ParticleTecture: Interactive Granular Soundspaces for Architectural Design

Joanne Jakovich
Key Centre of Design Computing and Cognition
University of Sydney
+61 2 9036 5496
Joanne@Jakovich.net

Kirsty Beilharz
Key Centre of Design Computing and Cognition
University of Sydney
+61 2 9351 4031
Kirsty@arch.usyd.edu.au

ABSTRACT
Architectural space is a key contributor to the perceptual world we experience daily. We present 'ParticleTecture', a soundspace installation system that extends spatial perception of ordinary architectural space through gestural interaction with sound in space. ParticleTecture employs a particle metaphor to produce granular synthesis soundspaces in response to video-tracking of human movement. It incorporates an adaptive mechanism that utilizes a measure of engagement to inform ongoing audio patterns in response to human activity. By identifying engaging features in its response, the system is able to predict, pre-empt and shape its evolving responses in accordance with the most engaging, compelling, interesting attributes of the active environment. An implementation of ParticleTecture for gallery installation is presented and discussed as one form of architectural space.

Keywords
Architecture, installation, interaction, granular synthesis, adaptation, engagement.

1. INTRODUCTION
The built environment contributes to our sense of space, shaping the way we perceive distance, orientation, movement, and size. Conventionally, Architecture as a discipline has dealt with design of spaces based on a Euclidean spatial model, overlooking the rich potential of perceptual models of space, e.g auditory space, kinesthetic space. While the finite geometry of the Euclidean model was well suited to the Modernist notion of building as 'functional machine', it is not easily transferred to the present paradigm of intelligent architecture and transformable, reactive spaces [13].

In the domain of installation art on the other hand, the creation of spaces that activate auditory, visual or even kinesthetic modes of perception is explored as a means to surprise, entertain or engage viewers [6]. Looking further to interactive art, extra-ordinary experiences of a space are achieved through realtime manipulation of the participant's perceptual expectations through interactive interfaces. In his pioneering 'responsive environment' of 1977, Myron Krueger observed the behavior of participants:

> "In the environment, the participant is confronted with a completely new kind of experience. He is stripped of his informed expectations and forced to deal with the moment in its own terms. He is actively involved, discovering that his limbs have been given new meaning and that he can express himself in new ways." [8]

In this and ensuing interactive spaces, *interaction* is a fluid medium in which any given perceptual modality might be confounded by stimuli in another modality: things seen may contradict things heard, or vice versa [7]. Furthermore, the point of intersection of any two modalities becomes a new hybrid modality that can likewise be designed for, and subsequently conceived as an alternative model of space.

1.1 Intermodal Mapping: Intermodal Space
We define 'intermodal space' as the conceptual model of space created through the conflation of two or more perceptual modalities. The conflation is enabled via a realtime cross-mapping process that transforms input from one modality to output in another. This produces a perceived hyper-space distinct from other perceptual models of space such as 'auditory space', 'visual space', or 'kinesthetic space'.

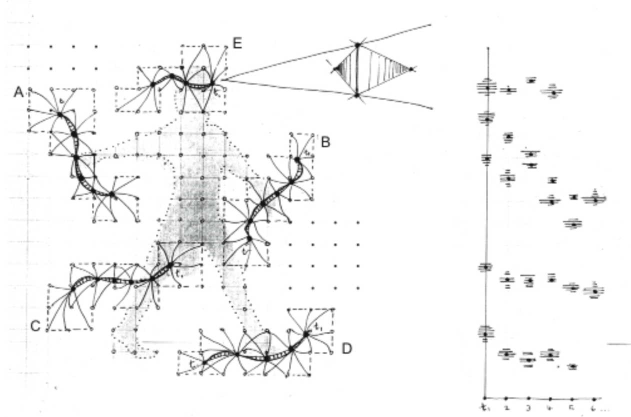

Figure 1. Conceptual diagram of an intermodal space model: Space as audio-kinesthetic perception. Gestural action of arms (A,B), legs (C,D) and head (E) is mapped to a granular synthesis soundspace, illustrated over 6 iterations.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
NIME07, June 7-9, 2007, New York, NY
Copyright remains with the author(s).

