

# ITERATIVE SYNAESTHETIC COMPOSING WITH MULTIMEDIA SIGNALS

Angus Forbes

University of California, Santa Barbara  
Media Arts and Technology Program  
angus.forbes@mat.ucsb.edu

Kiyomitsu Odai

University of California, Santa Barbara  
Department of Music  
kiyomitsu.odai@gmail.com

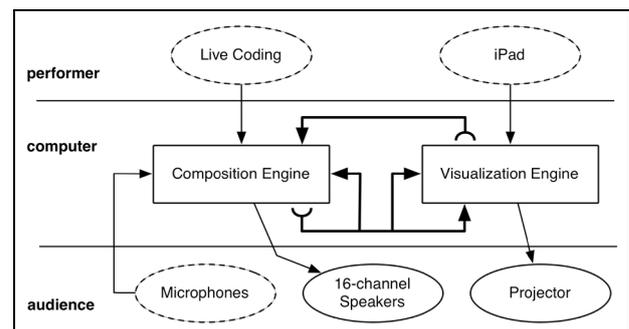
## ABSTRACT

This paper explores the use of orchestrated feedback as an organizational theme in an interactive multimedia composition entitled *Annular Genealogy*. The composition is performed by two players, each of whom use a separate digital interface to create and interact with the parallel iterative processing of compositional data in both the aural and visual domains. In the aural domain, music is generated using a stochastic process that sequences tones mapped to a psycho-acoustically linear Bark scale. The timbre of these tones and the parameters determining their sequencing are determined from various inputs, including especially the 16-channel output of the previous pass fed back into the system via a set of microphones. In the visual domain, animated, real-time graphics are generated using custom software to create an iterative visual feedback loop. This software runs on the iPad tablet and uses a custom fluid dynamics system, a vector visualization technique, and custom image processing filters to generate complex, evolving visual structures. Additionally, information from each of the domains is transferred into the other domain in real-time over a wireless network: sonic data is used to control the image processing parameters and visual information influences the generative parameters of the audio component. Interactive control of the composition is available through multi-touch interaction via the iPad tablet (for the visuals) and via the use of SuperCollider as a live coding environment (for the audio).

## 1. INTRODUCTION

*Annular Genealogy* is an interactive multimedia composition for two performers using multi-channel speakers, a projector, and a tablet computer. The performance is organized around a generative music composition and its visual analog. Both the audio and visual components are explorations of feedback processes that encourage the performers to interactively shape aleatory elements and transmute them into appealing, transient structures. The composition engine works with a stochastic sequencer that uses Brownian motion as a guiding metaphor. Similarly, the visualization engine depicts colored fluid energy as a representation of dynamic, ephemeral structures. In addition to exploring these feedback processes independently of each other, each engine also directly influences the other via networked communication: both the visual and audio processes broadcast data via OSC messages which then influence various parameters of the composition and/or visualization. Finally, even the physical interactions are fed into the generative system

as contact microphones are used to pick up the tapping and other ambient sounds made during the interaction. The ultimate goal of the performance is to bring various layers of feedback into a cohesive compositional experience. These feedback layers are interconnected, but can be broadly categorized as *physical* feedback, *internal* or *digital* feedback, *interconnected* or *networked* feedback, and *performative* feedback. Specifically, these layers include: the generation of new musical motifs being created from the processing of the output sound; the generation of visual forms from the processing of the output graphics; the vector positions that govern the displacement of the visual forms used as inputs to control music parameters; and the sequencing parameters controlling the generation of the composition used as inputs to control image processing parameters.



**Figure 1.** High-level overview of the *Annular Genealogy* project showing processes at three tiers: performer, computer, and audience member. The darker lines indicate the main feedback loops where the output of one process is piped in as the input to another.

In addition to having cybernetic properties of interlaced feedback systems, we can characterize the piece as being fundamentally synaesthetic. That is, the mixing of the mutual generative processes conflates the aural with the visual modality and vice versa [4]. Through the continuous interlinking of the two engines (via the performers and via the data sent over the network) a single interconnected multimodal signal is created. The output of this signal is represented simultaneously in multiple domains. Figure 1 shows a high-level chart of the relationship between the performers, the audience, and the visualization and composition engines. The performers input information to software in parallel using live coding and the multi-touch capabilities of the iPad. The outputs of each of the software engines then are fed back into themselves and into each other in various ways.

