Jøran Rudi
Norwegian Network for Technology, Acoustics, and Music (NOTAM)
Nedre Gate 5
0515 Oslo, Norway
joranru@notam02.no

Computer Music Video: A Composer's Perspective

Computer music and video can be combined and made interdependent in works of art in a multitude of ways. Works may either contain pre-rendered imagery, be generated on the fly from a set of rules, be passively dependent on external events, or actively seek interaction, requiring specific user activity for realization. Video imagery can be developed as digital models that change according to spatial or timbral relationships in the music, for example, and audio can be developed by mapping pixel values from a static or moving image to parameter values for synthesis and sound processing.

This article gives a brief overview of some issues in music representation and mapping, and it discusses the artistic strategies that I have employed in the cross-media approach to *When Timbre Comes Apart* (1992–95), *Concrete Net* (1996), *Planet (Terra)* (1988, 1996), and *Construction Drive* (2005).

Cross-Disciplinary Representation and Mapping

Visual representations of music, such as musical notation, spectrograms, and other mappings of sound data, may be considered cross-disciplinary. However, not all such forms of notations are intended as interesting visual objects in their own right, although several musical scores contain qualities considered as visual art, such as *Agglomaration* (1960) by Anestis Logothesis, *December 1952* by Earle Brown, and *Five Piano Pieces for David Tudor* (1959) by Sylvano Busotti. Musical notation, listening/analysis scores, and plots in time and spectral domains represent different aspects of music. Each representation type focuses on different elements or parameters of the musical idea or object, and these representations serve as tools for the composer/artist, facilitating different manipulations of the musical material.

Computer Music Journal, 29:4, pp. 36–44, Winter 2005

Mapping can be considered a subset of visual representation. In binary format, data can be ported with relative ease from one domain to another. However, the data may not make much sense when ported into the domain of another medium, as for example experiments with the resynthesis of graphics files in a phase vocoder such as used in the software program Ceres (originally authored by Øyvind Hammer in 1994; www.notam02.no/notam02/prod-prg-ceres.html). To represent music or sound visually in a cross-media approach, decisions must be made about which aspects are to be represented; in other words, the mapping process requires interpretation.

Cross-disciplinary representations may also include auditory display systems and sonification, although the goal for these might not necessarily be the artistic exploration of sound. Cross-disciplinary representations are not trivial from a timbral point of view, and it has been argued that they make most sense when informed by psychoacoustics and when aware of the implications they have for setting auditory scenes. An example of this is the use of sound in computer games, in which the feedback and graphical user-interface sounds provide auditory information on system status and user actions and the music changes dynamically according to position, action, and environment (Schütze 2003).

New possibilities for narrative arise when other forms of representation and expression are added to sound, and the composer must consider which links, if any, are desirable to establish between the forms of representation in the specific work. There are many interesting examples from interactive dance and dramatic performances in which different methods for extracting data from the dancers' movements have been employed (Siegel 1998; Bevilacqua, Ridenour, and Cuccia 2002; see also Todor Todoroff's 1998 work *En Jeu* and his 2000 collaborative work *In Between*). In these examples, the data have been linked to sound generation and processing, and in some instances to lighting and digital "scenography." Of

particular interest is video software such as BigEye from STEIM (www.steim.org/steim/bigeye.html) that calculates changes in pixel values and maps them to musical parameters. In the reverse direction, video imagery can be based on the mapping of parameters taken directly from the music, as in Roger Dannenberg's work *Uncertainty Principle* (2000).

Points of Departure, Intention, and Approach

Composition can be abstract storytelling, or the telling of abstract stories—perhaps both. Extending Varèse's definition of music as organized sound, music may also be conceptualized as contextualizing sounds, often from a hermeneutic perspective, where single components are meaningful only when considered as part of a whole, and where the context is complex and allows for multiple interpretations or understandings. In the compositions discussed later, several extra-musical ideas are interwoven with the goal of mirroring and addressing the complexities in nature and human interaction, which are context-dependent and difficult to predict. The objects of these compositions are often taken from ideas of action and consequence, namely how perceptions are structured and developed into patterns, and how resonance (in the sense of emotional recognition) emerges—in short, how our myths are formed. While myth today is often understood in terms of "a false statement," the Greek *mythos* refers to powerful truths about the world, often in the form of stories.

