Proceedings of the Second International Conference on Computational Creativity. México City, México: ICCC
Pages 27-32

Negotiated Content: Generative Soundscape Composition by Autonomous Musical Agents in *Coming Together: Freesound*

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

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

Abstract

Generative music systems have been successful in styles and genres where there are explicit rules that can be programmed into the system. Practices and procedures within soundscape composition have tended to be implicit, in which recordings are selected, combined, and processed based upon contextual relationships. We present a system – *Coming Together: Freesound* – in which four autonomous artificial agents choose sounds from a large pre-analyzed database of soundscape recordings (from freesound.org), based upon their spectral content and metadata tags. Agents analyze, in realtime, other agent's audio, and attempt to avoid dominant spectral areas of other agents by selecting sounds that do not mask other agent's spectra. Furthermore, selections from the database are constrained by metadata tags describing the sounds. Example compositions have been evaluated through subject testing, comparing them to human-composed compositions, and the results are discussed.

Introduction

Generative music systems have been successful in styles and genres where there are explicit rules that can be programmed into the system. Practices and procedures within soundscape composition have tended to be implicit, in which recordings are selected, combined, and processed based upon contextual relationships. Any generative system that attempts to create music based upon implicit rules will, therefore, require an awareness of the musical environment within which it is currently active.

Coming Together: Freesound is part of an ongoing exploration of musical metacreative systems[1] that generate music that would be considered creative if generated by a human (Whitelaw 2004). It can be considered a (generative) real-time composition system that creates soundscape compositions.

Generative Music

Generative music systems are those that create musical output that is different with each iteration. Although there is no direct requirement for such systems to be software-based (Galanter 2006) – for example, Riley's *In C* can be viewed as a generative system – the ability for algorithmic methods to control software synthesizers directly has made the sonification of generative systems much more practical.

Generative systems have varying degrees of autonomy. Fully algorithmic systems may only require the specification of parameters, and the system can then produce musical output, in, or out, of real-time (Collins 2008). Others may involve a composer interacting with the system during performance, an approach Chadabe terms interactive composing (Chadabe 1984). These latter systems have tended to be top-down, in the sense that a composer can control the system much as a conductor can control an orchestra; our approach is bottom-up, in which intelligent musical agents interact. The approach described in this paper is different from Chadabe's, in that it relies more upon intelligent decision making by the agents, rather than controlled random processes: as such, it can be seen as real-time composition.

Real-time Composition

Real-time composition (Eigenfeldt 2008) is the application of musical agents to interact in musically intelligent ways, during performance. Each agent has the potential to control an independent musical gesture — either pitch-based, or timbral — and the complexity of the interactions, along with the quantity of simultaneous gestures, cannot be controlled in any detailed way using existing performative actions. In other words, knowledge must be built into the agents on how to interact musically, and an environment created in which these agent interactions can result in artistically interesting and compositionally satisfying sonic artworks.

Real-time composition (RTC) is not improvisation, just as improvisation is not real-time composition (Lewis 2000). Although RTC has evolved from improvisatory interactive systems, the complexity desired by composers in RTC cannot be controlled through existing performative methods used in improvisational systems, nor through constrained random procedures (Eigenfeldt 2007). Imbuing

[1] http://www.metacreation.net/

multi-agents with musical knowledge and intelligence, and facilitating their interaction in real-time, allows for the creation of compositional environments during performance. As RTC systems model the composer's decisions, rather than an improvising performer, RTC is, first and foremost, a compositional medium, albeit one that is based within performance.

Soundscape Composition

Soundscape composition is a form of electroacoustic music characterized by the presence of recognizable environmental sounds and contexts, the purpose being to invoke the listener's associations, memories, and imagination related to the soundscape (Truax 2002). Four of its basic principles (after Truax) include:

- listener recognisability of the source material is maintained;
- listener's knowledge of the environmental and psychological context is invoked;
- composer's knowledge of the environmental and psychological context influences the shape of the composition at every level;
- the work enhances our understanding of the world and its influence carries over into everyday perceptual habits.