## 2. INTERACTIVE PERFORMANCE

By supplying a multi-touch and live coding environment as an interface to and influencer of the generative processes we add another layer of feedback where the performer is able to respond to and shape the multimedia output. That is, we conceive of the performers as participants in a compositional process rather than as on-the-fly creators of audio-visual output. The generative software serves to create some structures independently of the performers; it is the role of the performers to guide the generational processes toward more compositionally interesting output and away from output that is overly repetitive, monochromatic, garish, or otherwise less satisfactory. Likewise, the audio and visual engines, via the various feedback processes, continually push against the explicit control of the performers. Overall, the composition is defined by a network of nested feedback loops that link the performer and the algorithm to create an inherent aesthetic tension between the generative and the interactive, the performed and the composed, the random and the intended. Figure 2 shows a photo of a performance of the piece.



**Figure 2.** Photo of the authors performing a version of *Annular Genealogy*.

Our composition refers directly or indirectly to a number of previous installations. Compositionally, we were inspired by David Tudor's *Rainforest IV*, an early sound-art installation featuring an entirely analog feedback system. *Rainforest IV* is "a collaborative environmental work, spatially mixing the live sounds of suspended sculptures and found objects, with their transformed reflections in an audio system" [12]. In particular, *Annular Genealogy* extends the concept of using loudspeakers as "an instrument unto itself" (rather than a tool of replication) and of using feedback as a compositional source.

Another overt influence on *Annular Genealogy* is Iannis Xenakis's concept of "Stochastic Music." As described below, our stochastic sequencer is an integral part of the generative composition. In particular, Xenakis draws a parallel between his compositional methodology and such natural phenomena as "the collision of hail or rain with hard surfaces, or the song of cicadas in a summer field" [15]. Similarly, we use a circular, stochastic timeline that is elasticized by

Brownian randomness to create foreign sounds that nonetheless have the feel of natural phenomena. Xenakis, in describing his landmark orchestral composition, *Metastasis*, hypothesizes that by constructing acoustic spaces of constant expansion out of long passages of weaved glissandi "one can produce ruled surfaces by drawing the glissandi as straight lines" [15]. However Xenakis' straightforward mapping of these glissandi sweeps to a chromatic scale is somewhat problematic as it imposes a non-linear relationship between the input and output frequencies. Our approach to this linearity problem, described below, involves replacing the chromatic scale with the Bark scale, which is at least psycho-acoustically linear, and thus more effectively captures the intended naturalistic feel.

Other recent multimedia installations have also featured generative compositions that made use of feedback mechanisms between the audio and visual components. For instance, Karen Curley's *Licht und Klang* is an audio-visual installation that generates sounds via optical sensors that use the refractions of light through oil and water as inputs into sound generation software [3]. Various electro-acoustic ensembles have explored the use of networked feedback as a tool for improvised performance. Most famously, The Hub creates multimedia performances based on sets of rules that transform signals passed between performers and that are then presented in aural and visual domains [6]. A wide range of works have explicitly explored the notion of synaesthesia in installations. For instance, Jack Ox and David Britton's *21st Century Virtual Reality Color Organ* uses visual representations of sound waves as an element in creating an interactive landscape [10]. More recently, Daniela Voto's *Multisensory Interactive Installation* explores the interactive sonification of Kandinsky paintings [14]. Other works invoke Michel Chion's concept of synchresis to describe the "welding together" of auditory and visual phenomena [2]. For instance, Niall Moody's audiovisual instrument, *Ashitika*, generates simultaneous multimodal output from single gestures [9]. Our work similarly creates synaesthetic output based on a syncretic fusion of a mixed audio and visual feedback loop.

