Journal of the Japanese Society for Sonic Arts, Vol.7 No.2 pp.19–22

Composer's Note

STRATEGIES AGAINST MIMESIS: FRENETIC AUDIOVISUAL INTERACTION IN FUTURE CREATURES

Alexander SIGMAN
International College of Liberal Arts (iCLA)
Yamanashi Gakuin Daigaku
Kofu, Japan

ABSTRACT

Future Creatures (2013) was realized by the author and animation artist Eunjung Hwang. In this article, the intermedia collaborative model, as well as approaches to audiovisual interaction and sound-event structuring and layering are addressed. In so doing, special attention is paid to the implementation of concatentative synthesis techniques.

1. INTRODUCTION

In 2008-2009, the author met Korean New York-based animation and media artist Eunjung Hwang, while both were fellows at the Akademie Schloss Solitude artist residency program in Stuttgart, Germany. At that time, an animation/music collaboration was planned, but due to logistical difficulties, was delayed until 2013. Rather than creating an entirely new animation, Hwang proposed generating two "remixes" of visual material–one ca. 9:30 in duration, the other ca. fifteen minutes long–used to procuce an existing video. As such, the collaboration would be executed in serial (video, followed by audio), rather than in parallel. The video materials consisted of (and set in motion) images drawn from a book of animations, published in 2006.[1] Upon considering both video "remixes," the shorter version was selected for this project.

Figure 1. A copy of Eunjung Hwang's animation compilation *Future Creatures*, upon which the video images were based.

2. VIDEO STRUCTURE, OBJECTS, STATES, AND EVENTS

Following an eighteen-second introduction, the video divides into three primary segments, entitled "Bubbles" (3:11),"Sleep" (1:46), and "Home" (4:08), for a total duration of 9:23. Each main segment consists of multiple scenes and scene transitions. In turn, each scene is constituted by objects (creatures) that are associated with goal-directed events. Event completion results in the change of an object's internal state. Objects tend to interact with each other according to a predetermined set of rules. Furthermore, the final state of one scene predicts the initial state of the subsequent scene.

Despite the logical, algorithmic underpinning of the video's unfolding, the surface is populated by intricate and wildly contrasting layered compositions, frenzied event activity levels and scene-transitions, and a consistently high rate of change, a rapid and propulsive visual rhythm. This persistent state of controlled chaos is off-set by the recurrent appearances of objects/creatures and events, which contribute continuity and coherence to the fabric of the video.

3. AUDIOVISUAL INTERACTION: INTENTIONS AND APPROACHES

In contributing an audio dimension to the video, two principal objectives were established: 1) to avoid a strictly mimetic relationship between the visual and auditory and 2) to take into account and explore differences in visual and auditory perception via decoupling changes in the density and type of video and audio events, respectively.

[1] http://blog.eunjunghwang.com/?p=632 (accessed 31 July 2015)

Figure 2. Still from *Future Creatures* Part I, "Bubbles."

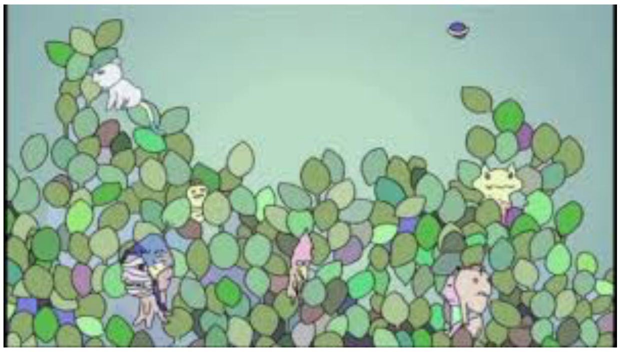

Figure 3. Still from *Future Creatures* Part II, "Sleep."