Electronic sound is the primary sound material in my compositions that are discussed in this article; as a tool, it has allowed great freedom to create narratives with different levels of abstraction by playing on the representational and symbolic values in the sounds themselves. Computer technology allows sounds to be manipulated away from (and toward) clarity. To make the context work in these compositions, it was important to intentionally blur clarity in the music, thus fostering ambiguity, avoiding one-dimensionality, and emphasizing choice. Blurring and ambiguity are familiar features in our everyday experience, as meaning is abstracted from complex auditory scenes (Bregman 1990) based on the combination, presence, or absence of sounds that we hear or expect to hear. Some sounds represent specific actions with enough clarity to be considered sound icons, whereas others are ambiguous and may refer to a diversity of actions, leaving the listener free to "connect the dots." Aesthetic and analytic issues such as these are frequently discussed in the framework of spectromorphology (Smalley 1986, 1997), as a further development of Pierre Schaeffer's writings in *Traité des Objects Musicaux* (1966).

Composing for audio and animation entails a relatively new art process that involves issues of audio-visual coherence. The process raises questions. How should the visuals relate to the sounds? Should they be referential or non-referential, literal or symbolic—should they be concrete and recognizable as representations we are familiar with, or should they be abstract? Should the visuals somehow be derived from the sound? Following the storyline and timeline, issues also arise pertaining to moving pictures: color, light, camera angles, and movement, as well as the editing to achieve speed and development.

In addition, there are thematic questions of poetry, politics, and program. Should the music be about something that can be clearly stated and easily understood, and should the work attempt to convey a sense of immersion, or should it make the illusion (of reality or representation) explicit and thus available for reflection? Such decisions are linked to the artistic intentions that define the genre of computer music video.

The weighing and balancing of contradictory pulls, visual and auditory, complicates the process of identifying the locus of artistic intention in computer music video, as the object of consideration may become pitted against the idea of some type of action, as in such game-type interactive compositions as my work in progress *Construction Drive.* For many composers, making music is a type of research, a personal exploration of ideas and how they may be given audible form. Combining music with visual media changes and recharges this notion of composition as research process, presenting opportunities to engage with recent and commercial tech-

nologies, such as DVD and computer game platforms. In regards to the latter, the currently available content for game platforms suggests that this is a largely unexploited niche for art, holding the promise of new audiences as well.

Ideas and Techniques in Four Computer Music Videos

In the works discussed below, several ideas are combined, drawn from previous personal experiences, specific sounds, numerical descriptions of natural phenomena, literature, visual art, and human motion. A consistent trait is nonetheless the desire to connect seemingly dissimilar ideas and to explore these connections through composition. The composer can tell several stories at the same time by modeling complex situations from many dissimilar elements. The totality is incomprehensible without an understanding of the parts, and the parts lose their significance if removed from the whole. In this complex weave, elements vie for attention, and what may considered particularly important parts of the compositions are also merely a small part of this larger sonic amalgam. The intention is that when those elements in the music that are most difficult to hear—those in which the surrounding sounds mask them and offer them the most resistance—are noticed, the listener will be rewarded by a sense of direction or continuity. This element of resistance requires concentration, and listeners may need to overcome their own resistance as well—not to succumb to boredom—because it may take awhile before the idea is uncovered.

The possibilities for these kinds of reactions are composed into the works, and as such, they enter into a tradition of works that employ a systematic stretching of the audience's patience. (I have personal, vivid memories of this type of forced "eartraining" from John Cage's reading/singing of James Joyce's *Finnegan's Wake.* It was a humbling experience to understand that my sense of time and compositional topic was insufficient for grasping the work at first.) By stretching the listener's patience and expectation of musical development in time, the composer demands that listeners be creative in finding their own patterns rather than waiting for the composer's "message." In any case, a sense of clarity might be deceptive on any level, as Czeslaw Milosz suggests in his 1951 novel *The Captive Mind:* "When someone is honestly 55% right, that's very good, and if someone is 60% right, it's wonderful and great luck, but if someone is 75% right, it is suspicious, and whoever says he is 100% right is a fanatic!" Uncertainty, then, is a compositional element that can be reinforced or quelled by either visual or auditory means—or both—in the computer music video.

