Interactive instrumental performance and gesture sonification

Professor Kirsty Beilharz
University of Technology Sydney
P.O. Box 123 Broadway 2007 Australia
kirsty.beilharz@uts.edu.au

Abstract.
This article describes a system for interactive performance that generates live musical accompaniment to an instrumental performer, using a Neural Network model and granular synthesis controlled by the gestures and breath of the performer. More broadly, the technologies for gesture-tracking can be applied to various interactive environments for real-time music augmentation. The use of pitch-tracking and musical parameters derived from the performer's output seed a generative computing model that procreates new material and perpetuates musical fragments over time to produce a contextual yet create response to the live performer. Electronic music and live visualisation of gesture augment the scope of an analogue traditional instrument, in this case a bamboo Japanese flute, *shakuhachi,* hence the title of the interactive music environment, *HyperShaku.* The design issues of relevance to inter-domain implementation include: integration of A.I. modules (Artificial Life, biologically-inspired and Evolutionary processes) in a system for real-time computational data processing; using A.I. modules for sonifying data – for automated and interactive generation of sound; and considerations for mapping gesture and other data to auditory display. The data mapping principles of this approach are applicable to a range of sonification contexts, using generative processes to synthesise and perpetuate sound in real-time, mediating between designer/user/interaction and representation.

Hyper-Shaku uses Evolutionary Looming to scale frequency as a consequence of input loudness and noisiness, a Neural Oscillator Network to perpetuate sounds with concordant pitch (frequency) derived from the live performer's auditory input, and gestural interaction to adjust parameters of granular synthesis and the generative processes.

Auditory display is an emerging modality for data representation, both for use alone in visually heavy contexts, where sonification presents an effective alternative to visualisation, and in bi-modal audio-visual display environments where sonification can reinforce other modalities, enhancing fidelity of representation and reasoning based on it. Due to the interplay of auditory cognition, memory and the inherently time-based representation of sound, sonification can provide superior recognition to visualisation for certain types of features, such as periodicity, discrete irregularities, subtle shifts over time, stream segregation and very fine increments of data represented using frequency. This example explores a creative-context application of data (motion and breath) sonification.

Keywords: generative design, interactive sonification, interactive music, multi-modal display, electronic music, gestural interaction

1. INTRODUCTION

The gestural data-capture and generative processes for automated computational sonification (or auditory display) that follow have been implemented in art installation contexts and music performance interactive environments. *Hyper-Shaku: Border-Crossing* is a gestural interaction environment (hyper-instrument) for creating real-time audio-visual augmentation of musical performance on the traditional Japanese *shakuhachi* (bamboo 5-holed, end-blown flute). The computational structures support a hyper-instrument performance environment whose purpose is to augment the human performer's delivery with electronic audio and visual display in real-time. This paper looks at processes of automated generative sound production, moderated by user interaction.

The performance context is pertinent in that immediacy, minimal latency and real-time responsiveness of the system are critical to its reception and usability. These features of real-time data representation are usefully applicable to other contexts where immediate response is important, such as live data analysis, data monitoring, diagnostics, manufacture of medical instruments and designing. The system transforms the modality of movement/gesture and physical/auditory input into an integrated auditory and visual output, thus considering issues of mapping kinaesthetic modality to audio-visual representation and interaction between the representational modalities resulting from a shared generative A.I. system for production of both sonification and visualisation. In addition, relevant criteria in making original music include creativity or inventiveness and aesthetic sound qualities.

2. RELATED WORKS AND BACKGROUND

The *Hyper-Shaku* environment brings together technologies applied in previous works involving intelligent sensor environments (sensate spaces) and computer vision. In *Emergent Energies* (by Kirsty Beilharz, Andrew Vande Moere, and Amanda Scott), sensor technologies were embedded in a responsive, sensate room that tracked mobility and activity over time. Ambient displays in architectural spaces have the potential to provide interesting information about the inhabitants and activities of a location in a socially reflective display. People can monitor information such as popular pedestrian paths, times of peak activity, locations of congestion, socially popular convergence points, response to environmental conditions such as temperature, noise and so forth, contributing to our understanding of social behaviour and environmental influences. Finding engaging and effective display modalities for the ever-increasing data collected by pervasive sensing and computing systems is especially poignant for the general public.

Biologically inspired generative algorithmic structures produce the representation with a consistent mapping relationship to data yet with a transforming, evolving display that is intended to enhance sustainable participation and perpetuate interest, without repetition. The residual, cumulative nature of the visual Lindenmeyer System employed in *Emergent Energies* (as compared with the ephemeral nature of transient audio-only display), allowed users to observe the history of interaction. Other works by the author using similar technology integration (wireless gesture-controllers, computer-vision motion triggering and real-time generative displays in Max/MSP and Jitter software) include *SensorCow* (a motion sonification system), *The Music Without* (gesture sonification while performing music), *Sonic Kung Fu* (a gestural interactive soundscape), *Fluid Velocity* (responsive visualization and sonification of sensor data captured from a physical bicycle interface) and *Sonic Tai Chi* (an Artificial Life visual colony and audio synthesis modified by spatial interaction).

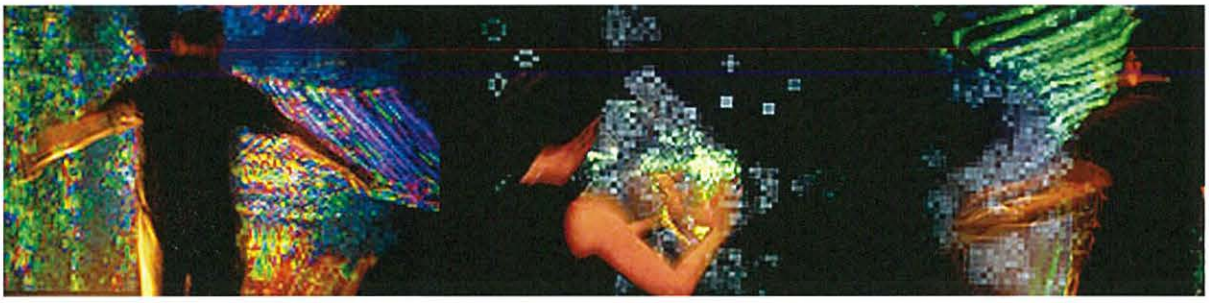

Figure 1. *Sonic Tai Chi* (BetaSpace, Sydney Powerhouse Museum, installation 2005-2006 developed with Joanne Jakovich) uses computer vision to capture movement data producing the visualisation and sonification. Generative Cellular Automata rules propagate particles and sonic grains in response to users' lateral motion.

