Daniel Bisig
Philippe Kocher

Early Investigations into Musical Applications of Time-Delayed Recurrent Networks

Topic: Sound Synthesis

Authors:

Daniel Bisig
Zurich University of the Arts
Institute for Computer Music and Sound Technology
Switzerland
http://www.icst.net/

Philippe Kocher
Zurich University of the Arts
Institute for Computer Music and Sound Technology
Switzerland
http://www.icst.net/

Abstract:

This publication presents the authors' preliminary investigations concerning the adaption of time-delayed recurrent networks into generative mechanisms for music creation. We argue that the dynamic properties of these networks renders them particularly attractive for the design of musical algorithms. The use of time-delays in a broad range of durations in combination with complex feedback mechanisms presents a system that unifies different aspects of algorithmic music creation in one single formalism. Not only do these networks present themselves as a means for sound synthesis, they are also able to generate rhythmical structures in different temporal scales.

As part of this research, the authors have developed and implemented different prototypes of network-based synthesis mechanisms. These prototypes are meant to serve as starting points not only for the assessment of the sonic capabilities of such networks but also for the identification of potential challenges concerning their integration into musical practice. The publication describes the algorithmic principles of these prototypes and provides an overview of some early acoustic results.

Based on this assessment, the publication addresses several challenges which lie beyond mere algorithmic issues and concern the adaptation of network-based synthesis systems for practical use. One category of challenges deals with the development of tools and strategies that assist musicians in the exploration of and experimentation with these highly non-linear synthesis techniques. We propose an automated evaluation mechanism that supports and guides the human search process by predicting the development of the sonic output for a given set of parameters. The second category of challenges deals with the provision of means that help musicians to acquire an intuitive understanding and appreciation of the generative principles that underly the synthesis methods. Here we argue in favour of a perception-driven form of comprehension that results from a tangible interaction with a physical interface whose affordances have been aligned with the algorithmic characteristics of the generative system.

Contact:
daniel.bisig@zhdk.ch

Keywords:
computer music, non-standard sound synthesis, recurrent time-delay networks

Early Investigations into Musical Applications of Time-Delayed Recurrent Networks

Dr. Daniel Bisig
Zurich University of the Arts
Institute for Computer Music and Sound Technology
Switzerland
http://www.icst.net/
e-mail: daniel.bisig@zhdk.ch

Philippe Kocher
Zurich University of the Arts
Institute for Computer Music and Sound Technology
Switzerland
http://www.icst.net/
e-mail: philippe.kocher@zhdk.ch

Abstract

This publication presents the authors' preliminary investigations concerning the adaptation of time-delayed recurrent networks into generative mechanisms for music creation. We argue that the dynamic properties of these networks renders them particularly attractive in the context of computer music. The use of time-delays in a broad range of durations in combination with complex feedback mechanisms presents a system that unifies different aspects of algorithmic music creation in one single formalism. Not only do these networks present themselves as a means for sound synthesis, they are also able to generate rhythmical structures in different temporal scales.

As part of this research, the authors have developed and implemented different prototypes of network-based synthesis mechanisms. These prototypes are meant to serve as starting points not only for the assessment of the sonic capabilities of such networks but also for the identification of potential challenges concerning their integration into musical practice. The publication describes the algorithmic principles of these prototypes and provides an overview of some early acoustic results.

Based on this assessment, the publication addresses several challenges which lie beyond mere algorithmic issues and concern the adaptation of network-based synthesis systems for practical use. One category of challenges deals with the development of tools and strategies that assist musicians in the exploration of and experimentation with these highly non-linear synthesis techniques. We propose an automated evaluation mechanism that supports and guides the human search process by predicting the development of the sonic output for a given set of parameters. The second category of challenges deals with the provision of means that help musicians to acquire an intuitive understanding and appreciation of the generative principles that underly the synthesis methods. Here we argue in favour of

a perception-driven form of comprehension that results from a tangible interaction with a physical interface whose affordances have been aligned with the algorithmic characteristics of the generative system.

