Interactive instrumental performance and gesture sonification

Kirsty Beilharz

Synopsis
Refereed articles
Information articles
Notes on contributors
Print friendly version

Introduction

In the broader context of Media and Cultural Studies, sonification fits into the category of non-visual visualisation, i.e. sonification is concerned with using sound to convey an informative representation of data. Auditory graphing, visualisation and graphing (in general) are all forms of representation that seek to expose new discoveries or reveal patterns, trends, information features through expression in a modality other than its original abstract textual source. Further, using motion, gesture and breath data views sonification as an instrument of art-making, expression and performance. This paper explores gestural interaction and generative processes of perpetuation as methodologies for live performance. Similar approaches to information-driven or gesture-driven mapping and augmentation could be applied in visual expression and the interactive generative systems produce user-centred live material. The generative models could form the basis of visual, auditory or other modes of new and continual material generation.

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.

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.

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.

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).

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.

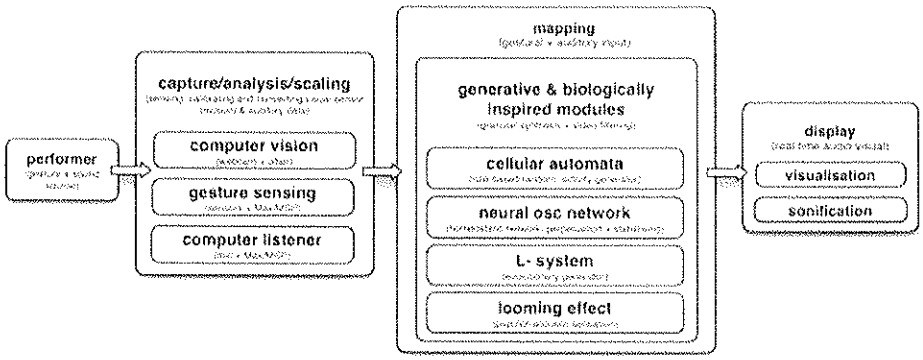


Figure 3. The modular approach to hyper instrument design using biologically inspired generative computation for real time gestural interaction that allows for individually customised and scalable performance scenarios.

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.

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.

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.

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

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.

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 changes in 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.

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: 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.

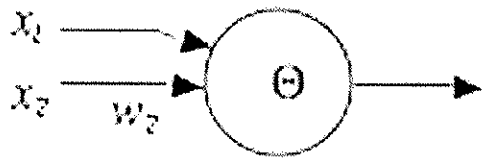


Figure 6. Neural Oscillator Network model using four synapse nodes to disperse sounds, audibly dissipating but rhythmic and energetic. Irregularity is controlled by head motion tracked through the computer vision. Transposition and pitch class arrives via the granular synthesis from pitch analysis of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

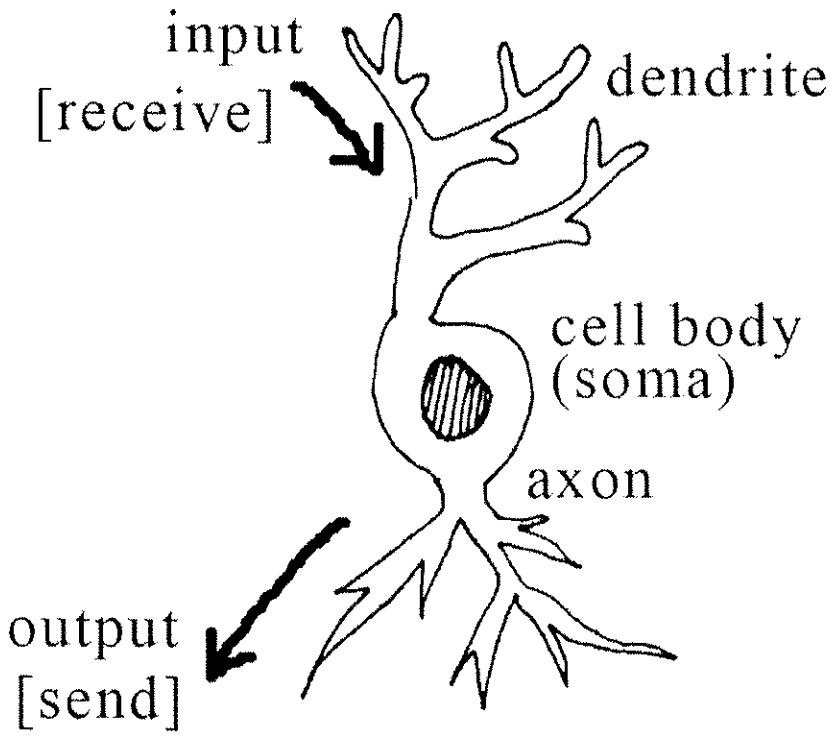


Figure 7. The Max/MSP Neural Oscillator Network patch (figure 6) is used as a stabilising influence affected by large camera-tracked gestures. It is modelled on individual neurons: dendrites receive impulses and when the critical threshold is reached in the cell body (soma), output is sent to other nodes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

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.