Figure 2. *Fluid Velocity* (Tin Sheds Gallery, installation 2006 developed with Sam Ferguson and Hong Jun Song) using motion sensors on the physical bicycle to modify the tentacle curviness, diameter, number, motion and angularity, as well as auditory filtering of the soundtrack, in response to the rider's activity.

These works utilise either gesture/motion data or social data captured through a range of sensing technologies. With the current ubiquity of computing and sensing devices and increasing integration of such information-gathering technologies in the workplace, manufacturing processes, recreational and entertainment contexts and designing or performing creative contexts, the importance of understanding the interplay between representational modalities and mapping real-time responsive information displays to data are crucial to giving meaning to our technical capabilities. Interactive and automated representation processes, like the sonification and visualisation in this system, provide methods for coupling structural activities and representation.

Designers, musicians and artists have long used manual and computational methods of transforming data into a multitude of representations, hence some techniques for rapid representation developed in those spheres can be useful to other domains where

information representation and deciphering are critical, either to enhance aesthetic information dissemination and comprehension, or to augment and reinforce data communication in strategic situations. Thus, sonification ranges in usefulness from augmenting artistic practice to the elaborate hierarchical systems of auditory alerts used in air-traffic control systems or for monitoring medical information in hospitals. An early instance of sonification for diagnostic purposes was the audification of the first University of Sydney computer, Silliac, whose boot sequence and correct functionality was ascertained by its technicians listening to the processing audification.

Sonification is a discipline that has arisen recently comparatively with visualisation, especially helpful in the representation of complex data, to assist recognition of trends (gesture) and afford optimised interpretation using an alternate sensory domain. Traditional applications of sonification range from representing medical, seismic, meteorological data to diagnostic, programming debugging and warning systems. Many of these advantages for using sonification to interpret abstract, numerical or text-based data (Bly 1982) also translate to auditory graphing of gesture input data. Principles for mapping, understanding auditory perception and cognition in the sonification field inform the process of auditory interpretation and representation used in the hyper-instrument model.

Some examples of information sonification for informative and aesthetic purposes include: Jon McCormack's *Eden* using Cellular Automata in response to audience interaction; Andrea Polli's *Atmospherics* gallery installation sonifying meteorological captured at different altitudes of the atmosphere during a cyclone; Fabio Cifariello Ciardi's *sMax: A Toolkit for Stock Market Data Sonification;* and Garth Paine's *PLantA* sonification of meteorological data captured live from a portable weather station.

3. TRANSFERABILITY / SCALABILITY

The system described here takes an interactive approach to autonomous creativity, using attributes of user interaction to modify the generative and responsive modules in the system because it is a collaborative environment. This idea of user intervention or affect could also be applied to autonomous systems of generativity or automated display for design domains. Many interactive feedback loop models are forms of action/reaction. An environment that transforms representation simply from one modality to another could be construed as a translation tool. There is already much to be learned by re-examining processes and structures in different modalities but interaction and modification sympathetic to the user adds an extra layer of control and subtlety. The significance of generative processes in an interactive music system are their capability of producing both a responsive, strict relationship between gesture and its auditory mapping while developing an evolving artefact that is neither repetitive nor predictable, harnessing the creative potential of emergent structures. The visualisation module expands conventional musical performance presentation. The use of a bi-modal representation can serve to reinforce and clarify. The different modalities can operate independently or, as here, activated by common A.I. processes and gestural triggers.

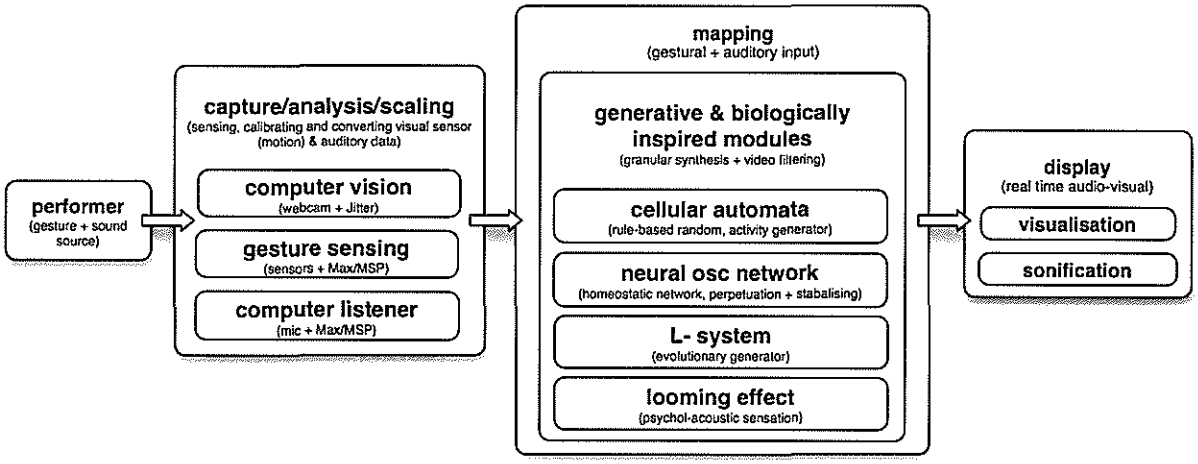

Figure 3. The modular approach to hyper instrument design using biologically inspired generative computation for real time gestural interaction that allows for individually customised and scalable performance scenarios.

The model presented in this paper is intended to be transferable and modular (figure 3). The interaction model and processing modules (technological method) can be applied in various performance situations, ranging from the performance of fixed, notated music, to improvisation or predominantly gesture-driven installation situations. A modular approach also allows different generative engines to be interchanged to varying effect and according to the aesthetic goals of the situation. The generative rules of the individual modules, such as the Cellular Automata rules, Lindenmayer System variables or Neural Oscillator Network thresholds, can be adjusted to significantly affect the action/reaction consequences and resulting artefact (Burraston & Edmonds 2005). Different modes of sensing and motion capture are appropriate for single or multi-user scenarios. Types of sensors (gyroscopic, acceleration, binary, proximity, etc.) and their calibration can be fine-tuned to suit the activity of the particular motion input or user collaboration involved. While this system uses sound attributes as input (loudness, noisiness and pitch) and visual tracking of motion for streams of input data, the generative modules may be triggered and affected by a different input configuration arising from motion, data streams or design activities.

4. THE GENERATIVE INTERACTION FEEDBACK LOOP

The modular framework is a system for sonifying gesture data. This section looks at the symbolic nature of the biologically inspired models for generation and the modularity of the generative systems forming variable interactions in the system.

