Exploring the Nexus of Holography and Holophony in Visual Music Composition

MICHAEL RHOADES

ABSTRACT

In this article the author explores the idea that, owing to their shared three-dimensional nature, holophons and virtual holograms are well suited as mediums for visual music composition. This union is ripe with creative opportunity and fraught with challenges in the areas of aesthetics and technical implementation. Squarely situated upon the bleeding edge of phenomenological research and creative practice, this novel medium is within reach. Here, one methodological pipeline is delineated that employs the convergence of holophony, virtual holography and supercomputing toward the creation of visual music compositions intended for head-mounted displays or large-scale 3D/360-degree projection screens and high-density loudspeaker arrays.

Holophony and virtual holography extend beyond more common 3D approaches. Although definitions for each tend to vary, here I consider holograms as stereoscopic projections of virtual three-dimensional visual objects that appear to occupy a common three-dimensional space with the viewer. Holophons are then their auditory counterparts. Holophons and holograms extend the range of expression beyond affording the potential for relationships that portray the realism of day-to-day perception into relationships between sight and sound that could not otherwise exist within our normal perceptions of reality. Potentially, one can allude to multiverses within such environments. To explain, although photographs are two-dimensional images consisting of width and height, they often allude to a third dimension of depth. Generalize this perspective to holophony and holography, and one can imagine an allusion to a fourth spatial dimension, thus significantly increasing the potential for audience immersion.

It is difficult for artists to engage holophons and holograms as a medium. This is in part because the practice of generating the required high-resolution stereoscopic visual material involves extensive computational power. For my work, this means engaging a supercomputing cluster for rendering. This is often a roadblock to many artists and composers due to the knowledge required and limited accessibility to supercomputers.

What does the relationship between holophony and holography consist of? How do we negotiate the manipulation of sight and sound such that they interact in an aesthetically feasible manner as a unified entity? I propose one approach to this challenge here.

VISUAL MUSIC

"Visual music depends upon a meaningful synthesis of visuals with audio" [1]. Visual music, in its current incarnation, is a relatively recent phenomenon, with humble beginnings in silent films in the early twentieth century [2]. Then, as now, the subject matter is most commonly abstract or nonrepresentational. Visual music generally parallels the Modernist perspective of art for art's sake [3]. The compositions usually mean nothing beyond that which is experienced.

I implemented several approaches in pursuing the development of interactions between the auditory and the visual for this project. Primarily, I employed a choreographic perspective wherein the objects and environments are related in terms of movement and spatial location. Aesthetic choices regarding the movements of musical objects, visual objects and the environments in which they interact determined relationships ranging from synchronous to contrapuntal to completely disparate. Another approach involved parametric relationships such as timbre, amplitude and frequency versus color, brightness and hue. Ultimately, sound existing as light—and light existing as sound—is the desired effect. The key to visual music expression is in the creation and interactions of these relationships.

CREATING THE MUSIC

Spatial orchestration is commonplace in the world of electroacoustic music. Further, composers are increasingly embracing 3D spatialization as the number of practical techniques and availability of venues increase. Often theoretical applications of techniques such as vector-based amplitude panning

Michael Rhoades (artist, composer, computer scientist, researcher), Virginia Tech, Blacksburg, VA 24061, U.S.A. Email: mrhoades@perceptionfactory.com. Web: www.perceptionfactory.com. ORCID: 0000-0002-2901-3337.

See www.mitpressjournals.org/toc/lmj/30 for supplemental files associated with this issue.

