Toward Certain Sonic Properties of an Audio Feedback System by Evolutionary Control of Second-Order Structures

Seunghun Kim[1], Juhan Nam[1] and Graham Wakefield[2]

[1]Graduate School of Culture Technology,
Korea Advanced Institute of Science and Technology (KAIST),
Dahakro 291, Yuseong-Gu, Daejeon, Republic of Korea
{seunghun.kim,juhannam}@kaist.ac.kr
[2]Digital Media, Visual Art & Art History,
York University,
4700 Keele St., Toronto, Canada
grrrwaaa@yorku.ca

Abstract. Aiming for high-level intentional control of audio feedback, though microphones, loudspeakers and digital signal processing, we present a system adapting toward chosen sonic features. Users control the system by selecting and changing feature objectives in real-time. The system has a second-order structure in which the internal signal processing algorithms are developed according to an evolutionary process. Genotypes develop into signal-processing algorithms, and fitness is measured by analysis of the incoming audio feedback. A prototype is evaluated experimentally to measure changes of audio feedback depending on the chosen target conditions. By enhancing interactivity of an audio feedback through the intentional control, we expect that feedback systems could be utilized more effectively in the fields of musical interaction, finding balance between nonlinearity and interactivity.

Keywords: Audio Feedback, Evolutionary Algorithm

1 Introduction

We use the term audio feedback to refer to systems of positive acoustic/digital feedback in which signals received by one or more microphones are amplified and played through one or more loudspeakers with sufficient energy gain to create a persistent loop. Such systems support unique features, such as nonlinearity, emergence, self-organization and openness to the environment, and efforts to leverage these properties in music composition and sound art are numerous [7, 12–14, 19].

Many prior work with audio feedback systems have been characterized by an emphasis upon the emergent interactions the medium supports, while intentional control of sonic behaviours is given less emphasis. Where performer interventions are supported these are generally limited to deliberate sounds and parameter changes, however due to the nonlinear dynamics of the medium the overall directions of tendencies remain unknown and thus unintentional. Our research aims to support intentional control through tendencies while preserving the attractive nonlinear characteristics of audio feedback, such as unpredictable yet rich transients, in order to open up new possibilities of musical application.

In this paper, we present a system adapting toward a certain chosen sonic properties such as average pitch, vibrato, tremolo, brightness, and spectral tonality: recognizable features of the feedback sound. The system allows users to select in real-time sonic characteristics as the target conditions of the system. Adaptation toward such conditions is achieved using a second-order structure which can organize and replace internal signal processing algorithms. This second-order structure uses follows an evolutionary process by selecting and changing the types and parameters of signal processing components. We evaluate a prototype through measurement of average and highest fitness curves over generations in several contexts.

2 Motivation

Although audio feedback is typically viewed as a problem to avoid in music and telecommunication industries, some performers have deliberately utilized feedback creatively. For example, *LIES(topology)* [19] is an improvisation performance based on the interaction between a feedback system and performers. The system is centred on a feedback delay network (FDN), in which several delay lines connected by a feedback matrix mediate the input and the output. Some feedback loops apply signal processing components, such as a ring modulator, frequency shifter, granulator, wave-shaper and reverberator. These form a complex network in which objects can be connected by a mixer or a ring modulator. Performers alter the topology by controlling amplitude changes of the recirculating signals and change the relations between the components by modifying parameters of the components.

Audible Eco-systemic Interface (AESI) [7] is a compositional work that interacts with its acoustic environment through sonic feedback, depending on ambient noise as its information source. The central idea is a self-feeding feedback loop, almost identical to the basic feedback structure. Features extracted from the received sound are compared with the original signal, and the difference is used to control parameters for sound synthesis, thereby adapting the system toward the room resonance.

Di Scipio notes how non-linear feedback systems may lead to new, emergent high-level behaviors, generated and maintained by a network of low-level components [8]. An implication is that specific performances cannot be formally defined or accurately predicted in advance. Instead, he directs attention to the technical conditions and sonic interactions of the system. As identities of feedback-based music systems are determined by the relations between low-level components, this results in fixed mappings from analysis parameters of the received signal to signal-processing parameters: the composer's role is to establish the mappings.