4.1. Gesture and Breath Data as Sonification

Sonification is the automated process of transforming data into an auditory representation. Mapping of gesture to visual and auditory display is considered as a type of sonification in which the contiguous data stream comes from coordinates of camera tracking, the rate of movement and distance/scope. Further gestural detail can be captured with more sensor types: gyroscopic, accelerometer and binary wireless sensing captors, for example, such as the WiSeBox (Flety 2005) that was used in the bicycle-activated 3D visual and auditory display, *Fluid Velocity. Fluid Velocity* for physical bicycle interface, visual projection and stereo audio production in the Tin Sheds Gallery, University of Sydney (Beilharz et al. 2006) used IRCAM WiSeBox WiFi transmission of data from captors located on the bicycle frame and handlebars to transform the 3D "creature" on screen and variable filtering and panning of the electronic sound (figure 2). The programming environment was Max/MSP and Jitter (Puckette & Zicarelli 1990-2005). Benefits of this approach are the scalability of generative modules and capture methods to broader sonification contexts, such as intelligent spaces.

Auditory representation of information has particular benefits. An obvious benefit is as an alternative to visualisation for people with visual disabilities (Wuensche & Lesser 1992) and it has specific attributes for general users. The ear is capable of gathering data from all directions and ranges without instantaneous re-adjustment or focus, i.e. "within an instant". Auditory perception and cognition are capable of segregating complex sounds comprised of superimposed inputs, deciphering layers of concurrent meaning. Identifying individual instruments in orchestral contexts or isolating conversation in a crowded room (so-called "cocktail party effect") are examples of this ability (Volpe 2002; Arons 1992). Potentially, for example, a multi-channel output of the granular synthesis process in Max/MSP from *Hyper-Shaku* could be spatialised to further emphasize the gestural spatial impact on the generative processes.

Sound is a very useful information carrier over time, i.e. our temporal acuity or resolution is finely honed, ranging from the ability to discern events with durations less than 30 microseconds (Wuensche & Lesser 1992) to perceiving very gradual transitions of pitch increment over long periods of time. Spectrum and pitch are based on acoustic sounds played by the *shakuhachi* in *Hyper-Shaku* with lasting influence over nodal dispersions in the Neural Oscillator Network. Auditory memory of related pitch is integral to understanding perpetuated signals stemming from one event and newly pitched relations emanating from a subsequent trigger-event. Cognitive pitch grouping occurs as similar pitches and sounds are related to a single interaction.

Volpe's *Algorithms for Aural Representation and Presentation of Quantitative Data to Complement and Enhance Data Visualisation* (Wuensche & Lesser 1992) contributes to the exploration of ways that using generic algorithms can transform various forms of data into effective aural representations.

Lodha and Wilson's *Listen* Toolkit (Lodha & Wilson 1996; Spieth et al. 1954) is an effective portable information sonification system, allowing users to map arbitrary data onto sound parameters in a variety of ways. The model described in this paper provides an equivalent portable solution for mapping interchangeable gestural (interaction) data from a variety of sensing mechanisms to interchangeable generative sound processes. In the comparatively mature field of visualisation, for example, the freely distributed *VTK Visualisation Toolkit* is a software library of visualisation applications. Synthesis of sound by using data to control various independent parameters of the generative process leads to performer-controlled parameters of the auditory representation, such as *timbre*, through fine controls of granularity, grain length, regularity, distribution,
pitch, durations and amplitude.

As well as converting numerical data into rational, direct mappings, the player's loudness influences psycho-acoustic phenomena by scaling intensity based on an interpretation of looming auditory motion, borrowed from Ecological Psychology. This sensation, caused by sound intensity rising steadily, naturally increases the perceptual significance of a gesture. In *Hyper-Shaku,* when such a gesture is found in the *shakuhachi* performance, pitch transposition and loudness are used to magnify this climax. The naturalness of this effect augments the performance in a plausible way as it substantiates the semantics of the performer's gestures, reiterated in a typical "leaf-shaped" note-shape and phrase-shape, that increases then decreases in intensity. This shape or gesture is inherent at many levels of traditional Japanese music, from the envelope of individual note dynamics, to phrase-shapes and the formal structure of whole pieces. *Jo-ha-kyu* or *jo-ha-kyu-zanshin,* as it is called in Japanese (start slowly, intensify, finish off/fade to infinity, "leaving the heart behind"), is also to be found as an aesthetic shape informing many traditional Japanese Zen art forms, such as body, music and vocal movement in *kabuki* theatre, dance, *shodo* (calligraphy), *taiko* (drumming) and in the *chado* (Tea Ceremony) (Kerr 1996; Kirkpatrick 2004). The eruption and subsidence gesture shape of the Neural Oscillator Network further reiterates this form as it magnifies, proliferates, then dissipates sonic events.

4.2. Gestural Modification of Generative and Biologically Inspired Design

In order to move away from an action-reaction predictable and repetitious response, integration of Artificial Life and biologically inspired models for procreating, populating, interpolating and generating endless new material create mediation between input (action) and output (reaction) in the interaction feedback loop ("soft" biomimicry).

Mapping retains a fixed relationship between stimulus and algorithmic process but non-linear and generative functions serve to create an unpredictable layer of interpretation between the performer (stimulus) and the outcome (representation). It is with this insertion of generative creativity (in the extra fabric of creative interpretation) that installation art works and creative performance environments are differentiated from directly-mapped data sonification.

The vast majority of the algorithms used for aural representation of data today consist of mapping scalar values onto a single sound parameter (Wuensche & Lesser 1992). Multivariate data representation by performing several mappings simultaneously is used for the data features extracted from the *shakuhachi* sound input. Its amplitude, pitch and noisiness, using Jehan's "noisiness" spectral flatness detector (Jehan 2001), are captured from the acoustic instrumental input. Amplitude is passed into the Looming detector that transposes pitch upwards when a *crescendo* (loudness increase) is detected exceeding a time threshold, giving the impression of sound approaching or "looming". Thus the Looming detector also affects all subsequent processes in the granular synthesis and Neural Oscillator Network that use pitch-transforming algorithms. Both pitch and noisiness attributes scale values in the granular synthesiser grain length and density (see figure 4). This multivariate approach maps several aspects of the input data to different processes.

