, Chadabe's (1996, 2004) concept of a 'process'-based rather than 'product'-based approach to creation is useful here, explaining how musical structures are a reflection of wider paradigms. Chadabe's 'non-linear' perspective allows for structures to be created by the dynamic interplay of the behaviour of participants, rather than being predetermined by scores or set forms. In addition, Dorin's (2001) conceptual work on generative arts illustrates how that in a 'process-based' approach, both form and content of new non-linear works can be the outcome of the creative process.

The LANGUAGE/KNOWLEDGE space has a continuum of choices (figure 2). A creator decides if a work is to be notation based, or will take Smalley's (1997) theoretical approach to *spectomorphology* in creating sound works, for example, or a combination of these. In addition, the creator may take the formalist view that language (sound or notation based) is self-contained with no affective/emotive attributes, or that it has an affective/emotive element(s) (Juslin and Sloboda 2001; Nussbaum 2007). With any creative approach, choices are made also if musical/sound language is something self-contained, or includes aspects of other artistic communicative elements, such as dance (see Bryant and Hagen 2003; Thaut 2005; Brown and Parsons 2008), or wider narrative storytelling. Finally, choices are made as to if a new work is

to be creator focussed, or based in reception studies and/or human physiology: see Landy (1999) or Weale (2006) on electroacoustic music reception, for example, and Patel (2008), Mithen (2005), and Brown, Merker and Wallin (2001) on evolutionary approaches to explaining human musical abilities and responsiveness.

A part of choices about language, and a reflection of those choices, is also sometimes a decision about knowledge. Many artificial intelligence (AI) approaches to music centre around four methods to applying artificial intelligence methodologies to notation (see Camurri's 1993 classification). In summary, a *symbolic approach* sees knowledge as being represented in an appropriate language, and machine manipulation of symbols gives new knowledge (see Cope 1991, for example). The *logic approach* assumes that using logical formalisms represents knowledge, and by manipulating this logic-based knowledge we can then infer the new information. In a *sub-symbolic approach*, connectionist researchers believe knowledge expressed by states and connections between simple processing units can be learned and replicated by machines. Finally, a *hybrid approach* begins from the assumption that knowledge is represented by an integration of a sub-symbolic system and a symbolic system.

In the AGENT VIEW space, technical decisions must be made as to the level of deployment of the technology, and the general approach to be taken: MAS, DAI, MABS. Subdivisions on the decision space in the agent view here (see figure 2, agent view modified from Consoli et al. 2006) should again be read as a continuum rather than binary oppositions, and a system might also contain a hybrid of approaches. The significant division here is the question as to if the system is self-contained (MAS) or can include external input (DAI), particularly human input. As previously noted, agents' adaptive learning ability and level of autonomy largely characterise the level of

agent ‘intelligence’. In addition, enhancing technical deployment, systems might be extended by drawing on other AI techniques such as machine learning.

2.3. Reaction and interaction

Theoretically, and as a basis to critique work, it is useful to have a test of the level of intelligence of agent-based systems. For this, and expanding Graugaard’s (2006) notion of interactivity (see figure 1) is Paine’s (2002) argument that in many machine/human interactive works (Rowe 1994; Winkler 1998) the human agent(s) *interacts*, but the machine largely *reacts*. To address this, he proposes a *conversational model* (Paine 2002: 297) of human and machine agency. Based on the analogy of a human conversation, he notes that the relationship should be unique and personal to participants, unique to the moment of interaction, vary with unfolding dialogue, and be maintained by both parties speaking the same language and addressing the same topic. Further, one party may know the beginning point of a conversation, and while there may be a pre-existing agenda, the terrain of the conversation might not be known in advance. An interactive process is then one of exchange and of sharing ideas, and the relationship between participants should deepen over time. A greater amount of cognition is then required on the machine’s part than is normally considered in many ‘interactive’ music systems (see Winkler 1998). Further, and independent of human–machine interaction, Paine’s model might also be drawn on as a test of interaction in machine-only systems in terms of the level of their ‘intelligence’.

While there are drawbacks in taking an extreme approach to using new musical/sonic languages in generating form and content using this model – as we have to understand something of a shared vocabulary, grammar or forms to have a conversation – the argument is useful as a counterpoint to avoiding excessively prescriptive approaches to music/sound art composition. The model may also be used to understand how common ‘musical’ languages might evolve from the dynamic input of different participants, and aid in understanding the range of current work in ‘interactivity’ when applying agent technology that addresses parts of Paine’s paradigm. Tempering Paine’s model, it is also worth noting that not all conversations are symmetrical in terms of knowledge, participation, and input quality.

3. FRAMEWORK FOR AGENT-BASED MUSIC/SOUND ART

Given the range and amount of academic and industrial work taking place that uses multi-agent systems, it is surprising how little attention the tools have attracted from composers or computer scientists interested in music/sound art. In contrast, less sur-

prising is the divergence of approaches that reflect the wide field of interest in the decision space (figures 1 and 2). Generally, technical or computer-science perspectives, rather than wider music/sound artistic disciplinary perspectives, influence many approaches taken.