Soundscape composition tends to keep a degree of recognisability in its sounds in order to retain a listener's recognition of and associations with these sounds (Truax 2002); Successful soundscape composition plays with the listener's associations between the recordings, and the expectations arising from these associations. Truax points out that these relationships are intrinsic to the composition: montages or collages of random environmental sounds are rarely successful:

"The problem here is that the arbitrary juxtaposition of the sounds prevents any coherent sense of a real or imagined environment from occurring. In addition, the lack of apparent semantic relationship between the sounds prevents a syntax from being developed in the listener's mind, hence it is impossible to construct a narrative for the piece" (Truax 2002).

Furthermore, generative systems have also tended to be limited to symbolic representations – i.e. MIDI – as opposed to audio (e.g. Assayag 2006). A generative soundscape system must combine audio recordings in ways that rely upon an understanding of those recordings spectral components, and semantic contexts.

Previous Work

Some work in generative soundscape creation has been done using virtual environments as a model (Eckel 2001, Serafin 2004, Birchfield et al 2005, Finney 2009, Janer et al 2009). These systems generate sonic environments in real-time in response to user actions and movements through a virtual space.

Misra and Cook (2009) provide a survey for potential synthesis tools and methods that are best implemented for specific types of sound types, including complex environmental scenes and compositions. The authors provide an example of a completed synthesized "sound scene".

Freesound radio (http://radio.freesound.org/) is a web-based system that allows users to collaborate and interact to create "sample-based music creations", using the freesound.org library as a source. An Editor interface allows users to program their own simple patches, while a Player interface allows users to vote on existing patches while bookmarking sounds and tags; this influences an evolutionary algorithm the creates new patches and remixes and mutates existing ones.

The system described here generates soundscape compositions during performance, a style that normally is composed as a fixed medium. It is one system within a series of systems under the title *Coming Together* (Eigenfeldt 2010, Eigenfeldt and Pasquier 2010). Each of these systems explore the potential for autonomous agents to negotiate content within a predefined compositional framework. The goal in each case is a computationally creative system which produces music in real-time, that would be considered creative if generated by a human. *Coming Together: Freesound* is designed to generate soundscape compositions using a database derived from freesound.org.

The systems was designed based upon an autoethnographic analysis of one of the author's own methods of soundscape composition. A recording will suggest a particular context and combination with other recordings based upon its spectral content and its semantic meaning; however, no effort is made to separate listener's degree of recognition and/or relationship to the sounds—in other words, acoustic ecology models are not employed. For example, a recording of urban traffic that contains a preponderance of low frequencies may suggest a combination with high frequency squealing of truck brakes, without worrying whether listener has a familiarity with urban soundscapes.

We consider this system to be a generator of a single composition, with an infinite variety of possible realisations. As such, the composition has a formal structure that is repeated with each generation (see Predefined Formal Characteristics).

Description

Decisions on how to combine recordings are made using a selection method that combines metadata tags and pre-performance audio analysis of available sound material. A database of 227 soundscape recordings, varying in length from 15 seconds to 3 minutes, was assembled from freesound.org. Metadata tags for each file were generated by hand by the composer. Up to four metadata tags were applied per recording, ordered by recognition. One example is "voices, inside, foreign, ambience", while another is "voices, animals, outside". The order of the tags is important, in that initial tags are perceived almost immediately during listening (i.e. "I hear voices"), while subsequent tags take longer to perceive and understand (i.e. "the voices are inside...they are speaking a foreign language").

Next, each file was analysed using a 24-band Bark analysis (Zwicker and Terhardt 1980), for maximum, mean, and standard deviation of each band. The database is randomly distributed between four agents prior to performance, with each agent receiving a unique combination of recordings.

Selection by Metatag Data

During performance, agents listen to the agent-generated sonic environment, and select material from their database based upon their perception of the current context. At different points of the composition, selection methods vary: during the first of four sections, agents select material based entirely upon metadata tags; during the final two sections, agents select material based entirely upon spectral regions; during the second section, both methods are used.

When an agent selects a recording (the initial selection is random), it places the associated metadata tags into a communal blackboard; agents access the blackboard, randomly selecting up to four tags, then rate their own database based upon similarity to this target. Scores are given based upon relative position to the request: a "hit" on the first tag scores 1.0, and each subsequent hit decreases by 0.2 (see Table 1). A Gaussian selection is made from the highest rankings, so as to avoid identical selections given the same request.