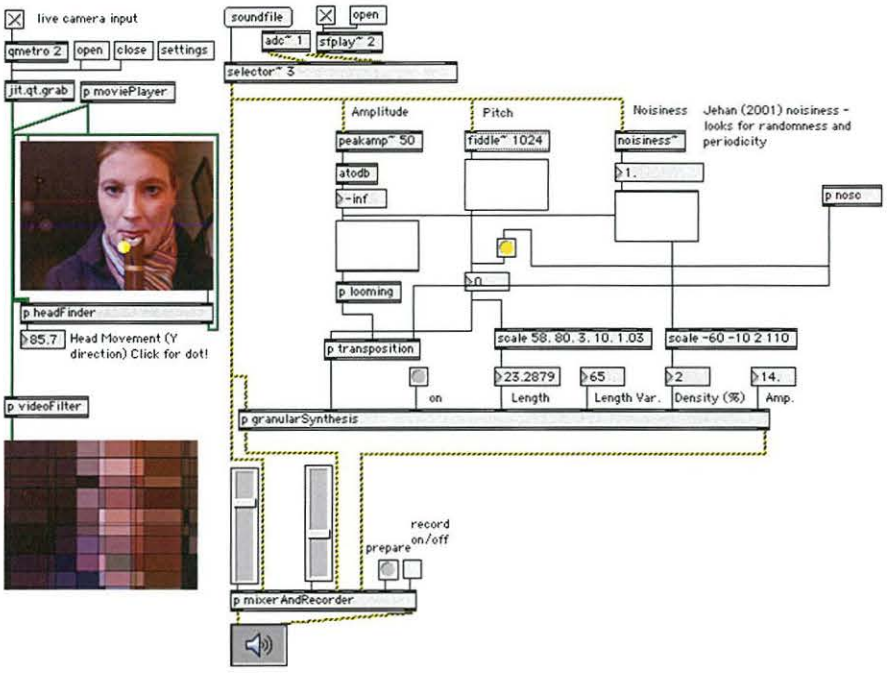

Figure 4. Gestural modification of the generative processes: sound, computer vision and motion sensor input detect gestural effects, used to send messages and input values to generator modules. Microphone acoustic input is used to control Looming (with loudness) and the granular synthesis (with loudness and noisiness measures). These gesture attributes also send messages to the Neural Oscillator Network and visualisation.

Other inputs also mesh with variable controls of these processes: the motion-tracked chin position (head movement while playing *shakuhachi)* from the web-cam changes Neural Oscillator Network and visual display scaling settings and optional motion captor data from wireless sensors can be used to transform additional controls in the granular synthesis (grain length and amplitude). It is intended that the interwoven gestural subtleties, mappings and representation processes provided by a multivariate technique should increase the intricacy of the relationship between the performer's gesture and computation.

Jon McCormack "On the Evolution of Sonic Ecosystems" (Adamatzky & Komosinski 2005) uses a multi-agent system that creates and hears sound to populate his virtual environment, *Eden*. While many Multi-Agent and Artificial Life systems are autonomously procreating, populating communities, *Eden,* like *Hyper-Shaku,* is a reactive Artificial Life artwork that modifies its processing in response to human user (even multi-user) interaction. In a hyper-instrument, the user plays a more dominant role than the user/interactor/audience in a public installation context. For the hyper-instrument, the user's influence and ability to intuitively and idiomatically control the artificial biological system is integral to its efficacy as a performative instrument (or tool). The notion of a musical instrument is arguably more finely honed than a general-purpose tool. We have associations of idiomatic gestures, refined technical competence, expertise and intimate rapport between the performer and instrument.

It is a goal of *Hyper-Shaku* to preserve the naturalness of this performer-to-instrument conduit. Thus gestural modification of the generative design aims to shape it in a way consistent with the semantic expression underlying the gesture. This requires some understanding of the physiology of performing and the physicality of playing the *shakuhachi.* Head motions are critical gestural attributes. Breath, intensity (velocity) and duration are also significant attributes that the listener observes and the performer aims to master. It is physically impossible, for example, to play loud notes without increased air velocity, which in turn changes the "airiness" or spectral distribution of the sound. These spectral (or noisiness) and pitch register changes are detected by the

Max/MSP patch to capture these breath attributes and further interpret them in the "hyper" augmentation through transposition and adjustment of granular synthesis parameters (such as grain length, distribution, number).

The modification of the generative systems is determined by three capture mechanisms in this example: visual gesture capture motion tracking using web-cam; computer listening and auditory analytical filters; and motion capture using wireless sensors (gyroscopic, acceleration, binary directional motion and others are available commercially). In this example, visual gesture capture and motion tracking is performed using Jean-Marc Pelletier's cv.jit algorithms (Pelletier 2005); microphone data is interpreted using Jehan's algorithms; and wireless sensor data captured, using Emmanuel Flety's WiSeBox and captors (Flety 2005). The sensing chosen for this example is selected for its portability, accessibility and unobtrusiveness.

4.3. Symbolic Representation in the A.I. Computation

When designers use models inspired by biological or Evolutionary phenomena, it is not simulation but constructive implementation of productive, creative organisms or methods that are sought. The representation is symbolic, metaphorical. This is important because scale, speed, and adaptation are "inspired" but not realistic. Scale and latency mean that biological inspirations like Neural Oscillator Networks are dramatically more extensive than computational implementations. Of more importance than the number of nodes and scale of the network for example, are the varying phasing and auditory outcomes that can arise from different configurations of nodal interoperation. Matsuoka's work on *Sustained oscillations generated by mutually inhibiting neurons with adaptation* (Matsuoka 1985) shows that oscillations generated by cyclic inhibition networks consisting of between 2 and 5 neurons receiving the same input, all exhibit oscillation, mostly periodic. Thus the main observable change in the output are the periodicity and phasing intervals (imagine sonic periodicity or pulsing and overlap/phasing). The significant principle is the *modus operandi* nodes, dispersing energy, distributed, perpetuating and stabilising, regulating activity. Thus our Neural Network is a "miniature".

If these implemented models are so unrealistic in scale, proportion, speed and latency, why are they suitable structures for Evolutionary art and biologically inspired design computation? Perhaps the usefulness and validity of biologically inspired processes lies in the variety of behavioral characteristics they demonstrate as well as the potential of their structure to produce innovative and novel outcomes that address unanticipated situations and non-programmable solutions.

The temporal and aesthetic reasons for using A.I. in a sonification system, rather than direct mapping, are the potential for generative systems to perpetuate and populate auditory (and visual) content and to combine elements of unpredictability, novelty, curiosity, engagement that come from new material concurrently with a regulated system of mapping input to display, thereby retaining informative representation.

4.4. Modular Generative Processes

The modular interlocking of different generative processes allows distinct generative systems to interact with each other. In order to avoid the audience "mastering" and understanding the systems at work too quickly, thus losing their engagement, the emergent characteristics obtained by combining interacting generative systems circumvents boredom with a second level of life-like complexity. "Decision-making" networks (derived from physics, biology, cognition with social and economic organisation strategies) are idealised models in the study of complexity and emergence, and in the behaviour of networks themselves (Adamatzky & Komosinski 2005; Harris et al. 1997; Kauffman 1993; Volpe 2002).