However, Kollias [14] criticizes that by doing so composer loses control over the overall sonic shape, as the system only determines microstructural sonic design. Although high-level behaviors emerge through a bottom-up organization, control of them is a distinct issue. By referring to Mitchel's work [16], he instead suggests that adaptive systems must preserve a balance between bottom-up and top-down processes. He proposed *Ephemeron* [14] [15], a feedback-based improvisation system, which is presented as a metaphorically living organism and consists of cells which are sonic units. This system is acoustically adaptive as the cells recognize environmental characteristics in an evolutionary process, yet also features both high- and low-level controls for producing music. Specifically, a composer proposes the emergence of certain sonic properties by design at a micro-structural level, while a performer controls the overall sonic result by modulating global parameters. This prevents the system's tendency toward a stable state. Nevertheless, direction of the tendency is unknown because the unpredictability of nonlinear dynamics hinders prediction.

Our motivation is to more deeply explore intentional control toward specified sonic properties without sacrificing attractive nonlinearities of the audio feedback itself. Intentional control means that people can observe desired tendencies in the feedback sound by setting and changing goal directions in real-time, enhancing interactivity through regulative processes. This would support idiosyncratic interactive applications that combine context-specificity, nonlinearity and interactivity; such as a sound installation sensitive to the acoustic environments according to audience's intentions, a generative improvisation system responding to its environment, or a sound generator that repeatedly generates new sound materials according to user-specified conditions.

3 System Design

Our system is built around three design ideas, outlined in the following subsections:

- Goal-directedness: sonic features are specified as target conditions, which can be controlled by users in real-time
- Second-order feedback structure: signal processing uses a modular approach to support diversity and dynamism
- Evolutionary process: the controller uses an evolutionary process for design of the second-order feedback structure

Figure 1 shows an overview of the system. The controller, which designs the second-order feedback structure, uses an evolutionary process. This removes the necessity of manually searching for the optimal structure for a target condition. We do not need to care about the specific design, only the final characteristics: we provide the goal, and the system evaluates several structures to achieve it.

3.1 Goal-directedness toward a specific sonic condition

A control method in cybernetic systems is presenting a goal and designing the system to follow it: it is referred to as goal-directed behavior [10]. An autonomous

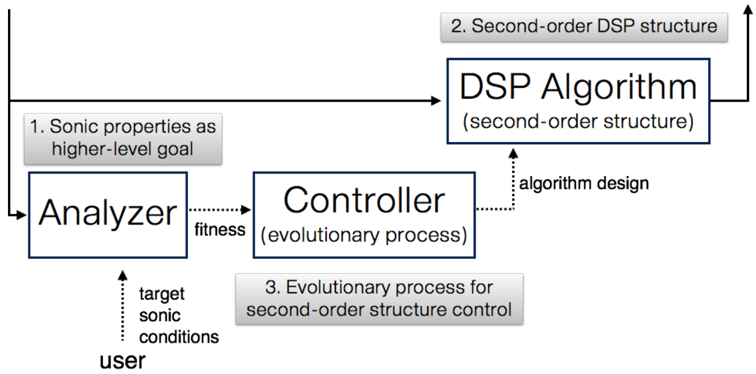


Fig. 1: Conceptual diagram of our system, which proposes 1) specific sonic features as target conditions, 2) second-order feedback structure and 3) evolutionary process for design of the second-order structure. Solid lines represent audio-rate signal flows while dotted lines sub-audio rate (data) signal flows.

system pursues its own purpose by trying to resist obstructions from the environment that may adversely affect getting close to the target state: goal-directedness regulates perturbation. By setting specific sonic features as its own purpose, an audio feedback system can adapt toward them and this implies the possibility of exterior/guided intentional control. For our purposes, this requires real-time feature extraction of the feedback sound to measure present state and its deviation from a target state.

