Creative Agents, Curatorial Agents, and Human-Agent Interaction in *Coming Together*

Arne Eigenfeldt
School for the Contemporary Arts
Simon Fraser University
Vancouver, Canada
arne_e@sfu.ca

Philippe Pasquier
School of Interactive Arts and Technology
Simon Fraser University
Surrey, Canada
pasquier@sfu.ca

ABSTRACT

We describe a multi-agent systems which composes in real-time, using negotiation as the active compositional technique. In one version of the system, creative agents' output is written to disk; during performance, a curatorial agent selects prior-composed movements and assembles a complete musical composition. The resulting score is then displayed to musicians, and performed live. A second version of the system is described, in which the real-time interaction is performed immediately by a mechanical musical instrument, and a human instrumentalist's performance data is included in system as being one of the agents (a human agent).

1. INTRODUCTION

Coming Together is a series of computational creative systems based upon the premise of composition by negotiation – within a controlled musical environment, autonomous multi-agents attempt to converge their data, resulting in a self-organised, dynamic, and musically meaningful performance. Previous versions have included soundscape composition – *Coming Together: Freesound* [1] – sonic ecosystem – *Coming Together: Shoals* [2] – and negotiation – *Coming Together: Beauty and Truth* [3].

In the latter system, agents spend a majority of the performance negotiating their parameters before agreeing upon a congruous phrase. The focus of that system is audible self-organisation; however, after one performance, an audience member suggested that after working so hard to align their phrases, it would be nice to hear the agents interact within the negotiated musical space for a longer period of time. The system described in this paper – hereafter referred to as *CT4* – can be considered a response to *Beauty and Truth*: exploitation of the shared musical space, rather being limited to exploration.

All the *Coming Together* systems involve some aspect of *a priori* structure around which the negotiation by the agents is centered. In *CT4*, the structure presupposes several discrete movements that together form a complete composition of a predetermined length. Characteristics of each movement – density, time signature, tempo – are generated using a fuzzy-logic method of avoiding similarity between succeeding movements (see Section 3.1).

Section 2 provides an overview of related work; Section 3 offers a detailed description of the system; Section 4 describes human-agent interaction; Section 5 describes the curatorial agent; Section 6 offers some conclusions and future directions.

2. RELATED WORK

Negotiation is an active research topic in the multi-agent community [4, 5], and our use of the term is admittedly idiosyncratic. Rather than referring to bargaining or auction protocols, we refer to the notion of a dialog between autonomous agents in order to achieve a collective course of action whose outcome satisfies the interests of the individual agents. Loosely implementing a BDI model [6], agents pursue collective musical goals of a similar volume, density, and onset distributions over a phrase. These goals are achieved through continual social interaction; furthermore, agent actions influence future decisions by other agents.

The models used here do not follow any known approach, and are built mainly by artistic, and thus heuristic, decisions. The resulting music can be considered emergent, in that each unique musical composition results from the complex interactions of the agents, a result which cannot be predicted given the initial conditions. Our work is built upon, and informed by, others who have explored emergent art, including John McCormack [7], Eduardo Miranda [8], Peter Beyls [9], and Joseph Nechvatal [10], to name a few.

3. DESCRIPTION

Coming Together was written in MaxMSP [11]. An earlier version of this system in performance can be viewed here: http://youtu.be/7HyU8nHs_pk

Two versions of *CT4* are described, used in two different musical compositions. The first, for the composition *And One More*, involves agents interacting in real-time, their output being sent via MIDI to a mechanical percussion instrument, the *Karmetik Notomoton* [12]. This version has nine different agents performing on eighteen different percussion instruments, and includes a live percussionist whose performance is encoded and considered an additional agent (see Section 4).

The second version, for the composition *More Than Four*, involves four agents, whose output is eventually translated into musical notation using *MaxScore*[1], for performance by four instrumentalists. Agent interaction is transcribed to disk prior to performance (for reasons described in Section 5); at the onset of the performance, a curatorial agent selects previous movements from the database, and chooses from those to create a musically unified composition.