Drawing on the framework presented in the previous section, four methods in using software agents in music/sound art seem apparent, grounded in similar technical and/or artistic perspectives (Whalley 2005). Put crudely, simulation and performance/reactive method are largely structure focused and often notation based, and generative and generative/improvisational methods draw largely on process-based paradigms and have broader view of language and knowledge. The examples given to illustrate each show something of the current scope of work, and the summary is necessarily partly descriptive. Aspects of computer science, music, emotion and narrative (see figure 1) are touched on here, comments on influence within the decision space made, and the ‘conversational’ test applied. A brief general critique of the scope of work is left to the next section.

3.1. The simulation approach

Technically, agent technology is generally well suited to *simulating* linear tonal music creation or realisation from motives within set forms or structures, because it deals with real-time data and multi-causal situations where the parts (‘players’) constantly adapt (react) to each other in the process of making a structural whole as a set goal. Given this, the emotional dimension of music and performance values are of less concern here. The method could also be implemented using other AI technologies, but the extent of real-time adaptation gives agent-technology some advantages.

Although not widely deployed, the linear principle here could also be partially used to make acousmatic music or sound art (Smalley 1997; Wishart 2002) where the structure and/or form are predetermined but the content be decided by the agents. A *simulation approach* may also potentially involve an agent-based closed system replicating known styles, without human interaction, regardless of the music AI technique used.

In terms of examples, the simulation approach to tonal music, a common area of MAS research to test out scenarios in software, is found in Nakayama, Wulfhorst and Vicari’s (2003) ‘A Multi-agent Approach for Musical Interactive System’. Here, a community of agents interact through musical events (MIDI), simulating the behaviours of a musical group. The outcome is an intelligent accompaniment system, where agents ‘listen’ and interact with each other. Agents are ascribed a basic level of knowledge

to play their instruments synchronously and satisfy their internal goals. The resulting Virtual Musician Multi-Agent System is analogous to instructing a beginning jazz band to accompany a singer based on an agreed structure and set of rules, into which individuals make musical choices as they interact with other players. While technically interesting, the work has limitations artistically: first by its narrow MIDI implementation (limiting the expressive potential of affective performance), and second because the structure is based on a logical or rule-based approach to music as language (GTTM and Camurri's 2003 classification), in contrast to a view that music is a medium for expression, or a language that can be expanded as part of the creative process.

A contrasting simulation approach is found in work by Gang, Goldman and Rosenschein (1999). Here, a system was developed for generating two-part counterpoint integrating sub-symbolic (machine learning, states/connections) and symbolic music processing (see Cope 1991 on recombincity). The work encodes musical knowledge, intuitions and aesthetic taste into different modules, captured by applying rules, fuzzy concepts and learning. Their work illustrates a hybrid AI approach made up of a connectionist module and an agent-based module. Again, the work is of technical interest, but many sound artists may see the aesthetic of tonal/notational stylistic variation as limiting.

A criticism that can be made here is that the level of agent deployment ('intelligence') is low in terms of agent autonomy and learning ability, making these self-contained systems largely reactive. Further, the division between the methods used to generate outcomes here and other straight rule-based systems is not clear cut at times, particularly when considering the continuum of possibilities that agent technology presents.

3.2. Generative approach

In contrast, a *generative approach* (see Dorin 2001 and Miranda 2003, who have also used the term) applied to either pitch/duration or sound art composition allows agent technology to create content and even structure based on the dynamic interplay of parts (process approach). This is a non-linear perspective, where the composer creates a range of agent behaviours, sometimes based on a logic-rule approach rather than predetermining forms, and there will be many possible outcomes and durations of works that result. Process then becomes the artistic focus, rather than a set product. These systems are generally technically self-contained once set in motion, and the language used is largely self-referential, neglecting the emotional perspective in terms of composition and performance value.

Deployment may also draw on other rule-based system such as those found in A-Life paradigms, used in combination with agent-based approaches. Some recent examples are cited to illustrate work here.

A largely *generative systems approach* is seen in 'Khorwa: a musical experience with autonomous agents' by Malt (2004), which is informed by A-Life/evolutionary systems. This installation also provides the opportunity to enter voice recordings. It is programmed in MAX/MSP rather than a common package for the development of agent systems. Subsequently, the software used for implementation limits agent applications to an extent, and Drogoul et al. (2002) might debate if there are any agents involved.

Although adding the 'Mobile Agent Systems' represented in by Kon and Ueda's work (2004) here, their platform remains largely aesthetically speculative. Mobile agents, in contrast to MAS systems that most practitioners implement in music/sound art, are active autonomous objects that can execute computation in a network, migrating from one node to another. Kon and Ueda's *Andante* is a prototype for the construction of a distributed application for music composition and performance based on mobile musical agents. Implemented in Java, two recent applications from this are *NoiseWeaver* and *Maestro*. Both are based on agents that generate and play stochastic music in real-time, either being controlled by a GUI or script. Kon and Ueda's infrastructure provides a shell for using DAI technology, but the current limitation is one of computational formalism (i.e. aiming to sonify the results of a successful technological system), and it lacks widespread artistic uptake.

Extending the generative approach is the work of Gimenes, Miranda and Johnson (2006) – also see Gimenes, Miranda and Johnson 2005 – where a society of agents is created to illustrate the development of musical styles. This model is used to study human beings' abilities and activities, toward understanding alternative routes in the historical evolution of musical styles from the perspective of music interaction and influences. To do this, agents in the system attempt to create their own worldview through aspects of music they learn, interaction with other agents, and generating new music. The system is a type of musical language lab where a user can define scenarios for evolution, and agents' abilities grow based on Dawkin's concept of memes. The system, rather than intended as a strictly creative generator, develops alternative stylistic approaches based on known musical language variations. In terms of intelligence, the strength of the work here is to implement some aspects of Paine's conversational model in a machine-based system that is beyond other machine reactive approaches, while also allowing aspects of language to evolve as part of the inter-agent dialogue.