We investigated feature extraction methods from several improvisation systems in which a single or multiple agents interact with external performers. For example, Murray-Rust et al. [17] and Wulhorst et al. [22] present artificial intelligence based music compositions, which use real-time acoustic feature extraction for the purpose of transmitting information to internal agents. These mostly focused on rhythmic and harmonic information from MIDI signals. Van Nort et al. [21] recognizes sonic gestures simultaneously by parallel operation of gestural spaces in different time scales, such as variety, brightness and pitch in short gestures and phrases in long gestures. Hsu [11] measures loudness, tempo and timbre from pitch/amplitude envelope and auditory roughness (interference between partials in a complex tone) of saxophone sounds. Ciufo [4] uses control methods based on both high-level and low-level audio analysis.

A difference between improvisation systems and audio feedback systems is that the former typically receive an external sound, such as instrument or voice, while the latter receive acoustic reflections and diffusions of their own sound. Since feedback sound is essentially different from instrumental sounds, note-based musical analysis is less comprehensive and candidates for target sonic properties must be broadened to encompass timbral characteristics. We selected the following sonic characteristics and corresponding measurement methods:

– Average pitch: average fundamental frequency

– Vibrato: standard deviation of the fundamental frequency curve
– Tremolo: standard deviation of the amplitude curve
– Spectral tonality (distinction between tone-like and noise-like signal): spectral flatness
– Brightness: spectral centroid.

Fundamental frequency is measured using a YIN algorithm [6], a popular pitch detection method based on a cumulative mean normalized difference function. Amplitude curve is then measured by local maximum points, positive zero deviation points of a waveform.

3.2 Second-order feedback structure

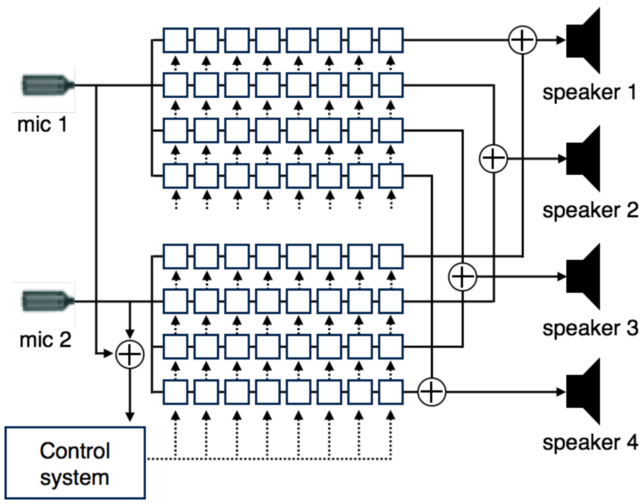


Fig. 2: The second-order structure connecting two microphones and four loudspeakers via each line consisting of eight DSP components organized by a controller

The system must search for digital signal processing algorithms having the capacity to achieve target sonic properties within unknown and possibly everchanging environmental conditions. We suspect that no single, compact signal processing algorithm will satisfy this criteria, however a modular approach can be an effective alternative. Defined as a mechanism in which a live algorithm is not fixed and can be replaced by another, the modular approach facilitates creativity by providing the basis for combinatoric search through a vast space of possibilities [2].

In the self-organized music model [1] the interpretation of the audio input (denoted P) continually provides information (denoted F) about the acoustic environment. Neural network and decision tree systems were used to derive F, and internal processing activities then provide behavioral options that can be instantly substituted for each other, using a modular mechanism [2]. Neuman [18] allows composers to generate musical structures in real-time by stochastic rewriting rules. Ciufo [4] presents an improvisation system pursuing flexibility through a modular matrix mixing technique. All signal processing modules such as ring modulation (RM), resonant filtering, and memory are connected to a two-dimensional signal matrix, enabling links between any input and any output.

We also applied a modular approach in our system to support switching of internal structure, types, and parameters of signal processing components, in order to grant greater dynamics and variety in shaping the feedback. As Figure 2 shows, the system connects microphones and loudspeakers via lines consisting of several signal-processing components, organized by a controller. Currently ten component types are possible, each with specific parameter ranges as follows:

– Lowpass Filter: 600∼1200Hz (cutoff frequency)
– Bandstop Filter: 600∼900 and 3600∼5400Hz (lower/upper cutoff frequencies)
– Bandpass Filter: 30∼200 and 430∼600Hz (lower/upper cutoff frequencies)
– Amplifier: 3∼100 (amplification degree)
– Frequency Shifter: -300∼300Hz
– Delay Line: 3∼1000 samples
– Sinewave Generator: 100∼600Hz
– Feedback: 0.5∼3 (amplification degree)
– Feedforward: 0.5∼3 (amplification degree)
– Bypass (no operation).