1. Introduction

Feedback and delay mechanisms play an important role in computer music as techniques for digital audio signal processing and sound synthesis. Delay times in the range of a few samples are used for the implementation of digital filters. The combination of time-delay and feedback forms the basis for the design of recursive filters. These filters can be employed for the realization of highly pronounced and computationally efficient filtering effects. Techniques that operate with larger time-delays are typically employed to simulate phenomena of room acoustics such as reverberation or discrete echos. For the purpose of sound synthesis, the method of digital waveguide synthesis represents a classical physical modelling approach. This technique simulates the propagation of a sound wave through a physical medium. The time it takes for the sound wave to travel a certain distance is implemented as a delay line and the reflection of the wave at material boundaries is implemented as feedback.

Feedback and delay mechanisms also play an important role for the design of artificial neural networks. In time-delayed recurrent networks, the connections among neurons form cyclic graphs and durations for signal propagation are taken into account. Due to the fact that these networks possess an internal memory, they are capable of processing and learning non-static input patterns. Apart from these networks' role as powerful machine learning tools, their behaviour is also of great interest from a dynamical systems perspective. Contrary to classical feed forward networks, these networks often exhibit chaotic and complex patterns in their activity propagation. The vast majority of musical applications employ neural networks for analysis tasks and imitation-based approaches to musical composition, see for example [1, 2]. Others use the complex temporal dynamics of these networks as generative mechanism for algorithmic composition, see for example [4, 5]. There exists very little research concerning the adoption of recurrent neural networks as mechanisms for signal processing and sound synthesis. Ohya describes a sound synthesis method based on a recurrent network that consists of continuous-time and continuous-value neurons whose interconnections possess both weight and delay [6]. A second and somewhat more recent example describes a neural network-based synthesis system that consists of two neurons that exhibit mutual inhibition and lock their internal oscillations to the frequency of an input signal [3].

This publication is motivated by the desire to continue and expand the promising but seemingly neglected research that deals with the adaptation and application of time-delayed recurrent networks for algorithmic composition, sound synthesis and signal processing. In particular, we aim to combine time-delay feedback techniques from signal processing and artificial neural networks into unified generative algorithms for computer music. We believe that an extension of well established methods from digital signal processing with network-based forms of non-standard sound synthesis enables interesting sonic possibilities. Furthermore, a unified algorithmic system which can operate over a diverse range of temporal scales possesses great promise as a coherent mechanism for the creation of musical structures on both the level of musical composition and sound synthesis. Finally, the behavioural complexity that these methods inherit from their neural network counterparts can serve as a starting point for the creation of autonomous musical systems.

2. Prototypes

As part of a brief pre-project that was conducted prior to an application for a publicly funded research project, the authors have realised three prototypes that incorporate different time-delay feedback mechanisms. These prototypes served as early experimentation platforms and helped us to gain first insights concerning the musical potentials and pitfalls of these systems. In particular, they helped us to assess the following topics:

- software architecture requirements for the implementation of diverse time-delays and arbitrary recurrent network topologies
- integration of feedback stabilisation mechanisms
- generation of non-periodic network activity patterns
- impact of parameter changes on the networks' behaviour and sonic output
- characteristics and diversity of the networks' sonic output
- comprehensibility of relationships between network properties and sonic output

The three prototypes exhibit three different approaches to sound synthesis. (1.) An extension of digital waveguide synthesis, (2.) a continuous signal propagation network that employs a gate control mechanism, and (3.) a synthesis method which is inspired by spiking neural networks.

2.2 Prototype 1

The first prototype takes digital waveguide synthesis as starting point. It uses signal feedback and delay lines, but unlike standard forms of digital waveguide synthesis, the delays vary over time and they do so independently from one another (see figure 1). This approach does not try to approximate the sound of a musical instrument nor does it model an existing physical property.

This prototype has been implemented in the programming environments CSound, Supercollider and Max. Until now, only systems consisting of a rather small number of delay lines (up to 4) have been tested. The variable delay times range between one single sample and 0.1s and the values are stored in breakpoint functions which are interpolated during audio playback.