Used in isolation, for example, Cellular Automata "produce trivial repeated patterns or plain "chaotic" randomness" (Wuensche 1999, p.54). In rare cases, Cellular Automata do exhibit emergent behaviours, often only realized in large data spaces or in manipulations such as Christopher Ariza's *Automata Bending: Applications of Dynamic Mutation and Dynamic Rules in Modular One-Dimensional Cellular Automata* (Ariza 2007) methods of random cell-state mutation and dynamic, probabilistic rule-sets. Similarly to *Sonic Tai Chi* (by Jakovich & Beilharz 2005), Ariza applies Cellular Automata values to musical parameters by extracting one-dimensional value sequences". One of the intentions behind a modular "plug-in" generative structure is to bring the unpredictability and emergence of A.I. systems to performance. The other is to observe and develop an understanding of ways in which different systems can interact and co-exist to produce new and unanticipated effects.

5. THE HYPER-SHAKU GENERATIVE ENVIRONMENT

Hyper-Shaku (Border-Crossing) is both a digital audio-visual creative environment and compositional outcome (artefact). Using gesture-data, it is a sonfication system, motivated by the design efficiency that a modular system affords future applications. The objective is two-fold: to develop a system of computer vision and sensors producing an augmented sound-scope and derivative visual projection for concert performance.

5.1. Design Motivation

The selection of input gesture and sound capture devices for *Hyper-Shaku* environment are motivated partly by pragmatic mobility, i.e. the necessity of a portable performable system that can be transferred easily to different venues with simple configuration, ubiquitous accessibility using available hardware and an unobtrusive, non-invasive interface that will not inhibit natural performance or require change in technique. Another design motivation was to create a generative responsive system that could accommodate and react to the sensitivity of the Japanese *shakuhachi*. It is renowned for the softness of its tone, yet exhibits versatility and pitch subtlety that the fine-tuning of the noisiness and loudness thresholds utilise. Finally, re-usable generative interaction for variable performance scenarios is a design motivation arising from the construction of various generative responsive interaction systems with overlapping technologies and computational processes. It seems efficient to re-use code and modular Max/MSP patches adaptable to different situations by adjustment for the instrument, sensor type and generative system variables. The wide variety of points of adjustment in this model enable it to produce a broad array of auditory and visual outcomes: related but differentiated.

5.2. Mapping Gesture (Motion Tracking) to Neural Oscillator Network Weightings

The technological problem of latency in real time generative systems (which are computationally complex) is addressed in the real time application, Max/MSP that can handle sonification and video effects efficiently for live rendition. *Hyper-Shaku* uses the homeostatic process of a Neural Oscillator Network. Within the Neural Oscillator Network, triggers in individual "neurons" (Max/MSP software model, Figure 7) instigate moments of excitement that infect other neurons. Over time, the effect of one neuron in the network influencing another develops stabilising homogeneity, as gradually the neurons resemble and emulate one another. The combination of the chaotic and vigorous process of the Cellular Automata and the stabilising, homeostatic nature of the Neural Oscillator Network provides a suitable "excitement versus stasis" balancing structure for the production of long background transitions, behind a foreground of dynamic activity responding to the live performance (agile responses to the wireless captors). Algorithmically, the output of each neuron in the network is determined by the weighted outputs of every other neuron. The critical threshold of perturbation (beyond which reorganisation is triggered) is an adjustable parameter in the Max/MSP model. The fundamental units of an artificial Neural Network (units/nodes/neurodes) are modelled after biologically inspired individual neurons: its dendritic tree collects excitatory and inhibitory inputs from other neurons (the "receives" in the Max/MSP model, figure 6), and passes these messages, as voltages, on to the cell body (soma) (figure 7). These voltages are added to the current voltage if excitatory and subtracted if inhibitory. When the threshold is exceeded, a signal is transmitted (the "sends" in the Max/MSP model) down an axon to synapses that connect the tree to dendrites of other neurons (Eldridge 2005; Franklin 1999). Neurons (also called a linear threshold device or a threshold logic unit) can be modelled formally (mathematically) for computational purposes (figure 5 and equation 1).

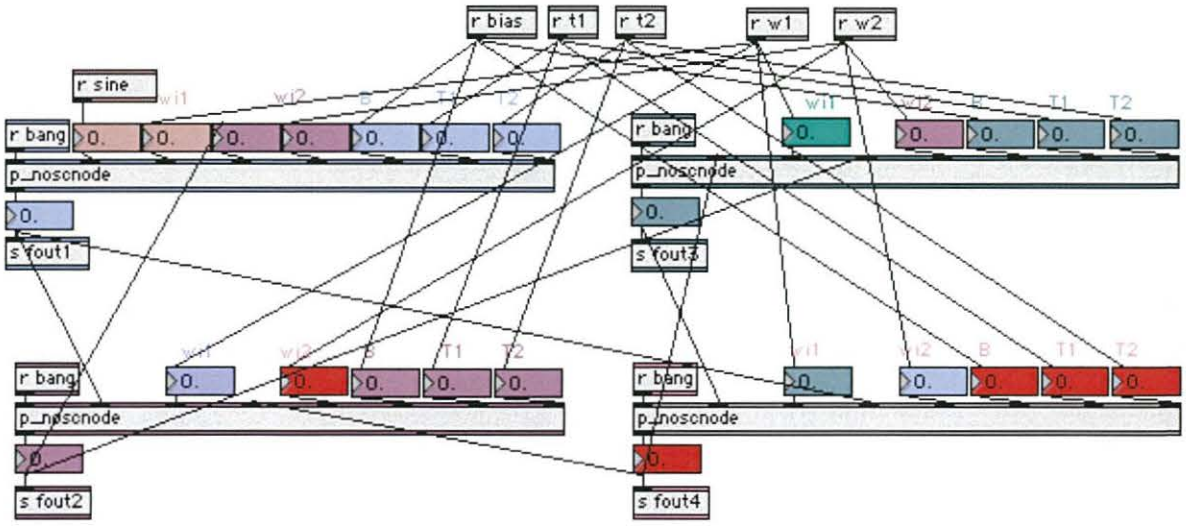

Figure 5. A symbolic simplification of a neuron with a formal model of threshold, see equation 1 following.

x_i -- the inputs
w_i -- the weights (synaptic strengths)
θ -- the threshold
y -- the output

$$y(t+1) = \begin{cases} 1 & \text{if} \quad \sum_i w_i x_i(t) \geq \theta \\ 0 & \text{otherwise} \end{cases} \qquad (1)$$

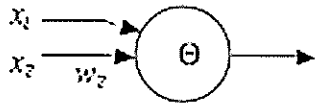

Figure 6. Neural Oscillator Network model using four synapse nodes to disperse sounds, audibly dissipating but rhythmic and energetic. Irregularity is controlled by head motion tracked through the computer vision. Transposition and pitch class arrives via the granular synthesis from pitch analysis of the acoustic *shakuhachi* and Looming intensity as a multiplier (transposition upward with greater intensity of gesture).