3.1 Score Generation

All the *Coming Together* systems balance the bottom-up self-organisation of autonomous agents with the top-down organisation typical of more traditional composition approaches. In *CT4*, the top-down organisation amounts to a user-defined performance duration, and its division into separate movements. Each movement is comprised of a unique musical environment – described below – and a progression from disorganization to mutually agreed upon musical parameters.

At the onset of the performance, an algorithm generates the number of movements within a composition, making sure no section is less than 30 seconds long, nor exceeds 33% of the complete composition's duration. A tempo is chosen for the first movement (between 60 and 132 beats per minute), and phrase length (between 7 and 22 beats). As phrases are contained within a single measure, the phrase length is also the time signature.

Phrases are constructed rhythmically using additive processes. For example, given a phrase of 10 beats, it will be grouped in the following pattern: 3 2 3 2 (see Figure 1). Once such a grouping is made, it remains constant for the phrase, and becomes a *tala*. Each movement has a single tala assigned to it.

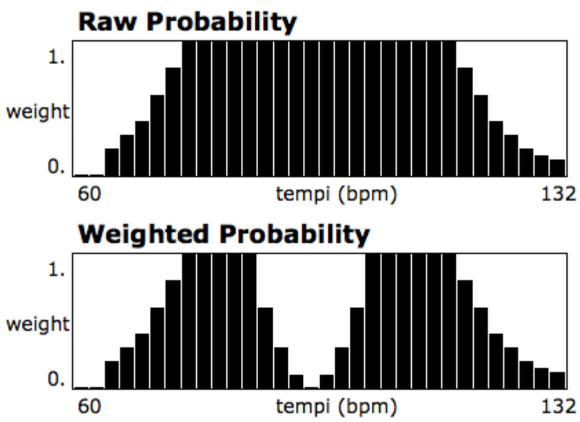

Figure 1. Tempo probability, given a tala of 12 (top), and after weighting, given a previous tala of 15 (bottom).

Fuzzy logic is used to weight probabilities within and between movements. For example, if a slow tempo is chosen for a movement, the probability for longer tala (phrase length) is weighted lower. Choices from previous movements are taken into account, and weighted negatively, in order to avoid immediate duplication (see Figure 1) and ensure contrast between movements.

Once the probability curve for a parameter has been generated, a roulette-wheel selection is made.

3.2 Fuzzy Decision Making

When performing, agents will often need to wait a certain number of phrase repetitions before making a decision, but the length of time is bounded: for example, "wait at least" four, but no more than eight, measures before making a change. This type of fuzzy decision-making occurs throughout the system; actions are delayed by a predefined time value (in this case, four repetitions), and once that time passes, the likelihood of triggering the action increases until the second specified time (in this case, eight measures), at which point there is a 100% chance that the action will have occurred.

The initial values (in this case, four and eight) are predefined; however, they can be globally altered by two scalers that are randomly chosen for each movement; as a result, some movements "move faster" or "move slower" than others. Two different scalers are used: the first modifies the initial wait (in this case, four measures); the second modifies how quickly the probability moves to 100% (in this case, the next four measures).

Similarly, binary choices – coin tosses – are weighted based upon previous selections in order to avoid complete randomness. In such cases, the coin is initially weighted equally: 0.5 and 0.5. Following a selection of "tails", or 1, the weighting changes to 0.6 and 0.4 for the next coin toss.

3.3 Initialisation

Once a score has been generated, agents are initialized at the beginning of each movement, which involves generating the following parameters:

- Onset distribution: whether onsets are randomly distributed over the phrase, closer to the phrase beginning, or phrase ending;
- Density: the relative number onsets within a phrase;
- Volume: initial MIDI velocity, between 1 and 127;
- Inner Density: the number of onsets between tala beats;
- Subdivisions: the number of onsets between inner-density beats;
- Downbeat: whether to ensure a downbeat;
- Grace Notes: whether to allow grace notes as onsets.

These values are generated randomly within a predefined range unique to specific agents (see Section 3.5), with the exception of density (see Section 3.4.2). Given these parameters, agents will generate a pattern. See Figure 2 for two example generations.

Once these values have been generated, agents broadcast the values globally to the agent community. Agents gather other agent's data for negotiation.