Figure 1 illustrates an intermodal concept for space based on the conflation of kinesthesia and hearing, or 'audio-gesture'. Through gesturing, a spatially distributed Granular Synthesis[1] lattice is activated and gestures are translated to sounds, reinforcing the relationship between one's location and the space.

In this paper we develop an approach for the design of aesthetic architectural spaces based on the perception of 'audio-gestures' within space. We incorporate a decentralized adaptive mechanism that utilizes a measure of engagement to inform ongoing audio patterns in response to gesturing. The following section describes precedent works that cross the fields of art, architecture and sound, focusing on methods of intermodal mapping and adaptation. Section 3 describes an integrated ParticleTecture system for the creation of space through audio-kinesthetic intermodal mapping. Section 4 demonstrates an example implementation of the system and section 5 discusses its outcomes which are the basis for future work.

2. PREVIOUS WORK

2.1 Hybrid Architecture

The 1958 Philips Pavilion was arguably the first physical space designed to extend spatial experience through the temporal manipulation of sound and light in a custom built environment [19].

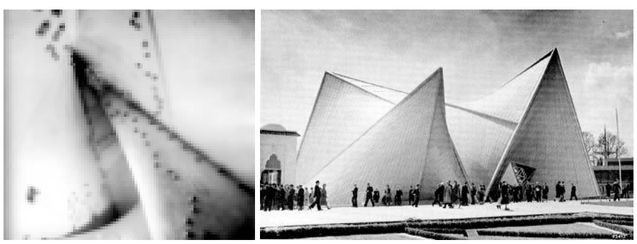

Figure 2. The Philips Pavilion (1958, Le Corbusier / Xenakis) was arguably the first multimodal architecture.

Contemporary hybrid spaces by both artists and architects explore the potential for new digital media to represent the elements that architecture had traditionally dealt with. Satoru Yamashiro (Figure 3), Ryoji Ikeda (Figure 5), and Garth Paine are among those exploring the role of the body and perception in interactive installation soundspaces [5, 14, 15]. The significance of the interactive form is that it enables an ongoing action-effect dialogue between the participant and the system producing a continually novel artifact [2, 14].

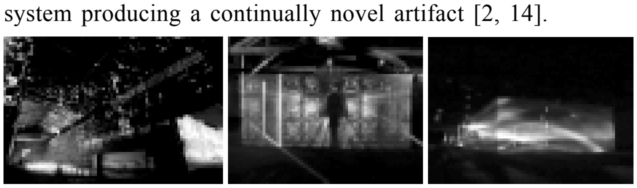

Figure 3. Installations by Satoru Yamashiro and the Responsive Environment Group (2004, 2006).

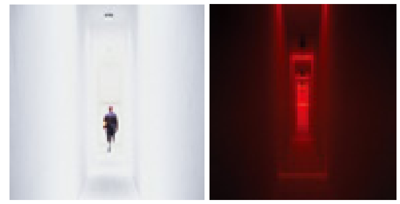

Figure 4. Installation environment Spectra II by Ryoji Ikeda radically extends perceptual expectations of space.

2.2 Approaches to Mapping & Adaptation

In interactive soundspace installation, the mapping is the procedure that translates gestural input to audio output. The mapping largely defines the esthetic expression of the space, and simultaneously plays a significant role in audience engagement. A simple one-to-one mapping might enable initial engagement, but potentially lack the complexity to maintain user interest beyond initial mastery. At the same time, a complex mapping may confound the user, yet motivate them to search for the underlying rules of the system [1, 2, 14].