 doi:10.1162/LMJ_a_01093

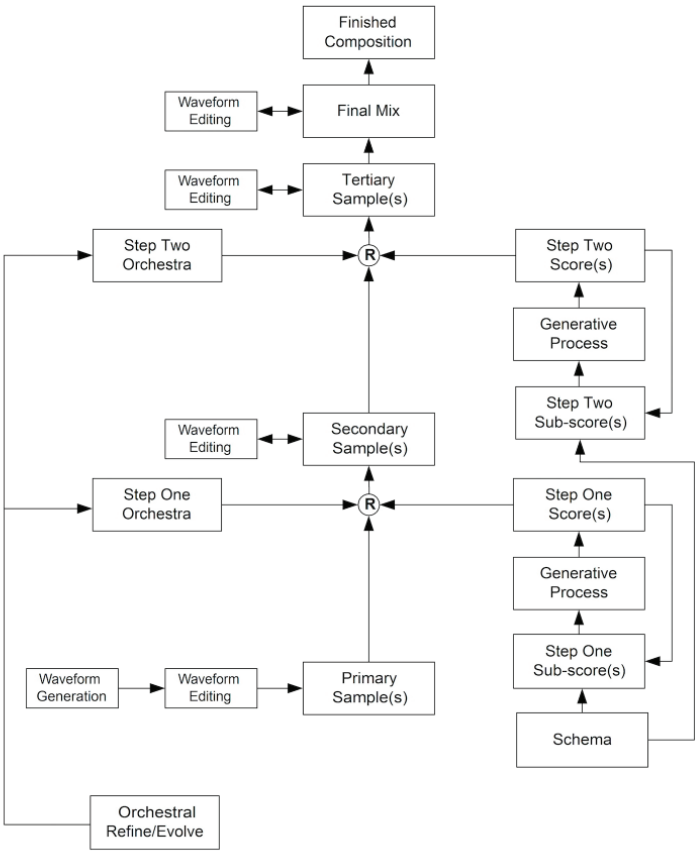

Fig. 1. Musical compositional workflow. (© Michael Rhoades)

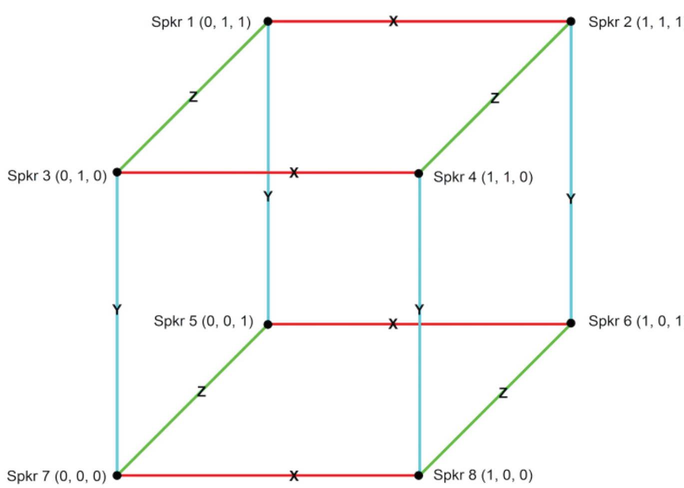

Fig. 2. 8-channel 3D Cartesian cuboid diffusion. (© Michael Rhoades)

(VBAP) and high order Ambisonics (HOA), diffused over high-density loudspeaker arrays (HDLA), are intended to provide the perception of holophony. However, in a practical perceptual sense, these events often seem relegated primarily to the periphery of the venue. Wavefield synthesis can offer a positive step in the direction of perceivable holophony and yet is generally impractical in medium to large venues due to the numerous loudspeakers required to accomplish it.

My research involves the implementation of a threefold hybrid approach consisting of convolution reverb (CR), HOA and my own version of VBAP (rVBAP). Although I did not test them empirically in this practice-based study, the use of these systems in a unified approach appears to contribute toward an increased perception of holophony.

Generative processes instantiated to compose the musical aspect of my work engage the use of Csound, Cmask and a reiterative score synthesis technique called Score Based Sampling; the latter originated at the Institute of Sonology in the Netherlands during the mid-1970s. The generative aspect of this process is largely based upon tendency masks that constrain stochastically generated values intended for the Csound numerical score [4]. Figure 1 is a diagrammatic representation of the bottom-up compositional workflow. I do not examine it here beyond stating that an aspect of this generative process includes the event-by-event implementation of distance and panning algorithms, based upon CR and rVBAP respectively, which determines the location and perceived movement of auditory events.

The distance algorithm implements convolution reverb based upon pconvolve, which is a Csound opcode. "Convolution is a mathematical procedure whereby one function is modified by another" [5]. Convolution reverb consists of mapping the reverberation signature of an impulse response (IR) recording onto another sample. This approach ascribes realistic reverberation characteristics to a sound event as opposed to more conventional forms of reverberation, where filters are employed to imitate generic qualities of an imaginary reverberant space [6]. CR provides an enhanced holophonic representation because reverberation is an important auditory cue in the perception of distance.