[1] www.computermusicnotation.com

Figure 2. Musical representation of a given tala (3+2+3+2), top, and two possible patterns by agents. Agent A is limited to inner beats – onsets on and between, the tala – while agent B has subdivisions as well.

3.4 Negotiation

"Every few phrases", agents will compare their parameter data to those of the ensemble – how often this occurs is dependent upon the overall time scaler described in Section 3.2. The specific process for volume is described below. Note that a similar negotiation occurs for all parameters.

3.4.1 Volume

Agents will attempt to converge on a uniform volume — in order to achieve a group dynamic — using averaging of all agent velocities. Convergence occurs for a given parameter – in this case volume – when the individual agent values for this parameter are the same. In order to avoid always moving towards a central MIDI velocity (i.e. 64), agents also compare the group mean to the group minimum and maximum velocities. If either of these extreme values is more than 13 velocity levels (i.e. 10% of the total range available) different from the mean, agents will converge toward the most extreme of these values; otherwise, it will converge toward the mean. The potential for converging upon extremes allows for dynamic volume changes over the course of a composition (see Figure 3 for an example).

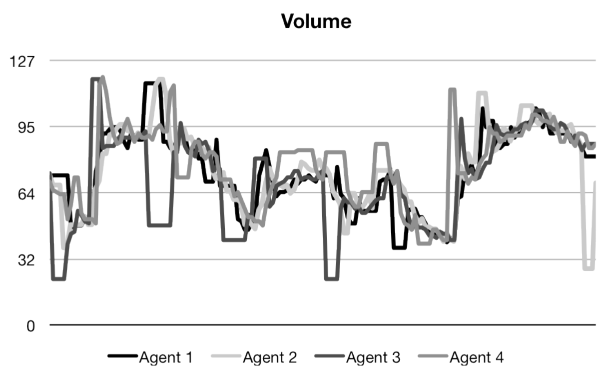

Figure 3. Volume convergence over the course of a movement. Despite the desire to generate a communal volume level, agents occasionally break away, which has the effect of generating macro-level volume changes.

Agents determine successful volume convergence by comparing their own velocity with the group mean. If they are the close (another fuzzy decision), the agent generates a random value between 0. and 1., and compares it to its *commitment* parameter (see Section 3.5). If the random value is greater, the agent will attempt to "break away" from the group. In this case, the agent will choose a random velocity from the available range, then wait between 10 and 20 seconds before once again averaging group volume. This pause essentially forces the other agents to converge on the new velocity, the musical result being a periodic shift in the ensemble dynamics.

3.4.2 Density

Density can be considered as the number of onsets over a given phrase. *CT4* considers two different approaches to density: textural density and cumulative density. *Textural density* is a globally defined parameter that is generated as part of the score, and is consistent within a movement. Agents attempt to match this global parameter based upon their *confidence* personality attribute (see Section 3.5), which determines whether an agent will turn on at the beginning of a movement. As such, textural density can be translated as the number of agents that will be active during a movement.

Cumulative density is the number of onsets agents play during a phrase, and is negotiated. The potential for contradiction can result in a complexity of textures and general unpredictability of responses; for example, low textural density and high cumulative density will result in a few very active agents, while a high textural density and low cumulative density will result in many agents with few onsets. Furthermore, if the textural density has been met, but the cumulative density begins to lower through negotiation, agents may decide to drop out, which will then result in the textural density having to be renegotiated.

3.5 Personality Attributes

Earlier versions of *Coming Together* used a large number of identical agents interacting within a musical space; however, the output of one version of *CT4* is sent to very unique mechanical percussion instruments [13]. As such, each agent is required to behave differently, and a method employed in an earlier system to differentiate agent actions was used [14]. Agents in *CT4* have "personality" attributes that determine how they will interact with other agents, and react to the environment.

For example, one agent is dedicated to performing the shakers, another a large frame drum: the former will almost always create subdivisions within its phrases, while the latter agent will rarely include inner-beat onsets, and instead assume the role of outlining the main beats of the tala. Furthermore, agents interpret density differently by scaling this parameter through a personality attribute: given a negotiated subdivision parameter, the shaker agent, for example, will interpret this value higher than other agents, resulting in a greater number of onsets.