When Timbre Comes Apart

In my work *When Timbre Comes Apart* (1992–95), meaning (timbre) is disassembled and reassembled, as the piece investigates what happens when parts are taken away from the whole and how the same elements can be used to reconstruct a new entity. The medium for this was timbre, the detailed manipulation and blending of harmonics in the continuum between pitch and noise, and "in-between timbres" that could be recognized and sounds where no source or specific identity was perceptible. The timbres were derived from additive and granulation synthesis algorithms, along with two sound recordings: one of my son (two years old at the time), and one of a large tam-tam, approximately five feet in diameter. The ratios between the partials of the synthesized timbre that is presented in the first half-second of the piece are used throughout the work as structuring numbers, and the concrete pitches of the partials serve as a guiding harmonic grid (see Rudi 1998).

At the time this piece was written, phase vocoding had become easy to do, because the price of computational power had dropped significantly. The first three parts of the work were synthesized with the Kyma System, as was the granulation in the last two parts. NOTAM's powerful phase vocoder program Ceres was the principal software for the timbral manipulation. The music was written first, and the animation was developed directly from the music. At the time, graphical modeling was an exploding field owing to the hardware-based

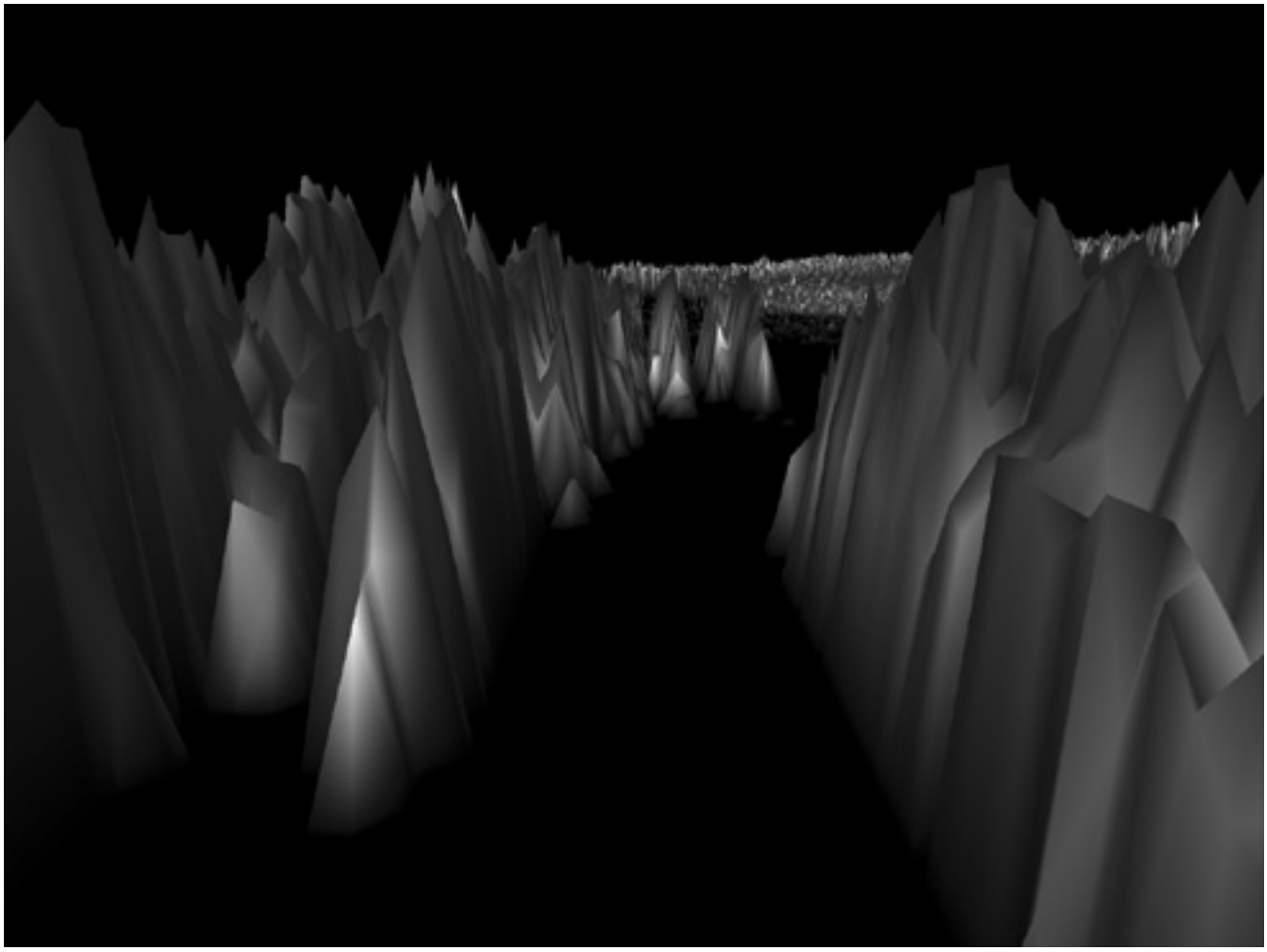


Figure 1. Three-dimensional sonogram from a section of When Timbre Comes Apart *(1992–95). Time runs from front to back, and frequency is distributed left to right, with higher frequencies to the left.*