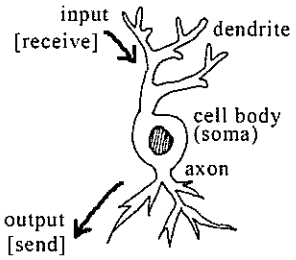

Figure 7. The Max/MSP Neural Oscillator Network patch (figure 6) is used as a stabilizing influence affected by large camera-tracked gestures. It is modelled on individual neurons: dendrites receive impulses and when the critical threshold is reached in the cell body (soma), output is sent to other nodes in the Neural Network. The "impulses" in the musical system derive from the granular synthesis pitch output.

The NOSC pre-sets (determinants of irregularity) especially of rhythm and range, are controlled by head motion tracked through the computer vision. Transposition and pitch class arrives via the granular synthesis following pitch analysis of the acoustic *shakuhachi* and Looming intensity as a multiplier (transposition upward with greater intensity of gesture). Different weights are influenced by registration (frequency) in the *shakuhachi*. Its natural pitch range has been divided into the lower octave, low upper octave and high upper octave (and upwards) to trigger different weighting thresholds in the NOSC. When a weight is "full" the impulse is passed on to another node. The audible outcome is the variety of agitation, pitch and conformity depending on the gestural and dynamic intensity of the musical input.

In the NOSC module, the calibration of the input affect on parameters of weight, float, node numbers, connections and feedback patterns are adjustable to fine-tune the network behaviour and aesthetic result (i.e. gestural modification of the generative processes). The network connections and directionality of transfer between nodes can significantly impact on the phasing effects and periodicity audible in the network, e.g. whether connections are mono-directional, bi-directional and the number of nodes can produce varying phased outcomes. As Williamson (1999) states, "Neural Oscillators offer simple and robust solutions to problems such as ... dynamic manipulation ... [but] the parameters are notoriously difficult to tune". His paper, *Designing rhythmic motions using neural oscillators* (Williamson 1999) offers an analysis technique that alleviates the difficulty of tuning.

5.3. Mapping Loudness to Looming Effects

In Ecological Psychology (sometimes called Evolutionary Psychology), auditory Looming refers to a phenomenon in which the magnitude estimation of rising intensity in ascending tones is more often over-estimated than equivalent reduction in intensity in falling tones indicating the greater importance that our cognition attributes to rising intensity (Neuhoff 2001; Neuhoff & Heckel 2004). This is thought to be founded in a primordial awareness that approaching or Looming pitches rise (Neuhoff 1998), similar to the physical phenomenon of the Doppler Effect, indicating something significant or dangerous is approaching. From an Evolutionary perspective, the perception of changing acoustic intensity is an important task (Neuhoff 1998). Rapidly approaching objects can produce increases in intensity and receding objects produce corresponding decreases.

Looming perception is a multimodal process that can be carried out by the visual system, the auditory system, or both (Lee et al.1992). From the *shakuhachi* acoustic input (captured with condenser microphone), the measure of loudness is sent to the Looming sub-patch in which change over time, i.e. intensity change mapped across time to evaluate increase or decrease in intensity, provides the multiplier for the Looming effect. Significantly, increasing intensities are used to generate a nonlinear upwards pitch transposition that exaggerates perceived looming. As this effect, emulating looming danger, is a pitch-distorted transformation, an aberration, it is used sparingly for maximum effect, only when dramatic dynamic events occur. This threshold is an adjustable parameter of customisation in the sub-patch. It needs to be adjusted depending on the natural loudness of the base instrument in the hyper-instrument environment and resulting signal range.

5.4. Tracking Pitch and Mapping Noisiness (Breathiness) to Granular Synthesis

Pitch is gathered from the *shakuhachi* (see programming architecture, figure 8). If it is combined with a significant increase in loudness over time, it is transposed by the Looming effect, otherwise its pitch is retained. This data passes into the multiplier of pitch, or pitch controller of grains in the granular synthesis engine. Registration affects grain density to produce a variety of sound qualities ranging from "grainy" or "pixilated", fragmentary-sounding to "globular" and recognisably pitched clusters (Roads 2001; Xenakis 1971; Xenakis 1990). Amplitude and noisiness are multivariate controls that, along with Looming-induced pitch transposition, affect feedback-like distortion through to percussive sounds in the granular synthesis to reinforce and substantiate intensely noisy moments in the *shakuhachi* performance such as the explosive attacks or breathy outbursts *(muraiki)* occasionally required in traditional Japanese solo *shakuhachi* music *(honkyoku)* or specified by contemporary composers. These are violent, forceful musical gestures that correspond with textural aberrations in the synthesis.

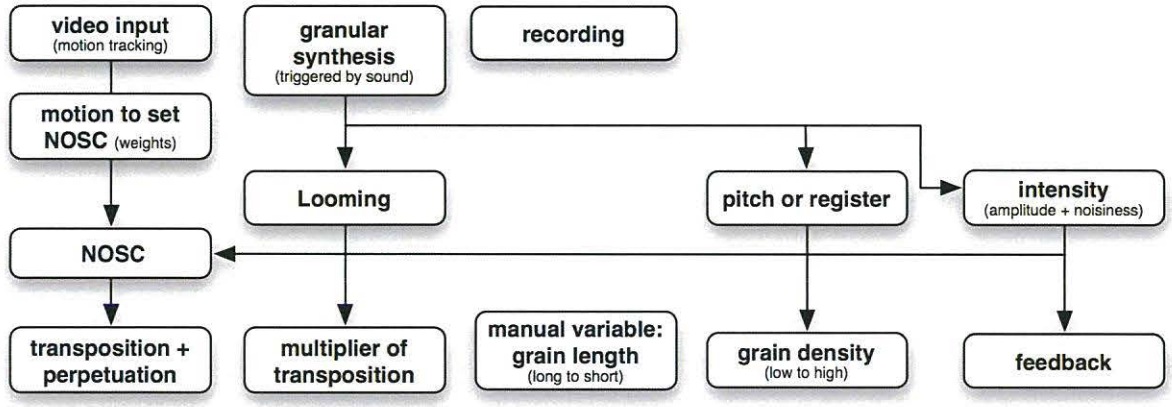

Figure 8. This diagram shows the programming architecture: data flows from input gesture and sound and affects the Looming and granular synthesis engine. The synthesized sound is fed through the Oscillator to perpetuate and generate new material that derives its pitch origins from the real time *shakuhachi* pitch analysis and oscillator weights from computer vision-sensed motion (gesture). Loudness and noisiness are used in the scaling of the Ecological Looming effect and to control variables in the grain density and transposition in the granular synthesis.