To implement CR within the compositional process, the 100 percent reverberant event is mixed in varying ratios with the original nonreverberant event. Envelopes are often implemented to modulate between the reverberant and nonreverberant mixes to create the perception of dynamic distance relationships. Used with overall amplitude envelopes and effective IR files, convolution reverb can powerfully enhance the perception of distance.

The procedural flow of the Csound orchestra relies upon the stochastically generated score file to determine the parameters of the rVBAP panning algorithm to be implemented. In my Csound orchestra, there are 51 different movement patterns, each of which can be applied to any combination of up to eight loudspeakers. Amplitude panning involves the modulation of an event between two or more loudspeakers. Figure 2 diagrams a basic three-dimensional Cartesian cube existing in the positive quadrants of each dimension. Each axis of the cube ranges from 0 to 1. An

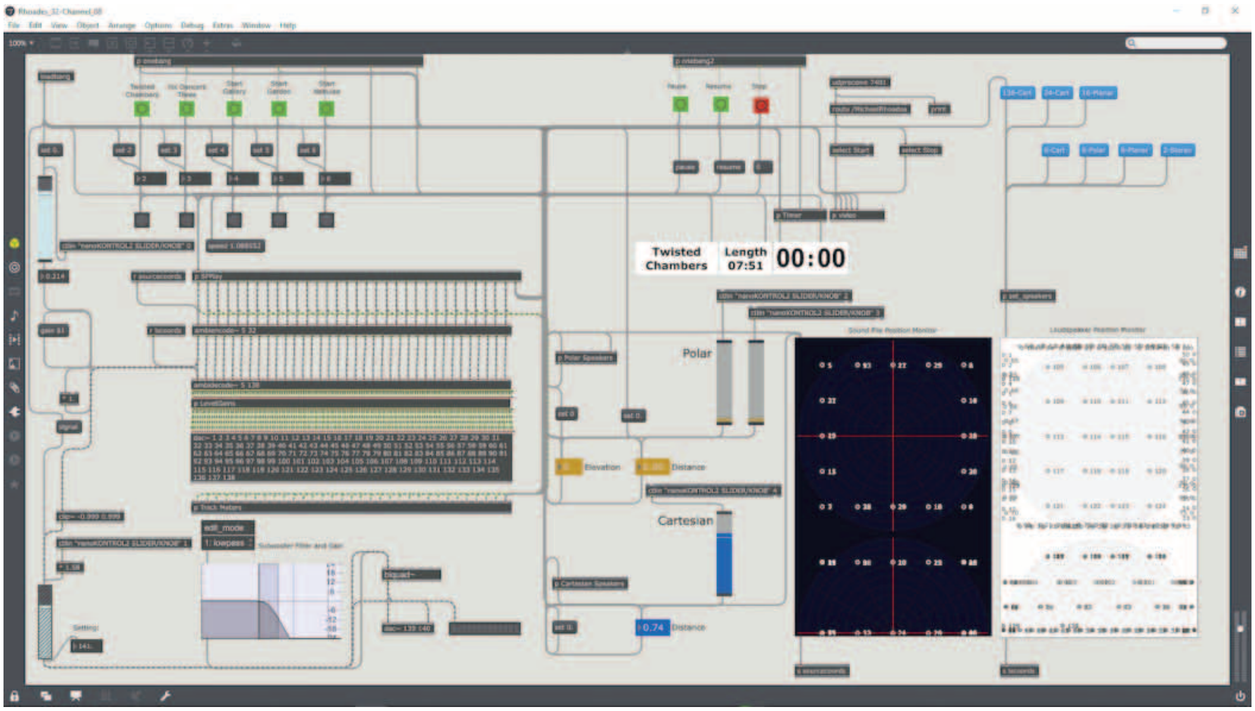

Fig. 3. MAX patch utilized for HOA diffusion. (© Michael Rhoades)

8-channel loudspeaker system is mapped onto the diagram. Referencing this set of coordinates determines the perception of the location of a sound event in the Csound numerical score [7].

For simplification, I diagrammed an eight-channel example; however, for this project I utilized a 32-channel approach. Although similar to the 8-channel configuration, 32-channel diffusion allows for a higher spatial resolution, which affords more specific perception of the location of sound events [8].