Finally, Eigenfeldt (2008) developed an ensemble based on a multiple-agent system programmed in MAX/MSP that generates pitch/duration polyphonic rhythmic patterns. Agents here require little input or supervision once set in motion. Based on the Kinetic Engine, the system uses networked computers and individual software agents to emulate what drummers improvising in a percussion ensemble might do. The agents here can develop within and between one performance and another (learn), evaluating their own input in comparison to ‘personality parameters’. Although intended as a creative method in contrast to the previous example, it also illustrates how a *conversational approach* to composition might best be applied when provided with aspects of pitch/duration-based language in the first instance.

From a technical perspective here, the theoretical advantage of deploying agent-based systems over other algorithmic rule-based systems, such as cellular automata, is the ability of agents to evolve rules in real-time (adaptation) and learn from experience. The extent to which this is applied to rule-based languages, in part, allows the deployment of parts of the conversational model of interaction.

3.3. Performance/reactive approach

A *performance/reactive approach* is generally represented in the application of MAX/MSP to live performance (Rowe 1994; Winkler 1998), MAX/MSP’s deployment in many human/machine interactive installations, or other software-based systems used to create various types of net-based music with real-time human input (Weinberg 2005). The term ‘reactive’ is used here to satisfy Paine’s (2002) concept of ‘interaction’.

In terms of agent deployment, these systems tend to be focused on increasing control over aspects of structure (performance ‘command and control’), have low levels of agent autonomy and learning ability, and are generally based on traditional notation approaches to self-contained musical language. They extend *simulation approaches* by being open systems (DAI) that allow for human input. Two recent examples that treat language as something self-contained illustrate this method.

Spicer, Tan and Tan (2003) in ‘The Learning Agent Based Interactive Performance System’ apply the idea of an artificial performer to aid managing a complex system. Using a tonal MIDI-based pitch/duration paradigm, musical agents communicate and collaborate to produce compositions, often in conjunction with human input where the agents make up an ensemble of virtual performers. The method allows a human performer to ‘play the system like an instrument’. The performer can alter target values so that the program will change to arrive at the user’s desired state, and the system incorporates a delta-learning

rule that allows agents to respond and adapt to user input. The machine agent is based on an expert system model.

Spicer (2004) has also explored a further performance/reactive approach in ‘AALIVENET: An agent based distributed interactive composition environment’. This allows human performers in different locations to interact with each other via agents that they control. The system uses a client/server autonomous agent structure, client machines generating local variations of music based on updated information from human performers. The server acts as an information hub that passes on performers’ intentions to client machines. The client machines will converge on a state that reflects all performers’ intentions, but the music made at each site will differ in detail.

Both systems are limited in that the language used (pitch/duration tonal music generation) restricts the output to a small range of music/sound art possibilities. Further, the way that MIDI or synthesis is deployed in these systems neglects performance expressiveness.

3.4. Generative/improvisation approach

The most extensive deployment of agent technology, and an area that takes a broader view of music/sound art language although still largely self-referential, has taken place the area of a process-focused *generative/improvisation approach* (also see Brown 2002 on evolutionary systems approaches). This allows real-time improvised human input, but rather than the machine agent simply reacting, there is a human *and* machine adaptation (autonomy and learning) to the generative input of the machine and the input of human agency. Again in theoretical terms, if the adaptive relationship between human and machine demonstrates balanced listening/dialogue, and little of the final output is prescribed, a conversational model of interaction is then possible.

The scope of work here is best illustrated by citing examples. Two recent deployments of the *generative/improvisation approach* lean on a sound art perspective. Both use improvised human input and an embedded multi-agent system that adapts in an ongoing dialectic, as either net or installation based. Chen and Kiss (2003) used a multi-agent system as part of an installation in *Quorum Sensing*. This work metaphorically reconstitutes an ecosystem with synthesised sounds and images. It makes use of a multi-agent system that responds to the input of visitors to the site, the source images and sounds containing objects that can be perceived, created, destroyed and modified by the software agents.

Whalley (2004) also uses an embedded multi-agent system in the PIWeCS project, a Public Interactive Web-based Composition System. This is an

internet-based structure that increases the sense of dialogue between human and machine agency through integrating intelligent agent programming with MAX/MSP. Human input is through a web interface. The ‘conversation’ is initiated and continued by participants through arrangements and composition based on short performed samples, and the system allows the extension of a composition through the electroacoustic manipulation of source material by human or machine. A limitation of the method in terms of the conversational test is in using predetermined material, rather than letting the machine decide on some of the source content.

Both of these generative/improvisation systems enhances machine agency, attempting to add a sense of machine cognition to interactive works. However, the degree of autonomy of the agent systems and extent of interaction differs between them. Chen and Kiss (2003) allow greater autonomy over agent decision-making regarding content, but their system has no memory to allow an evolving dialogue. Whalley’s (2004) MAS is built on an expert system, and the memory function allows for the machine accumulation of user patterns, facilitating a conversation that is unique to each session, and user/machine dialogue to develop by mutual understanding over time, in line with the conversational model.