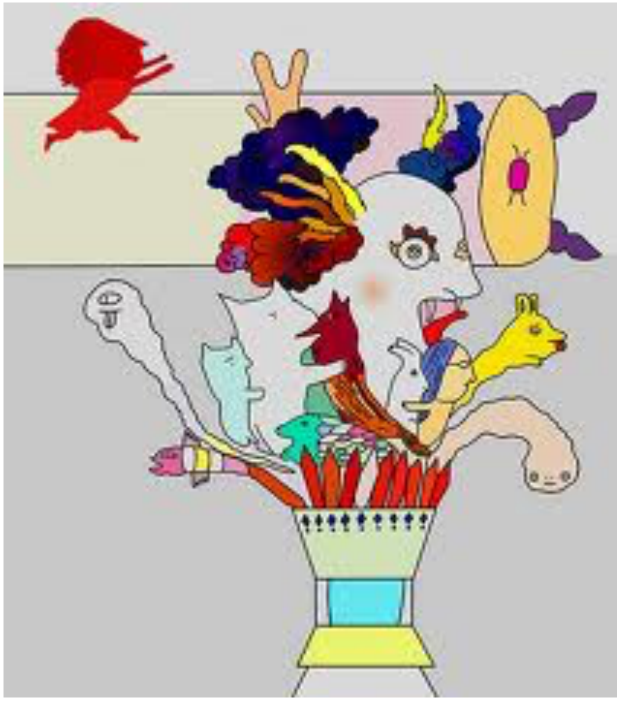

Figure 4. Still from *Future Creatures* Part III, "Home."

With the above aims in mind, several approaches were considered, ranging from a more generalist/statistical "aerial view" interaction model, in which each scene would be assigned a unique sound-complex that would reflect the atmosphere and intensity of the visuals, to a deterministic "microscope view" method, in which each video *object* would be associated with a unique and dynamic audio object. The "aerial view" model was rejected, as it would relegate the music/audio to a mere background function. By contrast, the one-to-one video-to-audio object mapping approach would be problematic, in that a) the tendency towards cross-modal surface imitation would be too great, thereby establishing trivial audiovisual relationships; b) operating on such a local level would quickly lead to surface saturation; and c) perhaps most importantly, exclusive attention to local morphology would result in a sacrifice (or at least compromise) of global shaping and control.

As such, the selected interaction strategy fell somewhere in between these extremes.Not only would connections be established between the visible and audible *surfaces* of the work, but *structural* similarities between the visual and auditory domains would also be achieved through the application of the video's layering processes, density distributions, and goal-directed procedures to the organization of audio data. Rather than maintaining a static audiovisual correspondence or demanding a perpetual focus on minute details on the part of the viewer, it was intended that a "flickering" effect be achieved, such that significant changes in a complex field of sonic micro-events would incite a perspectival shift within the visual field, and vice versa. Consequently, the hierarchy of perceptually salient events within a given scene would be subject to frequent fluctuations, triggered by cross-modal cues. But what would be the optimal means of implementing this interaction strategy?

4. CONCATENATIVE SYNTHESIS TECHNIQUES AND SOURCE MATERIALS

4.1. What is Concatenative Synthesis?

Concatenative synthesis is a sub-type of granular synthesis, in which segmented recordings are analyzed for acoustic features, gathered into corpora, and conjoined (concatenated) into extended sequences according to a given set of criteria or models. In CataRT, IRCAM's real-time concatenative synthesis system,[2] for instance, source file segments (samples) in a corpus may be distributed and activated in a two-dimensional acoustic features descriptor space (with descriptors specified by the user), mapped to

[2] imtr.ircam.fr/imtr/CataRT (accessed 04 August 2015)

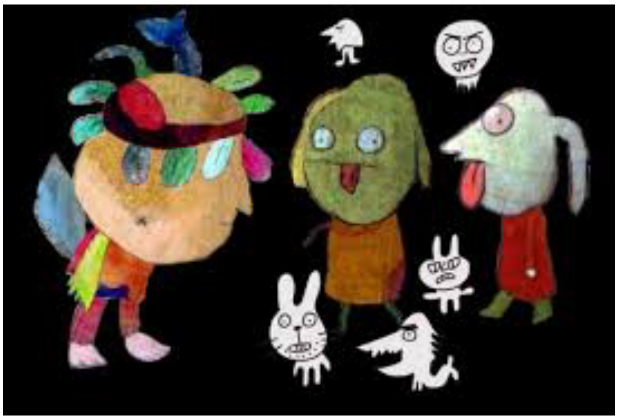

Figure 5. Another still from *Future Creatures* Part III, ”Home.”

live instrumental or vocal input, or used for the resynthesis of a ”target” soundfile. Thus, a corpus of samples of Bach chorales may be used to resynthesize a Beethoven symphony (or vice versa), John Coltrane samples may be triggered by live saxophone input, a Lachenmann orchestra piece may be articulated by a corpus of dog barks, and so on.