Each component also includes an amplifier for gain control, which is smoothly ramped through zero to avoid clicks when a DSP structure is switched.

3.3 Evolutionary process for second-order structure control

Our system uses an evolutionary process to design the second-order feedback structure. Evolutionary algorithms are a well-established method to explore huge parameter spaces, including tasks in music composition and sound synthesis [5]. In [2] and [3] genotypes correspond to continuous-time recurrent neural networks (CTRNNs), which improvise through interactions with live performers. A CTRNN is a network of simple artificial neurons in which each neuron interconnected via weighted synapses and processes a floating-point value or maintains a state. The fitness of each genotype is measured by the absolute difference between corresponding pairs of values in input and output sequences; repeated sequences receive zero score. Individuals of high fitness values in a population are chosen to generate individuals in the next population, with mutation. In [3] each network controls parameters for sound playing, such as parameters for a FM synthesizer and playback position for a granular sample player. Fastbreeder [9] is a genetic programming synthesizer. It grows code for sound synthesis by choosing from automatically generated functions. Syntaxis [20] is an example using a genetic algorithm in an invariable structure of a feedback system: individuals in a population correspond to bandpass filters and the fitness is measured by their deviations from resonant frequencies of the feedback sound. Filter banks thereby gradually evolve to fit the resonant peaks. Use of such genetic algorithms produces musically interesting results in which target behaviors and other behaviors not specified by the goals coexist.

Our genotypes consist of pairs of genes to specify the types and parameters of signal-processing components. The manifestation of these genotypes as phenotypes forms a signal-processing algorithm. Eight genotypes in a present generation are manifested, and the sound generated as a result of each genotype re-enters the system via microphones after being diffused and reflected in a room. The controller evaluates the fitness of each genotype through a function measuring deviation of features of the incoming sound from the currently chosen target conditions. Individuals in subsequent generations are determined according to the fitness criteria of previous generations, with small mutations, so that the feedback sound gradually evolves toward the objective.

Reproduction incorporates the possibility of mutations in the parameters or type of a component. Based on informal experimentation we began with default mutation rates of 0.08 for the parameter change and 0.05 for the component change in each gene. Accordingly, the parameter change happens rather frequently, but this responds to a necessary condition of the task. Our system is subject to real-time/real-space bottleneck: the population size is limited because the evaluation of each individual must take place in real-space over a reasonable duration. It is analogous to the *fitness bottleneck* familiar to aesthetic selection and interactive evolution [**?**, **?**], yet different in that it is not the human that is a limiting evaluation factor. Nevertheless, an interactive installation for an audience may need to privilege fast adaptation over stability and accuracy, suggesting the use of higher mutation rates.

4 Results

We implemented this design as software authored using openFrameworks for the analyzer and controller (Figure 3) and Max/MSP for the second-order DSP structure (Figure 4). In an installation, four ESI nEar05 monitor speakers were installed at each corner of a room, and SM57 and SM58 microphones were placed in the middle facing toward the floor to avoid direct sound paths.

This installation was evaluated to show how sonic behavior of the system would change according to the target conditions. For example, Figure 5 presents flatness states (spectral centroid values) when the minimum or maximum spectral flatness was selected as the target condition, which respectively drives the sound to a pure tone or a white noise. Dash-dotted lines represent the target conditions: spectral flatness is measured as 0.8 and 0.1 on average when the system plays a white noise and a pure tone instead of the feedback sound. Both cases start at below 0.1, but we can observe that it increases to almost 0.8 with some fluctuations when targeting noise sound (high flatness). Similarly, Table 1