Three other systems expand the general approach here in terms of agent ‘intelligence’.

A partly speculative system is Edwards, Murray-Rust and Smaill’s (2006) MAMA (Musical Acts – Musical Agents) a generative architecture for interactive musical agents. The Musical Acts method is based on Speech Acts theory (simply defined as: in the act of saying something, we do something). The aim is to make a collection of agents that can improvise music (notation based) with each other or with human input, embedded in a language through which agents can represent and reason about music. Their system provides an agent-based ‘Musical Middleware’ for composers and musicians, with some constraint on input, and a set of general style libraries geared towards creating music. Using a logic-based approach to AI based on Music Acts, this allows for computational efficiency, rather than each agent having to listen to and interpret each other, and allows for the possibility of analysing musical acts and gestures from human input. At least, the system here acknowledges that there are different types of improvisational approaches to form and content, and that conversations start from clusters of known language and grammar. The theoretical limitation is that, in prescribing parts of the language, one prescribes something of the gambit of possible conversational outcomes.

A recent system introduced by Beyls (2007) is found in ‘Interaction and Self-Organisation in a Society of

Musical Agents’. Again partly speculative, particularly regarding human input, this is a distributed architecture where musical agents interact according to mutual affinities in a virtual world. Here, agents continuously exchange information in their neighbourhoods while self-organisation takes place. The ‘society’ may function on a continuum between autonomy or man–machine interaction, providing ‘an adaptive musical playground’. Within this, agents associate spontaneously, creating temporary emergent structures, the result of perpetual self-production in line with the theory of autopoiesis (a self-maintaining system that produces and replaces its own components and distinguishes itself from its environment). The great potential of the system, based on biological models and implemented through an array of AI techniques, is that it puts forward the possibility of human/machine interaction based on non-ideomatic improvisation. Despite being embedded in MIDI streams and pitch/duration methods, the system potentially presents a flexible means to implementing more of the conversational model; as fewer musical elements are prescribed (although behavioural boundaries can be specified), the artistic input becomes a more integral aspect of an evolving two-way dialogue, and an independent musical ‘personality’ can potentially emerge from agent input.

Finally, Miranda and Gimenes (2008) have introduced iMe (Interactive Music Environments) in ‘An A-Life Approach to Machine Learning of Musical Worldviews for Improvisation Systems’. This extends work described in the *generative approach* (above) by Gimenes et al. (2006) and develops techniques beyond genetic algorithmic methods (see Biles’ 2003 GA taxonomy) by using adaptive learning agents that interact with ongoing human input, generating music autonomously and in real-time. The system again takes a cultural view (memes) seen as being reflected in a musical style/worldview established by patterns and artefacts produced by human behaviour resulting from constrained choices. Like previous work, the research question raised is again how different musical influences lead to musical worldviews based on interaction between agents, but this time adding influence by external systems and human input. Agents’ perception and abilities can change over time based on their ability and experience (memory, structure of memory, and forgetting) of musical input: the system allowing one to see how different musical elements and the balance between them defines a musical style. The work here allows for creating mechanisms/agents with cognition through which new styles can emerge, rather than relying on fixed rules from known styles, extending Cope’s (1991) fixed pattern-based work, and extending Rowe’s (1994) work to foster a greater sense of interaction in line with the conversational model.

4. SCOPE, CHALLENGES, POINTERS

4.1. Scope and decision-making space

As noted, generally and in light of Graugaard's (2006: 125) overview (figure 1) and its augmentation for discussion here (figure 2), the balance of work in applying intelligent agent technology to music/sound art is largely grounded in computer-science technical perspectives, or in biologically influenced technical perspectives. Further, although AI techniques have been applied to linear and non-linear idioms' form and content, work has largely been embedded in tonal pitch/duration views of language/knowledge – the few explorations that move beyond conventional western musical language replication or stylistic variation allowing for more artistic flexibility in terms of sonic languages, even if the language is seen as being self-referential. Until recently, the technical focus has also favoured machine-based (closed) or controlled systems, rather than a real-time machine and human agency dialectic.

Given the current thrust of research work, technical approaches are likely to continue and to be extended, neglecting other possibilities that might be usefully explored as part of this process. Two areas on Graugaard's model that need further exploration are as follows.

Beyond the narrative structures associated with linear tonal music or other pitch/duration languages, an area that begs consideration or taxonomy is the narrative dimension of interactivity (see figure 1), to discuss new structures that might emerge from using new sound languages and/or non-linear processes: whether predetermined in full or part, or constructed solely as an outcome of the interplay of agent deployment. The research question here is one of the relationships between current agent work and generic dramatic narrative structures that are found in other time-based art forms that are a part and reflection of human experience.

A starting point for this discussion that might be useful for future agent-based work includes Booker's (2004) *The Seven Basic Plots: Why We Tell Stories*. Booker argues that regardless of place, time or culture there are reoccurring narrative forms that are expressions of the human condition. In these terms, people retell old stories in new ways, regardless of language variations. Similarly, system dynamics modelling (Senge 1992) provides a set of narrative structures that are a long-standing part of both time-based arts and real-world situations that could be drawn on (see Whalley 2001) to extend discussion, Senge arguing that there are reoccurring dramatic narrative archetypes that can be coupled together in different ways to reflect narrative structure and the tension/relaxation that dynamically results from them.