Adaptive mapping is a computational method that utilizes feedback to continuously transform a mapping, enabling continuously novel output in response to the behavior of the participant [21]. Foundational models of adaptive mapping include neural networks [3], interactive genetic algorithms [9], multi-agent systems [10], cellular automata [12] and self-organizing markov models [20]. These implement a decentralized approach to mapping that enables both varied complexity and emergence based on simple rules. In the following ParticleTecture system, the goal of adaptation is to address varying degrees of participant engagement through incremental modification of a distributed gesture-to-soundspace mapping.

3. PARTICLE*TECTURE* SYSTEM

In the ParticleTecture system (Figure 5), physical space is analyzed via digital video feed. Bodies in motion are translated to a video analysis engine called the Gesture Pixelspace. Following, the extracted gestural patterns activate a spatialized granular synthesis engine called the Sonic Grainspace. The resultant soundspace proliferates through the space as a real-time response to the gestural motion enhancing audio-kinesthetic spatial awareness. The mapping methods, adaptation method and index of engagement are described in the following.

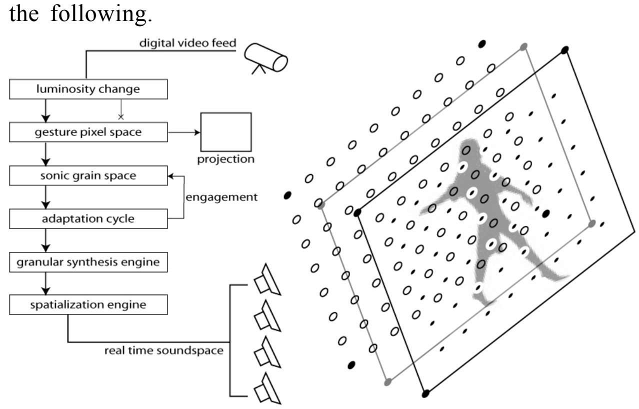

Figure 5. The ParticleTecture system overview. Video feed from the space is analyzed in the Pixelspace and gestural

[1] Granular Synthesis is a method of sound synthesis that utilizes 'grains' or particles of microsounds of duration less than 100 milliseconds to form longer, complex, layered sounds. [17]

activity is translated to the Grainspace to synthesize a granular synthesis soundspace.

3.1 Mapping Method

The following describes an original mapping method within which adaptation and engagement in the ParticleTecture system is structured. In order for a mapping to be significant, it needs to operate on representations that are appropriate to the application. In soundspace installation, there are several fundamental considerations that influence how gesture and sound are represented: gesture is a continuous motion of the limbs in relation to the body and of the body in space [14]; gesture occurs in space over time [11]; sound occurs in space over time; (but) the speed of sound is approximately 300 times that of gesture; gesturing may simultaneously occur in two or more places (e.g. two or more participants); sound is not music, and a soundspace is not an instrument, hence musical structures (rhythm, melody, etc) are not relevant [15]; the cognitive association of sound to a gesture requires a minimal latency from input to output [18]. A two-part mapping framework that addresses these considerations for soundspace installation is developed below. It is designed to integrate a cellular, decentralized model of adaptation.

3.1.1 Sonic Grainspace

The sonic grainspace is a method for sound representation that builds on granular synthesis theory [17, 22]. In granular synthesis, 'grains' are microacoustic events of duration less than 100 milliseconds that can be used to produce complex sounds by the arrangement of many grains of varying pitch and/or waveform [17, 22]. In the proposed sonic grainspace (hereafter referred to as grainspace), grains are reactive units with a fixed location in a virtual lattice space. Grains activate, at which point they display a sound, when triggered by gesturing. The capacity for the grains to change configuration in relation to other grains enables significant patterns of continuity to form within the lattice. This capacity is specified through a process of adaptation discussed below.