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.

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.

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.

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.

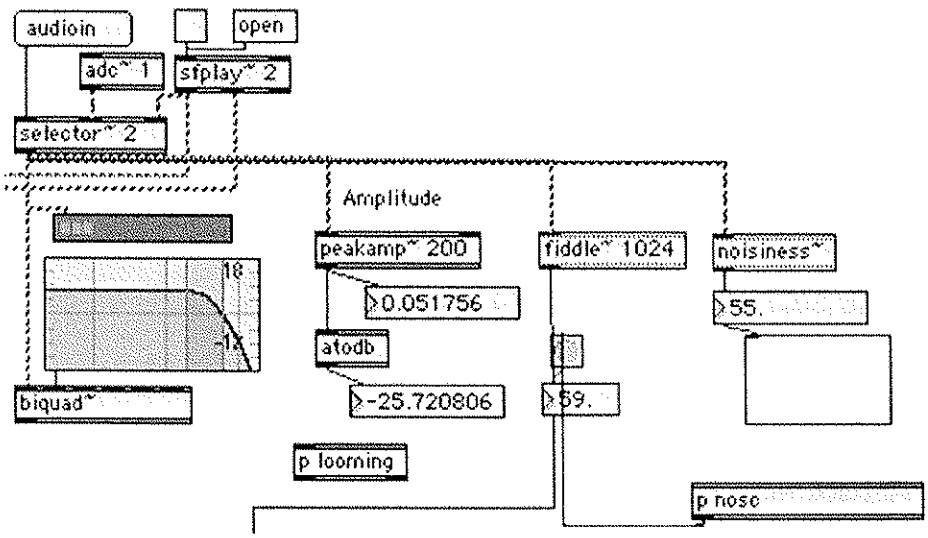


Figure 10. Microphone input for the integrated biologically inspired generative systems triggered by gestural interaction (computer vision and listening). Its loudness and noisiness values are extracted for processing.

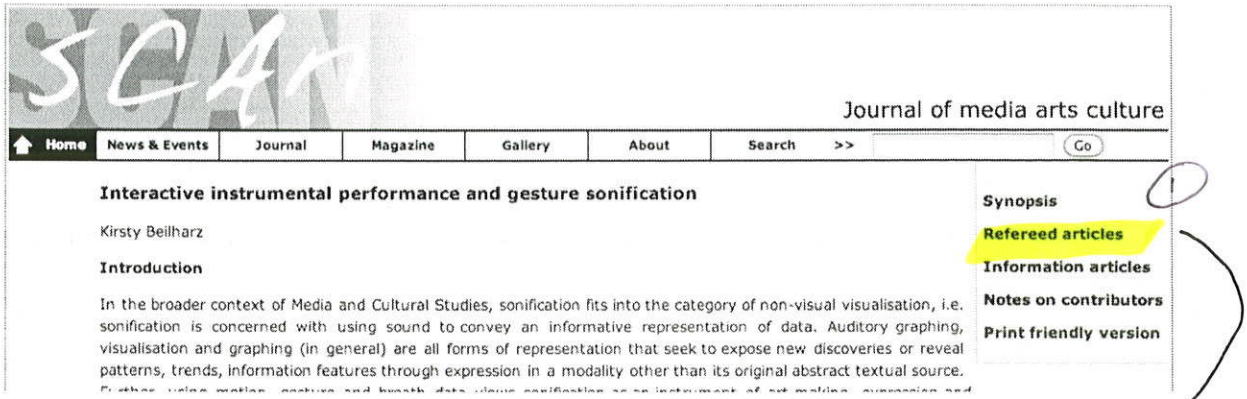
Conclusion

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.

