Computer music animations

JØRAN RUDI

Norwegian Network for Technology, Acoustics and Music (NoTAM), University of Oslo, P.O. Box 1137 Blindern, N-0317 Oslo, Norway
E-mail: joranru@notam.uio.no
URL: http://www.notam.uio.no/~joranru
Fax: +47 22 85 79 74

1. INTRODUCTION

Electroacoustic music, since its inception, has been situated in a cross-disciplinary no-man's-land, with areas of interest spread in many directions; from ideas of musical structures ordered through traditional pitch-classes to research on physical modelling and analysis of sound and compositional structures through the use of neural networks. Much of the material found in computer music has been developed through processes of appropriation, both from sound recordings of physical events and from the application of various principles found in the natural sciences.

The binary system serves as a common denominator, and porting data from one domain to another is now a relatively trivial task, making it easy to consider the same data structures from different perspectives. Visual modelling of mathematical principles is common, and industrial use of visualisations (in, for example, the oil industry) is common in decision-making processes.

Principles of cross-disciplinary mapping can also be used in the visualisation of music, taking the modern multimedia experience one step away from the older notions of 'Gesamtkunstwerk', where music, images, text and movement merge. The choices made in the mapping process are important, because a mapping will never be a neutral transfer of information. Interpretation will play a significant role in determining the final results, and while mapping provides a great opportunity for crafting expressions, there are also certain responsibilities involved in not misrepresenting data or assigning importance where there is little basis for it. Particular care must be taken in academic circumstances within, for example, sonification of data, where inherent qualities of the chosen sounds themselves might suggest *interpretations* of the contexts in question. In artistic expressions, one would do well to remember that porting into a new domain will *add* something to the original data, due to the particular characteristics of the domain itself.

There is a broad correspondence between different kinds of artistic expressions – on stage or canvas, on screen or through speakers – and artistic gesture; time and content seem to gather energy from the same types of motions and relationships. However, a good measure of caution is needed when moving into new artistic domains, in order to avoid some of the problems evident in many of the multimedia installations, where striking lack of competence in one of the aesthetic domains often reduces the overall quality of the works.

My personal work for the last four years has revolved around cross-disciplinary mapping of data from the musical domain into the visual domain, resulting in what I have labelled *concert videos* – referring to the performance medium. The following is a description of four works based on the kinds of considerations described above. Similarities and differences between the works are outlined, as well as the different musical parameters that may be traced in each work.

2. *WHEN TIMBRE COMES APART* (1992–5)

Technically, this work was realised first as music, and the musical data was formatted through an FFT analysis of the same type used to make sonograms. This type of analysis was chosen for the nice 'spikes' produced by the windowing function. The data set was then made available for the Silicon Graphics program *Explorer*, which was used to create a computer model where the amplitude was mapped to 'altitude', and the frequency was mapped so that the lower frequencies appear to the right in the image, the higher to the left. The model was coloured, in order to separate the different sections of the piece, as well as to enhance the different character of each section. The colour was mapped so that the same amplitudes in a given section received the same colour. The model of the music was later 'filmed' with a scripted moving camera, and the result was an experience of flying over/under/through the music as it played. Both the model and camera movement were programmed by

Roger O. Nordby at the University Center for Information Technology (USIT) at The University of Oslo.

As with most of my compositions, *When Timbre Comes Apart* (*WTCA*) takes its point of departure in a set of ratios and other numbers describing a chosen phenomenon. The numbers are used to create a huge collection of material, from which the composition itself is created. Frequently, the same numbers are used to structure the overall form of the music, as well as movement on the level of micro detail.

The ratios are derived from an analysis of a bell timbre which is presented in the first ten seconds, and are used in various ways to structure nearly the whole piece. The first part is realised through synthesis only, and the ratios have been used to generate several 'generations' of sine tones. These frequencies have been organised in a variety of combinations and structures, and form the core of the first nine minutes of the piece.