The significance of the grainspace is twofold. Firstly, the composition of sound is directly related to the path the gesture takes through the lattice, which implies the quality and density of a complex sound is a product of the speed and path of a gesture. Secondly, the mechanisms that control the sound output, and their adaptive capacity, are not globally specified in what is usually termed the 'mapping'. Rather all operations that effect the global sound generation occur within and between these units. The way the soundspace evolves over time is a function of the interactions between grains in the lattice as activated by human gesture.

3.1.2 Gesture Pixelspace

The gesture pixelspace is introduced as a method for representing information about gestural activity captured by video sensing in a soundspace installation. This research adopts the definition of gesture as human movement that is characterized by the smooth, continuously changing relationship of the limb(s) to the body [14]. The gesture pixelspace (hereafter referred to as pixelspace) is a representation of the history of pixel fluctuations from a digital video feed that updates in real time. It is based on computer vision and image processing techniques that abstract information about motion in video sequences by comparing changes in pixel values over frames [16]. The pixelspace represents gestural patterns (current and past) in individual cells, and in the connections betweens cells. There are no predefined classes of gesture; rather cells build up information about what degree of pixel change is significant through experience and local exchange. Cells store information about both the current optical flow and the experienced optical flow.

This approach differs from the prevalent paradigm of gesture 'recognition', which aims to mathematically abstract motions of gestures into geometric (shape, contour, velocity) representations, and classify incoming gestures accordingly. In contrast, the pixelspace accumulates information about the frequency, speed and direction of movement within each individual cell. This information can be used to estimate patterns of movement in the future. In this way the spatial location of gesture patterns is given importance over shape, which implies that participant actions are significant within the spatial context of the soundspace installation rather than in isolation. In computational applications, this approach is demonstrated by the use of cellular automata [12], neural networks [3], and multi-agent systems [10], in which the macro behavior of a system is not specified but is an emergent feature of interactions on a local scale. Thus, low-level abstraction of gestural patterns is achieved without higher-level concerns for segmentation and classification. In addition, the small units of information about gesture are directly transferable to the grainspace (Figure 6) without interpretation or recognition.

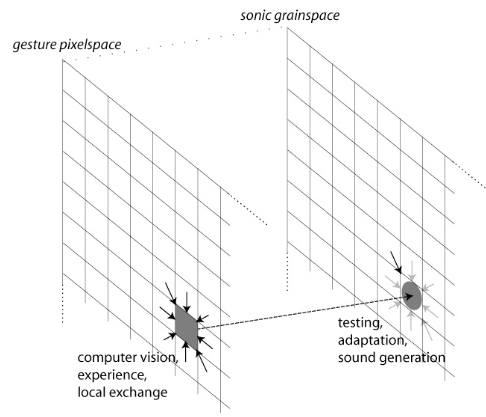

Figure 6. The gesture pixelspace processes incoming streaming video and sends activated pixel information to the sonic grainspace, where testing, adaptation and sound generation occurs.

3.2 Adaptation Method

This section develops an adaptation method in which each sound grain undergoes several stages of evaluation, at which point if it fails, it has the capacity to adapt based on input from successful neighbors. This draws on the work of Miranda [12] and McCormack [10] who used combinations of cellular automata and genetic algorithms to develop adaptive music and art systems.

3.2.1 Cell Architecture

In the proposed decentralized system all operations are performed on the cellular level. A single cell encompasses the sequence of processes from video input (pixel) though to audio output (grain), connecting the pixelspace to the grainspace, and implementing an adaptation cycle in the adaptive feature generator (described below). Each stage represents a decision process that enables or disables the flow of data. This ensures that information only flows through cells activated by gesture, distinguishing this system from a cellular automata composition system [12].

