

Coming Together – Negotiated Content by Multi-agents

Arne Eigenfeldt
School for the Contemporary Arts
Simon Fraser University
Vancouver, Canada
00 1 604 460-0552

arne_e@sfu.ca

ABSTRACT

In this paper, we describe a software system that generates unique musical compositions in realtime, created by four autonomous multi-agents. Given no explicit musical data, agents explore their environment, building beliefs through interactions with other agents via messaging and listening (to both audio and/or MIDI data), generating goals, and executing plans. The artistic focus of *Coming Together* is the actual process of convergence, heard during performance (each of which usually lasts about ten minutes): the movement from random individualism to converged ensemble interaction. If convergence is successful, four additional agents are instantiated that exploit the emergent harmony and rhythm through brief, but beautiful melodic gestures. Once these agents have completed their work, or if the original “explorer” agents fail to converge, the system resets itself, and the process begins again.

Categories and Subject Descriptors

H.5.5 [Sound and Music Computing]: Methodologies and techniques.

Keywords

Generative music, multi-agents, emergent art.

1. INTRODUCTION

The use of agents in multimedia applications has proven to be a useful avenue of investigation [1]. Instead of attempting to create complex top-down systems for intelligent control or creation, imbuing individual agents with the capacity to make decisions and interact with other agents to create complex, dynamic, and emergent systems provides a bottom-up method that has been successfully used in games [2], sound art [3], and interactive environments [4].

We have successfully used agents to create complex emergent rhythms within a virtual percussion ensemble [5] and a generative music performance system [6]. In these cases, however, agents were actively influenced by a performer in an effort to maintain compositional control over musical structure. In *Coming Together*, agents are completely autonomous; although working within a framework that predefines certain goals [see Section 4], the content is generated and determined by the actions and interactions of the agents themselves. However, rather than

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Conference '10, Month 1–2, 2010, City, State, Country.

Copyright 2010 ACM 1-58113-000-0/00/0010...\$10.00.

generating a finished — and fixed — artwork, the agents interact in realtime: it is the process of convergence that is the artistic focus of the software and the work itself.

An individual agent’s desire is to generate a repeating phrase that has a musically meaningful relationship to all other agent phrases. By musically meaningful, we mean harmonic, melodic, and rhythmic associations that could have been produced by a human composer (using a post-tonal language found in contemporary concert music and post-bebop jazz). Agents move towards this collective goal by listening to one another’s musical output (MIDI data or audio) as well as messaging one another, altering their own musical output based upon formed beliefs as to other agent states. The end result is a negotiated solution that demonstrates a unique melodic, harmonic, and rhythmic environment that has been arrived at collectively.

If successful convergence occurs, four additional “quark” agents are instantiated that exploit the resultant musical environment. These new agents are so named because of their similarity to those elementary particles of matter; their lives are brief (in comparison to the duration of the composition thus far), active (in comparison to the explorer agents), and can only be created and observed under very precise conditions. The composition is subtitled “Beauty and Truth”, which are the original (and far more poetic) names for Top and Bottom quarks [7]. In *Coming Together*, the quark agents have very fleeting, but beautiful lives in which they demonstrate the “truth” of the environment in which they exist. After a brief interlude by these agents, the system is reinitialized, and a new composition begins.

Since the initial state of each agent is randomly generated, and its progression and negotiation is not deterministic, the resulting music cannot be predicted, nor can a successful convergence be guaranteed. Agents monitor their own progression over time by examining specific relationships for success or failure (see Section 4): if enough failures have occurred, the agents acknowledge a global defeat, and the system resets itself, and begins again.

2. RELATED WORK

Negotiation is an active research topic in the multi-agent community [8, 9, 10], and our use of the term is slightly 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 [11], agents pursue a collective musical goal achieved through continual interaction; individual agent actions influence future decisions by other agents.

The models used here do not follow any known model, 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 [4], Eduardo Miranda [12], Peter Beyls [3], and Joseph Nechvatal [13], to name a few important ones.