The introduction of natural sound in the middle of the piece makes it difficult to discuss 'pure' ratios. The recorded material used is approximately 3.5 s, and most of the material stems from a recording 1.25 s long. The timbres of the piece are composed without regard for the references embedded in the recorded material (voice sounds from my, at the time, two-year-old son, where he toys with language when looking at a screensaver of the domestic Mac), but with a strictness in relation to the ratios. The processed sounds appear in many layers, where the treatment of each layer has varied according to the ratios. A principal idea has been to 'comb' noise-rich sounds, in order to create movement in and out of the (in)harmonic spectrum of the bell presented at the outset of the piece. Several kinds of granulation have been applied, and the numbers determining grain duration and density have been derived from the ratios. There are close to forty such granulations. The effect is most noticeable in the region 10′30″ to 11′30″, where the sounding timbre slowly changes from a 'jumping' quality to a 'shimmering' quality.

The noise components are stronger towards the end of the piece, and the 'combed' noise sounds come to their conclusion approximately fifteen minutes into the piece. The piece slowly disappears with sine tone clusters similar to those heard at the beginning of the work.

An overview would show that the first part of the work is presented as blocks, while in the second half of the piece, the blocks yield to richer spectra with a more detailed and careful balancing of the elements. The music continually refers to earlier developments in the piece, sometimes as direct quotes, sometimes as reinterpretations of the structural idea.

The sounds for this work have been realised and/or processed on Macintosh and Silicon Graphics Indy computers, using various software, as well as a KYMA system. The programs used were SVP, Soundhack, SoundDesigner II, KYMA, Lemur/LemurEdit for the Mac, and Ceres on the SGIs. The mix was done using Yamaha DMC 1000 mixers, Lexicon Nuverb and Studio Vision sequencer, running on a Mac Quadra 900.

The visual idea, as mentioned above, was based on using the data set from an FFT analysis. An FFT analysis provides data on the frequency components present in every moment of a sound. This data set was considered as a sonogram, a two-dimensional (2D) representation of time, frequency and amplitude. A model of the data set was created in the program *Explorer* on a Silicon Graphics Indigo II, and the material quality as well as lighting and lighting angle was set there. A number of small C-programs were written to generate the splines needed for smooth camera movement, and notation included both camera angle, focus, horizontal placement and vertical placement. The images, 25 s^{-1}, were shot onto a SONY CRV-disk, and transferred to Betamax tapes for the final sync/mix with the music.

By making a three-dimensional (3D) model of the sounding spectrum of the music, it became possible to focus the camera freely in relation to the musical material. I chose, however, to focus the camera on the spectrum that one 'was about to hear', in other words, the camera was right 'above' the current sound. The movement and focus of the camera was set with the intention of augmenting the musical development, either by focusing on the strong-sounding parts of the spectrum, or by showing weaker, yet significant and connected elements. It was important to never have the camera work *against* the music, and with a visual model displaying two minutes of material, the camera revealed what was coming next, depending on angle and distance of the perspective. It can be argued that this kind of preparation takes something away from the normal experience of temporality in music, but the expectation that is constructed is in many ways similar to what we encounter in verbal communication, where grammar and word categories form the foundation for pattern recognition. An example of this care in not working against the music can be found in the first 3′30″ of the piece, where the static character of the music is paralleled by a slow and nondramatic camera movement. Camera movement was also used to open and close sections, thereby assisting the listener/viewer in finding and structuring the time flow of the work. To reveal the 'virtuality' of the concert video, the camera also moved under the computer model, with sudden zoom effects and abrupt changes of perspective attaining the same result.

The images could not have existed without the music in this video, but due to the independent nature