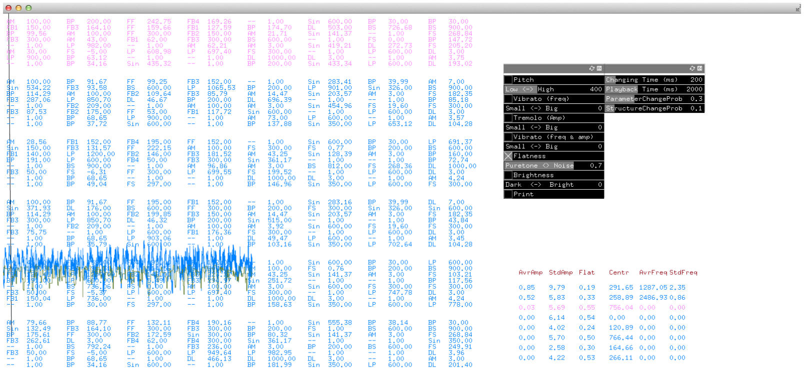
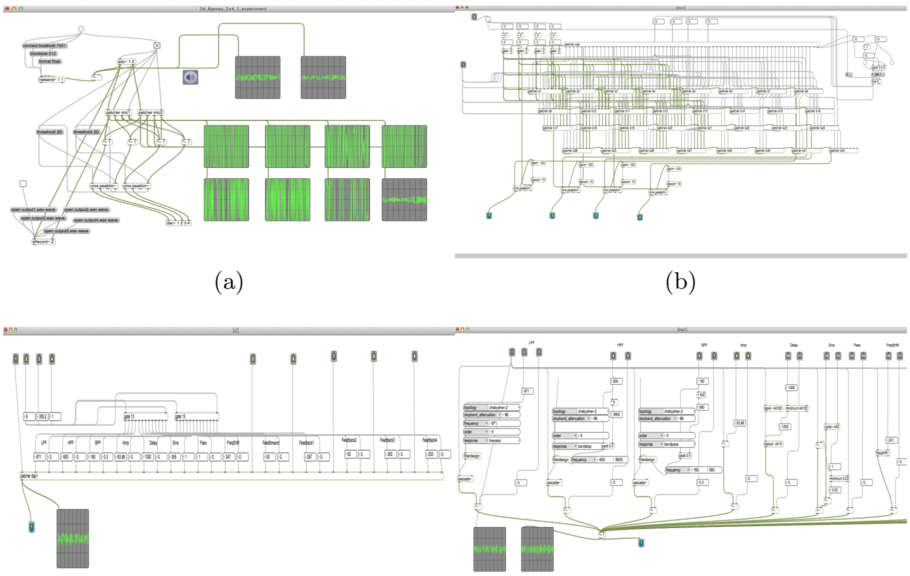


Fig. 3: The analyzer and the controller, authored with openFrameworks. Each 8 by 8 block represents a genotype of the present generation. Reading across the row gives the path from one of the microphones to one of the loudspeakers; with two microphones and four loudspeakers this makes 8 rows. The current best candidate is shown in pink at the top. The user can select target conditions and control system parameters using the black panels at the right-hand side. Numerical values resulting from the evaluations are given in the bottom-right corner.

Table 1: Comparisons of the sonic features of the feedback sounds in the first population and 15th population

		Initial population		15th population	
	Target Condition	Mean	Optimum	Mean	Optimum
Average fundamental frequency (average pitch)	1000 (Hz)	167	472	528	845
	100 (Hz)	279	128	121	100
SD of fundamental frequency (vibrato)	1.2	0.25	0.70	0.91	1.09
	0	0.34	0.08	0.12	0.07
SD of local maximum amplitude (tremolo)	40	5.1	17.8	24.4	30
	0	17.8	8.3	10	7.3
Spectral flatness (spectral tonality)	0.8	0.03	0.27	0.64	0.68
	0.1	0.02	0.08	0.07	0.1
Spectral centroid (brightness)	1300 (Hz)	10	351	769	1030
	0 (Hz)	281	129	116	75



(a) (b)

(c) (d)

Fig. 4: (a) Main patch for the second-order signal processing structure, authored with Max/MSP, and (b) subpatches for operating branch lines between the microphones and the loudspeakers, (c) designating each DSP component and parameter from every candidate based on genetic information and (d) operating the DSP algorithms and linking the feedback loop.