Lastly, three attributes determine the agents' response to the global textural density parameter:
- Confidence: whether an agent plays at the start of the movement;
- Responsiveness: whether an agent joins in, or cuts out, if global density is different than ensemble density;
- Stability: if an agent is on, whether it wants to stay on.

3.6 Autonomous Variation

The ability for autonomous agents to determine both their own variations and the amount of variation has been integral to all of the first author's systems since 2006 [15]. In *CT4*, whenever an agent adjusts a parameter that results in an audible change to its phrase, the agent broadcasts this as a variation. Agents collect all these variation points, and create a vector of the previous eight phrases, which is then averaged. This value is subtracted from an ongoing measure counter that results in a "variation desire" parameter. When this parameter exceeds a preset value, it suggests that a greater overall variation is desired by the agents.

At this point, if the progression through the movement is less than 80% of the scored duration[2], a harmonic change is invoked (harmonic organization is not discussed here). If the agents have passed 80%, the movement can begin its transition towards an ending (see Section 3.7).

In practice, while the agents are actively negotiating and moving towards convergence, enough variation occurs so as to keep the variation-desire low; however, once the agents have converged, variation will ostensibly cease for a time (before an agent may choose to break away from the group); at this point, the variation-desire may rise high enough to trigger a harmonic change or the end of the movement. However, if the agents are not able to converge for any reason, and the variation-desire remains low, and movement durations can exceed the scored duration.

3.7 Transitions

Once convergence has occurred over enough parameters so as to trigger progression toward the ending of the movement, an ending transition in generated. This transition essentially alters the horizontal structure to a vertical one, where currently active agents emphasize the tala beats. This simple change serves as a unifying texture to end all movements.

4. HUMAN-AGENT INTERACTION

And One More – the composition which incorporates a mechanical percussion instrument – also includes a live human percussionist. The human agent's performance is encoded, and is considered part of the overall society of agents.

The model in this case is not of the agents reacting to the human input directly in an interactive way. Just as the virtual agents do not interact with, or even communicate, their musical phrases with one another, the human musical input is only considered in a general way. The human performer's onsets are counted over the duration of a phrase, and a vector is created for the previous four measures. This vector is averaged, providing the system with the general density of the human performer. The current phrase's onsets are also compared to the previous three measures, so as to interpret whether the performer is more, less, or similarly active compared to the recent past.

The human agent's density is broadcast to the virtual agents, who treat it as any other agent data towards their convergence. A heuristic decision was made so as to allow the human agent to have a greater level of influence over the agent community – and thus the resulting self-organisation – by multiplying the number of times the human density is broadcast. For example, as there are nine virtual agents, broadcasting the human performer's parameter eight additional times gives it an eightfold increase in influence.

Human influence is thus not only limited (even with the described multiplication), but also leisurely, since convergence takes place over several phrases. This was considered to be artistically limiting, and counter to how the human performers wanted to interact with the system. During rehearsals, a serendipitous interaction occurred in which the virtual agents seemed to stop playing in reaction to the live performer. This direct interaction, although contrary to principles of self-organisation, was thought to be artistically interesting, and was incorporated into the system as a potential interaction. The longer the human performer doesn't play, a tension parameter increases; when the performer then plays, if the tension parameter is high enough, the virtual agents can temporarily stop playing. This interaction is limited through a fuzzy rule: it cannot occur too often (which would limit its effectiveness), nor can it occur in every movement.

5. CURATORIAL AGENT

More Than Four involves four agents whose output is translated into musical notation, and performed live by humans on two vibraphones and two marimba. Performance data, rather than being sent out as MIDI information, is converted into notation directions for *MaxScore*, creating a continually updated traditionally notated score (see Figure 4).

Several major changes were required in order to implement this extension. When generating performance information for virtual performers, the data does not have to be complete before performance can begin: only the immediately required onset data. In other words, the complete phrase does not have to be generated before the agent can begin playing, since the incomplete information can be generated while the agent is playing. However,

[2] 80% is a heuristic value arrived at through continual listening to the system's output.