Although rVBAP is adequate for providing the perception of three spatial dimensions, from my perspective, the quality of the holophonic effect varies from event to event. In order to enhance it further, the final aspect of my approach involves the diffusion of sound events through the implementation of HOA. After the 32-channel Csound composition is complete, it is diffused for final playback using MAX (Fig. 3). The final link toward the perception of holophony is approached in this study through the implementation of a patch that was developed from a basic MAX patch created by Tanner Upthegrove, a media engineer at the Cube at Virginia Tech, in which the ICST Ambisonics plugin is utilized.

HOA creates spherical sound fields to which audio events are directed. Each loudspeaker receives a weighted sum of all Ambisonics channels, which increases the size of the listening sweet spot and the quality of sound localization [9]. Fifth-order Ambisonics is appropriate for 32-channel diffusion. For this work, I conceived 32 spherical sound fields to be arranged in a cuboid pattern in the venue. One of the numerous variations I added to the original MAX patch was to make the sound fields adjustable in their proximity to the center of the venue. Positioning the spherical sound fields toward the center of the venue increases the overall presence in a composition, thus potentially increasing the perception of holophony. I added control over the amplitude levels for each ascending row of loudspeakers. In cases where there are two or more rows in a venue, this allows for precise control toward a balanced relationship between them, which is essential for a relatively equal perception from the listening position. This approach enhances the diffusion of the height dimension, which is the most difficult to perceive [10].

CREATING THE VISUAL

For this work I implemented a stereoscopic approach to 3D visuals to create virtual holographic content. The techniques involved primarily rely upon Maya, Adobe After Effects, and Adobe Premiere. The DomeMaster3D plugin for Maya provides a stereoscopic virtual camera rig that calculates 360°-horizontal-by-180°-vertical imaging. The right eye and the left eye perspective for each frame of video is rendered at a cumulative resolution of 7000 × 7000 pixels. After an extensive testing and rendering process, the resultant stereoscopic image sequences are stacked into a single image sequence, using Adobe After Effects, in an over/under configuration as demonstrated in Fig. 4. This stacked sequence is then imported into Adobe Premiere for final editing and rendering. Figure 5 presents a diagram of the top-down visual compositional process.

As with the audio compositional process, the generative approach to the creation of the visual aspect of this work involves mining for the elusive, and simultaneously

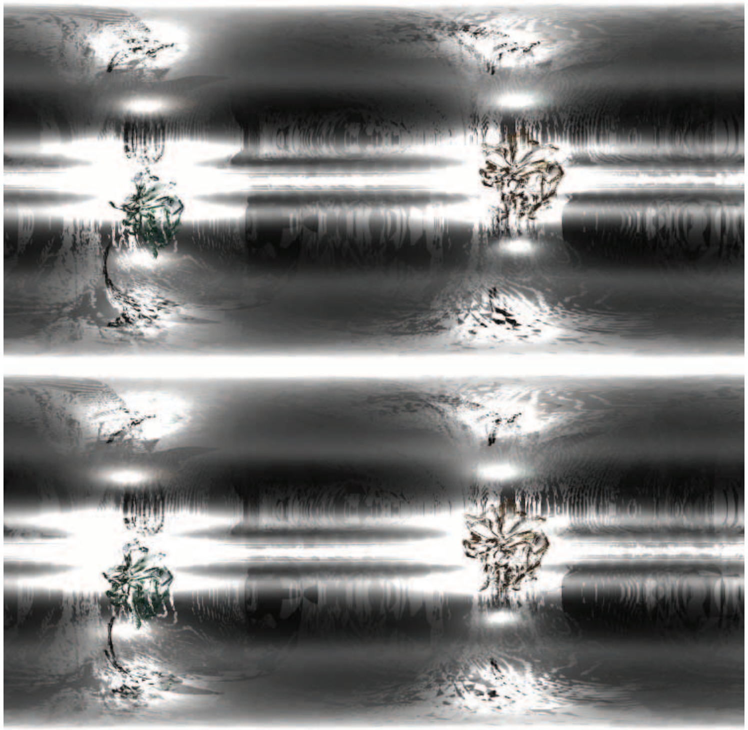

Fig. 4. An example of a stacked stereoscopic image. (© Michael Rhoades)