graphic engine of Silicon Graphics (SGI) computers, and animation of a modeled spectrogram was easily executed. The SGI featured the excellent modeling package Explorer, and the camera movement along the model was scripted using a number of splines to achieve smooth camera movement with little or few "jumpy" movements and to eliminate abrupt beginnings and endings of camera trajectories. The model was normally viewed with a variable time-limited horizon to not reveal too much of the piece at any time, and the camera was moved back and forth, over and under the model.

The narrative of *When Timbre Comes Apart* aimed to point out important sound elements and compositional perspectives, elements that perhaps would have otherwise passed unnoticed. An example of this is shown in Figure 1, where the camera is located between partials that almost disappear in the total sound at that time. Another example of this is seen the closing of the first part, where the camera looks back while ducking below the model. The intention is to strengthen a sense of closure for the listeners. The first part is composed to be intense and long, in order to allow the listener a sense of release at the onset of the second part. Other examples of how narrative was strategically employed are the "zooming effect" leading into the fourth and fifth parts, where the camera is kept close to the

model, and when the listener experiences crashing through the representation at the beginning of the fourth part as a mirroring of the strong onset sound.

The narrative was made possible by extensive camera movement, changing focus and perspective, looking forward, backward, sideways, over, and under the model. The educational aspect of this method of pointing to essential compositional elements was further emphasized in another work *And the Birds . . . ?* (1997), written for NOTAM's educational software program called DSP (Sack 2003). The even flow of sound samples determined the speed of the camera along the time axis in both of these works, with the listening point situated directly under the camera. The even movement gave a "drive" to the works that was stronger than the music would have provided on its own. The dramaturgy of the animation followed the dramatic development of the music through the use of focus, distance to the model (which influences the perception of speed), and color. Movement under the model was used to open and close compositional sections, and color was used to separate the different parts of the composition. To me, for example, "orangey red" suggested a "warmer" timbre than "metallic green," and other colors were used to visually suggest different timbral qualities and musical expressions in the other parts of the compositions. As a perceptual tool, the spectrum was displayed opposite from what is common in three-dimensional representations, with the higher partials closer to the camera than the lower partials.

Given that the images and the sound are two types of representation of the same data in this work, *When Timbre Comes Apart* is a true cross-domain mapping. However, the camera was neither neutral nor still, always aimed at an important feature of the music to influence the listener's appreciation of it. One could say that the camera tried to be loyal to the distribution of interesting and salient events in the composition.