3. DESCRIPTION

Coming Together was written in MaxMSP [14], and its code is available on the author’s website. Examples of the work can be found here:

- MIDI - <http://tinyurl.com/3xypqqv>;
- AUDIO - <http://tinyurl.com/27t8csq>.

Currently, *Coming Together* can operate in two states: MIDI, in which musical information (pitch, volume, duration) is sent as a symbolic representation to a virtual synthesizer, which translates this information into sound; and AUDIO, in which the resultant sound generated by the synthesizer is part of the agent’s data stream, analyzed by other agents. See Section 3.2 for a more detailed description. Agents currently exist within a single computer; however, each agent can operate autonomously within its own computer, with data exchanged over a network.

Agents begin generating their phrases by playing a single note, whose pitch and volume is randomly generated. This note is repeated, and will eventually become the first note of a longer phrase, whose initial length, speed, and rhythmic organization will be random as well. In both MIDI and AUDIO modes, this information is sent to the synthesizer and broadcast to the ensemble (see Figure 1).

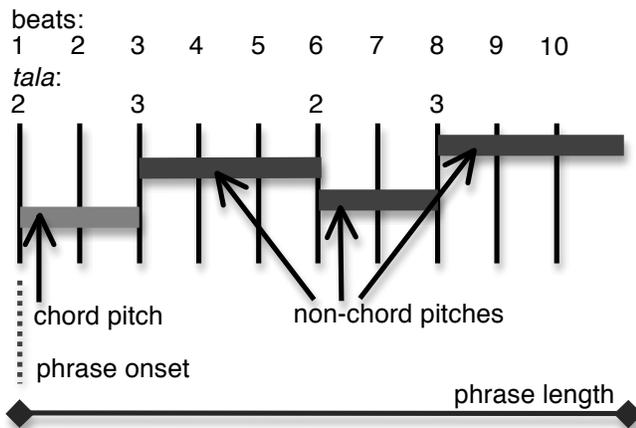


Figure 1. Parameters within an agent phrase. This phrase has a time signature of 10, with a tala of 2+3+2+3. As there are four notes in the phrase, the density is 1.0

3.1 Listening and Negotiation

Agents store other agent’s performance information. Importantly, the information transmitted and stored is performance data (i.e. what the audience hears) versus agent intentions. Agents are not privy to other agent’s goals, desires, or plans, and must form their own beliefs based upon the *results* of these internal states.

3.1.1 Harmony

The first pitch played in an agent’s phrase is considered its chord pitch. Since agents always play a note on the first beat of their phrase, and phrases will eventually align in time, these notes will

eventually form a chord, or vertical construct: the only guaranteed chord in the eventual environment. Agent’s attempt to find suitable locations within the pitch field (of five octaves) by moving their chord pitches relative to other agent pitches. This pitch goal is constrained by other agent chord pitches by an avoidance of immediate proximity (i.e. small intervals) and remoteness (i.e. very large intervals). Given the other three agent chord pitches, agents calculate a probability vector for neighboring pitches (within a distance of six steps on either side of the current pitch), and select from this probability vector using a roulette-wheel selection method. No predefined harmonic constructs are privileged (i.e. major chords or scales); instead, unique harmonies emerge that exhibit many of the tendencies of such constructs while not replicating them exactly.

Initially, agents exhibit a fair amount of pitch movement as they search for their ideal pitch goals. This movement is counterbalanced by an increase in pitch inertia as the composition progresses, which amounts to gradually weighting the current pitch higher within the probability vector. The practical result of such increased inertia is less negligible pitch movement and a gradual acceptance of emergent harmony, without which convergence is less likely to occur.

Agents continually evaluate their current pitch location in relation to potential movement within their immediate environment, and calculate a *pitchGoal* coefficient. This coefficient reaches 1.0 when the current location is weighted as ideal and no other neighboring pitches have a probability greater than zero.