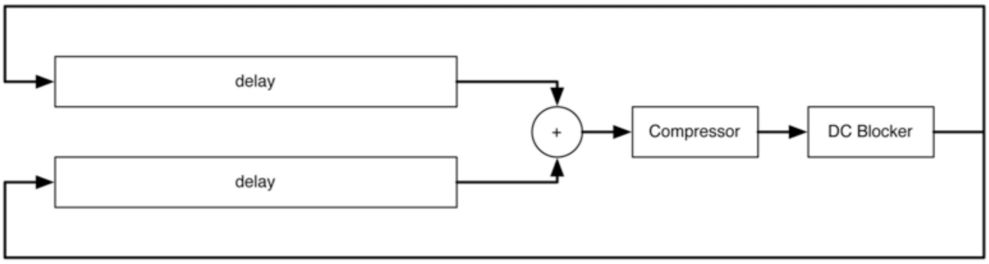

Figure 1. Digital waveguide-based time-delayed feedback system. Two independent delay lines with a summed feedback.

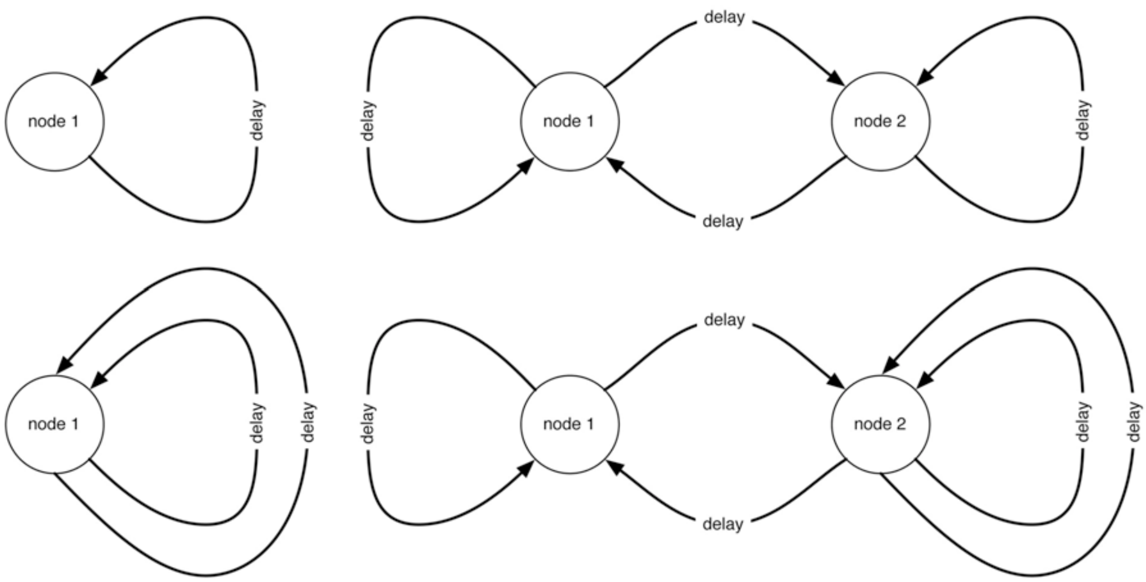

Figure 2. Examples of recurrent network topologies employed for prototypes 2 and 3. From left to right and top to bottom: single neuron single connection (1n1c) network, two neuron four connection (2n4c) network, single neuron two connection (1n2c) network, two neuron five connection (2n5c) network.

2.2 Prototype 2

This prototype consists of an audio signal graph in which gates and delay lines are connected (see examples in figure 2). At the nodes which act as gates, all incoming signals are summed and averaged by calculating the root mean square of the signals' amplitude (see figure 3). The gate opens if this average falls between a lower and upper threshold and the gate's responsivity is determined by an attack and decay time. These nodes and their gating behaviour are loosely inspired by neurons and their action potential threshold. Between the nodes, the signal is time-delayed, and, if desired, attenuated and low-pass filtered. The system is initially excited with a short audio signal, e.g. a brief tone or a noise burst.