Perhaps most importantly, however, a fruitful area for further investigation, largely ignored in the recent

agent-technology exploration process, is the affective/emotional dimension of music/sound art given in Graugaard's (2006) model: the gap between interpreting music/sound art language in the technical way that practitioners tend to do, and what it means in affective terms to a wider range of people. The significance of this is that emotional interpretation is an essential part of human communicative intelligence, carried by languages that are sometimes shared by both practitioners and lay people. To point to a possible future toward integrating technical and human reception perspectives in continuing work, the proposal in this paper is then to first attempt to include an affective dimension in agent-based work to readdress the imbalance.

4.2. Affective music: research challenges and departures

A starting point in integrating affective music approaches is the assumption that music/sound art composition and performance is a semiotic system that signifies meaning to elicit emotional responses (Gabrielsson 2001; Mithen 2005). This is not to imply that the mechanism by which this happens is a straightforward one-to-one system, or that meaning cannot be multiple or change with the context of reception or time, or that the chain from composition to performance to reception is always stable. The argument is simply that there is a robust enough theoretical and empirical basis to attempt this in music composition and performance in western music practice (see Juslin 2001; Scherer and Zentner 2001), and that there is a successful history of practically implementing aspects of the approach in media such as film music in western culture (Cohen 2001). Recent work in musical and emotional mapping also provides a basis to implement this technically (see Gabrielsson and Lindstrom 2001; Schubert 2001; Scherer 2004). Similarly, there is now significant theoretical or empirical work on emotion and meaning in music that reinforces the basis of this approach (Peretz 2001; Patel 2008: 300–51), and this is supported by research on musical embodiment and rhythm previously cited (Bryant and Hagen 2003; Thaut 2005; Brown and Parsons 2008).

The technical research challenge is then how to implement an affective approach within the interactive/agent-based paradigm (see Nakamura, Numao and Takagi 2002 as an early starting point), linking composition, performance and reception. Recent work here has already begun to be explored that can be drawn upon.

4.3. Pointers: affective composition, performance and reception linking

Affective computing is a growing field of interest in computer science research, and understanding how

music and emotion matching in new media might be incorporated into technical systems is an up-and-coming field in multimedia research. Recent explorations are found in Li and Shan (2007), Homer et al. (2007), and Wang, Zhang, and Zhu (2004). Cai, Wei-Ying, Wang, Zhang and Zhang (2007), for example, present a system to automatically suggest music when users read web documents, matching music to a document's content using songs drawn on in terms of the emotions expressed. These explorations attempt to update, in non-linear new media, what film composers were sometimes skilled at in older linear media. The concern is largely with mapping existing music to trigger emotional response.

Mapping emotion to real-time musical expression in *performance* is also a continuing part of computer music research that might be usefully drawn. A recent example, and one that uses agent technology, is found in Coutinho, Miranda and Silva (2005). This addresses the issue from an A-Life standpoint combined with neuroscience's perspectives on emotion. The method partly uses an agent-based system to model different emotional influences, and then plays MIDI files expressively according to an adaptive process.

The area of emotional embodiment has begun to be addressed in *authoring* systems: two musically conservative examples include Whalley's (2002) 'Kansei in Non-Linear Automated Music Composition: Semiotic Structure and recombincity'; and the work by Hashimoto, Kurihara, Legaspi, Moriyama and Numao (2007) in 'Music Compositional Intelligence with an Affective Flavor'. Whalley (2002) proposes a software-based automated composition system operated by a single user to create music for closed system dramatic narratives where the dramatic parameters are known but the dramatic shape and outcomes are not predetermined. The concern is with a system that will address Kansei (emotion-based; see Hashimoto 1998) approaches to narrative structure, musical generation and performance. The model outlined allows for music creation from controlling an emotional 'flight simulator' interface that represents affective states, rather than dealing directly with the composition process, allowing non-composers to 'recompose' or explore a work in different ways.

Extending this is the recent offering from Hashimoto et al. (2007), noting that human feelings have received little attention in automated music generation by intelligent music systems. Their work aims for a method that is musical compositional intelligent and directly linked with the listener's affective perceptions. The system induces a model describing the relationship between feelings and musical structures, learned by the use of the inductive logic programming paradigm in FOIL (see Quinlan n.d.), 'coupled with the Diverse Density weighting metric over a dataset that was constructed using musical score fragments that were hand-labelled by the listener according to a semantic

differential scale that uses bipolar affective descriptor pairs'. From this, a genetic algorithm based on the acquired model and following standard notational theory generates variations of original musical structures. Although the range of emotional impressions from the music is restricted (favourable-unfavourable, bright-dark, happy-sad, and heartrending-not heartrending), their system is at least able to classify and generate impressions with reasonable accuracy.

Finally, a more successful integration of the affective approach linking composition, performance and reception is found in the recent work of Brown, Livingstone and Muhlberger, 'Controlling Musical Emotionality: An Affective Computational Architecture for Influencing Musical Emotions' (2007). Their system, using western tonal music, demonstrates an affective computing architecture where music is dynamically modified to predictably affect induced musical emotions. The advance over other systems, based in research on musical/emotional mapping from reception studies, is that the system aims to reliably control both perceived (what is sent) and induced (what is received) musical emotions. This is a rule-based system used to modify a subset of musical features for both composition and performance. It is interactive in that it leverages the listeners' affective sense by adapting the emotionality of the modification made to the music in real-time towards assisting the listeners to reach a desired emotional state. This provides a basis, in an adaptive environment, to implement what film music composers partly attempt to do instinctively in a linear film-music environment.