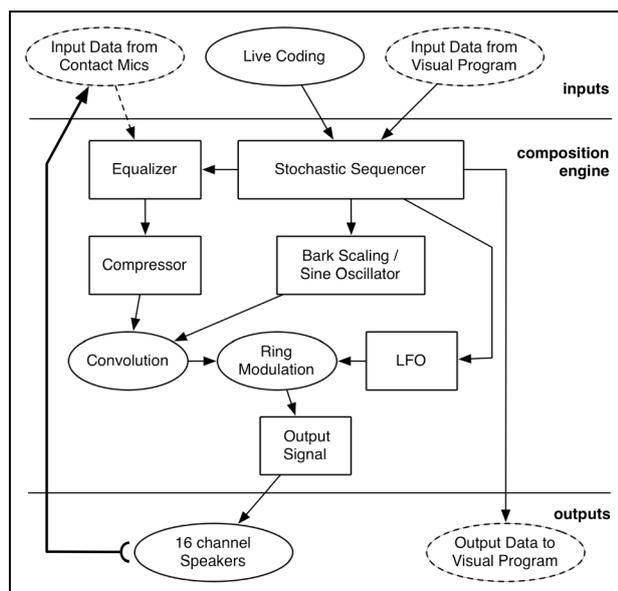
## 3. COMPOSITIONAL DETAILS

*Annular Genealogy* is made up of two distinct software engines, one governing the aural domain and the other the visual. The visual and aural components are related by the structural mechanisms of generating and processing feedback and by the thematic focus on generating organic structures that continually devolve and transform into new structures. Both the aural and the visual engines represent the movement of energy through a system. These software engines are completely decoupled, but influence each other via the output of different multimedia data transmitted via OSC messages. In this way, each of the engines becomes part of a component of each other's feedback loop. In this section, we describe the individual iterative feedback

processing for the composition and the visualization engines, and also indicate where output is sent to and received by the engines.

### 3.1. Composition Engine

The generative composition is largely generated through the receiving and processing of feedback from various sources: from the performer, from parameters received from the visualization engine, from the audience interaction, and via the piping of the composition data back into the composition engine itself. The external inputs are directly provided by the output of the visualization engine, performer interaction via live coding of a SuperCollider script, and contact microphone inputs that capture ambient sound. Moreover, the composition process is based on the continual recycling of audio data that is iteratively fed back into the microphones. Figure 3 shows a more detailed diagram outlining the main components of the composition engine.



**Figure 3.** Overview of the interconnected components that comprise the composition engine. The darker line indicates the main feedback loop where the output audio signals are recapture by microphones to be used in the generation of sound.

The composition engine is written entirely in SuperCollider 3, and consists of various interrelated components, described below: the interactive timbre generator, the stochastic sequencer unit, and the Bark scaling unit. The interactive timbre generator controls the overall quality of sounds by convolving the output of a compressor unit generator with a sine tone oscillator. In addition, a parametric equalizer and a ring modulator further ornament the signals before and after the convolution. The stochastic sequencer unit controls

the timing and frequencies following a series of compositional heuristics. These frequency values are then piped into the Bark scaling unit, which defines the mapping of the frequency values according to the perceptual linearity of human ears (described below).

#### 3.1.1. Interactive Timbre

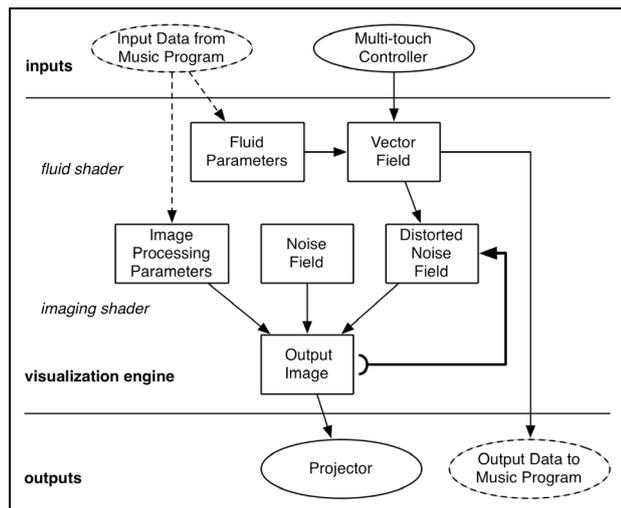
The main input signal is captured by a number of contact microphones, passing through a 6-band parametric equalizer that manipulates and enhances the timbral variety of the signal. This signal is then densely compressed, generating a thick feedback loop: the compressor amplifies the softer signals so that they are loud enough to be fed back to the loop, and at the same time, it squashes those above the set-up threshold in order to make the output sounds controllable. The feedback loop further functions as a distortion box by convolving sine tones with the output signal from the compressor. At the final stage of the system, the signal is ring-modulated with the low-frequency oscillator (LFO). The process is then iteratively repeated in real-time. The output through the speakers is again picked up by the contact microphones and becomes the main component of the input signal for the next pass of interactive timbre unit's feedback loop. Interactive control of this feedback loop is available through the live coding environment and, additionally, simply by making sounds that will be captured by the microphones. In our original performance, for instance, contact microphones were attached to the iPad controller to use the percussive tapping of the visual performer as another input into the composition.