In the pixelspace, the current gesture features of the individual cell are determined. Real time digital video data is tested for luminosity change in the short term and then long term to determine the significance of gestural movement. Local exchange between neighboring cells provides information about the direction and speed of the flow of movement. The current gesture features (direction, speed and significance) are forwarded to the grainspace. In the grainspace, the sonic features are based on the parameters of granular synthesis [17]: the fundamental frequency of the waveform (referred to from this point as pitch); the waveform (or timbre); the duration of the grain (in milliseconds); and the amplitude (or loudness). Sonic features are assigned to specific gesture features using a set of heuristics: waveform is mapped to direction, pitch to speed and duration to significance. In the grainspace, a sequence is undertaken where each pixel's features are compared against its associated grain's features. If the mappings are compatible, then the sonic features are implemented and a sound is produced. If not, the grain initiates an adaptive cycle in the adaptive feature generator.

3.2.2 Adaptive Feature Generation

The adaptive feature generator employs a genetic algorithm based on Miranda [12] to incrementally adapt the feature sets of grains using neighborhood selection. The fundamental goal of a grain is to increase its chance of emitting sound by adapting its ability to map to incoming gesture patterns. Because human gesturing is not consistent or predictable, it is useful for the grain to be able to generate new feature sets to increase its chance of compatibility.

Over time, grains that are incompatible adapt their features to be more susceptible to activation by the gesture patterns. Within a single cell this may involve a constant push and pull behavior between two extremes, but on the macro level patterns of similarity and continuity between neighboring grains form. Clearly, if the grainspace evolves new features more rapidly than input from the pixelspace is received, then the degree of compatibility of the grains is not verified. Patterns of similarity and continuity might emerge, but based on limited input. Similarly, the way the grainspace evolves new sounds requires evaluation, since it may produce increasingly incompatible sounds. This is amplified by the tendency for participants to have varied preference towards certain sound qualities, which cannot be predicted by the grainspace. Therefore adaptation operators need to correspond to some relevant performance indicators. The following section develops a set of indicators, called the index of engagement that can be used by the adaptive feature generator to moderate the process of evolving new grains.

3.3 Engagement

Artists employ adaptive mapping systems based on the intuition that a system should be able to adapt to the patterns of its participants in order to be engaging [2, 14]. In research, an empirical approach to analyzing this assumption has not been established. A means for comparing computational adaptation against participant responses is required in order to confirm the relationship between adaptation and engagement. This section presents a definition and method for analyzing engagement within the computational art system as a quantitative measure for cross-reference with qualitative responses from participants.

Engagement is usually concerned with the degree of novelty, interest or satisfaction experienced by the participant in response to the interactive art system. Here, we take into account the symbiotic relationship between the participant and art system to establish a measure of engagement based on a computational analysis of interaction. As described by Paine [14] interaction can be likened to an ongoing dialogue. He writes: “each party constantly monitors the other’s response and uses this information to make alterations to their own response strategy” (p.297). Using this metaphor, it is evident that engagement is a property of not only one speaker, or the other, in isolation, but of the mutual effect of successive responses. The computational index of engagement developed in this research observes the symbiotic changes in the pixelspace and grainspace as representative of the ongoing dialogue of interaction.

The computational index cannot measure the same qualities that the human survey reports, namely it is impossible for a computer to measure satisfaction, intrigue and similarly subjective and internal notions. However, the computer can observe behavioral consequences that might equate to engagement, such as level of activity, repetitive actions, preferred sounds, relationships between action and time, that is sustainability of threads over time, length of commitment, and so on. Developing the index commences with identifying these dimensions that can be measured or sensed by the computational system and used as the basis for adaptive decision-making.

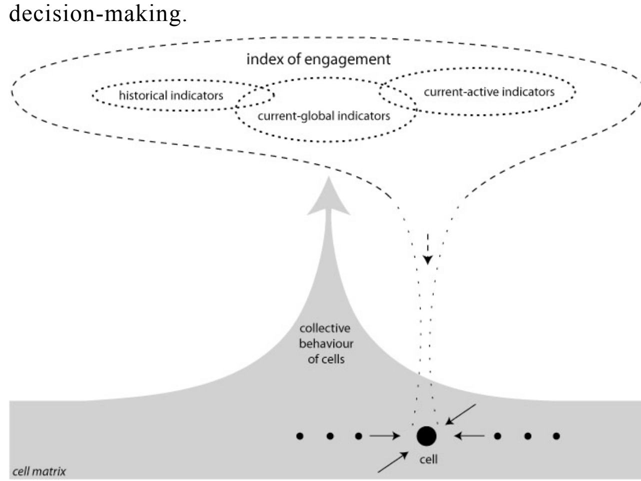

Figure 7. The index of engagement is a composite of historical indicators, current-global indicators and current-active indicators of engagement patterns.