The implementation of this prototype has been realised in the programming environments Max, Pure Data and SuperCollider. The scenarios tested so far consisted of only a small number of nodes (mostly one or two), and a larger number of recurrent and time-delayed connections. The parameters to control the behaviour of each node and connection are as follows:

- lower and upper thresholds to open the gate
- attack and decay time of the gate envelope
- signal propagation time delay
- signal attenuation factor
- cutoff frequency for low pass filter

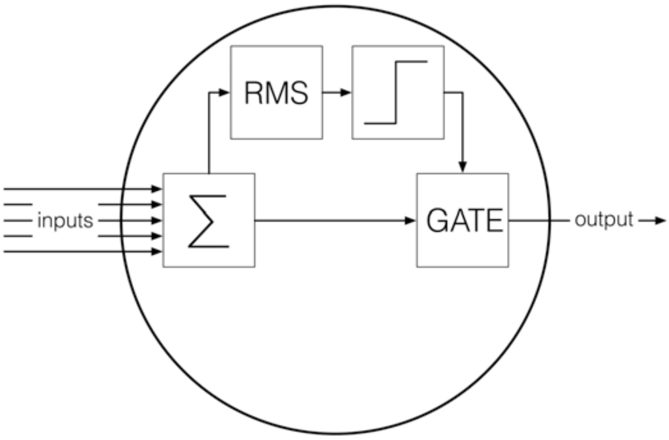

Figure 3. Prototype 2 node characteristics. The boxes within the node represent from left to right: input signal summation, root mean square average, threshold function, gate with attack and decay time.

2.3 Prototype 3

This prototype realises an audio signal propagation mechanism that resembles neuronal spiking. All audio signals that arrive at a neuron are summed together and subsequently passed through a sigmoid activation function in order to calculate the neuron's activity level. If this activity level is in between a lower and upper boundary, the neuron triggers a spike in the form of a sound grain and resets its activity level (see figure 4). The grain signal is time-delayed and attenuated when it travels across a neuronal connection.

The neural simulation and DSP engine have been implemented from scratch in C++. The software integrates the Jack audio routing library to connect neural activity to an audio hardware output. Any number of neurons can be rendered audible at the same time in this way. The network exposes a number of parameters to real-time control:

- signal content of audio grains
- amplitude envelope of audio grains
- slope of sigmoid activation function
- lower and upper activity thresholds for triggering audio grains

- amplitude of triggered audio grains
- duration of refractory period before triggering subsequent grains
- maximum number of overlapping grains
- signal propagation time delay
- signal attenuation factor

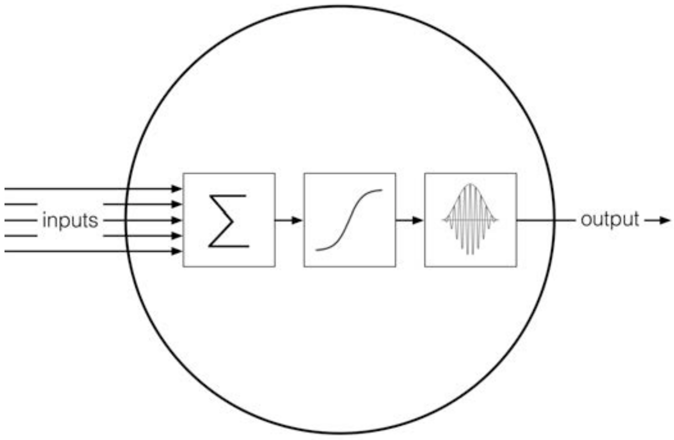

Figure 4: Prototype 3 node characteristics. The boxes within the node represent from left to right: input signal summation, sigmoid activation function, signal grain generation.

3. Results

3.1 Implementation, Software Architecture

Prototypes 1 and 2 have been implemented in an existing audio programming environment. In these cases, the general observation can be made that a graphical programming paradigm, as can be found in Max or Pure Data, is more suitable than text based environments such as Supercollider or CSound; since the visual interfaces of the former resemble flowcharts they are more readily able to represent the layout of a network.

Prototype 3 has been developed from scratch in C++. As expected, this approach significantly facilitates the close integration between network simulation and signal processing aspects of the software. Furthermore, it allows to overcome any constraints imposed by other software concerning the flexibility of the signal flow, in particular the usage of feedback at any point, and the application of arbitrary feedback times.