#### 3.1.2. Stochastic Sequencer

The sequencer unit triggers the data for the following elements: the frequencies and durations for the enveloped sine tones, the six passing center frequencies and their bandwidth as Q values for the parametric equalization, LFO frequencies for the ring modulation, panning values for the spatialization, and the amplitude of the output sounds. All the values for those sequenced parameters are generated through stochastic processes that are based on four modes of increasing randomness. In *mode 0*, a sequence of values is created through a simple rising motion (which is not random). In *mode 1*, we use Brownian motion, where each following number is either incremented or decremented only slightly from the current number. In *mode 2*, we use interpolated randomness, where a random number is averaged with the current number, and thus more closely related to the current number than a purely random number. Finally, in *mode 3*, we use a non-interpolated, completely random number that is not related to the current value. Each parameter is then modified by the value resulting from the current mode. Both *mode 0* and *mode 1* are utilized to update the panning amongst the 16 speakers. We use *mode 1* to update the amplitude, band-pass

frequency, and EQ bandwidth. The LFO frequency is updated in *mode 2*.

The sequencer runs indefinitely; the duration of the output frequencies are contoured proportionally using a meandering Fibonacci series that creates asymmetrical cycles. The duration of each note is also further adjusted using Brownian randomness (*mode 1*). In using this mode to update the majority of parameters serves to create the perception of naturalistic sounds. The output frequency is updated via a more complex amalgamation of the outputs of all modes, and then passed on to the Bark scaling unit.



**Figure 4.** Overview of the interconnected components that comprise the visualization engine. The darker line indicates the main feedback loop where the output image becomes the input data for further processing.

### 3.1.3. Bark Scaling

The Bark scale is a non-linear frequency scale that was psycho-acoustically designed originally by Eberhard Zwicker in order to translate frequencies into values that sound perceptually linear (to human ears). The mapping from an input frequency to a corresponding output value in the Bark scale is governed by the following equation:

$$b = 13 \tan^{-1}(0.00076f) + 3.5 \tan^{-1}(f/7500)^2 \quad (1)$$

Here  $f$  is an input frequency value and  $b$  is the output Bark number. The non-linear frequency response of human ears is described by the concept of *critical banding*. The width of each of these critical bands remains more or less constant up to 500 Hz, and then jumps by approximately 20 percent thereafter [16]. The Bark scale is a sequence of these critical bands that are discretely enumerated. We utilize the Bark scale as a compositional tool in order to generate a perceptually linear sweep of pitches and to define an evenly distributed pointillistic texture, and to create a solid mass of sounds.

## 3.2. Visualization Engine

The dynamic visualization is also created through a series of feedback processes, using an interactive application called *Fluid Automata* (initially created by one of the authors as the visualization engine for the original performance) [5]. Inputs into the visualization engine are provided by the composition engine via OSC messages, from performer interaction, and via the iterative real-time processing of the output image. The engine has three main processing layers: the interface layer, which translates multi-touch gestures into fluid vectors; the fluid dynamics layer, which translates fluid energy across a discretized grid of cells; and the image processing layer, which interprets the fluid energy as distortions of a texture map, and blends this distorted texture map with a noise field made up of randomly colored pixels.

The engine was coded in Objective-C using the OpenGL ES 2.0 graphics library and the GLSL shading language and runs entirely on the iPad tablet device. Figure 4 shows a more detailed diagram outlining the main components of the visualization engine.

### 3.2.1. Interactive Fluid Dynamics

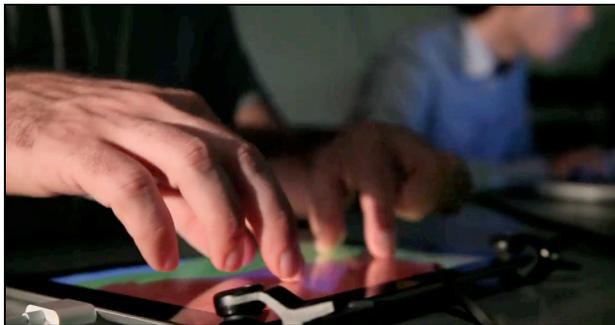
In keeping with the goal of creating artificial, yet naturalistic elements, we used a fluid system as a primary metaphor for creating evocative, transient image structures; a visual analog to the stochastic elements created by the composition engine. A number of interactive art projects use fluid simulation as a component of the work. A method invented by Jos Stam to create stable fluid systems first made it possible to represent realistic looking fluids at real-time frame rates [11]. Many interactive artworks have made use of this technique. For instance, Memo Atken has created a series of demos based upon Stam's method, showcasing the use of mobile devices for interaction with the fluid system [1]. Other similar fluid simulation methods use shader programs that are optimized for real-time interaction in video games [7]. Since the goal of the project was not to represent reality, but rather to generate new creative possibilities, we chose to create our own fluid engine. By so doing, we had a much greater control over the types of visual structures that were created. Moreover, we were able to ensure that our system was robust and also to create visual structures not based on simulation equations that aim to empirically describe the physical world.