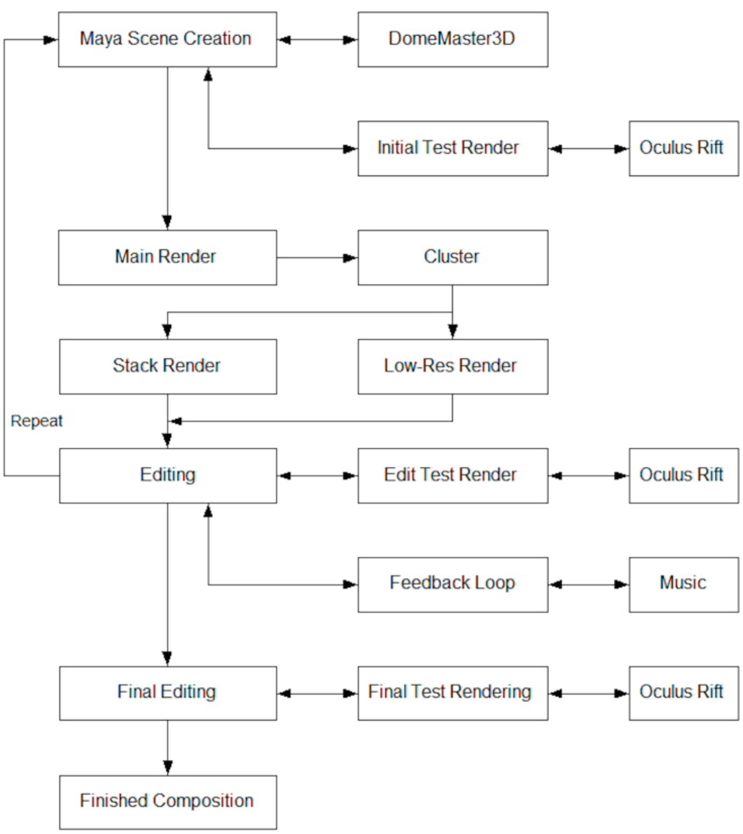

Fig. 5. Visual compositional workflow. (© Michael Rhoades)

tantalizing, properties of emergence. "In general, emergence might be described as the effect of the outputs being greater than the sum of the inputs. Something happens—something beyond what one would expect the combination of algorithms we set into motion should be able to produce" [11]. To create an environment conducive to emergence, I employ numerous refractions and reflections in the creation of each scene. The geometry of each, although often quite simple, achieves complexity through this process. Together these elemental components contribute dynamically toward the indeterminate behaviors of the renders.

Generally, I begin the process by creating a container for the scene based upon one or more basic three-dimensional geometric objects such as cubes, rectangles, cones and cylinders. This environment is often reflective. As such it could be viewed as a nonlinear architectural artifact. Into it I place abstract or nonobjective objects, of varying sizes, geometric configurations, textures and colors. These objects exist and move within the environment, as do their reflections. Often these and other aspects of the scene, such as planes and secondary containers, are translucent, thus providing for refracted reflections of the objects and the environment. Since it is difficult to anticipate the results of the subsequent renders given these variables, each aspect of the scene increases the opportunities for emergent phenomena. Visual objects are intended to occupy the space adjacent to the viewer and to intersect it, affording an immersive holographic experience.

However, this process comes at a massive computational cost. For the Mental Ray renderer in Maya to calculate the huge number of positions and associated parameters for each photon in each stereoscopic frame of each scene requires a lengthy rendering process, which necessitates a specialized computational approach.

HIGH-PERFORMANCE COMPUTING RENDERING

Rendering high-resolution stereoscopic image sequences at 30 frames per second for each scene is impractical for most computer workstations. Therefore, I began experimenting with supercomputing clusters, which implement parallel processing. I found building a ROCKS supercomputing cluster effective for the task. With it one can configure the CentOS operating system specifically for the task at hand, thus eliminating any unnecessary computational overhead. The Sun Grid Engine (SGE) scheduler that accompanies the ROCKS bundle