Figure 4. Three phrases of output from *CT4*, shown as musical notation, for two vibraphone and two marimba.

notation requires an entire measure to be complete before it can be notated and interpreted by musicians. Musicians do not read music note to note, but instead chunk notes into sequences [16]. As such, a one measure delay was required in order notate the agent interactions in the previous measure. Other practical issues – including score rendering and communication over the local area network – forced a further one measure delay.

During an extended workshop-rehearsal period, a temporary solution to latency issues was used in which the agent data was written to disk in advance as separate movements; during rehearsal, these movements were recalled individually for read-through performance. Although this was obviously recognized as taking the system out of real-time – the two-measure delay between agent interaction and audible musical result by performers had already done this to an extent – this process was seen as a possible avenue of exploration: selection of movements for performance could be done by an artificial agent that analyses the performance data within the movements, and uses this information to create a single composition.

Research into artificial music critics has been previously undertaken [17, 18], however such critics are involved in the creative production; in our case, the creation is complete, and it is the selection for presentation that is undertaken, which can be seen as an act of curation [19]. Curation has traditionally involved the selection, preservation, maintenance, collection, and archiving of art objects; recently, digital curation has emerged, which has expanded this endevour to digital data. Furthermore, audio curating is a growing field, which involves the selection and presentation of soundart [20].

Score generation in *CT4* attempts to create a variety of environments for the agents to interact within, by ensuring varied tempi, durations, tala, and durations. Due to the self-organization within the system, no meaningful control can be placed over the musical results during generation. However, once the interaction is complete and written to disk, the curatorial agent can analyze the actual musical data, searching for generated musical motives in terms of rhythmic and melodic motion, as well as larger patterns of textural density and harmonic motion.

In selecting movements to combine, the curatorial agent selects a movement from the database at random, then rejects all those movements that have the same time signature, tempo, and rhythmic grouping. It then evaluates the remaining movements, rating them in terms of cumulative density, pitch range and variation, volume, overall length, specific instrumental presence, and harmonic movement, with those closest to the selected movement rated higher. Finally, it looks at the selected movement for recurring pitch and rhythmic patterns, and rates the database for similarity.

A Gaussian selection is made from the highest rated movements to choose the next movement, so as to avoid the same selection every time, given the initial movement. The user requests an overall duration: when this duration is reached, the composition is considered complete.

The original conception of *More Than Four* was to generate hundreds of movements, from which the curator agent would assemble a composition immediately prior to the performance, in front of the audience. However, during the extended rehearsal process, it became clear that the performers much preferred to have some familiarity with the material beforehand; although the musicians were professional percussionists of the highest caliber, the unusual rhythmic groupings that could extend to 24 beats proved extremely challenging. As such, a week before the performance, approximately thirty movements were selected, from which the curator agent would assemble approximately five for a twelve minute performance.

The curator agent is creating the overall form of the resulting music, a musically high-level act that is an open challenge in the area of computational music systems.

6. CONCLUSIONS

Two versions of a computationally creative music production system were described. These systems are the continued exploration of a model of composition the authors' consider composition by negotiation, a process investigated in the other systems within the *Coming Together* series. Many of the approaches described build upon earlier systems by the first author that, like the ones described here, are created for artistic production, rather than purely scientific research.

Future directions include more detailed live performance analysis for human-agent interaction, and more detailed musical analysis of generated material by the curatorial agent.

Lastly, all of our recent computational creative works have undergone aspects of evaluation [1, 21]; both systems described will be similarly evaluated.

Acknowledgments

This research was funded by a grant from the Canada Council for the Arts and the Natural Sciences and Engineering Research Council of Canada.

7. REFERENCES

[1] A. Eigenfeldt, P. Pasquier, “Negotiated Content: Generative Soundscape Composition by Autonomous Musical Agents in Coming Together: Freesound,” in Proceedings of the Second International Conference on Computational Creativity, Mexico City, 2011, pp. 27-32.