	Metadata tags	Scores	Rating
file a	Voices animals outside	1.0 0.0.	1.0
file b	footsteps inside water	0.0.8 0.	0.8
file c	inside office ambience	0.8 0.0.4	1.2

Table 1. A request – voices inside foreign ambience – and three metadata tagged files, showing their scores based upon relative positions to the request, and a cumulative rating for each file.

Selection by Spectral Content

Agents generate beliefs about the spectral content of the current environment by analysing each individual agent's audio separately over five second windows, using the same 24 band Bark analysis (see Figure 1).

Figure 1. Spectral bands in agents 1-3 over separate five-second windows.

Note that this analysis will be different than the information agents use to select their recordings: beliefs are generated over discrete time windows, while the data selection is made of the recording's statistical data. Agents can thus never really assemble an accurate understanding of their continually changing environment, a compositional decision that ensures variability.

Combining the other agent's spectra, the agent generates a cumulative spectrum which represents its belief for that period in time. An inverse spectrum is calculated to determine low spectral regions, which is used to generate a request to its own database (see Figure 2).

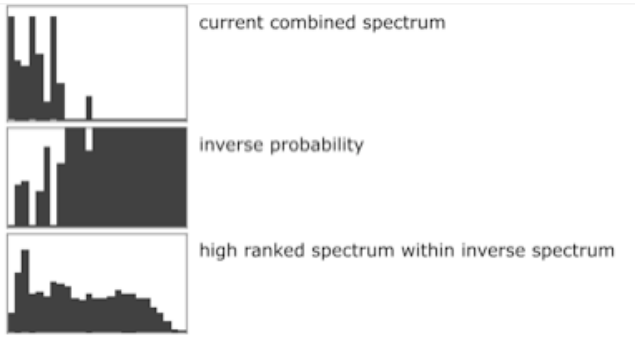

Figure 2. Generating a request using inverse spectrum, and the returned result.

In all four sections, agents attempt to converge their selections using either contextual or spectral relationships. As the composition progresses, convergence is further facilitated by lowering the bandwidth of the agent's resonant filters, projecting an artificial harmonic field upon the recordings that are derived from the spectral content of the recordings themselves. Finally, in the last section, each agent adds granulated instrumental tones at the resonant frequencies, thereby completing the 'coming together'.

Predefined Formal Characteristics

Although limited performance control exists over the environment — overall duration can be set at initialization, and agent volumes are controlled in real-time — certain aspects of the environment's evolution in time are predefined:

- the four sections define agent interactions, while the relative length of these sections within the overall duration is generated randomly at initialization;
- the global evolution of certain parameters (filter bandwidth, duration of files, delay between files) use preset tendency masks, the ranges of which were set by listening to the system and deciding the best balance between variety and guaranteed success;
- the overall increase in resonant filtering, which can be considered the defining audio feature of the composition;
- the initial selection of recordings from freesound.org.

Freesound demonstrates characteristics of a sonic ecosystem (Bown 2009) in that the environment is carefully designed, yet the interaction between its components remains nondeterministic, yet not random. Musically, *Freesound* generates surprising and varied compositions. For example, in the initial section, agents react to one another's metadata tags, and the resulting relationship between selections is clear; during the second section, an agent may select a sound based upon spectral analysis, yet the metadata tags for this recording, with potentially little contextual

relationship to the other sounds, will enter the blackboard, and influence the further selection of recordings.

Validation

In many generative music systems, success is determined solely by its creator: if the system produces output with which the designer is artistically satisfied, then the creator could claim is to be successful. Any argument as to its artistic merits could be deflected, suggesting the criticism is one of the creator's artistic sensibilities, rather than any failing of the system. However, such arguments are obviously moot in assessing the true success of computationally creative systems.

Collins discusses various methods of analysis of generative systems (Collins 2008), while Colton provides a set of criteria for assessing whether a computational system is actually creative (Colton 2008). Finally, Boden's segregation of creative systems into H– and P– creativity are also useful (Boden 2003). By these measures, *Coming Together: Freesound* is not a creative system (it is not aware of its own creativity in that it cannot adjust its behaviour based upon prior output), and it is limited to P–creative output (it will only generate soundscape compositions within predefined style; however, those will be original).