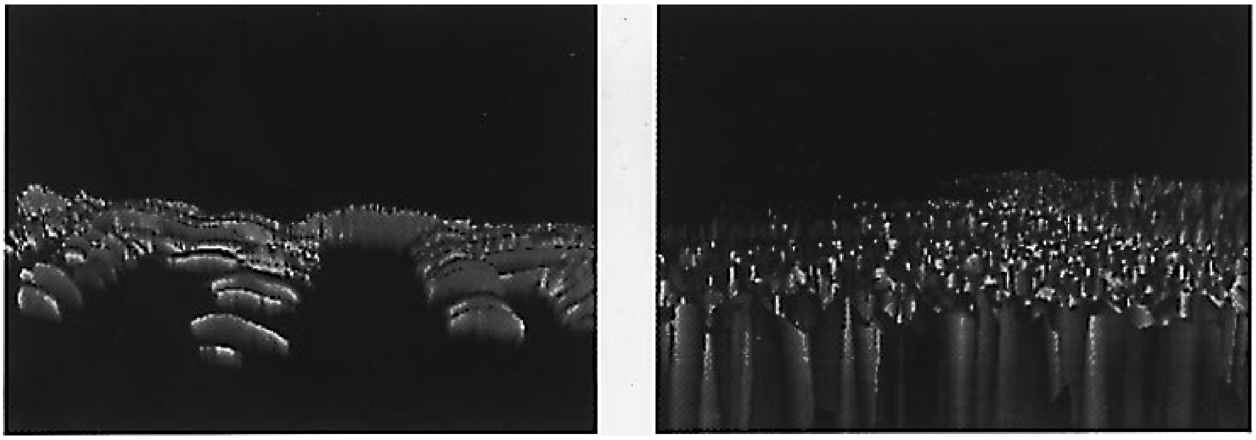

Figure 1. Examples of the images associated with *When Timbre Comes Apart*. Further examples may be found at http://www.notam.uio.no/~joranru/wtca.html

of the interpretation they represent, and the additional perspectives they bring forth for consideration, they should not be considered as a subset. It seems natural to say that the images and the music are reciprocal explanations of one another.

3. *PLANET (TERRA)* (1987–95)

The music was composed as a study at New York University in 1987, from sound material either synthesised or processed beyond recognition. This kind of play with source bonding is common within computer music, and crafting the music was easily executed on the Fairlight CMI. The (almost randomly) chosen material was subordinated completely to the program content of the piece, which explored the creation of structured order from timbral and rhythmical disjunctive elements resulting, nonetheless, in a construction of familiar sound icons.

The images are based on an FFT similar to the one used in *WTCA* described above. Here, the analysis data is mapped onto a sphere, easily lending itself to associations with geographical features on an imaginary planet. The graphic (spiky) nature of the FFT data is thus employed directly. To parallel the gradual building of meaning that takes place in the music, a second sphere was placed around the first one. This second sphere 'shrinks' as the music is playing, gradually revealing the 'landscape' underneath and thereby the content of the piece. The louder-sounding partials come into view before the softer-sounding ones, and the spectrum is fully displayed approximately halfway into the piece.

The model is coloured to resemble nature, and the camera movement is a swirling motion around the planet, starting from far away and ending with a dive into the 'sea' at the end of the music. There are few points of strict correlation between the music and the images; the idea was rather to provide an image of speed and change than to focus on singular musical elements, as was the case in *WTCA*. The model and animation were both programmed by Henrik Sundt at NoTAM, using Inventor and C++.

4. *CONCRETE NET* (1996–7)

Concrete Net represents both a continuation of the work described above and a departure from mapping the spectrum of the music. *Concrete Net* maps certain parts of the fundamental *idea* for the music, and images and sounds were developed simultaneously. Many of the visual objects are directly linked to the sound objects through physical modelling. The work is based on several ideas of quite different character. One idea for the sound material was taken from the long steel wires mounted by farmers in the western parts of Norway in order to transport hay from mountain farms to the valley floor – directly into their barns for winter fodder. The sounds of these wires have intrigued me since I first heard them, and the string sounds in the piece are recreated through physical modelling.

A wire is a component in a network, and information-sharing networks both isolate those unable to take part, as well as connect those with access. Access to modern communication technology is rapidly becoming a significant segregating element in society. J. G. Ballard writes intelligently about this in his novel *Concrete Island*, and recordings of excerpts from this book read aloud appear in the music. Structuring elements on both micro and macro levels are found in the distances between objects in our solar system, which also suggested the context for the computer models and the animation. The asteroid belt is represented visually by clouds of space debris – percussive sounds in the music.