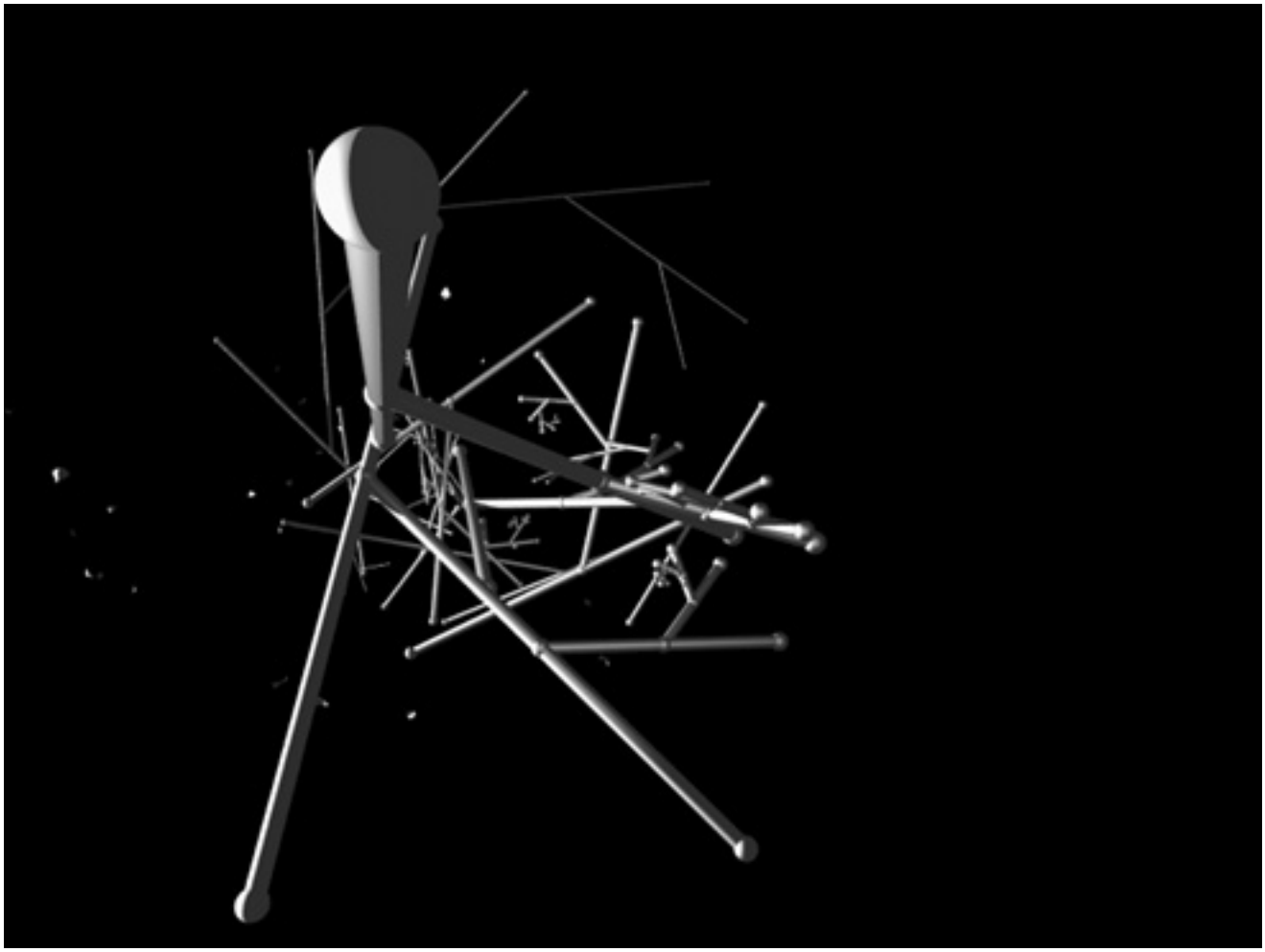
Concrete Net

In *Concrete Net* (1996), the camera had another function. *Concrete Net* is about memory and reflection, working through and experiencing the remains of something that has happened in the past. The listener is "awakened" in the beginning of the piece, which then passes through three parts with significant sound events, and then winds down to the end of the piece during a long fourth part of reflection. The sounds in the piece emerge from a long fascination with sounds from long steel wires, sounds from steel scraps, and the 1974 novel *Concrete Island* by J. G. Ballard. The novel is about a highway network and its excluding nature once one has exited it and is no longer moving along smoothly. As long as the principal character in the book follows conventions and behaves according to the network rules, the network is a good thing, but once off the highway, it is virtually impossible to get back on. As in much of Ballard's writings, there are vivid descriptions of an interesting subculture with specific survival strategies—an alternative network. Networks are thus both inclusive and exclusive, and in the novel, it is virtually impossible to get back on the highway once one has exited. This duality was explored in this musical piece through readings of selected passages from Ballard's book, deliberately obscured through signal processing where vocal samples excite physical models of strings. The physical models were visualized and adapted in the video part, and traffic sounds were used as sound material, filtered and pitched to create an underlying, changing harmony.

To construct the "space" of the piece, nine sets of ratios were used, measurements of the distances between the planets in our solar system. Each set comprised measurements made from one planet to the others, and the sets were quasi-transposed to be in the same range. These ratios were used as parameter values for all signal processing in the piece and to "slice" the vocal recordings to excite different physically modeled strings with different signals. (If all of the strings had been excited with the same signal, the result would have been similar to that of an analog spring reverberator; see Rudi 1998.)