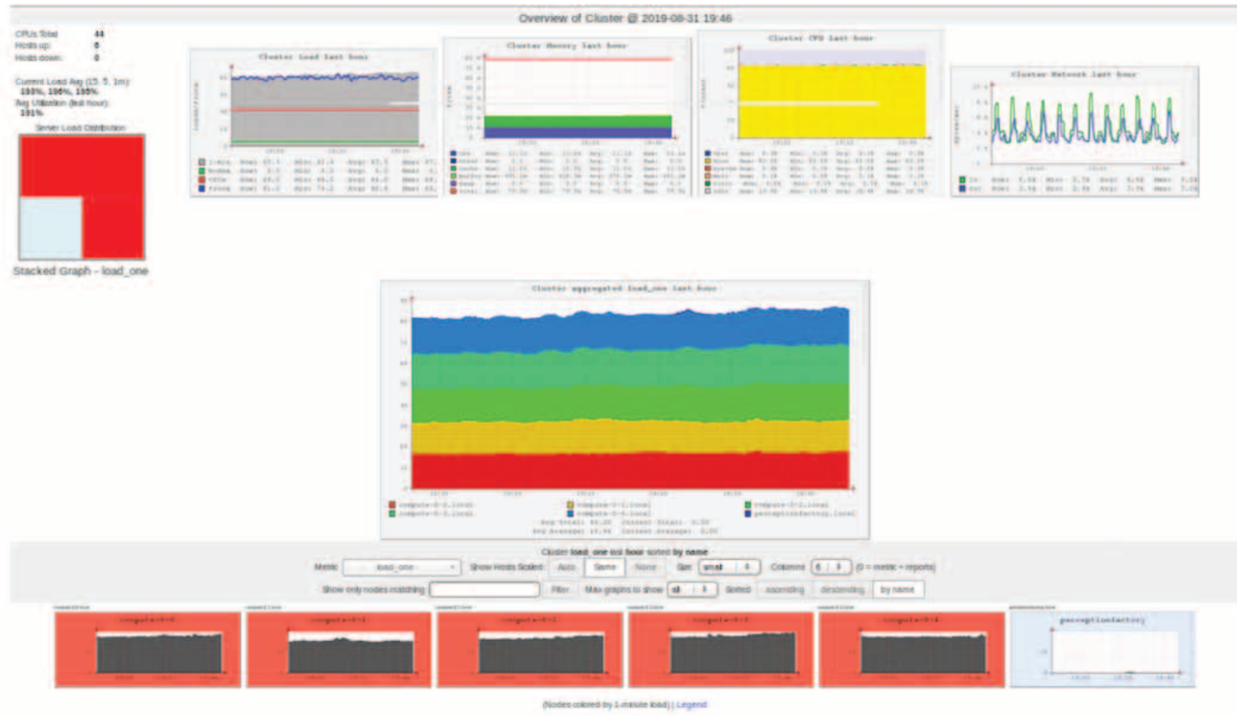

Fig. 6. Ganglia, web-based cluster hardware monitoring. (© Michael Rhoades)

allows each node of each CPU to be specifically assigned a rendering task. Although not empirically tested for this practice-based study, informal testing demonstrated a ~300 percent increase in rendering speed with the implementation of HPC clusters as compared to standard render farms. Figure 6 is a screenshot of the hardware monitoring system for the cluster. The left five rectangles at the bottom represent CPU usage for each compute node.

Optimizing rendering times is extremely important when creating visual music compositions. However, even utilizing the HPC cluster, I was not able to render every conceivable scene for this project. Often, the scenes were too complex or too large in file size to render in an efficient fashion. Also, features like ambient occlusion and final gathering, each of which add immensely to the quality of a rendered image, had to be forgone due to rendering-time limitations. In this manner, the cluster becomes a partner in the compositional process.

THE NEXUS

I have thus far presented the major constituents in the process of creating the holophonic and holographic visual music compositions separately as a matter of communicative convenience. In practice the out-of-time process occurs nonlinearly.

One of the primary activities that guides and catalyzes the process is that of a four-part creative feedback loop. The musical, visual and HPC processes, in addition to the composer's sensibilities, all contribute toward a highly effective feedback loop, which is the crux of the creative process involved in this project. As demonstrated by Fig. 7, each aspect of the processes involved influences and informs the other. In this manner a confluence of workflow culminates in a uniquely situated artifact.

As stated above, I implemented a choreographic perspective as a major factor in determining the relationships between the audio and visual events. Since each is based upon movement, this choice was clearly beneficial. However, another more elusive influence is that of synesthesia. Although I am not a synesthete, I nonetheless experienced general "feelings" about the relationships between sight and sound throughout the project. These underlying intuitions, although difficult to describe clearly, were perhaps the most powerful influence in determining the relationships between sight and sound.