The form of the music is thought of as a short excerpt from a longer journey. Only the first six

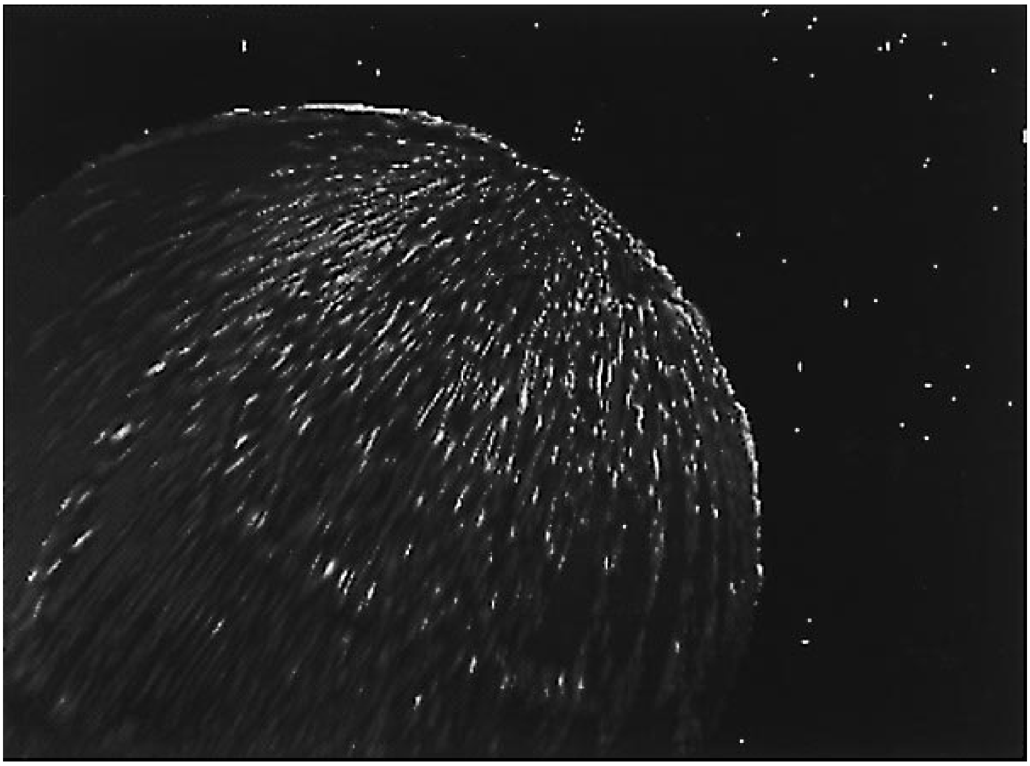

Figure 2. An example of the images associated with *Planet (Terra)*. Further examples may be found at http://www.notam.uio.no/~joranru/planet.html

minutes contain external movement where gestures and phrases are distinguishable on the 'surface' of the music. The material is often repeated or referred to later in the work, but pulled into the texturally dense second half of the piece – the material has become a part of memory. Thus a significantly less dramatic presentation is also called for in the movement of the images; the camera slowly performs a scan of the surroundings. The piece ends with a return to the same quiet with which it began – before alertness was provoked by the literal crash into the system of sounds that comprise the piece.

The recorded sound sources are traffic noise, falling aluminium scrap pieces and voice from a reading of text fragments from Ballard's book, *Concrete Island*. The orchestral-sounding glissandi result from granulated traffic noise, as do the deep rumbling sounds. The parameter values in the granulation are rescaled and adjusted versions of the ratios mentioned above, and the granulations are executed in KYMA from Symbolic Sound. The bird-like sounds over the green planet under the pink sky are recorded voice sounds, spectrally separated through use of the same numbers and mixing phrases from the recording; they are executed in Ceres, NoTAM's phase vocoder.

The floating cloud of space debris is represented through time-stretched and pitched versions of falling aluminium scraps, and the factors used stem from another reworking of the same ratios, also executed in Ceres, and mixed with Mix, NoTAM's screen-based mixer for SGI.

There are nine different sets of golden strings, the string lengths having been determined by distances in the solar system. Visually, the strings rotate with slightly different speeds and angles. The sound results from physical modelling executed in the IRCAM program Modalys, and the strings are 'welded' together almost at the centre (~0.58) so that the shortest string is connected to the next-to-shortest string and so on. The strings are excited by recorded sections of readings from Ballard's book, spectrally separated (used earlier in the piece as 'bird sounds'), with each slice exciting one string each. Each string in every set is excited by a different 'sound slice', and when one string is excited, all strings resonate. The output is quite a complex timbre, and the sonic result shares some characteristics with predigital spring reverbs, and relates in true fashion to the steel wires that can be heard in Western Norway.