Although a composition by the system was selected for performance at an international soundscape concert (Sound and Music Computing 2010 Barcelona) – thereby seemingly validating its output at an artistic level – further validation was sought through subject testing.

Test Compositions

One composition was generated by the system, and recorded. Another composition was generated using the same parameters (database, methods of processing, overall duration) but without the contextual linking through metadata tags, nor the spectral combinations: in other words, a random selection of soundfiles from the database.

It should be pointed out that soundscape composition consists of a continuum of aesthetics, between transparent recording (sometimes called phonography) and more acousmatic, in which recordings are treated much like any other sound object recording and ripe for processing. As such, randomly selecting soundscape recordings for playback does arguably result in appropriate soundscape composition.

Two additional soundscape compositions were created by a human, a composer who has received national awards for his soundscape compositions: one composition was limited to the same parameters (database, methods of processing, overall duration, static spatial distribution of four gestures in four channels) as the system, while another was to be freely composed, restricted only by the duration and the selection of material from the same database.

At the same time, the composer was asked to create a journal of his compositional decisions. This will potentially allow a comparison between two different working methods – the commissioned composer and the system designer – and whether *Coming Together: Freesound* could be expanded to include alternative methods of creative decision-making.

The four 8 minute compositions were played in a random order to discrete test groups that consisted of 39 novice listeners (sound design students unfamiliar with the genre of soundscape composition), 8 expert (composers and graduate students of a soundscape class) and 11 semi-expert (electroacoustic composition students).

The groups were unaware that two of the compositions were machine generated, and were asked to rate each composition on a seven point scale on twelve questions, grouped into four sections.

Results

All the comparative claims made in the text have been proven statistically significant using a paired two-sided t-test. $P<0.05$, often an order or more less.

Soundscape characteristics The first four questions sought to discover how accurate a soundscape composition was produced (1 Disagree, 7 - Agree):

1. Listener recognisability of the source material is maintained;
2. Listener's knowledge of the environmental and psychological context is invoked;
3. Composer's knowledge of the environmental and psychological context influences the shape of the composition at every level;
4. The work enhances our understanding of the world and its influence carries over into everyday perceptual habits.

Question	Human-limited	Random	System	Human-free
1	6.05 (0.73)	5.41 (1.17)	4.82 (1.39)	4.44 (1.45)
2	5.53 (1.16)	4.54 (1.48)	4.95 (1.32)	5.1 (1.21)
3	5.27 (1.04)	4.69 (1.39)	4.95 (1.23)	5.26 (1.02)
4	4.32 (1.55)	3.97 (1.32)	4.31 (1.58)	4.26 (1.63)

Table 2. Experimental results for novice listeners. Mean levels, with standard deviation in parentheses, for success within the genre of soundscape composition.

Question	Human-limited	Random	System	Human-free
1	5.52 (1.5)	6.1 (0.7)	5.62 (1.02)	3.81 (1.47)
2	5.67 (0.86)	4.67 (1.32)	4.67 (1.56)	5.1 (1.26)
3	5.76 (1.09)	4.33 (1.65)	4.62 (1.4)	5.43 (1.12)
4	4.95 (1.47)	4.14 (1.56)	4.4 (1.57)	4.43 (1.6)

Table 3. Experimental results for expert and semi-expert listeners. Mean levels, with standard deviation in parentheses, for success within the genre of soundscape composition.

In almost all cases, both groups of listeners rated the system as a better generator of soundscape composition than random. The expert listeners could distinguish that the freely composed human composition was slightly more acousmatic, while the random composition used the least amount of processing, and thus remained truest to the first goal of soundscape - recognisability of source (question 1).

Compositional success The next five questions rated the success of each work on a comparative scale between two descriptors:

5. Boring - Interesting;
6. Predictable - Surprising;
7. Mechanical - Organic;
8. Sterile - Emotional;
9. Uncommunicative - Communicative.

Question	Human-limited	Random	System	Human-free
5	4.42 (1.37)	3.64 (1.38)	4.62 (1.71)	4.9 (1.76)
6	3.73 (1.17)	4.19 (1.29)	4.56 (1.07)	5.08 (1.55)
7	4.86 (1.18)	4.51 (1.39)	5.03 (1.16)	3.9 (1.71)
8	3.92 (1.3)	3.46 (1.46)	4.82 (1.45)	4.49 (1.37)
9	4.45 (1.11)	3.62 (1.44)	4.49 (1.3)	4.42 (1.45)

Table 4. Experimental results for novice listeners for compositional success.