3.3.1 Grain Engagement

The collective behavior of the grains over time is the basis for the index of engagement. A grain’s individual activity status is the fundamental unit for the index. A grain’s goal for adaptation is to be engaged. On the cellular level ‘engaged’ has a simple meaning: to become activated or operational. That is, if a grain is engaged it emits a sound. An engaged grain is one whose grain features are aligned to incoming gesture features.

3.3.2 Index of Engagement

A set of indicators collectively contributes to a computational index of engagement (Figure 7). These monitor how individual incremental adaptive processes influence large-scale interaction. The core assumption used to develop the index is that the output of the grainspace exerts some influence on the incoming patterns in the pixelspace, and that the timing of this influence is significant. That is, it is assumed that gesturing in the participant environment is in response to real time changes in audio output. Based on this assumption, observations are made by comparing the changes in the pixelspace with changes in the grainspace. There are three scales in which change is monitored: its change since initialization of the soundspace (historical); change within a short-term period (e.g. 12 frames) across the whole cell lattice (current-global); and change within a short-term period only within cells that are activated by gestural motion (current-active). Indicators contributing to the index include: the rate at which gesturing patterns evolve compared to that of the audio response; the degree the grains remain scattered (white noise) or clusters emerge (perceivable sound); the temporal relationship between overall gesturing activity and ordering of sound.

4. IMPLEMENTATION

4.1 Motion-activated Pixelspace with Game of Life Grainspace

'Sonic Tai Chi' (Figure 8) is designed for spatial interaction by the general public in a gallery setting. It is installed in an approximately 25m-square room with a rear projection screen, stereo speakers hidden in the ceiling and camera concealed below the screen.

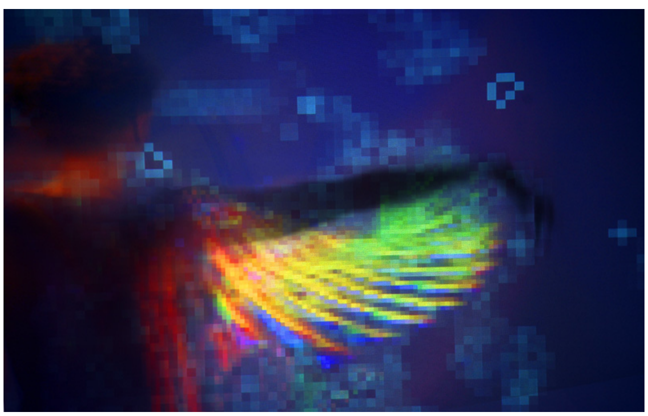

Figure 8. *Sonic Tai Chi* by Jakovich and Beilharz (Beta_space, Sydney Powerhouse Museum, installation 2005-2006) uses computer vision to capture movement data that produces the granular sonification. It uses generative Cellular Automata rules to propagate particles in response to users' lateral motion.

Sonic Tai Chi implements several stages of pixel evaluation using Pelletier's Computer Vision 'cv.jit'[2] objects in Jitter. The objects use the Horn-Schunk method to calculate the optical flow of pixels captured using a simple Internet conference camera. Optical flow refers to the amount of luminosity change in the pixel over a given number of frames. In each stage, pixels with a luminosity change over a given threshold are selected for further evaluation. The collection of pixels with the greatest luminosity change represents areas in the soundspace with the greatest movement, typically humans. This enables a non-intrusive method for determining gestural patterns without the need for gloves or markers (Figure 9).