An accepted limitation of current work here is it is largely based in western tonal languages. An area that remains to be explored in affective music/sound art is then taxonomy for mapping affective emotional responses to a wider range of sound art languages that could be drawn upon. This is partly a reflection of the general neglect of this area in research based in sound art.

5. TOWARDS A MODEL FOR FUTURE WORK

In recent music/sound art works that use agent technology, the generative improvisation approach demonstrates the greatest potential to extend and broaden music making that is process-based. This is because the method moves beyond the limitations of non-adaptive rule-based systems, and combines the best of both human and machine agency to control both performance and composition.

In addition, agent technology has the potential to further enhance music/sound art creative practice through allowing machines to take on greater 'intelligence' in the interactive process by increasing the level of autonomy, adaptability and learning ability of agent technology deployment. Through this technical underpinning, the conversational model of human/machine

interaction can be more fully explored. The questions then concern what the conversations are about, and what musical language is used to enact them.

In addressing ‘what’, the conversational model is best embedded in machine agency and adaptable rules based on affective taxonomy that are part of human communicative intelligence, expressed through an affective music paradigm that links music composition and performance with reception research. The theoretical or empirical basis for this is covered in the previous section, and some recent technical deployments described to demonstrate how this might be implemented in agent-based work.

In line with the conversational model, the music/sound art ‘language’ used in a work needs to be embedded not only in affective reception rules but the associated language used to convey affective meaning that is familiar to the intended audience, in order facilitate dialogue between human and machine agency. The music/sound art language chosen must also be able to be varied to convey reoccurring ideas in new ways to sustain interest, and be expanded within the reception paradigm to enhance education, experimentation and engagement. The machine agency, in balancing familiar and unknown languages, should then allow each interactive session to develop something of its own language. Machine agency can then lead or follow in the interactive process with human agency, acknowledging that not all conversations are symmetrical in terms of knowledge and participation.

In terms of research ‘stretch’ of the proposal here, two areas require attention in line with in Graugaard’s (2006) model of interactivity. The first is the narrative dimension, with some possible directions for future work being outlined in the prior discussion. Second, but beyond the scope of the paper here, is human gesture input sensing/capture/mapping beyond the simple computer/mouse input method of reflecting 2D affective human intension (see Schubert 2001). The field of HCI research could be drawn on here to allow for more extensive communicative input to music/sound art based agent systems by non-musicians.

In these terms and as a proposed first step, a hybrid agent system based on human/machine generative improvisation that includes an affective conversational model based in known and evolving languages, at least, presents opportunities to draw together and extend multidisciplinary work in agent-based systems applied to music and sound art, redressing the current lack of balance between technical and human reception perspectives in the field.

REFERENCES

Anumba, C., Ren, Z. and Ugwu, O. (eds.) 2005. *Agents and Multi-Agent Systems in Construction*. London: Taylor & Francis.

- Aronson, J., Liang, T. and Turban, E. 2004. *Decision Support Systems and Intelligent Systems* (7th ed.). Englewood Cliffs, NJ: Prentice Hall.
- Balogh, Z., Budinska, I., Dang, T. T., Hluchy, L., Laclavik, M. and Nguyen, T. 2002. *Agent Platforms Evaluation and Comparison*. Institute of Informatics, Slovak Academy of Sciences. Pellucid 5FP IST-2001-34519. Viewed online: <http://ups.savba.sk/~giang/publication?tr-2002-06.pdf> (accessed 17 April 2009).
- Bigus, J. P. and Bigus, J. 2001. *Constructing Intelligent Agents In Java* (2nd ed.). New York: John Wiley & Sons.
- Biles, J. 2003. GenJam in Perspective: A Tentative Taxonomy for GA Music and Art Systems. *Leonardo* **36**(1): 43–5.
- Beyls, P. 2007. Interaction and Self-Organisation in a Society of Musical Agents. *Proceedings of ECAL 2007 Workshop on Music and Artificial Life*. Lisbon, Portugal.
- Booker, C. 2004. *The Seven Basic Plots. Why We Tell Stories*. London: Continuum.
- Brown, A. 2002. Opportunities for Evolutionary Music Composition. *Proceedings of Australasian Computer Music Association Conference*. Melbourne, Australia: 27–34.
- Brown, A., Livingstone, S. and Muhlberger, R. 2007. Controlling Musical Emotionality: An Affective Computational Architecture for Influencing Musical Emotions. *Digital Creativity* **18**(1): 43–53.
- Brown, S. and Parsons, L. 2008. The Neuroscience of Dance. *Scientific America* (July): 58–63.
- Brown, S., Merker, B. and Wallin, N. 2001. *The Origins of Music*. London: The MIT Press.
- Bryant, G. and Hagen, E. 2003. Music and Dance as a Coalition Signalling System. *Human Nature* **14**(1): 21–51.
- Cai, T., Wei-Ying, M., Wang, C., Zhang, C. and Zhang, L. 2007. MusicSense: Contextual Music Recommendation using Emotional Allocation Modelling. *Proceedings of the 15th International Conference on Multimedia*. Augsburg, Germany: 553–6.
- Camurri, A. 1993. Applications of Artificial Intelligence Methodologies and Tools for Music Description and Processing. In G. Haus (ed.) *Music Processing*. Madison, WI: A-R Editions, 233–66.
- Chadabe, J. 1996. The History of Electronic Music as a Reflection of Structural Paradigms. *Leonardo Music Journal* **6**: 41–4.
- Chadabe, J. 2004. Electronic Music and Life. *Organised Sound* **9**(1): 3–6.
- Chen, C. and Kiss, J. 2003. Setting Up a Self-Organized Multi-Agent System for the Creation Of Sound and Visual Virtual Environments within the Framework of a Collective Interactivity. *International Computer Music Conference*. Singapore: 11–14.
- Cohen, A. 2001. Music as a Source of Emotion in Film. In P. Juslin and J. Sloboda (eds) *Music and Emotion: Theory and Research*. Oxford: Oxford University Press, 249–72.
- Consoli, A., Ichalkaranje, N., Jarvis, B., Phillips-Wren, G., Tweedale, J. and Sioutis, C. 2006. Innovations in Multi-agent Systems. *Journal of Network and Computer Applications* **30**: 1089–115.
- Cope, D. 1991. *Computers And Musical Style*. Madison, WI: A-R Editions.