The system was rated higher by novice listeners than the randomly generated work in every case, and was even considered better than the human-composed limited work in terms of interest, and surprise. Furthermore, it was considered the most organic, the most emotional, and the most communicative of all four works.

Question	Human-limited	Random	System	Human-free
5	5.95 (0.89)	4.38 (1.6)	4.48 (1.72)	6.1 (1.04)
6	5.43 (1.12)	4.67 (1.43)	5.14 (1.01)	6.14 (0.96)
7	4.95 (1.32)	5.14 (1.24)	4.62 (1.24)	4.62 (1.72)
8	5.62 (1.12)	3.48 (1.57)	4.81 (1.25)	5.71 (0.9)
9	5.55 (1.1)	4.05 (1.66)	4.65 (1.63)	5.71 (0.9)

Table 5. Experimental results for expert and semi-expert listeners for compositional success.

Expert listeners judged the system to be better than random in every instance except mechanical vs. organic; however, the system was judged similar to the freely composed human work in that aspect.

Skill level The next two questions assessed the skill level of the composer, on a comparative scale between two descriptors:

10. Student-like - Professional
11. Poor craftsmanship - high craftsmanship

Question	Human - limited	Random	System	Human-free
10	5 (1.19)	4.05 (1.25)	4.69 (1.22)	5.26 (1.18)
11	5.32 (1.02)	4.57 (1.07)	4.86 (1.13)	5.38 (1.26)

Table 6. Experimental results for novice listeners for skill level.

Questions	Human-limited	Random	System	Human-free
10	5.76 (0.89)	3.86 (1.74)	4.14 (1.59)	6.14 (1.01)
11	6.05 (0.74)	4.25 (1.33)	4.52 (1.44)	6.29 (0.96)

Table 7. Experimental results for expert and semi-expert listeners for skill level.

Here, both sets of listeners were able to discern the human-composed from the machine-composed music. Although the expert listeners rated the system less successful than the novice listeners, they also rated the random composition much lower. In all instances, the system was considered more skillful than randomly assembled soundscapes.

Subjective Reaction Finally, the last question asked whether the listener disliked or liked the composition, on a comparative scale between "Did not like it" and "Liked it a lot".

12. My feelings towards this soundscape composition.

Question	Human - limited	Random	System	Human-free
12	4.42 (1.27)	3.57 (1.41)	4.36 (1.51)	4.92 (1.44)

Table 8. Experimental results for novice listeners for listener subjective reaction.

Question	Human-limited	Random	System	Human-free
12	5.76 (0.94)	3.67 (1.53)	4.05 (1.69)	5.95 (0.8)

Table 9. Experimental results for expert and semi-expert listeners for listener subjective reaction.

Again, both sets of listeners preferred human-generated soundscape composition to machine-generated. Interestingly, the variation in responses was higher to the machine-generated works than the human-composed works, and the spread of these differences is higher for the expert listeners than for the novice.

Qualitative Results

Respondents were allowed to add any further comments on each of the works. One expert listener admitted to having a difficult time distinguishing between the success of the system piece and the limited human piece, only slightly preferring the latter for the sole reason that the signal processing was more closely correlated to the material itself – something that would be extremely difficult to automate.

Conclusions and Future Work

Listeners did prefer human-composed soundscape compositions to machine-generated. Interestingly, the freely composed human work was consistently rated higher than the piece that imposed the same restrictions in which the system operated: the type of processing, and the limited spatial distribution. This suggests that the compositional decisions that define *Coming Together: Freesound* may, in fact, be limiting its artistic success.

One aspect that differentiated both machine-generated compositions from the human-composed was the static nature of the overall amplitude envelope. This is a very high-level parameter that would require subtle changes in volume based not only upon the overall density and amplitude, but the recent past. This action is actually managed by the composer during performance, carefully balancing

levels, and, for example, bringing down levels of more static recordings in favour of more dynamic ones. Creating such intelligent, autonomous high-level actions is currently being investigated, with the potential for a high-level "listener" agent analysing the cumulative result, and communicating its suggestions to the four generative agents.