The fluid system is discretized into a grid made up of a 15 by 15 cells. The fluid system is defined by directional energy vectors that move energy from each of the cells into their adjacent neighbor cells. New energy is added into the system via the iPad's multi-touch interface as the performer drags one or more fingers across the screen. The energy is divided into three streams of directional momentum: forward momentum, and left and right orthogonal momentum. Energy flows from the current cell into the neighbor

cells at each iteration of the screen refresh rate, (approximately 60 times a second). A small amount of energy is removed at each step, and after some time (if no new energy is added), the fluid system will have no energy. A more detailed description of the system can be found in [5]. Despite the straightforwardness of this algorithm, particular ratios of forward to orthogonal momentum create pleasing patterns of vortices and waves. Moreover, other parameters controlling the rate of movement between cells create effects that look somewhat like cracking ice, or drifting sand, or are extremely turbulent, or imply some other unfamiliar, yet naturalistic effect. Figure 5 is a close-up photo of a performer using the iPad interface to manipulate the fluid energy.

### 3.2.2. Iterative Image Processing

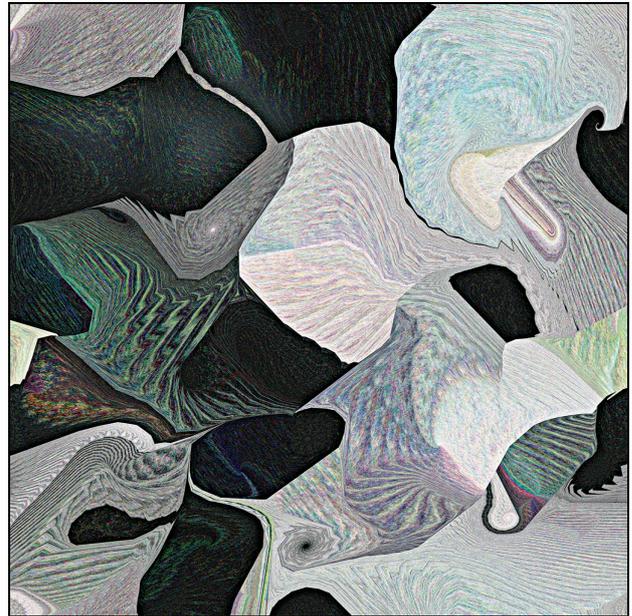
The main image processing scheme is based on a feedback loop whereby a high-resolution background image is perpetually blended together with a distorted version of itself. The characteristics of the distortion are based directly on the current state of the fluid system. Through the feedback loop of blending the previous output with a colored noise field, the individual colored pixels in the noise field are “smeared” in the direction of the current fluid vectors. This system is similar to a vector visualization technique first described by Jarke J. van Wijk, called Image Base Flow Visualization (IBVF). IBVF has been extended for use in a variety of scientific visualization applications, including animated and 3D flows [13].



**Figure 5.** Photo of multi-touch interaction during a performance of *Annular Genealogy*.

The image is further processed by a number of image processing kernels that control the brightness, contrast, saturation of the image, as well as the blending factor weighting the averaging of the distorted image with the original noise field of random colors. The parameters governing these simple kernels are updated in real-time by the output data from the composition engine (via OSC messages). An enormous amount of color variation is possible through the adjustment of these parameters. Because the performer has no control over these parameters during the performance, care was taken so that combinations of parameters did not lead to

unfortunate cases, such as the screen turning completely black when brightness and saturation were set too low. Figure 6 shows a detail of the projected output image of the visualization engine.