- Coutinho, E., Miranda, E. R. and da Silva, P. 2005. *Evolving Emotional Behaviour for the Expressive Performance of Music. Lecture Notes in Computer Science*. London: Springer-Verlag: 497.
- Detlor, B. and Serenko, A. 2002. *Agent Toolkits: A General Overview of the Market and an Assessment of the Instructor Satisfaction with Utilizing Toolkits in the Classroom*. Working Paper #455. School Of Business, McMaster University, Hamilton, ON.
- Dorin, A. 2001. Generative Processes and the Electronic Arts. *Organised Sound* 6(1): 47–53.
- Drogoul, A., Meurisse, T. and Vanbergue, D. 2002. Multi-Agent Based Simulation: Where are the Agents? *Proceedings of MABS'02 (Multi-Agent Based Simulation)* Bologna, Italy, LNCS, Springer-Verlag.
- Edwards, M., Murray-Rust, A. and Smaill, D. 2006. MAMA: An Architecture for Interactive Musical Agents. *Frontiers in Artificial Intelligence and Applications* 141: 36–40.
- Eigenfeldt, A. 2008. Emergent Rhythms through Multi-agency in Max/MSP. In R. Kronland-Martinet, S. Ystad and K. Jensen (eds.) *Computer Music Modeling and Retrieval: Sense of Sounds*. Copenhagen: CMMR, 368–79.
- Gabrielsson, A. 2001. Emotion in Strong Experiences with Music. In P. Juslin and J. Sloboda (eds.) *Music and Emotion: Theory and Research*. Oxford: Oxford University Press, 441–9.
- Gabrielsson, A. and Lindstrom, E. 2001. The Influence of Musical Structure on Emotional Expression. In P. Juslin and J. Sloboda (eds.) *Music and Emotion: Theory and Research*. Oxford: Oxford University Press, 223–48.
- Gang, D., Goldman, C. and Rosenschein J. 1999. NegNet: A Connectionist-Agent Integrated System For Representing Musical Knowledge. *Annals of Mathematics and Artificial Intelligence* 25: 69–90.
- Gimenes, M., Miranda, E. R. and Johnson, C. 2005. A Memetic Approach to the Evolution of Rhythms in a Society of Software Agents. *Proceedings of the 10th Brazilian Symposium on Computer Music (SBCM)*. Belo Horizonte, Brazil.
- Gimenes, M., Miranda, E. R. and Johnson, C. 2006. The Development of Musical Styles in a Society of Software Agents. *Proceedings of the International Conference on Music Perception and Cognition*. Bologna.
- Graugaard, L. 2006. Implicit Relevance Feedback in Interactive Music: Issues, Challenges, and Case Studies. *Proceedings of the 1st International Conference on Information Interaction*, Copenhagen, Denmark. *ACM International Conference Proceeding Series* 176: 119–28.
- Hashimoto, S. 1998. KANSEI as the Third Target of Information Processing and Related Topics in Japan. *Proceedings of International Workshop on Kansei – Technology of Emotion*, Tokyo, Japan: 101–4.
- Hashimoto, Y., Kurihara, S., Legaspi, R., Moriyama, K. and Numao, M. 2007. Music Compositional Intelligence with an Affective Flavor. *Proceedings of the 12th International Conference on Intelligent User Interfaces*. Honolulu: 216–24.
- Homer, H., Chen, H., Lin, Y. and Su, Y. 2007. Music Emotion Recognition: the Role of Individuality. *Proceedings of the International Workshop on Human-centered Multimedia*. Augsburg: 13–22.
- Juslin, P. 2001. Communicating Emotion in Music Performance: A Review of Theoretical Framework. In P. Juslin and J. Sloboda (eds.) *Music and Emotion: Theory and Research*. Oxford: Oxford University Press, 309–77.
- Juslin, P. and Sloboda, J. (eds.) 2001. *Music and Emotion: Theory and Research*. Oxford: Oxford University Press.
- Kon, F. and Ueda L. 2004. Andante: Composition and Performance With Mobile Musical Agents. *International Computer Music Conference*. Miami, FL 1–6 November: 604–7.
- Landy, L. 1999. Reviewing the Musicology of Electro-acoustic Music: A Plea for Greater Triangulation. *Organised Sound* 4(1): 61–70.
- Li, C. and Shan, M. 2007. Emotion-based Impressionism Slideshow with Automatic Music Accompaniment. *International Multimedia Conference Proceedings of the 15th International Conference on Multimedia*, Augsburg, Germany, 839–42.
- Malt, M. 2004. Khorwa: A Musical Experience with Autonomous Agents. *International Computer Music Conference*. Miami, FL 1–6 November: 59–62.
- Mithen, S. 2005. *The Singing Neanderthals. The Origins of Music, Language, Mind and Body*. London: Weidenfeld & Nicolson.
- Miranda, E. R. 2003. A-Life and Musical Composition. A Brief Survey. *IX Brazilian Symposium on Computer Music*. Campinas, Brazil.
- Miranda, E. and Gimenes, M. 2008. An A-Life Approach to Machine Learning of Musical Worldviews for Improvisation Systems. *Proceedings of 5th Sound and Music Computing Conference*. Berlin.
- Nakamura, K., Numao, M. and Takagi, S. 2002. CAUI Demonstration – Composing Music Based on Human Feelings. *Proceedings of American Association for Artificial Intelligence*. www.aaai.org.
- Nakayama, L., Wulfhorst, R. and Vicari, R. 2003. A Multi-Agent Approach for Musical Interactive System. *AAMAS'03*, Australia: 584–91.
- Nussbaum, C. 2007. *The Musical Representation. Meaning, Ontology, and Emotion*. London: MIT Press.
- Paine, G. 2002. Interactivity, Where to From Here? *Organised Sound* 7(3): 295–304.
- Patel, A. 2008. *Music, Language and the Brain*. Oxford: Oxford University Press.
- Peretz, I. 2001. Listen to the Brain: A Biological Perspective on Musical Emotions. In P. Juslin and J. Sloboda (eds.) *Music and Emotion: Theory and Research*. Oxford: Oxford University Press, 105–34.
- Quinlan, R. n.d. FOIL <http://www.rulequest.com/Personal>.
- Rowe, R. 1994. *Interactive Music Systems: Machine Listening and Composing*. London: MIT Press.
- Schubert, E. 2001. Continuous Measurement of Self-Report Emotional Response to Music. In P. Juslin and J. Sloboda (eds.) *Music and Emotion: Theory and Research*. Oxford: Oxford University Press, 393–414.
- Scherer, K. 2004. Which Emotions Can be Induced? What are the Underlying Mechanisms? And How Can We Measure Them? *Journal of New Music Research* 33(3): 239–51.
- Scherer, K. and Zentner, M. 2001. Emotional Effects of Music: Production Rules. In P. Juslin and J. Sloboda