The research instrument discussed here is a contribution in itself. As this system is a musical metacreation, validation and evaluation of such a system's output is itself a challenging research area. Our future will investigate and try to evaluate the methodologies to do so. One particularly challenging aspect is that the system is capable of generating numerous pieces, with possibly varying levels of success. Designing methodologies to measure that variability is an inherent challenge of the area.

Acknowledgements

This research was funded by a grant from the Canada Council for the Arts, and the Natural Sciences and Engineering Research Council of Canada. The authors would also like to thank Barry Truax for his suggestions, James O'Callaghan for his compositions, and Alireza Dovoodi for his data analysis.

References

Assayag, G., Bloch, G., Chemellier, M., Cont, A., Dubnov, S. 2006. OMax Brothers: a Dynamic Topology of Agents for Improvisation Learning. Workshop on Audio and Music Computing for Multimedia. ACM Multimedia 2006. Santa Barbara, 125-132.

Birchfield, D., Mattar, N., Sudaram, H. 2005. Design of a Generative Model for Soundscape Creation. Proceedings of the International Computer Music Conference. Barcelona.

Boden, M. 2003. The Creative Mind: Myths and Mechanisms (second edition). Routledge.

Bown, O. 2009. A Framework for Eco-System-Based Generative Music. Proceedings of Sound and Music Computing 2009. Porto, 195-200.

Chadabe, J. 1984. Interactive Composing. Computer Music Journal 8(1): 22-27.

Collins, N. 2008. The Analysis of Generative Music Programs. Organised Sound 13:237–248. Cambridge University Press.

Colton, S. 2008. Creativity versus the Perception of Creativity in Computational Systems. Proceedings of the AAAI Spring Symposium on Creative Systems 2008. Stanford.

Cope, D. 1996. Experiments in Musical Intelligence. Middleton: A-R Editions.

Eckel, G. 2001. Immersive Audio-Augmented Environments: The LISTEN Project. Proceedings of the 5th International Conference on Information Visualisation. 571–573.

Eigenfeldt, A. 2007. Real-time Composition or Computer Improvisation? A composer's search for intelligent tools in interactive computer music. Proceedings of the Electronic Music Studies 2007. http://www.ems-network.org/IMG/pdf_EigenfeldtEMS07.pdf

Eigenfeldt, A. 2008. Intelligent Real-time Composition. eContact 10.4 Live-electronics-Improvisation-Interactivity in Electroacoustics, http://cec.concordia.ca/econtact/10_4/.

Eigenfeldt, A. 2010. Coming Together - Composition by Negotiation. ACM Multimedia 2010. Firenze.

Eigenfeldt, A., Pasquier, P. 2010. A Sonic Eco-System of Self-Organising Musical Agents. Proceedings of EvoApplications 2011. Turin.

Finney, N. 2009. Autonomous Generation of Soundscapes Using Unstructured Sound Databases. Master's thesis, Universitat Pompeu Fabra.

Galanter, P. 2006. Generative Art and Rules-Based Art. vagueterrain03. http://vagueterrain.net

Janer, J., Finney, N., Roma, G., Kersten, S., Serra, X. 2009. Supporting soundscape design in virtual environments with content-based audio retrieval. Journal of Virtual Worlds Research, 2(3).

Lewis, G. 2000. Too Many Notes: Computers, Complexity and Culture in Voyager. Leonardo Music Journal 10: 33-9.

Misra, A., Cook, P. 2009. Toward Synthesized Environment: A Survey of Analysis and Synthesis Methods for Sound Designers and Composers. Proceedings of the International Computer Music Conference 2009. Montreal, 155–162.

Serafin, S., Serafin, G. 2004. Sound design to enhance presence in photorealistic virtual reality. Proceedings of the 2004 International Conference on Auditory Display, Sydney, 6–9.

Truax, B. 2002. Genres and techniques of soundscape composition as developed at Simon Fraser University. Organised Sound 7(2): 5-14.

Whitelaw, M. 2004. Metacreation. Art and Artificial Life. Cambridge, MA: MIT Press.

Zwicker, E., Terhardt, E. 1980. Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. Journal of the Acoustical Society of America 68(5): 1523-5.