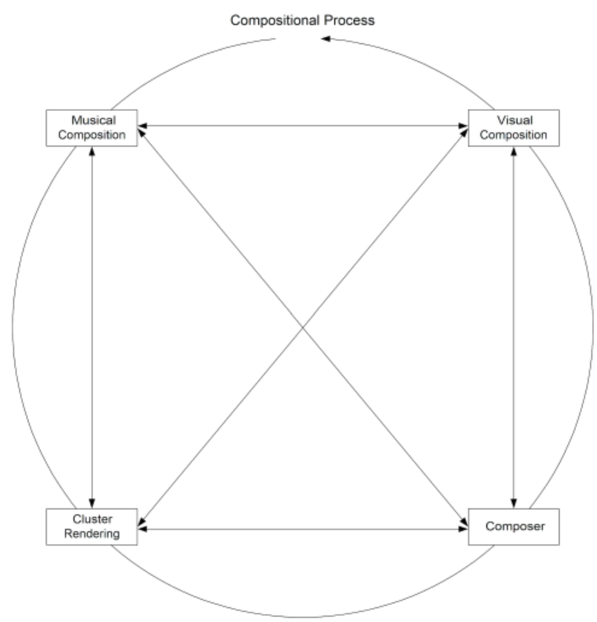

Fig. 7. The quadrilateral reciprocative feedback loop. (© Michael Rhoades)

NOVEL AESTHETICS

The incorporation of holophony and holography into visual music compositions is accompanied by unique aesthetic choices. The examples listed below represent a few. Here the objects referred to may be visual or auditory in nature.

- The "cross-eyed effect": when objects are placed so close to the cameras within a scene that the viewer's eyes cross when viewing it
- The "monocular effect": when an object occludes one of the viewer's eyes
- The "claustrophobic effect": when either the space or several large objects are located in close proximity to the listener
- The "syncing effect": when the movement of the audio is contrary to the video as a form of perceptual decoupling
- The "dizzying effect": when the viewer loses equilibrium
- The "strobe effect": when a scene rapidly alternates between bright and dark

Additionally, there are other aesthetic considerations when creating audio/visual material for immersive environments, which are referred to as "frameless environments" [12]. For instance, pacing of material in a planetarium is normally required to proceed much slower than in a 2D environment [13]. This is also pertinent in a holophonic and holographic environment. The audience can be overwhelmed by the realization that they cannot see everything occurring on the screen.

This is a brief summary of the challenges of the aesthetics involved in this research. I call them aesthetic here since the concern is with practice. Interesting questions arise that would in other mediums be considered irrelevant. For instance, is it useful to employ the dizzying effect to produce a disorientation that is then resolved in the subsequent section? Considering this simple question, one can see each of these effects as potential formal elements. Used judiciously, they can contribute toward varied and novel visual music experiences.

With these and other phenomenological elements, user experience plays a part. During the practice of conducting this research, as I became accustomed to viewing material using a head-mounted display (HMD), I was much more accepting of that which would have previously been disorienting. These visual music compositions specifically, and holophonic and holographic art in general, may favor those accustomed to virtual reality exploration.

PRESENTATION

Venues for presentation of this leading-edge creative outcome are currently somewhat limited in number. However, HMDs are being developed at a rapid rate, due in part to corporate competition, increasing in quality and decreasing in price. They provide novel venues for individuals and groups. Numerous film festivals and conferences are embracing VR theaters. Examples include Tribeca, SIGGRAPH and Cannes. It is further viable for individual composers to install their own equipment in VR theaters for the presentation of their holophonic and holographic content.

Finally, universities are developing and deploying systems that utilize 3D/360-degree stereoscopic projection that can be combined with HDLAs. The Cube and Cyclorama at Virginia Tech is one example. The Allosphere at the University of California in Santa Barbara is another. It is not unreasonable to posit a future where these types of systems are ubiquitous.

CONCLUSIONS

Owing to their mutually 3D/360-degree nature, the intersection of holograms and holophons forms a symbiotic union as an expressive medium for the composition of visual music. In this medium exists the potential to place an audience within a fully immersive environment of creative expression. In the work I describe here, relationships between holograms and holophons were based upon choreography, audio/visual timbre and indirect synesthesia. These relationships contain the potential to transcend our daily sensorial perceptions and achieve expressions that could not otherwise exist.