3.1.2 Melody

When any agent’s *pitchGoal* reaches 1.0, a message is transmitted to the ensemble, which triggers the possibility of adding non-chord pitches. (This is an example of a heuristic decision that guarantees melodic initiation). These non-chord pitches occur within the agent’s phrase, and can be heard as either melodic elements, or elaborations of the chord pitch.

The number of unique non-chord tones an agent plays is dependent upon the location of the agent’s chord pitch: lower pitches have fewer non-chord tones associated with them. The number of actual, non-unique pitches within an agent’s phrase is determined by the overall density, a value determined through negotiation between agents (see Section 3.1.5). Actual non-chord pitches (which, obviously, determine the overall harmony of the music as well) are, like chord pitches, constrained by probabilistic relationships based upon other agent chord and non-chord pitches.

3.1.3 Phrase Length

Agents listen for the amount of time between other agent’s chord pitches — this interval is the length of an agent’s phrase. Agent’s calculate the mean of all other agent phrase lengths, and gradually move their own phrase length toward that mean by altering the tempo of their own phrase.

For example, an agent may have 8 beats within its phrase (as determined by the agent’s time signature, see Section 3.1.4), and a tempo of 60 beats per minute. This would result in a phrase length of 8000 milliseconds. If the group mean for phrase length were 6000 ms, the agent would gradually increase its tempo to 80 beats per minute. This adjustment would alter the group mean, and cause further adjustments by other agents to their own tempi.

Phrase lengths eventually converge, but phrases will not necessarily align since their start point will still differ. As such, when all phrases are of equal length (i.e. an agent’s phrase length is the same as the group mean), agents decide on a single agent

with which to align their phrase start points, then adjust their phrases start points until all phrases have aligned: this is considered onset convergence.

3.1.4 Time Signature

Once phrases are of equal length and have aligned their onsets, the concurrence of chord pitches will have aligned as well. However, individual agents most likely will have a different number of beats within their phrases, as the time signature — the number of beats within a phrase — was randomly chosen by each agent. In order for time signatures to converge, agents must determine the number of beats within every other agent's phrase. This is not as easy as simply counting events between chord pitches, since phrases are rhythmically constructed using a *tala*, or a repeating rhythmic cycle. *Tala* are created as collections of one and two beats, or two and three beats: for example, a ten beat cycle could be created as 1 2 3 4 5 6 7 8 9 10, which could also be written as 2 + 3 + 2 + 3. Chord pitches will only fall on the first beat of the *tala*; non-chord pitches will only fall on accented notes, or the first beat of a *tala* grouping (see Figure 1).

In order to determine the time signature, agents collect the entry delay differences between all the pitches — both chord and non-chord pitches — of a given agent. Knowing the agent's phrase length in milliseconds, potential time signatures are tabulated for each non-chord tone based upon the millisecond entry delays. As several time signatures might be possible (especially when given few non-chord pitches to fill out the *tala*), agents rate their certainty in a given result. When one result scores significantly higher than others, an agent accepts that result as its belief in another agent's time signature.

When an agent believes it has determined all other agent's time signatures, it sends a message to the group — a random value between 0. and 1. — that acts as a vote. When all agents have voted, the votes are tabulated, and the ensemble assumes the time signature of the agent with the highest vote.

3.1.5 Volume and Density

Agents will attempt to converge on a uniform volume — in order to achieve a group dynamic — by averaging all agent volumes. In order to avoid consistently moving towards a midpoint, agents also compare the group mean to its current minimum and maximum. If either of these extremes is more than 10% above or below the mean (a difference that can be heard as uniquely distinct from, rather than variation of, the mean), agents will converge on the most extreme of these values; otherwise, it will converge on the mean. This potential for converging upon extremes allows for dynamic volume changes over the course of a composition. Once a volume goal has been determined, agents individually move towards that goal.