It has become clear that the computation of signal propagation within highly recurrent networks demands a different audio buffer update mechanism than is provided by most existing audio environments. For example, the custom software implementation, as realised for prototype 3, has allowed us to realise signal processing procedures that operate with heterogenous audio buffer sizes rather than a global uniform buffer size. This aspect forms an essential prerequisite for a computationally efficient implementation of arbitrary recurrent network topologies.

3.2 Network Dynamics

As a result of the recurrent network topologies, some of the systems described above are highly prone to enter runaway conditions, i.e. to blow up. In order to ensure that the systems remain stable, the amplitude of the signal has to be carefully kept under control.

In prototype 1, an automatic amplitude control (i.e. compressor) and a dc blocker are inserted into the signal flow before the signal is fed back into the delay lines. In prototype 2, the desired stability is achieved by setting the parameters for each node's gate function in such a way that the gate closes not only when the signal falls under a certain limit but also when the signal exceeds an upper threshold. In case of prototype 3, the signal spiking mechanism provides an inherent protection against runaway situations. Here, the neuronal activity level that can be reached by the network remains bounded due to the fact that the output signal, which consists of grains of fixed amplitude, entirely supersedes the input signal.

It has turned to be a non-trivial task to configure the networks, especially in prototypes 2 and 3, in such a way that they maintain activity over prolonged periods of time rather than to quickly die away. This behaviour is influenced by a number of mutually interdependent parameters, of which the most important ones are the minimum and maximum thresholds, the delay times, the gate attack and decay time in case of prototype 2 and the grain amplitude, maximum grain overlap and refractory period in case of prototype 3.

3.3 Sonic Characteristics

Tests with prototype 1 have shown that this type of network is able to easily produce a large variety of timbres and, because of the changing delay times which result in doppler shifts, different pitches and glissandos. In particular, by means of delay times that vary rapidly or over a large range of values, a richness of musical gestures can be produced. This quality is desirable from an artistic standpoint, but its complex dynamics is hardly predictable for the user. Experiments with different network topologies, i.e. more than two delay lines, have shown no essential changes to the sonic characteristics of the output. This raises the question under which conditions the network topology has a significant impact on the resulting sound.

The experiments conducted with prototype 2 started with networks consisting of a single node and an increasing number of time delayed feedback connections. Due to the large number of combinations of control parameters, a systematic evaluation of these systems under aesthetic or artistic criteria becomes unfeasible. For this reason, we conducted an explorative search for an "attractive" sets of parameters. It has been found that the choice of delay times has a significant impact on the musical quality of the output. Very small delay times are applied to achieve sonic continuity, as well as timbral effects of a strong comb filter quality, larger delay times produce rhythmic textures. The attack and decay times of the gate cause time lags which allow the signal to overshoot for a short period of time. The dynamics of this behaviour gives rise to pulsating rhythms in the sonic output.

With respect to prototype 3, sonic experiments have been conducted with audio grains that are 32 samples long and employ a hamming window as amplitude envelope. As was expected, regular periodic signals could easily be generated in case of the 1n1c and 1n2c networks by choosing a refractory period that exceeds the duration of a grain signal. By modifying the time delays, these periodic signals become audible either as rhythmic patterns or pitched sounds. On the other hand, the ease with which multiple levels of interleaved period patterns can be achieved by setting the refractory periods to a smaller value than the grain duration, constitutes a rather surprising result. These periodic patterns manifest as cyclically alternating intervals in pitch shifts and sound texture changes. These behaviours could be achieved for all tested network types. In case of the 1n2c and 2n4c networks, the aperiodic output signals that result from irrational ratios between signal propagation delays almost always give raise to sonically very diversified transition phases. In particular, the presence of multiple recurrent connections to a single neuron proved to be a very useful setup to obtain this kind of sonic result. In case of the more complicated 2n4c network on the other hand, the acoustically rewarding transition period is usually of very short duration and quickly approaches an invariable and noisy output. It seems clear, that the difficulty of designing a network with the aim of achieving a sonically interesting behaviour increases very rapidly as the network grows in size.