A collection of procedural methodologies, each aimed toward this target, shows the viability of the practice. A quadrilateral feedback loop allows each of the major constituents involved in the process to exert significant influence over the outcome. Further, the indeterministic approach of mining emergent quanta in this fertile landscape is ripe with challenge and opportunity.

Fundamental to the production of this work is the challenge of rendering high-resolution images, which requires more than a powerful computer workstation to accomplish. High-performance computing offers a viable solution. The use of parallel computing makes it possible to render the stereoscopic images within a reasonable time frame.

Creating audio/visual art with this medium presents novel aesthetic challenges. Expressions that in two dimensions are unquestionably acceptable require further consideration in three dimensions. For instance, an object moving toward the viewer in two-dimensional space simply becomes larger. However, due to the effects of stereopsis, this type of event causes the eyes to cross. The speed at which such an event occurs may determine its viability. Certainly, a modicum of discomfort is perhaps acceptable, or even desirable, as a formal element, especially when it later resolves. Certainly, tension and release move an audience through a composition [14].

Venues for the presentation of this work are increasingly available. For instance, VR theaters are common attractions at many contemporary film festivals. Head-mounted displays and proprietary 3D/360-degree projection systems are being developed, deployed and utilized. It is reasonable to surmise they will eventually, in some form, become ubiquitous.

This novel creative territory is ripe for exploration although, as with every novel expressive medium, courage is required to take the leap. The associated challenges fostered by this work render the journey as exciting as the destination.

References and Notes

This paper serves as a synopsis of the written aspect of Michael Rhoades's PhD dissertation. For a detailed delineation of the material covered here, please visit www.perceptionfactory.com/exegesis.pdf.

1 Julie Watkins, "Composing Visual Music: Visual Music Practice at the Intersection of Technology, Audio-Visual Rhythms and Human Traces," *Body, Space & Technology* **17**, No. 1, 2018, 51–75, p. 71. DOI: 10.16995/bst.296. See also www.perceptionfactory.com/visual-music.

2 Emmanouil Kanellos, "Visual Trends in Contemporary Visual Music Practice," *Body, Space & Technology* **17**, No. 1, 2018, pp. 22–33. DOI: 10.16995/bst.294.

3 Watkins [1].

4 For more information see Michael Rhoades, "Azimuth—Algorithmic Score Synthesis Techniques," *Journal SEAMUS* **18**, No. 1, 2–11 (2005).

5 Iain McCurdy, "05 H. Convolution," in J. Heintz et al., *CSound* (Floss Manuals, 2015): https://flossmanual.csound.com/sound-modification/convolution (accessed January 2020).

6 McCurdy in Heintz et al. [5].

7 For details see Michael Rhoades, "Hadronized Spectra (The LHC Sonifications)," *eContact!* **16**, No. 3 (2013): www.econtact.ca/16_3/rhoades_LHCsonifications.html (accessed January 2020).

8 Eric Lyon et al., "Genesis of the Cube: The Design and Deployment of an HDLA-Based Performance and Research Facility," *Computer Music Journal* **40**, No. 4, 62–78 (2016). DOI: 10.1162/comj_a_00394.

9 Florian Hollerweger, "An Introduction to Higher-Order Ambisonic" (2005): https://decoy.iki.fi/dsound/ambisonic/motherlode/source/HOA_intro.pdf.

10 Hollerweger [9].

11 Rhoades [7].

12 Ka C. Yu, Dan Neafus and Ryan Wyatt, "Filmmaking for Fulldome: Best Practices and Guidelines for Immersive Cinema (Part I)," *Planetarian* **45**, No. 4, 26–39 (2016).

13 Yu et al. [12].

14 Brian Evans, "Foundations of a Visual Music," *Computer Music Journal* **29**, No. 4, 11–24 (2005). DOI: 10.1162/014892605775179955.

Manuscript received 4 December 2019.

MICHAEL RHOADES *is a doctoral candidate in the interdisciplinary PhD program at Virginia Tech conducting research and practice in the areas of holophony, holography and supercomputing. His defense and graduation are scheduled for the fall of 2020.*