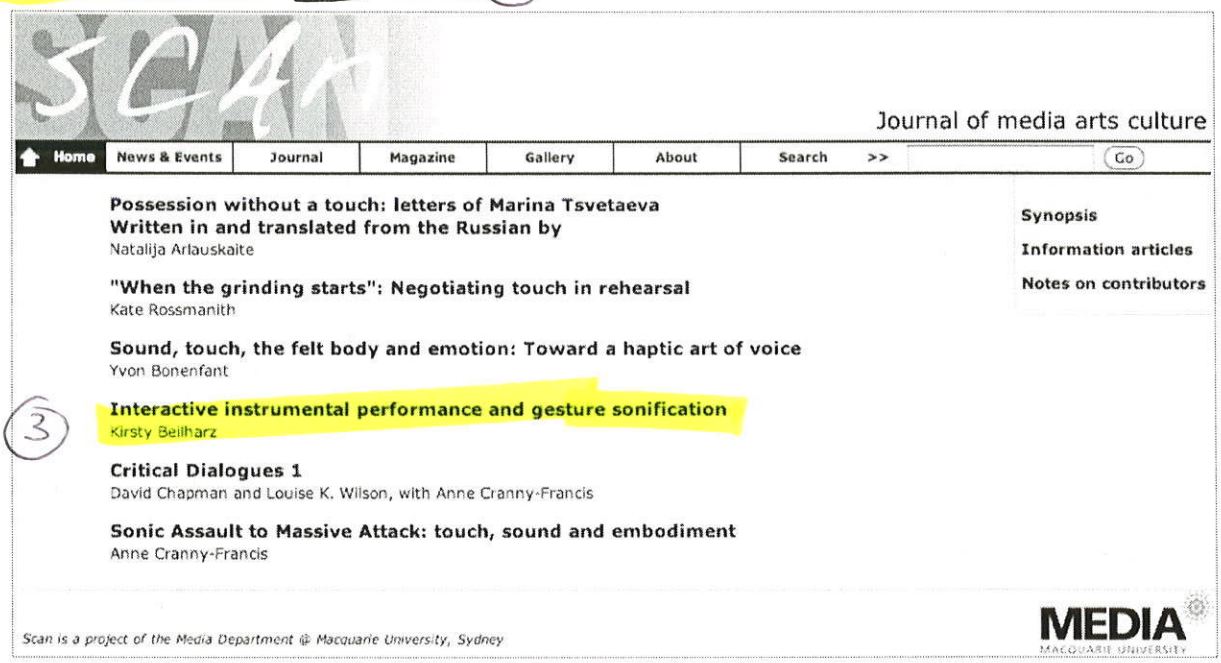
Acknowledgements

Press

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



http://scan.net.au/scan/journal/display_refereed.php?j_id=15 (2)



Scan is a refereed on-line journal (ISSN 1449-1818) devoted to the media arts and culture, hosted by the **Media Department** at **Macquarie University**, Sydney.

Its approach is inter-disciplinary, as is its subject matter. **Scan** draws on media studies, cultural studies, media law, information and technology studies, fine arts and philosophy. **Scan** considers developments in new media, digital art, screen arts, music and audio arts, as well as the culture enveloping these practices and technologies.

Scan is concerned with both the aesthetics and the political economy of media arts, as practised in both new and traditional media forms. **Scan** will be published electronically three times a year. Each issue will be thematic, comprising 6-10 articles, with a maximum word-length of 6,000 words.

Current Issue:

Vol 6 Number 3 December 2009

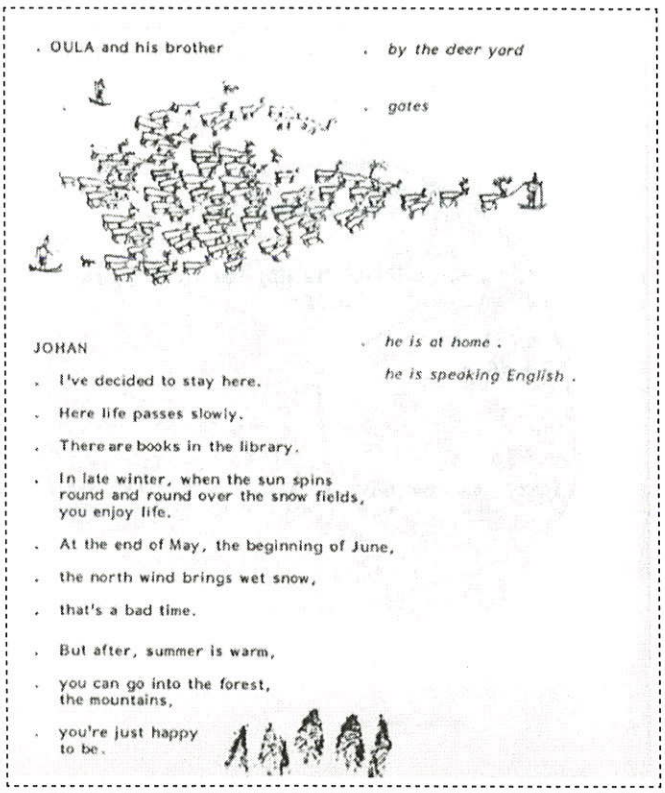
Authorship and the Documentary
edited by Maree Delofski & Virginia Madsen

Synopsis

Refereed articles

Information articles

Notes on contributors



Previous Issues:

Vol 6 Number 2 September 2009
Sound.Music.Design
edited by Norie Neumark

Vol 6 Number 1 June 2009
Reading Between the Panels (Part II)
edited by Steve Collins

Vol 5 Number 3 December 2008
Biopolitics of the senses: touch, sound and embodied being
edited by Anne Cranny-Francis

Vol 5 Number 2 September 2008
Reading Between the Panels (Part I)
edited by Steve Collins & Can Yalcinkaya

Vol 5 Number 1 May 2008
Screenscapes Past Present Future
edited by Chris Chesher, Peter Marks, Kathy Cleland

Vol 4 Number 3 December 2007
Mobile Media/Public Spaces
edited by John Potts

Vol 4 Number 2 August 2007