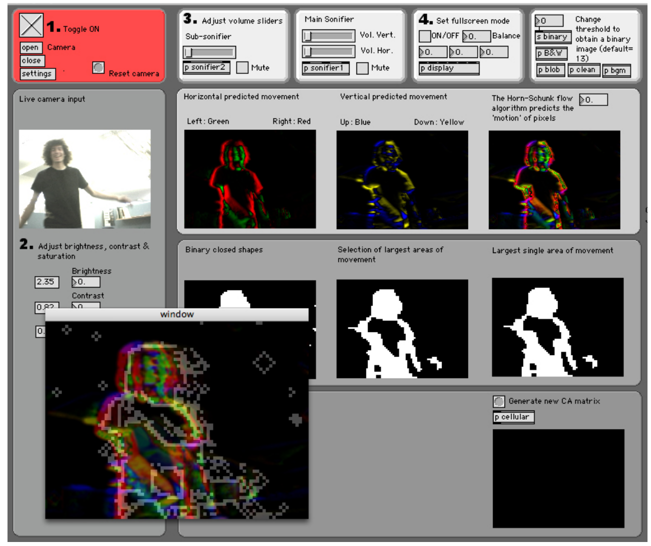

Figure 9. Gestural motion abstracted from streaming video in real time using optical flow estimation, implemented in Max/MSP and Jitter. The window in the lower left corner visualizes cells in an interactive Game of Life grainspace. Each live cell emits a sound.

In the grainspace, the pixels activated by the optical flow measure are interpolated into a medium-resolution cellular automata matrix. The matrix is based on John Conway's Game of Life (GoL) algorithm for cellular automata in which a cell (or grain) is either alive if it has two or three live neighbors, or dead if it has less, and can come alive again if it has exactly three live neighbors.[3] In the GoL grainspace, all live cells (grains) emit a sound. However, a fundamental difference between the regular GoL and the GoL grainspace is that cells in the grainspace can also come alive again by human gestural interaction. Grains can be triggered into rapid proliferation by moving the body in one horizontal direction across the room, and towards stasis by moving in the opposite direction. This can affect neighboring grains, causing an iterative cycle that lasts longer than the gesture itself. In Figure 9, the window in the lower left corner visualizes live cells in the grainspace using grey squares. It illustrates how cells remain alive even after the active gesture (represented by multicolored silhouette) has passed through. The Game of Life grainspace implementation demonstrates a simple and efficient method for enabling adaptive intermodal mapping.

In the public sphere, the audience is transient, covering a range of ages from children to adults. The immediacy and clarity of feedback determined the length of time users interacted with the system. The significance of the generative process used in the Sonic Tai Chi environment is its capability for producing both a perceptible relationship between gesture and the audio-visual mapping while developing an evolving artifact that is neither repetitive nor predictable.

[2] http://www.iamas.ac.jp/~jovan02/cv/

[3] http://www.bitstorm.org/gameoflife/

5. FUTURE WORK

We have presented ParticleTecture as a system for specifying and generating new spatial experiences overlaid in ordinary architectural spaces. The implementation 'Sonic Tai Chi' demonstrates the relationship created between gestural motion and granular synthesis sound, based on a gridded particle metaphor. An adaptive mechanism that uses a measure of engagement to inform ongoing audio patterns in response to human activity is also outlined.

Future work includes further development and testing of the computational index of engagement in an architectural installation context, measuring its efficacy by looking both at change and the relevance of this adaptation to the participants' indication of engaging features. We hope to imbue the ParticleTecture system with an index that adequately produces innovative and engaging spaces over time matching human expectations of engagement. Finally we hope that the design of engaging interactive, multi-modal experiences may become a fundamental aim of architectural design.