**Figure 6.** Detail snapshot of a projected output image from the visualization engine.

## 4. CONCLUSION

*Annular Genealogy* is an innovative interactive composition that invokes a synaesthetic experience by emphasizing the use of syncretic feedback on multiple levels, the physical, the digital, the interconnected, and the human. Physical feedback occurs as multimedia information is put into the performance space and then recaptured by physical sensors. For instance, this happens continuously as the contact microphones reintroduce the output signal as an input signal into the composition engine. Digital feedback occurs internally in the software engines as output data is immediately used to update aspects of the digital heuristics. In the visualization engine, this occurs frame by frame as the previously generated texture is used as the new input texture to be distorted by the fluid vectors. Interconnected feedback occurs when the output data is sent over a wireless network to external software processes. This included the fluid vectors created on the iPad updating the stochastic sequencer running in SuperCollider on the laptop, and the sequencing data influencing the image processing parameters. Human feedback occurs as the performers use their instruments to influence the software, via multi-touch and live coding.

The original interactive performance was shown at the *Bits & Pieces* media arts exhibition in Santa Barbara, California. The trickiest part in developing the original composition was to find an appropriate balance of automatically generated composition/visual via feedback and performers input. That is, it took time to

find an aesthetic balance between the human and the computational. The iterative generative system can quickly fall into patterns that become either repetitive or overly chaotic. Finding the creative “cusps” teetering between these two extremes was the most rewarding aspect of the performance.

While some of the results of interconnecting multiple feedback layers are unpredictable, the performers nonetheless begin to have an intuition as to how their actions will update the overall composition. For example, while there is no direct mapping of how the visualization data will update the compositional structures, after some experience using the iPad interface, it becomes clear that certain gestures during certain kinds of passages generate a particular shaping of the composition. We also found it interesting to reconceive the performers role as a “guider” of aesthetics, rather than as a creator. A possible direction for future versions of the piece would be to more explicitly highlight the effect that an interaction has as it is transmuted from one medium to the other.

Although the focus of the piece has been on interconnecting different feedback loops, one obvious feedback cycle that we did not attend to is the incorporation of the audience as a co-performer. Though adding extra elements to the piece increases the difficulty of maintaining aesthetic balances in the visual and aural domain, a careful integration of audience interaction in future versions of the piece will incorporate audience members as participants rather than as passive viewers.

## 5. REFERENCES

- [1] Atken, M. “MSA fluid demos”. <http://www.memo.tv>. 2009.
- [2] Chion, M. *Audio-Vision: Sound on Screen*. Columbia University Press, 1994.
- [3] Curley, K. “Licht und klang”. <http://vimeo.com/5333718>. 2009.
- [4] Cytowic, R. E. *Synaesthesia, a union of the Senses*. MIT Press, Boston, 2002.
- [5] Forbes, A.G., Höllerer, T., and Legrady, G. “The fluid automata project.” *Proceedings of the International Symposium of Electronic Art*, Albuquerque, USA, 2012.
- [6] Graham-Lancaster, S. “The aesthetics and history of the hub: The effects of changing technology on network computer music”, *Leonardo Music Journal*. 39-44. 1998.
- [7] Guay, M., Colin F., Egli, R. “Simple and fast fluids”. *GPU Pro 2*. Ed. Wolfgang Engel. 433-444. CRC Press. 2011.
- [8] Loy, G. *Musimathics Volume 1*. MIT Press, Cambridge, MA, 2006.
- [9] Moody, N. and Fells, N. and Bailey, N. “Ashitaka: an audiovisual instrument”. *Proceedings of the Conference on New Interfaces for Musical Expression (NIME)*. 2007.
- [10] Ox, J. and Britton, D. “The 21st century virtual reality color organ”, *IEEE Multimedia*. Vol. 7, No 3, 6-9 (2000).
- [11] Stam, J. “Stable fluids”, *Proceedings of the 26th annual conference on Computer Graphics and Interactive Techniques*, p.121-128, July 1999.
- [12] Tudor, David. “Tudor compositions: rainforest”, <http://davidtudor.org/Works>. 1998.
- [13] Van Wijk, J. J. “Image based flow visualization”. *ACM Transactions on Graphics*. Vol. 21, No. 3, 745-754 (2002).
- [14] Voto, D. “Multisensory interactive installation”, *Sound and Computing*. Vol. 5, 24-26 (2005).
- [15] Xenakis, I. *Formalized Music*. Pendragon Press, Hillsdale, NY, 1992.
- [16] Zwicker, E. “Subdivision of the audible frequency range into critical bands”, *The Journal of the Acoustical Society of America*. Vol. 33, 1961.