Given the strictly non-real-time application of concatenative synthesis in the context of *Future Creatures*, the AudioGuide program,[3] designed by composer Ben Hackbarth, was employed. Written in Python, AudioGuide efficiently generates audio output, and interfaces with CSound, Max/MSP, and Pure Data (Pd). In the AudioGuide interface, the user is provided numerous affordances regarding time-varying audio-descriptors in performing the analysis of segments in a database, such that one segment corpus/target pairing may produce multiple contrasting resynthesis results. With respect to segment-to-target mapping, segment *density* and *alignment* may be varied.

4.2. Why Concatenative Synthesis in Future Creatures?

In order to achieve the audiovisual interaction strategy described above, there were three essential conditions to fulfill: 1) the ability to generate heterogeneous layers of micro-events; 2) goal-directedness and 3) control over event density. In addition, as a response to the representational nature of the video (anthropomorphic characters and urban/suburban/rural settings for the scenes), the potential for sound-source recurrence and identifiability, as well as the capacity to situate segments along a sound iconicity-abstractness gradient, were desired. The use of concatenative synthesis in general, and AudioGuide in particular, proved highly effective in achieving these goals. While variations in segment density, duration, and number of segment corpus sources could occur via any type of granular or stochastic synthesis, the resynthesis of target soundfiles, and the broad range of acoustic descriptor and mapping parameter settings that could be applied are unique to concatenative synthesis. The targets–of varying durations at different levels of scale–would fulfill the essential role of shaping the sonic events (collections of segments) constituting each scene. As a given soundfile could function as a target and undergo segmentation, it would be possible to construct a ”figure/ground” relationship between soundfiles, such that a soundfile that at one moment contributes to the surface material (segments), would later constitute a scene’s structural backbone.

[3] http://www.benhackbarth.com/audioGuide/ (accessed 05 August 2015)

4.3. Segment and Target Sources

The sound sources deployed in *Future Creatures* belong to four principal categories:

1) urban and industrial environment mechanical and electronic sounds (e.g., construction equipment, car alarms, and bells); 2) animal sounds; 3) human voices (both natural and synthetic); and 4) K-Pop samples (a reference to both the country of origin of the animation artist and the country of residence of the author at the time of composition).

In total, six targets–each compressed and stretched to occupy durations ranging from 18 seconds to four minutes–and 32 segmented soundfiles were utilized. In several cases, concatenative synthesis procedures were applied iteratively, such that a soundfile generated via segmentation and concatenation would itself be subsequently segmented and mapped to a target. The selected target, as well as descriptor, density, and segment alignment parameters would be subject to modification at each successive iteration.

5. EXTENSIONS

5.1. The fcremap Series

Following the completion of *Future Creatures*,[4] the author began to compose a set of pieces for solo instruments and live electronics based upon the video’s visual and sonic materials. Entitled the *fcremap* series (fc = ”Future Creatures”; remap = remapping of the video onto a different medium and performance context), the pieces therein all metabolize still image data derived from the video through the use of image/sound processing and conversion software. RGB color values are interpreted as spectral data, whose control parameters may be adjusted by the user. The resulting data has been applied to synthesis and filtering processes in the electronics, as well as a means of generating harmonic structures for the instrumental parts. Another commonality amongst the *fcremap* works is the projection of the electronics through small tactile transducers affixed to the instruments. Besides obviating the need for a PA system, using the instrument as a resonator corporealizes the electronic sounds, as well as tightly integrating electronic and instrumental sources. These performative, three-dimensional, and physical ”remixes” of the original elements stand in stark contrast to a two-dimensional digital video representation.