[2] A. Eigenfeldt, P. Pasquier, “A Sonic Eco-system of Self-Organising Musical Agents,” in EvoApplications 2011, Part II, C. Di Chio et al. (Eds.), LNCS 6625, pp. 283-292.

[3] A. Eigenfeldt, “Coming Together - Negotiated Content by Multi-agents,” in Proceedings of the ACM Multimedia 2010 International Conference, Firenze, 2010, p.1433-1436.

[4] C. Bartolini, C. Preist, N. Jennings, “A Software Framework for Automated Negotiation,” in Proceedings of SELMAS’2004. LNCS 3390, Springer-Verlag, pp.213-235.

[5] M. Krainin, B. An, V. Lesser, “An Application of Automated Negotiation to Distributed Task Allocation,” in IEEE/WIC/ACM International Conference on Intelligent Agent Technology, IEEE Computer Society Press, 2007, pp.138-145.

[6] A. Rao, M. Georgeff, “BDI-agents: From Theory to Practice,” Proceedings of the First International Conference on Multiagent Systems (ICMAS'95), San Francisco, 1995, pp. 312-319.

[7] J. McCormack, “Eden: An Evolutionary Sonic Ecosystem,” in Advances in Artificial Life, 6th European Conference, ECAL 2001, Springer, pp.133-142.

[8] E. Miranda, A. Biles, Evolutionary Computer Music. Springer-Verlag, 2007.

[9] P. Beyls, “Interaction and Self-Organization in a Society of Musical Agents,” in Proceedings of the European Conference on Artificial Life, Lisbon, 2007.

[10] J. Nechvatal, “Towards an Immersive Intelligence,” in Leonardo, 34(5), 2001, pp.417-422.

[11] D. Zicarelli, “An Extensible Real-Time Signal Processing Environment for Max,” in Proceedings of the International Computer Music Conference. San Francisco: International Computer Music Association, 1998.

[12] A. Kapur, M. Darling, J. Murphy, J. Hochenbaum, D. Diakopoulos, Trimpin, “The karmetik notomoton: A new breed of musical robot for teaching and performance,” in Proceedings of NIME 2011, Oslo, pp.228-231.

[13] A. Eigenfeldt, A. Kapur, “An Agent-based System for Robotic Musical Performance,” in Proceedings NIME, Genoa, 2008, pp.144-149.

[14] A. Eigenfeldt, “Emergent Rhythms through Multiagency in Max/MSP”, in Computer Music Modeling and Retrieval. Sense of Sounds, Lecture Notes in Computer Science, Springer Berlin, 2008, pp. 368-379.

[15] A. Eigenfeldt, “Kinetic Engine: Toward an Intelligent Improvising Instrument” in Proceedings of the Sound and Music Computing Conference, Marseille, 2006, pp. 97-100.

[16] J. Sloboda, “The Eye-Hand Span--An Approach to the Study of Sight Reading,” in Psychology of Music, Vol 2(2), Oct 1974, pp.4-10.

[17] B. Manaris, P. Roos, P. Machado, D. Krehbiel, L. Pellicoro, J. Romero, “A corpus-based hybrid approach to music analysis and composition,” in Proceedings of the 22nd Conference on Artificial Intelligence, Vancouver, 2007, pp. 839-845.

[18] P. Todd, G. Werner, “Frankensteinian Methods for Evolutionary Music Composition,” in Musical Networks—Parallel Distributed Perception and Performance, N. Griffith and P. M. Todd, eds., Cambridge, MA: MIT Press, 1999, pp. 313-339.

[19] E. Scime, “The Content Strategist as Digital Curator”,http://www.alistapart.com/articles/content-strategist-as-digital-curator/, retrieved April 1, 2012.

[20] M. Griffin, “We Built Ourselves a Ghetto: An Analysis of Curatorial Practices in Electroacoustic Music”, presented at Toronto Electroacoustic Symposium (TES 2011), http://cec.sonus.ca/events/TES/2011/tes2011_abstracts.html#griffin

[21] A. Eigenfeldt, P. Pasquier, A. Burnett, “Evaluating Musical Metacreation”, in Proceeding of the International Conference of Computation Creativity, Dublin, 2012, pp. 140-144.