6. ACKNOWLEDGMENTS

Our gratitude to the Creativity and Cognition Studios and Beta_space at the Powerhouse Museum, Sydney. This research is supported by a University of Sydney Postgraduate Award and a Cooperative Research Centre (CRC) for Construction Innovation Scholarship.

7. REFERENCES

[1] Bevilacqua, F., Muller, R., and Schnell, N. MnM: a MaxMSP mapping toolbox. In Proc. 2005 Conference on New Musical Interfaces for Musical Expression (NIME05), (2005), 85-88.

[2] Edmonds, E. Logics for Constructing Generative Art Systems. Digital Creativity 14, 1 (2003), 23-38.

[3] Fels, S. and Hinton, G. Glove Talk II: an adaptive gesture-to-formant interface. In Proc SIGCHI conference on Human factors in computing systems, ACM Press (1995), 456-463.

[4] Holland, J. Hidden Order: how adaptation builds complexity, Addison-Wesley, Mass., USA, (1995).

[5] Ikeda, R. *Formula*. Forma, London, UK.

[6] Jakovich, J. and Beilharz, K. Multimodal Spatial Emergence in the Design of Sensate Spaces: Physical spatial interaction in reactive soundspaces. In Duarte, JP, Ducla-Soares, G & Zita Sampio, A (eds) *Digital Design: The Quest for New Paradigms (eCAADe 2005)* (Lisbon). 427-432.

[7] Jakovich, J. and Beilharz, K. From Audience to Inhabitant; Interaction as a medium in architecture. In *Engage: Interaction, Art and Audience Experience*. Creativity & Cognition Studios Press, Sydney, Australia, 2006, 40-49.

[8] Krueger, M. Responsive Environments. In Packer, R. & Jordan, K. (eds) (2001) *Multimedia: from Wagner to virtual reality*. New York, Norton & Company Ltd, 1977. 104-120.

[9] Mandelis, J. and Husbands, P. Don't just play it, grow it!: Breeding sound synthesis and performance mappings. Proc. 2004 Conference on New Musical Interfaces for Musical Expression (NIME04), (2004), 47-50.

[10] McCormack, J. Evolving sonic ecosystems, Kybernetes. The International Journal of Systems & Cybernetics 32, 1/2 (2003), 184-202.

[11] McNeill, D. and Levy, E. Conceptual representations in language activity and gesture. In Jarvella, R.J. and Klein, W. (eds). Speech, Place, and Action, John Wiley & Sons Ltd. (1982), 271-296.

[12] Miranda, E. Composing Music with Computers, Focal Press, Oxford, UK (2001).

[13] Oosterhuis, K. *Architecture goes wild.* 010 Publishers, Rotterdam, The Netherlands, 2002.

[14] Paine, G. Interactivity: Where to from here? Organised Sound 7, 3 (2002), 295-304.

[15] Paine, G. Gesture and musical interaction: Interactive engagement through dynamic morphology. In Proc. 2004 Conference on New Musical Interfaces for Musical Expression (NIME04), (2004), 80-86.

[16] Pelletier, J. A shape-based approach to computer vision musical performance systems. In Proc. 2004 Conference on New Musical Interfaces for Musical Expression (NIME04), (2004), 197-198.

[17] Roads, C. Microsound, MIT Press, Cambridge, Mass. USA, (2001).

[18] Rokeby, D. The construction of experience: Interface as content. In Dodsworth, C. (ed). Digital Illusion: Entertaining the future with high technology, ACM Press, NY, USA (1998).

[19] Trieb, M. *Space calculated in seconds.* Princeton University Press, Princeton, New Jersey, 1996.

[20] Visell, Y. Spontaneous organisation, pattern models, and music. Organised Sound 9, 2 (2004), 151-165.

[21] Wessel, D. Instruments that learn, refined controllers, and source model loudspeakers. Computer Music Journal 15, 4 (1991), 82-86.

[22] Xenakis, I. Formalized music : thought and mathematics in composition, Indiana University Press, Bloomington, USA (1971).