The video part is quite simple, rendered in Persistence of Vision (POV-Ray), a freeware raytracing program. There are three main events in the piece—all represented by different and distinct models—unlike *When Timbre Comes Apart*, in which the

Figure 2. Image taken from Concrete Net *(1996), at the entrance point to the section of physically modeled strings. There are nine sets of strings.*



spectrum served as a model for the entire piece. The images are abstract, taken from space imagery: a planet, debris, and a star. The planet is the first object encountered, followed by a cloud of fog and metal debris, which is in turn followed by nine rotating string constructions (see Figure 2). In between the objects are smaller lumps of fog and metal scraps, signifying closure and transition, to prepare the listener and to recollect the past musical experience. Following the string constructions is a long stretch with visual noise in the form of metal scraps, placed mainly to give the viewer a sense of movement and direction. The camera movement is simple, pointing forward in linear fashion until the end of the string constructions, and slowly rotating from there toward the end of the piece. For the camera rotation to have the desired effect of visual recapitulation, the trajectory needed to follow a curve, and the effect is that the listener sees restatements and reinterpretations of events that took place earlier in the piece as they move further and further away, eventually disappearing behind the star. The sound in this part is also formed of restatements and reinterpretations. The slowness of the rotation brings calm and enables reflection on the material and the reminders of the events, and it returns the mood of the piece to a tranquil state after having worked through the material several times.

The camera is quite neutral throughout this piece, with the exception of the beginning, when the camera vibrates to emphasize the sudden change in timbre that occurs after breaking through the atmosphere of the planet. The piece is composed to be unbalanced, with all important events occurring in the beginning. The subsequently relatively still, rotating camera in the fourth part encourages careful listening, as it becomes clear that the camera, having shown what is coming, does not point out new elements in the composition. The idea was to create space for the listener's reflection by creating a distance to the previous events, mentally and physically, thus reinforcing the original intent of the piece related to notions of inclusion/exclusion.

Planet (Terra)

Another work that belongs to this group of works is *Planet (Terra),* an animation of music that was realized in 1987–1988. The music originally attempted to depict developing order in a chaotic sound world. The piece is short (about 3 min 30 sec), merely suggesting an idea rather than expanding on it, as is the case in the two previously discussed works. The animation was based on mapping of the spectral data from the piece, modeled onto the surface of a sphere. This sphere was enclosed in another sphere that had same base color as the first, and this sphere shrinks as the music plays, allowing the mapped spectrum to emerge. The spectrum was colored to resemble a geological map, and the landscape materializes as a mythical receding of water that bares mountains, valleys, and plains, as is shown in Figure 3.