The images were all generated on SGIs using Povray, a shareware ray-tracing program. C-programs generate a description of form and the appearance of the 3D objects, and these descriptions are sent to Povray, which calculates shadows, reflexes and ultimately renders the image. Colours and shapes are determined by subjective decisions based on aesthetics, and camera movement is linear with an even speed. The concert video thus appears to the viewer as a journey with only sounds and the objects associated to them present.

The generation of each image takes up to twelve hours, depending on the complexity, and with 25 pictures s^{-1} and a duration of thirteen minutes, the computation time is considerable. The rendering has been running continually on three machines for months, with the material being dumped onto a SONY CRV disk in batches for final sync/mix with the music.

The images for this piece are based on models of the computer instruments that produce the music and although not all sounds are represented visually, it is quite easy to follow the changes in the music through the images. Ballard's text is presented early on, and placed visually in a space clearly separated from the rest of the images. The recordings have undergone spectral separation based on ratios that will be used later in the work, and the section thus functions as a 'reverse echo' or something that has come 'back from the future'. This section is followed by percussive sounds in the music, modelled by a large cloud of elements crafted through mathematical descriptions of curved planes. Later in the piece, nine string sets appear and the visual dimensions are the same as those used in making the physical models that generate the sound in that section. The dimensions are transposed versions of the distances in our solar system, and the excitation is stimulated by the recordings of Ballard's texts; each string is excited by one of the 'sound slices' that appeared untreated earlier in the work. The overall sound quality of the strings is tweaked to resemble the steel wires mentioned above. Much more can be said about the sounds, the processing and the distribution of the spectra along the timeline, although this exceeds the scope of this article.

Where the two previously discussed animations work with the sounding spectra, *Concrete Net* maps the compositional idea to both visual and aural domains, as a structure consisting of musical parameters necessary for creating concentrated listening and time travel through sound. The camera movement is steady, with an even flow, and follows the music as travel unfolds through musical instruments laid out in a linear fashion. The objects and animation are both programmed by Øyvind Hammer at NoTAM.

5. *AND THE BIRDS . . . ?* (1997)

This piece was written for NoTAM's educational CD-ROM *DSP*, using only the signal processing and synthesis tools provided as part of the program. The piece is well documented on the CD-ROM, with the developmental history of each sound explained in detail to provide students insight into the compositional ideas on both micro and macro levels. The