4. Discussion and Outlook

Based on these preliminary explorations into the sonic capabilities of time-delayed feedback networks, it is clear that these systems possess great musical potential but are at the same very challenging to work with. For a given network architecture, the anticipation of the sonic effects of parameter changes is difficult due to a number factors. (1.) The effects are often highly non-linear and mutually inter-dependent, (2.) they can differ considerably depending on the network's topology and its current state, (3.) they often not only affect the immediate audible results but also influence the long term sonic evolution of the network. Accordingly, these networks are particularly difficult to employ for purposes of musical improvisation and live experimentation. For the same reason, it is even more challenging to make musically informed decisions concerning the design of novel network architectures. Trial and error approaches to network design are time-consuming and likely very frustrating, in particular when dealing with larger networks than those presented in the prototype experimentation scenarios. Therefore, it is very important to establish design heuristics that are based on a basic understanding of the relationships between network properties and sonic output.

We believe that a systematic approach to the development of novel network-based generative algorithms should interrelate a thorough mathematical assessment of the networks' dynamic properties with a musical evaluation and classification of the networks' sonic output. Such a combined approach is instrumental in identifying promising network algorithms and for their subsequent adaptation and customization into generative algorithms for sound synthesis and composition.

In order to foster the musical usefulness of these networks, the aforementioned activities need to be complemented by research that specifically addresses issues relating to musical practice. This research should include the development of strategies and tools that assist musicians in exploring these systems and in acquiring an intuitive understanding of their underlying algorithmic principles. In that regard, we are currently planning to focus on the following aspects. (1.) The design and implementation of a semi-automated network evaluation system that guides musicians in their search for potentially interesting network configurations and behaviours. (2.) The development of physical interfaces that render network properties and behaviours accessible for tangible interaction and thereby help in the acquisition of an intuitive comprehension of this generative system. Both the guidance system and the tangible interfaces are meant to strike a balance between exposing the algorithmic principles of the networks' architectures and providing means for intuitive and perception-driven forms of interaction and musical experimentation.

The guidance system is intended to fulfill a dual function of predicting long term sonic developments and of searching for alternative network customizations that might exhibit musically interesting behaviours. The first functionality allows a musician to listen ahead of the current sonic output of a network and thereby helps to anticipate and influence the network's future musical development. The second functionality is likely to implement a local evolutionary search within the parametric and topological neighborhood of the currently employed network. Those alternative network architectures that best meet basic user-defined fitness criteria are then integrated into a constantly updated repertoire of networks that serve as promising candidates for further musical experimentation. By employing only low-level behavioural and acoustic criteria for the fitness function, the aesthetic evaluation of the musical output remains entirely up to human evaluation.

The research direction that deals with the development of physical interfaces is based on the premise, that a tangible interaction with a complex generative system helps musicians to acquire an understanding of how algorithmic principles and sonic output interrelate. The planned interface design is based on an approach that aligns perception-driven interaction with the topological and behavioural characteristics of network-based generative processes. For instance, the topological characteristics of networks lend themselves to a translation into spatial representations. In such a representation, the individual network nodes can appear as physical objects that individually emit sound and whose interconnectivity and time-delayed signal exchange mechanisms are represented by the spatial distances among each other. Such an installative setup not only renders the characteristics of the networks' acoustic signal propagation perceivable in a spatial manner, it also translates these characteristics into tangible affordances for an interactive manipulation of the network's architecture. By manually moving physical objects through space, a position tracking system can map these changes back into the network simulation in order to change the network's topology and signal propagation delays accordingly. We believe that such a mutually constituted relationship between a network-based generative system and its representation as tangible interface and installation links musical experimentation and engagement with perception-based forms of comprehension.