6. VISUALISATION

One of the differentiating features of *Hyper-Shaku* in the genre of hyper-instruments, is that it augments the musical performance with automatically generated real time visualisation. While this process is comparatively low priority alongside the auditory generation and performance, visual display adds to the multimedia projectable scope of the work and adds a layer of representation to the gestural interaction, changing focus from conventional instrumental performance on the performer alone to performer and machine. The visualisation can operate in two modes. The simplest involves a live video filter interpolated image of the performer processed immediately on the web cam input (figure 9). For this purpose an abstract Jitter filter that divides the image into rectangular shapes with dynamic colour mapping is implemented. Precise scale, size and dimension of these shapes can be adjusted in the object properties to achieve different appearances. The intention of this method is an ambient visualisation that clearly connects with the performer, likely to be recognised by an audience relating motion and colour to the source. It does not require much imagination to understand how the image is produced and it is extremely kinetic, sensitive to even the slightest motion but has the advantage of retaining a sense of "visibility" of the performer. Amount of head movement is used to scale the grid size of the video filter resulting in appearance changes that resemble changing "resolution" or clarity/abstraction according to amount of motion *(y*-distance). This is a simple bi-modal reinforcement in which the visualisation utilizes a visual input of the gesture directly, the same gesture capture that sends data metrics to the sound processing and A.I. modules of the system. Thus it is not independent or generative.

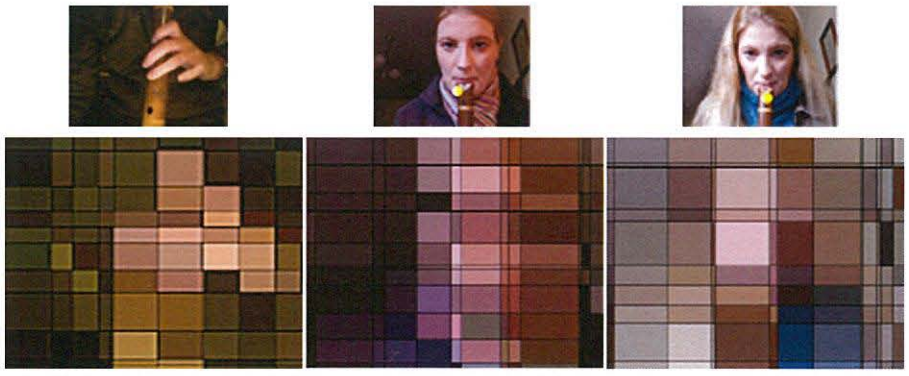

Figure 9. Visualisation using a real-time video filter in Jitter. Amount of motion scales the "resolution" or grid-size of the pattern.

A contrasting approach to visualisation takes the output following all generative processing and uses these values to map onto an abstract visualisation system. It is far more discreet and hence, perhaps, engaging or mysterious, than the first method, though its connection to the performer is more obscure or removed. Choice of method must depend on the implementation context. The latter method highlights obvious parallels between particle systems and granular systems, for example, by mapping granular characteristics to particle system flow characteristics, such as density, distribution, speed (rate of motion) and particle size to granular synthesis. This second approach is bi-modal using different data to activate the auditory processing and visual processing (though working from the same capture source). This second approach results in visualisation that closely matches the auditory outcome whereas the previous method results in visualisation that matches the visual input. Depending on the magnitude of effect of the generative modules, those states can be quite different.

7. INTERACTION MODULES (INFORMATION CAPTURE)

Following is an overview of the technologies used for computer vision gesture capture, computer listening audio capture and the on-stage hardware configuration minimum requirements for performance.

7.1. Computer Vision – Gesture Capture

The physical nature of playing the *shakuhachi* makes it especially suitable for motion triggering since pitch inflection is achieved by moving the head and angling the chin relative to the instrument, in addition to fingering and upper body movement typical when performing an instrument. Traditional live music-processing approaches analyse and synthesise real time musical response from the musical (audio) content of a performer. The approach of this project, in contrast, focuses on the gestural/spatial and theatrical nature of *shakuhachi* performance. Jean-Marc Pelletier's (Pelletier 2005) "cv.jit" objects (computer vision externals library for Max/MSP + Jitter software) include optical flow tracking, statistical calculators, image transformations and image analysis tools, all drawing data from video such as the simple web-cam used in this interface. When a region from the camera-view is selected, the optical flow tracking follows the coordinates of that region within the frame. The *y* value is extracted to track up and down head motion of bend, "head-shake" and *vibrato* actions. Only vertical gesture data is extracted, not affected by the player's movement from side to side.

7.2. Computer Listening — Audio Analysis

Audio input from the performer is captured using a condenser microphone positioned close to the player. The ratio of breathiness or noisiness to sound and of overtone spectrum in the sound contributes to the noisiness analysis filter object (Jehan 2001). Thus the positioning of the microphone and performer's distance from it can influence the results. In the prototype (figure 10), an alternative input mechanism for playing sound files allows any sound file to be loaded and played through the responsive generative system for the purpose of testing, calibration, fine-tuning and choosing pre-sets on manual controls. The "adc" object is where the user selects and configures the DSP (Digital Signal Processing – audio handling hardware). An external microphone, built-in microphone or input from external sound interface can be used in preference to the computer's inbuilt sound card. Output destination (e.g. to external speakers) is also selected in the DSP options in Max/MSP.

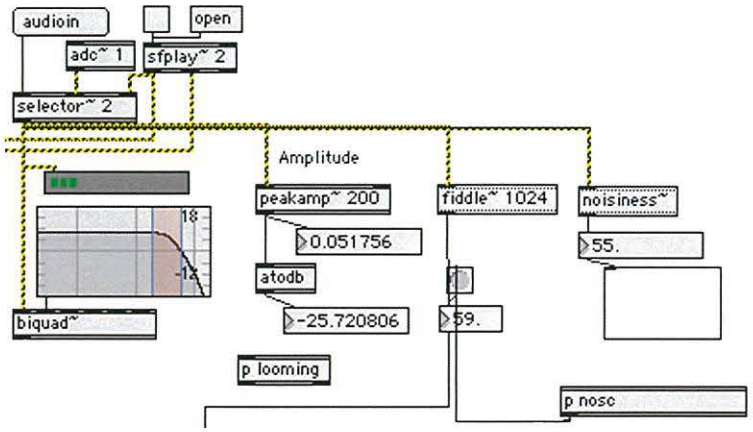

Figure 10. Microphone input for the integrated biologically inspired generative systems triggered by gestural interaction (computer vision and listening). Its loudness and noisiness values are extracted for processing.