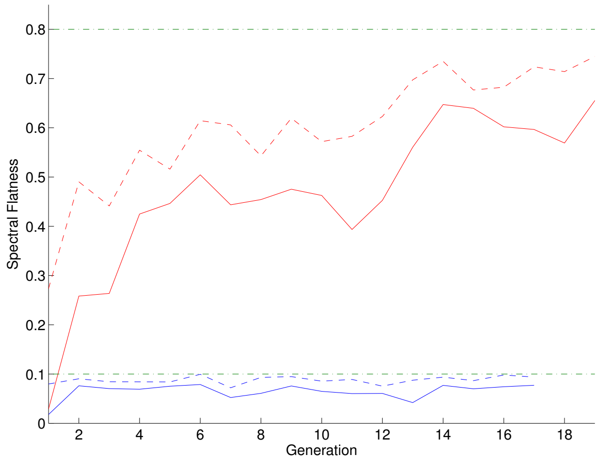


Fig. 5: Average (solid lines) and optimal flatness values (dotted lines) of individuals in each generation when the maximum or minimum flatness is selected as a target condition, which is represented as a green dash-dotted line. Red and blue lines are measured when the target state is set as maximum and minimum value, and target states are described as dashed lines.

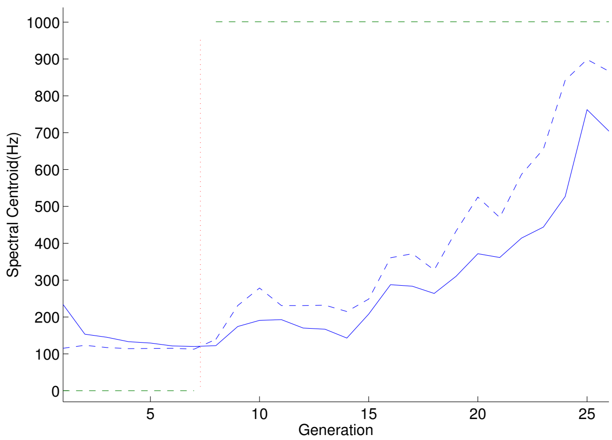


Fig. 6: Average (solid lines) and optimal centroid values (dotted lines) of individuals when a target condition is selected as minimum brightness until the 7th generation, and changed to high brightness as 1000Hz spectral centroid.

compares the average and optimal values of individuals at the first and 15th population (after approximately 4 minutes), when one of the five sonic properties is chosen to be a target condition. Even though variations of the features are limited to certain ranges, we could find that the features of the feedback sound could be driven to a certain degree by these target conditions with the evolutionary process to design the second-order structure. These sounds samples are available at **http://sites.google.com/site/asuramk88/research/feedbackevocontrol**.

Figure 6 presents the sonic behaviors when the target condition changes during performance. We could observe adaptation of sonic behavior toward the current target condition as the sound tends to be brighter since the 15th generation, which is about eight generations after changing the target condition to high brightness (7th generation). One might expect the possibility of an installation in which the feedback sound is controlled by the audience who can set and change the target conditions in real-time.

5 Conclusion

This work presents a system for high-level intentional control of audio feedback, which uses a second-order structure in which the internal signal processing algorithms are developed according to an evolutionary process. Users control the system by selecting and changing feature objectives in real-time. A prototype was implemented and evaluated to observe changes of sonic behaviors depending on the target conditions. The results show the possibility of intentional control, which could result in enhancement of interactivity in the overall sonic behaviors of feedback systems and lead to sound/music applications which feature a balance between nonlinear emergence from the relations between low-level components and regulation over the overall sonic shape.

The target features are currently closely-tied direct audio analysis features, however in a broader motivation we hope to extend our system to support higher-level target features of sound streams that may have more readily musical application, such as stability, continuity, tension, and contrast. We noted that varying the mutation rates according to purpose follows a trade-off of the real-time/real-space bottleneck, such as increasing rates for interactive installation or decreasing for sound library generation. However we are also keen to investigate whether varying rates at per-gene or per-operator level, possibly also adaptively, may improve adaptation. Finally, analysis of the timbre space occupied by a second-order signal processing structure may also enhance intentional control.