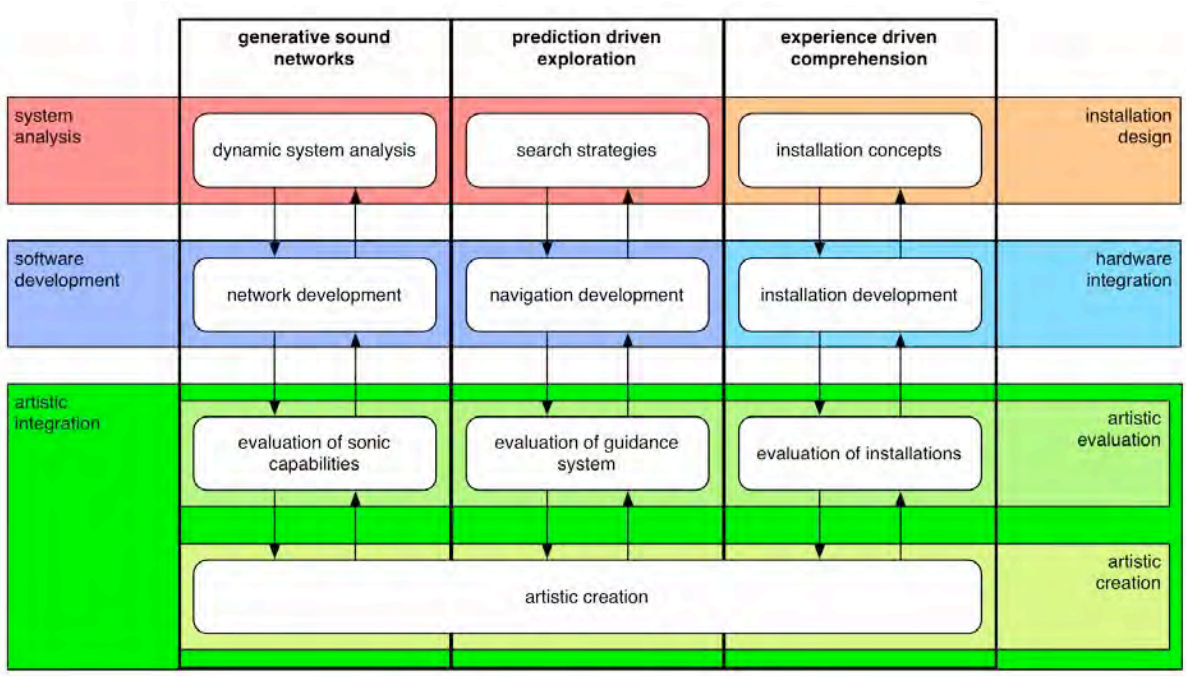

Figure 6. Research strands. The research directions that are planned for the submitted project.

Based on these aforementioned results and considerations, we have devised and submitted a proposal for a publicly funded research project. This proposal will run for the duration of two years and incorporates the previously discussed aspects into three main research strands (see figure 6). These research strands tightly integrate mathematical analysis and algorithmic developments with musical experimentation and evaluation. All activities converge into a residency program during which invited artists and musicians from the fields of algorithmic composition, electroacoustic performance and sound installation art will be able to adapt and integrate the project results into their own artistic creations.

We believe, that such an integrated approach which combines mathematical, engineering and musical expertise to address different aspects that pertain to musical uses of a complex system is essential for developing and transferring a novel generative system into artistic practice. Accordingly, we hope that this project can serve as inspiration for other research projects within generative art.

4. References

[1] Giuseppe Buzzanca. Artificial Neural Networks in Real-Life Applications, chapter Music and Neural Networks, pages 239–264. IGI Global, 2005.

[2] Douglas Eck and Juergen Schmidhuber. A first look at music composition using LSTM recurrent neural networks. Technical Report IDSIA-07-02, IDSIA, www.idsia.ch/techrep.html, March 2002.

[3] Alice Eldridge. Neural Oscillator Synthesis: generating adaptive signals with a continuous-time neural model. 2005.

[4] Hernán Kerlleñevich, Pablo Ernesto Riera, and Manuel Camilo Eguia. SANTIAGO - A Real-Time Biological Neural Network Environment for Generative Music Creation. In Applications of Evolutionary Computation, volume 6625 of Lecture Notes in Computer Science, pages 344–353. Springer Berlin Heidelberg, 2011.

[5] Kevin McCracken, John Matthias, and Eduardo Reck Miranda. Neurogranular synthesis: Granular synthesis controlled by a pulse-coupled network of spiking neurons. In EvoApplications (2), volume 6625 of Lecture Notes in Computer Science, pages 354–363. Springer, 2011.

[6] Kenichi Ohya. Sound Variations by Recurrent Neural Network Synthesis. In Proceedings of the International Computer Music Conference (ICMC), pages 280–283, Ann Arbor, MI, USA, 1998.