- (eds.) *Music and Emotion: Theory and Research*. Oxford: Oxford University Press, 361–92.
- Senge, P. 1992. *The Fifth Discipline. The Art and Practice of the Learning Organization*. New York: Random House.
- Smalley, D. 1997. Spectromorphology: Explaining Sound-shapes. *Organised Sound* 2(2): 107–26.
- Spicer, M. 2004. AALIVENET: An Agent Based Distributed Interactive Composition Environment. *International Computer Music Conference*. Miami, FL 1–6 November: 201–204.
- Spicer, M., Tan, B. T. G. and Tan, C. 2003. The Learning Agent Based Interactive Performance System. *International Computer Music Association Conference*. Singapore: 95–9.
- Thaut, M. 2005. *Rhythm, Music, and the Brain: Scientific Foundations and Clinical Applications*. London: Routledge.
- Wang, M., Zhang, N. and Zhu, H. 2004. User-adaptive Music Emotion Recognition. *7th International Conference on Signal Processing*. Beijing: 1352–5.
- Weale, R. 2006. Discovering How Accessible Electro-acoustic Music Can Be: The Intention/Reception Project. *Organised Sound* 11(2): 189–200.
- Weinberg, G. 2005. Interconnected Musical Networks: Toward a Theoretical Framework. *Computer Music Journal* 29(2): 23–39.
- Weiss, G. (ed.) 2000. *Multi-Agent Systems, a Modern Approach to Distributed Artificial Intelligence*. Cambridge, MA: MIT Press.
- Whalley, I. 2001. Applications of System Dynamics Modelling to Computer Music. *Organised Sound* 5(3) 2000: 149–57.
- Whalley, I. 2002. Kansei in Non-linear Automated Composition. Semiotic Structure and Recombinicity. *11th International Symposium on Electronic Arts*. Nagoya: 62–5.
- Whalley, I. 2004. Adding Machine Cognition to a Web-Based Interactive Composition. *International Computer Music Conference*. Miami, FL 1–6 November: 197–200.
- Whalley, I. 2005. Software Agents and Creating Music/Sound Art: Frames, Directions, and Where to From Here? *Proceedings of International Computer Music Conference*. Barcelona, 5–9 September: 691–5.
- Winkler, T. 1998. *Composing Interactive Music: Techniques and Ideas Using MAX*. Cambridge, MA: MIT Press.
- Wishart, T. 2002. *On Sonic Art*, ed. S. Emmerson. London: Routledge.