Acknowledgments. This work was supported by the BK21 plus program through the National Research Foundation (NRF) funded by the Ministry of Education of Korea.

References

1. Tim Blackwell and Michael Young. Self-organised music. *Organised Sound*, 9(02):123–136, October 2004.

2. Oliver Bown. Experiments in Modular Design for the Creative Composition of Live Algorithms. *Computer Music Journal*, 35(3):73–85, September 2011.
3. Oliver Bown and Sebastian Lexer. Continuous-time recurrent neural networks for generative and interactive musical performance. In *Applications of Evolutionary Computing*, pages 652–663. Springer, 2006.
4. Thomas Ciufo. BEGINNER'S MIND : AN ENVIRONMENT FOR SONIC IMPROVISATION. In *Proceedings of the International Computer Music ...*, 2005.
5. Palle Dahlstedt. A mutasynth in parameter space: interactive composition through evolution. *Organised Sound*, 6(02):121–124, 2001.
6. Alain De Cheveigné and Hideki Kawahara. Yin, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, 111(4):1917–1930, 2002.
7. Agostino Di Scipio. 'Sound is the interface': from interactive to ecosystemic signal processing. *Organised Sound*, 8(03), April 2004.
8. Agostino Di Scipio. Listening to Yourself through the Otherself: On Background Noise Study and other works. *Organised Sound*, 16(02):97–108, June 2011.
9. David Griffiths. Fastbreeder.
10. Francis Heylighen and Cliff Joslyn. Cybernetics and second order cybernetics. *Encyclopedia of physical science & technology*, 4:155–170, 2001.
11. William Hsu. Strategies for Managing Timbre and Interaction in Automatic Improvisation Systems. *Leonardo Music Journal*, 20:33–39, December 2010.
12. Seunghun Kim and Woon Seung Yeo. Musical control of a pipe based on acoustic resonance. In *Proceedings of the 2011 conference on New interfaces for musical expression, Oslo, Norway*, pages 217–219. Citeseer, 2011.
13. Seunghun Kim and Woon Seung Yeo. Electronic pipe instrument using audio feedback. In *Proceedings of the Sound and Music Computing*, 2012.
14. Phivos-angelos Kollias. Ephemeron : Control over Self-Organised Music. 14(December):138–146, 2009.
15. Phivos-Angelos Kollias. The Self-Organising Work of Music. *Organised Sound*, 16(02):192–199, June 2011.
16. Melanie Mitchell. Complex systems: Network thinking. *Artificial Intelligence*, 170(18):1194–1212, 2006.
17. David Murray-Rust, Alan Smaill, and Manuel Contreras Maya. Virtualatin-towards a musical multi-agent system. In *Computational Intelligence and Multimedia Applications, 2005. Sixth International Conference on*, pages 17–22. IEEE, 2005.
18. Israel Neuman. Generative tools for interactive composition: Real-time musical structures based on schaeffer's tartyp and on klumpenhouwer networks. *Computer Music Journal*, 38(2):63–77, 2014.
19. D Sanfilippo. Turning Perturbation Into Emergent Sound, and Sound into Perturbation. *interferencejournal.com*, pages 1–16, 2013.
20. Massimo Scamarcio. Space as an Evolution Strategy. Sketch of a Generative Ecosystemic Structure of Sound. In *Proceedings of the Sound and Music Computing ...*, 2008.
21. Doug Van Nort, Jonas Braasch, and Pauline Oliveros. A system for musical improvisation combining sonic gesture recognition and genetic algorithms. In *Proceedings of the 6th Sound and Music Computing Conference (SMC), Porto, Portugal*, pages 131–6. Éditions du Seuil, 2009.
22. Rodolfo Daniel Wulfhorst, Lauro Nakayama, and Rosa Maria Vicari. A multiagent approach for musical interactive systems. In *Proceedings of the second international*

joint conference on Autonomous agents and multiagent systems, pages 584–591. ACM, 2003.