Once agents have successfully converged upon a group volume, an agent flips a five-sided coin (i.e. a 20% chance) to determine whether it will “break away” from the group. If successful, the agent chooses a new random volume, then waits between 10 and 20 seconds before once again averaging group volume. This pause essentially forces the other agents to converge on the new volume, the musical result being a periodic shift in the ensemble dynamics.

A similar process governs ensemble density: the number of notes an agent plays within its phrase. An agent's actual density is constrained by the time signature (the number of beats within a phrase) and the *tala* (the organization of beats within the phrase). Thus, a phrase of 8 beats with a *tala* of 2+3+3 has only three possible notes, whereas a phrase of 18 beats with a *tala* of

2+3+2+3+2+3+3 has seven possible notes. Density, then, is the ratio of actual notes to potential notes within a phrase. Because agents cannot determine another agent's maximum number of notes, agent's broadcast their current density. As with volume, this value is averaged, and agents converge — and diverge — their densities throughout the composition.

Both processes contribute to compositional variation that balances the overall convergence of other musical parameters.

3.2 Timbre Analysis

In audio mode, timbre is an additional negotiated parameter. During performance, agents analyze other agent's audio¹ to determine frequency and timbral content. The agent uses this information to select an instrument within its current range that has a similar timbre. Because most instruments have limited ranges, this negotiation is not one of timbral matching, but of timbral similarity.

3.2.1 Non-realtime analysis

Thirty-seven sample-based instruments — consisting of World music instruments divided roughly evenly between pitched percussion, strings, and winds — were analyzed using Jehan's pitch~ external [15]. Each instrument was sampled multiple times at every note of its range to derive peak amplitude vectors consisting of the first eight partials. These vectors were averaged within octaves, and divided into five separate databases based upon MIDI note ranges: i.e. notes 36-47; 48-59; 60-71; 72-83; 84-96.

3.2.2 Timbre selection

At initialization, each agent chooses its first chord pitch (see Section 3.1.1). An array of available instrument ranges is checked, and the instrument in which the new chord pitch is most central is selected.

During performance, chord pitch movement may extend beyond the range of a given instrument: pitch movement is not constrained by instrument range. Therefore, the instrument range array is continually checked: if the chord pitch moves too close to the current instrument's boundaries, a new instrument is selected in which the chord pitch is more central to its range.

3.2.3 Realtime analysis

During performance, agents derive other agent's peak amplitude values for the first eight partials. This vector is compared, using a Euclidean distance function, to the database of available timbres in the agent's current range. The closest match is selected.

4. Monitoring Progress

4.1.1 Determining Failed Convergence

The successful progression through the composition requires certain aspects of the system to converge: for example, agents cannot synchronize their phrase beginnings until the phrase lengths have converged. Some convergences are guaranteed — phrase lengths — while others are not: determining the number of beats in other agent phrases. In such cases, the agents may be perennially waiting for certain conditions to appear, unable to progress any further until another agent achieves a goal that may never be reached. Therefore, successive individual failures by agents are tallied. These include unsynchronized onsets, uncertainty in other agent's time signature, continued chord pitch

¹ Audio is generated using Native Instrument's *Kontakt 4* sample player and the *Ethno World* sample library.

instability, and an inability to achieve continued increase in success (see Section 3.2.2). If any of these counters progress beyond a predefined point, agents admit failure, and collectively give up. The system then reinitializes itself, and the entire process begins again. A short illustrated example of collective agent failure can be seen here: <http://tinyurl.com/296u7n3>.

4.1.2 Determining Successful Convergence

At the same time, successes are also monitored, including:

- phrase lengths converging;
- time signatures converging;
- phrase start points converging;
- pitch goals;
- chord pitches stabilizing;
- non-chord tones stabilizing.

A composition is considered to have successfully converged once all agents meet the first three conditions, and the latter three conditions are met by at least two agents. Successful convergence is marked by the instantiation of quark agents.