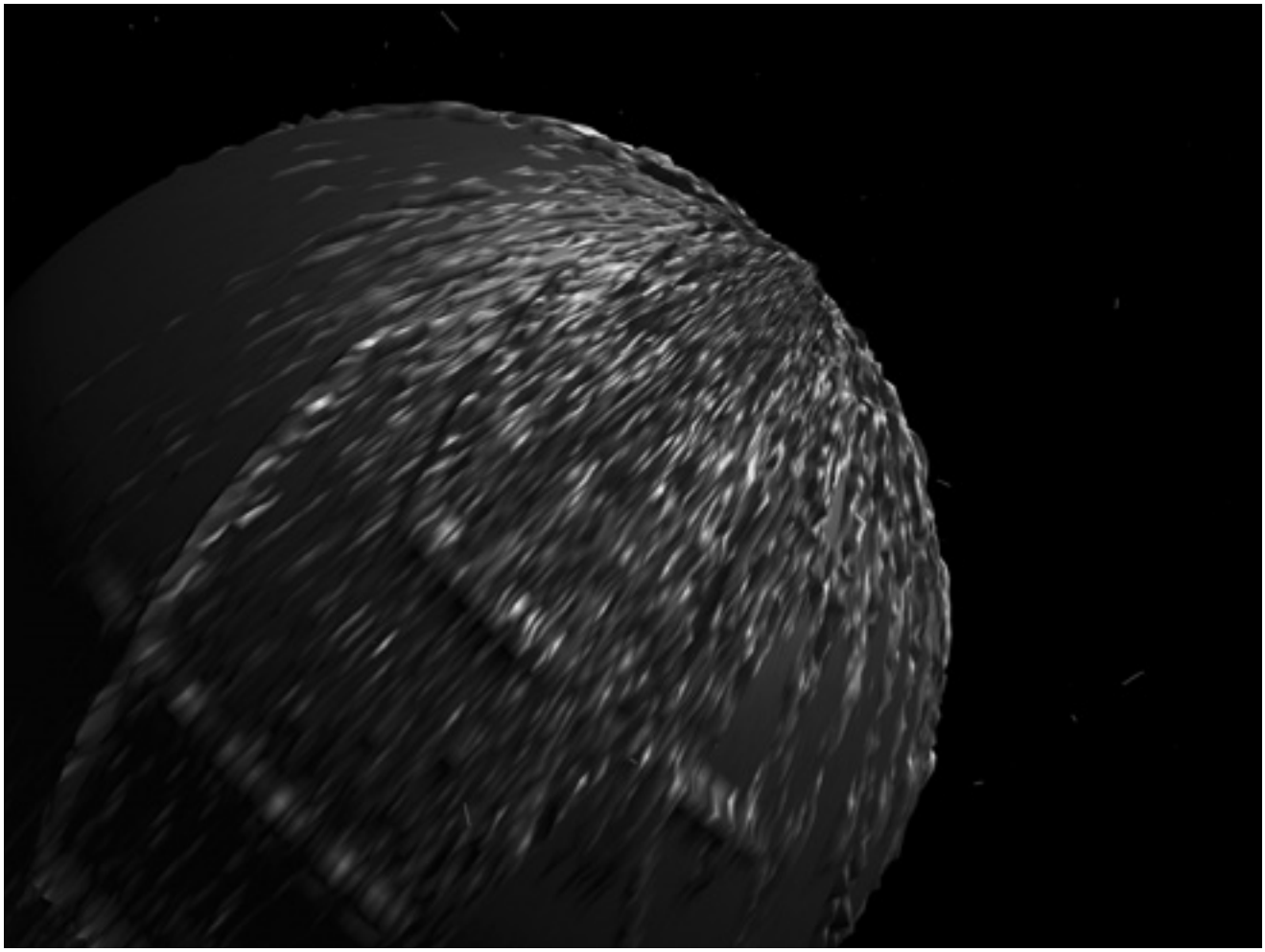
A piece of music consists of many parts, and in relation to the above discussion of contextualization, the listener is not really able to grasp the whole piece until all parts have been played. The idea is developed over time, and each moment has only limited significance when taken out of context. The model and animation in *Planet (Terra)* attempted to address this concept through a gradual revelation of the entire spectrum (the content) and through a circling camera movement—constantly moving, focusing, and refocusing different parts of the sphere. The piece, mapped around the sphere, is considered from all angles, becoming increasingly clearer as the piece plays and the model becomes more exposed. The camera movement is independent from the time in the music: it speeds up and slows down, adding excitement and calm in keeping with the ideas in the music. In this work, the camera and the music indirectly complement one another; rather than following the idea or development of the music, the camera is used to make more of a philosophical point about communication and understanding. When the image has been viewed from all sides, the camera self-destructs; the idea has been exposed and its job done.

Construction Drive

Construction Drive (see Figure 4) is a computer music game in the final stages of completion. The player navigates in a virtual world, and the graphics are based on a well-known painting. Apart from the associations engendered by the painting itself, the image has been used directly in two ways. The painting has been mapped onto a modeled, three-dimensional landscape, with geological formations that largely follow lines suggested in the image. Geometric and amorphous shapes in the painting have been made three-dimensional and situated as objects in the landscape according to their original placement. A substantial part of the sounding material stems from recordings of readings of Guillaume Apollinaire on themes related to the artistic style of this particular painting. Together, the sounds and images in the game are concerned with the nature of art and the riddle of perception. The player navigates by driving through the landscape, which is strewn with sounds that are performed as the player enters their active zone. This is a compositional technique that has also been employed in a game for the Microsoft Xbox (Schütze 2003).

In such interactive works as games, the composer relinquishes control of the distribution of sound events in time throughout the piece. In many computer games, composers employ sequencing software like Microsoft's Direct Music, building soundtracks from general MIDI patches and recorded sound, and changing orchestration through cues from user ac-

Figure 3. A view of the planet model showing the quasi-geological map after the waters have receded, taken from Planet (Terra) *(1987–88).*