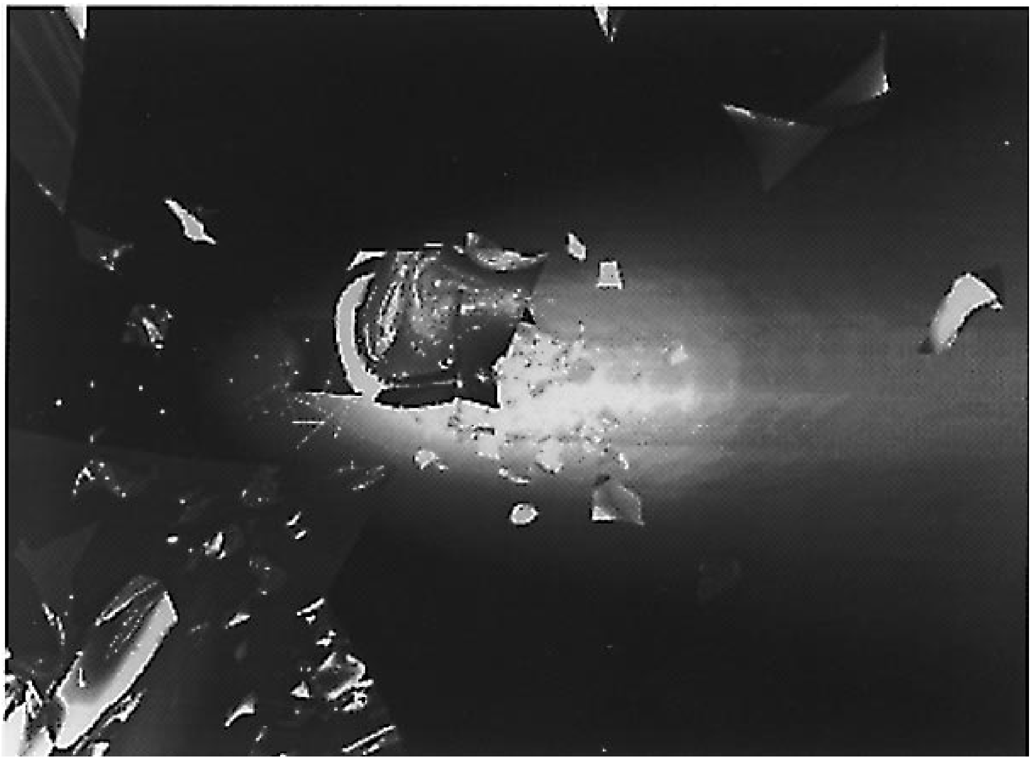

Figure 3. An example of the images associated with *Concrete Net*. Further examples may be found at http://www.notam.uio.no/~joranru/cn.html

purpose of the music was also to demonstrate *all* the DSP programs on the CD-ROM.

The construction of the piece was kept simple and the animation, which drew on ideas used in *When Timbre Comes Apart*, was tailored especially for educational purposes. The animation is a camera movement through a visual model of the spectrum based on an FFT. 'Street signs' are placed in the model at appropriate places, marking the transition from one compositional element to the next by displaying a drawing of something associated with the sound quality. There are also steps taken to make the animation more entertaining in a traditional sense, through the use of virtual strobe light and a modelling of wave movement for shifting the spectrum in a specific section. To most students, visual representations of music other than notes are novel and 'seeing' the music as it is being played is an effective means of teaching the physics of sound within music, as well as physics and mathematics in general. The model and animation were both programmed by Roger O. Nordby at the University of Oslo.

6. CONCLUSION: MODELLING AND INTERACTIVITY

Musical visualisations are common, and there are many theoretical perspectives on the combinatory nature of computer music animations that could be explored. The field of work distinguishes itself through a commitment to representation of measurable data, and it diverges from more 'popular' approaches where images are used in a more associative manner. The use of images would then aim to *add* something to the music, while computer music animations aim to show aspects of what actually comprises the music.

My main interest in making the works described in this article has been to investigate different approaches in mapping data from one domain to another. The intention has been to create artistic expressions and also to provide audiences with visual cues to the music in the form of parallel representations of both sound and image. (Naturally, the idea of parallelism is approximate – as soon as another form of expression appears, something is added into the context, which becomes altered.)

Other well-known and also well-documented approaches to visualising musical parameters can be found in forms of physical modelling, like the CORDIS-ANIMA system (Cadoz, Luciani and Florens 1994, Cadoz, Florens and Luciani 1995), and in the work at NCSA with CAVE-based applications for animation of sound data, often appearing as installations and pieces for performance by Insook Choi and Robin Bargar.

The combination of computer modelling of visual elements with the physical modelling of sounding instruments is integral in the development of virtual reality; the integration of virtual instruments and virtual acoustics has been a vital area of research for many years (Huopaniemi, Karjalainen, Valimaki and Huotilainen 1994, among others). However, musical interaction in virtual reality is a complicated issue, and for interactivity to move beyond the simple activation of a curious audience there are issues of quality and depth that need to be addressed.

Interactive musical art works, where the audience participates in the realisation, need to be open-ended and provide the user with good signal processing and composition tools, structured as objects designed for different levels of interaction. Modelling of different musical parameters on several levels is also interesting, and I like to think of the animations described above as small investigations in the direction of an art where many of the differences between composer/artist and audience become dissolved – an art where the audience input provides form to the artistic abstraction represented in the algorithm.

REFERENCES

Cadoz, C., Florens, J. L., and Luciani, A. 1995. Musical sounds, animated images with cordis-anima and its multimodal interfaces. Demonstration at *International Computer Music Conference*.

Cadoz, C., Luciani, A., and Florens, J. L. 1994. Physical models for music and animated image. The use of CORDIS-ANIMA in "ESQISSES", a music film by ACROE. *Proc. 1994 Int., Computer Music Conf.*, pp. 11–18. Aarhus, Denmark.

Huopaniemi, J., Karjalainen, M., Valimaki, V., and Huotilainen, T. 1994. Virtual instruments in virtual rooms – a real-time binaural room simulation environment for physical models of musical instruments. *Proc. 1994 Int. Computer Music Conf.*, pp. 455–62. Aarhus, Denmark.