At the time of writing (August 2015), two compositions in this cycle have been completed: *fcremaperc* for percussion, [5] and *fcremapno* for prepared piano.[6] Both

[4] The final version may be viewed here: https://vimeo.com/70344676 (accessed 05 August 2015)

[5] Here is a recording of the US premiere, as performed at Spectrum in New York by percussionist Eric Derr: https://soundcloud.com/atsigman/fcremaperc-2014 (accessed 05 August 2015).

[6] Frederick Croene’s January 2015 performance: https://soundcloud.com/atsigman/fcremapno-january-2015-version (accessed 05 August 2015).

are ca. 9:30 in duration (i.e., the approximate duration of *Future Creatures*), and date from 2014. While the former preserves the tripartite structure of the video, the latter unfolds as a single continuous movement. In *fcremaperc*, transducers are attached to a metal sheet, snare drum, and a re-tuned zither; in *fcremapno*, onto the piano frame.

5.2. Interactive Audio-Centric Game Version

Taking the rule-based "narrative" of *Future Creatures* as a point of departure, the next step will entail creating a game adaptation of the video in the near future. In the game, players interact directly with objects (characters), changing their internal state via the execution of events. These changes of state correlate with the manipulation of associated audio parameters. As opposed to more conventional game scenarios, in which activities are purely focused on earning points or ascending levels, this game will be somewhat pedagogical in nature: through engagment with the game environment, players effectively learn how to play a collection of virtual micro-instruments. Needless to say, designing such an environment and enabling such a mode of engagement requires fine-grained audiovisual correspondences, and therefore a different multimedia interaction model than was the case in the video version of *Future Creatures*. Far from being conceived as an *improvement* upon the video's interaction model, it is simply considered more *appropriate* for the (active and open-ended) game context than the (passive and self-contained) video representation.

6. AUTHOR'S PROFILE

Alexander SIGMAN

Alexander Sigman's award-winning instrumental, electroacoustic, multimedia, and installation works have been featured on major international festivals, exhibitions, institutions, and venues across Europe, Asia, Australia, and the US. In June 2007, Sigman was Composer-in-Residence at the *Musiques Demesurees* festival in Clermont-Ferrand, France. Subsequently, he was awarded residency fellowships by the Akademie Schloss Solitude (Stuttgart, Germany), the Djerassi Foundation, and the Paul Dresher Ensemble Artists Residency Center. In 2013-2014, he undertook a musical research residency at IRCAM.

Nominal/Noumenal, Sigman's first portrait recording, was released on Carrier Records in 2012. This album features performances by Les Percussions de Strasbourg, Ensemble SurPlus, and ensemble ascolta. His violin and electronics piece *VURTRUVURT* is included on violinist/violist Miranda Cuckson's solo CD *Melting the Darkness*, released in November 2014 on the Urlicht Audiovisual label.

Also active as a researcher, Sigman has published articles and book chapters on a variety of subjects in international peer reviewed journals, books, conference proceedings, and exhibition catalogues. He has presented lectures at numerous conferences and institutions, including Harvard University, New York University, the Aspen Institute, the American Musicological Society, the Musikhochschule Stuttgart, the University of Leeds, Aristotle University Thessaloniki, IRCAM, Nagoya City University, WOCMAT (Taiwan), the Korean Electro-Acoustic Music Society's 2013 Annual Conference, and the International Computer Music Conference (ICMC).

Sigman completed his doctorate in Music Composition at Stanford University in 2010. Prior to Stanford, he obtained a BM in Music and a BA in Cognitive Sciences from Rice University. Further postgraduate studies were undertaken at the University for Music and the Performing Arts Vienna, as well as the Institute for Sonology of the Royal Conservatory in The Hague (Netherlands). He is currently Associate Professor of Music at the International College of Liberal Arts (iCLA) of Yamanashi Gakuin University in Kofu, Japan. Between 2012-2015, he was Assistant Professor of Composition at Keimyung University in Daegu, South Korea.