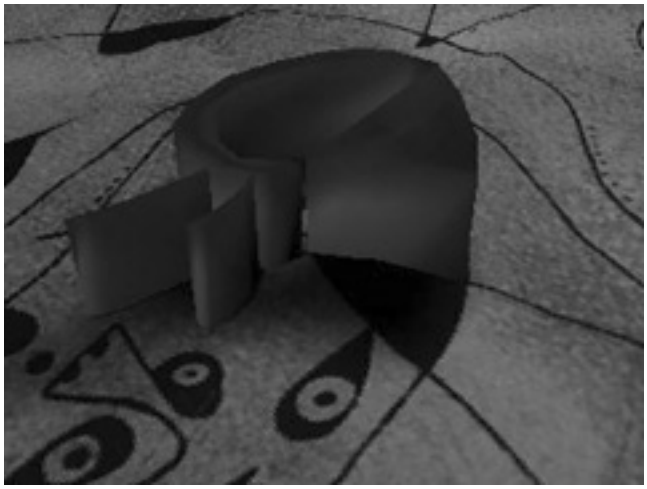


tions, game events, and locations in the game environment. This is an effective means of maintaining a high degree of real-time control of the soundtrack, but is better-suited as a tool for creating a computer game music track with mood changes and effects than computer music, owing to the limited number of available signal-processing routines and the high dependence on available RAM for loading material that can be processed.

In *Construction Drive,* the sequence of sounds resulting from player movement can be different every time, likewise the spacing of events. This type of situation is challenging for the composer. It is nonetheless possible to group sounds as different areas of timbral and thematic spaces, enabling control over the coherence in the piece. The visual elements reinforce the possibility of creating a thematic space, although the video narrative is nonlinear, given that the user has significant freedom to choose where in the landscape to drive. An important consideration has thus been to secure "good enough" musical results for the players to encourage them to return, and certain concessions have also been made to the game paradigm, which demands that there be a "point," or reason for playing.

As with any visual media, the eye "performs" the image through the act of looking, and the geological underlay for the surface, volumes, and colors of the

Figure 4. A preliminary example from Construction Drive *of a three-dimensionally modeled shape from the painting.*



modeled objects suggest connections and leads the gaze in certain directions, encouraging initial stages of interpretation. The color palette of the video reflects that of the painting, slightly filtered, and is intended to play a thematically supporting role by suggesting associations to the painter's style, vocabulary, and object types. The intention of the work is to tell a story about art's construction, identity, and context.

Conclusion

Integration of video and computer music expands the composer's opportunity for crafting multifaceted narratives, and it changes the art-making process through the use of new types of tools. The composer can use video to augment the musical idea or to develop a storyline not easily conveyed through sound alone. As discussed, the artistic idea may also be made apparent through the use of video as a tool for meta-composition, or as an educational tool to bring salient (but not necessarily prominent) elements to the listener's attention. In other words, in computer music video, imagery and audio are able to articulate and mirror their respective and mutual compositional and artistic characteristics and intents.

Although video art requires unique sensitivities and technical skills, it entails much of the same understanding of composition and dramaturgy as music. The works discussed have primarily exploited video as a complementary means of expressing musical ideas, by creating an additional layer of resonance for the listeners and greater opportunity for reflection or appreciation of complex interconnections of meaning and suggestion. It is perhaps in this superimposition of layers of meaning and interpretation that the most rewarding artistic aspects lie in the genre of computer music video.

References

Ballard, J. G. 1974. *Concrete Island.* London: Granada Publishing.

Bevilacqua, F., J. Ridenour, and D. J. Cuccia. 2002. "3D Motion Capture Data: Motion Analysis and Mapping to Music." *Proceedings of the 2002 Workshop/Symposium on Sensing and Input for Media-centric Systems.* Santa Barbara, California.

Bregman, Albert. 1990. *Auditory Scene Analysis.* Cambridge, Massachusetts: MIT Press.

Milosz, C. 1951. *The Captive Mind.* New York: Vintage.

Rudi, J. 1998. "Computer Music Animations." *Organised Sound* 3(3):193–198.

Sack, B. 2003. "DSP: Digital Sound Processing." *Computer Music Journal* 27(1):92–93.

Schaeffer, P. 1966. *Traité des Objects Musicaux.* Paris: Editions du Seuil.

Schütze, S, 2003. "The Creation of an Audio Environment As Part of a Computer Game World: The Design for *Jurassic Park—Operation Genesis* on the Xbox As a Broad Concept for Surround Installation Creation." *Organised Sound* 8(2):170–180.

Siegel, W. 1998. "The Challenges of Interactive Dance: An Overview and Case Study." *Computer Music Journal* 22(4):29–43.

Smalley, D. 1986. "Spectro-Morphology and Structuring Processes." In S. Emmerson, ed. *The Language of Electroacoustic Music.* London: Macmillan, pp. 61–93.

Smalley, D. 1997. "Spectromorphology: Explaining Sound-Shapes." *Organised Sound* 2(2):107–126.