8. CONCLUSION

This paper presented a contextual example of a computational creative system in order to illustrate ways in which gestural data can be mapped onto a Neural Oscillator Network system, utilising audio and gesture input data to trigger and scale values in granular synthesis and Looming processes to produce immediate multi-modal representation. Although the implementation here is targeted towards musical performance, the versatility of generative sonification and the ubiquity of gestural source data are widely applicable in other contexts. The system described is modular in design, in which different generative processes; different gesture, sound and data input capture methods; can be substituted in order to apply the real-time generative mechanism to other situations.

ACKNOWLEDGEMENTS

This research was supported by an Australian Research Council Grant DP0773107 and the University of Sydney Research & Development Grant Scheme. The author gratefully acknowledges the support of a Matsumae International Foundation Research Fellowship and Professor Koichi Hori, University of Tokyo (RCAST) A.I. Lab. Thanks to research assistant and programmer, Sam Ferguson.

REFERENCES

Adamatzky, A., & Komosinski, M. (2005) Artificial Life Models in Software, New York: Springer

Ariza, C. (2007) "Automata Bending: Applications of Dynamic Mutation and Dynamic Rules in Modular One-Dimensional Cellular Automata" in Computer Music Journal 31(1), 29-49, MA: MIT Press

Arons, B. (1992) "A Review of the Cocktail Party Effect" in Journal of the American Voice I/O Society, 12, 35–50

Beilharz, K., Ferguson, S., & Song, H.J. (2006) Fluid Velocity (Interactive Installation), Sydney: Tin Sheds Gallery, University of Sydney

Bly, S. (1982) Sound and Computer Information Presentation (Ph.D Thesis) University of California

Burraston, D. & Edmonds, E. (2005) "Cellular Automata in Generative Electronic Music and Sonic Art: A Historical and Technical Review" in Digital Creativity 16(3),165-185

Eldridge, A. (2005) NOSC (Neural Oscillator Nodes) Open-source Max/MSP (patches & abstractions), Sussex, UK: Creative Systems Lab, Evolutionary and Adaptive Systems Group, Department of Informatics, University of Sussex

Flety, E. (2005) WiSeBox (Hardware & Software Sensor Controller), Paris: IRCAM

Franklin, S. (1999) Artificial Minds, Cambridge, MA: MIT Press

Harris, E.S., Wuensche, A., & Kauffman, S. (1997) "Biased Eukaryotic Gene Regulation Rules Suggest Genome Behaviour is Near Edge of Chaos" in Working Paper, Santa Fe, NM: Santa Fe Institute

Jakovich, J. & Beilharz, K. (2005) Sonic Tai Chi, (installation), Sydney: BetaSpace, Sydney Powerhouse Museum

Jehan, T. (2001) Perceptual Synthesis Engine: An Audio-Driven Timbre Generator (Master's Thesis), Cambridge, MA: MIT (Media Lab)

Kauffman, S. (1993) The Origins of Order, Oxford, UK: Oxford University Press

Kerr, A. (1996) Lost Japan, Melbourne: Lonely Planet, Australia

Kirkpatrick, B. (2004) Churchill Fellowship in Japan Studying Shakuhachi, Sydney, Australia: Churchill Memorial Trust

Lee, D.N., Vanderweel, F.R., Hitchcock, T., Matejowsky, E., & Pettigrew, J.D. (1992) "Common Principle of Guidance by Echolocation and Vision" in Journal of Comparative Physiology, (Sensory, Neural, and Behavioral Physiology) 171, 563–571

Lodha, S.K., Wilson, C.M., & Sheehan, R.E. (1996) "Listen: Sounding Uncertainty Visualisation" in Proceedings of the Visualisation '96 Conference, Washington, DC: IEEE Computer Society Press

Matsuoka, K. (1985) "Sustained Oscillations Generated by Mutually Inhibiting Neurons with Adaptation' in Biological Cybernetics 52, 367-376, Berlin/Heidelberg, Germany: Springer-Verlag

Neuhoff, J.G. (1998) "Perceptual Bias for Rising Tones" in Nature 395(6698), 123–124

Neuhoff, J.G. (2001) "An Adaptive Bias in the Perception of Looming Auditory Motion" in Ecological Psychology 132, 87–110

Neuhoff, J.G., & Heckel, T. (2004) "Sex Differences in Perceiving Auditory "Looming" Produced By Acoustic Intensity Change" in Proceedings of the 10th Meeting of the International Conference on Auditory Display, Sydney, Australia

Pelletier, J.-M. (2005) CV.jit (Software Library for Max/MSP Computer Vision), Gifu: IAMAS, Japan

Puckette, M., & Zicarelli, D. (1990-2005) Max/MSP + Jitter (software), California: Cycling '74 / IRCAM

Roads, C. (2001) Microsound, Cambridge MA: MIT Press

Spieth, W., Curtis, J., & Webster, J.C. (1954) "Responding to One of Two Simultaneous Messages" in Journal of the Acoustical Society of America 26(1), 391–396

Volpe, C.R. (2002) Algorithms for Aural Representation and Presentation of Quantitative Data to Complement and Enhance Data Visualization (Ph.D Thesis - Computer Science), Troy, NY: Graduate Faculty of Rensselaer Polytechnic Institute

Williamson, M.M. (1999) "Designing Rhythmic Motions Using Neural Oscillators" in Proceedings of the International Conference on Intelligent Robot and Systems, Vol.1, pp. 494-500

Wuensche, A., & Lesser, M.J. (1992) The Global Dynamics of Cellular Automata, Reading, MA: Addison-Wesley

Wuensche, A. (1999) "Classifying Cellular Automata Automatically: Finding Gliders, Filtering, and Relating Space-Time Patterns, Attractor Basins, and the *Z* Parameter" in Complexity 4(3), 47-66

Xenakis, I. (1971) Formalized Music: Thought and Mathematics in Composition, Bloomington, IN: Indiana University Press

Xenakis, I. (1990) "Sieves" in Perspectives of New Music 28(1), 58-78