5. Quark Agents

Quark agents can be considered the exploiters of the parameters negotiated by the initial explorer agents. Each quark agent is related to an explorer agent, and shares its specific data. When the system is judged to have converged enough so as to instantiate the quark agents, a duration for this final section is calculated that is between 5 and 10 percent of the composition length thus far. When this added duration is eventually reached, all agents - explorer and quark - simply fade out, and the entire process reinitializes itself.

Quark agents use the chord pitches from all agents and the non-chord pitches from their related explorer agent. Quark phrases are between 4 and 16 notes, sampling a melodic shape collectively chosen by the quark agents.

Volume is shaped across the phrase by a logarithmic curve, based upon the explorer agent's current volume (and thus influenced by the current group dynamic).

Rhythms are subdivisions of the ensemble pulse (which has now converged), with a roulette-wheel selection from subdivisions of 2, 3, or 4. The weighting for the subdivisions are dependent upon the actual negotiated tempo, with higher subdivisions more likely at slower tempi. The explorer agent's *tala* is emphasized by the quark agent by potentially changing subdivisions on *tala* beats.

6. CONCLUSION

Coming Together is a series of realtime compositions in which autonomous agents negotiate to achieve unique musical creations without user input or specification of data. Although each composition follows a similar pattern and generates a related result, the converged harmony, melody, and rhythm will be different for each work. The research in this paper uses musical content as its basis, but the content could easily be substituted with image, moving image, or other media. The premise of intelligent media agents generating content using negotiation allows for novel new directions in new media creation.

7. ACKNOWLEDGMENTS

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

8. REFERENCES

- [1] Whalley, I. 2009. Software Agents in Music and Sound Art Research/Creative Work: current state and a possible direction. *Organised Sound*, 14:156-167.
- [2] Binkley, T. 1998. Autonomous Creations – Birthing Intelligent Agents. *Leonardo*, 31:5 333-336.
- [3] 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.
- [4] McCormack, J. 2001. Eden: An Evolutionary Sonic Ecosystem. LNCS, *Advances In Artificial Life*, LNCS, Springer Berlin / Heidelberg, 133-142.
- [5] 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.
- [6] Eigenfeldt, A. and Pasquier, P. 2009 A Realtime Generative Music System using Autonomous Melody, Harmony, and Rhythm Agents. *12th Generative Art Conf. GA2009*, Milan.
- [7] Close, F. E. 2007. *The New Cosmic Onion: quarks and the nature of the universe*. Taylor & Francis.
- [8] Bartolini, C., Preist, C., and Jennings, N.R. 2005. A Software Framework for Automated Negotiation. *Proceedings of SELMAS'2004*. LNCS 3390, Springer-Verlag, 213-235.
- [9] Krainin, M., An, B., and Lesser, V. 2007. An Application of Automated Negotiation to Distributed Task Allocation. *IEEE/WIC/ACM International Conference on Intelligent Agent Technology*, IEEE Computer Society Press, 138-145.
- [10] Martínez, E., Kwiatkowski, I., and Pasquier, P. 2010. Towards a Model of Social Coherence in Multi-Agent Organizations. *Proceedings of the 9th Inter. Conference on Autonomous Agents and Multiagent Systems*, Toronto.
- [11] Rao, A. S. and Georgeff, M. P. 1995. BDI-agents: From Theory to Practice. *Proceedings of the First International Conference on Multiagent Systems*, San Francisco.
- [12] Miranda, E. R. and Biles, A. 2007. *Evolutionary Computer Music*, Springer-Verlag.
- [13] Nechvatal, J. 2001. Towards an Immersive Intelligence. *Leonardo*, 34(5): 417-422.
- [14] Zicarelli, D. 1998. An Extensible Real-Time Signal Processing Environment for Max. *Proceedings of the International Computer Music Conference*, Ann Arbor.
- [15] Jehan T., and Schoner, B. 2001. An Audio-Driven Perceptually Meaningful Timbre Synthesizer. *Proceedings of the International Computer Music Conference*, Havana.