Visualization by Ear: Auditory Imagery for Scientific Visualization and Virtual Reality

Article *in* Computer Music Journal · December 1992

DOI: 10.2307/3681076

1 author:

Gary S. Kendall
Queen's University Belfast

Gary S. Kendall
Computer Music Studio
School of Music
Northwestern University
Evanston, Illinois 60208, USA
gary@music.nwu.edu

Visualization by Ear: Auditory Imagery for Scientific Visualization and Virtual Reality

The title of our conference, "Dream Machines," is intentionally ambiguous and suggests at least two rather disparate, literal interpretations—one, the machines we dream of, and the other, the machines that produce our dreams. These two conflicting interpretations would seem easily resolved in a meeting of computer music practitioners who have struggled to create art with distinctly inartistic hardware and who have coveted machines more responsive to their artistic goals. Then again, this is a meeting of people who are focused and centered in auditory experience: "audiles" is the dictionary's term. I think that all of us have been motivated to be here and devote our lives to the art and science of sound because we have experienced a dream of sound, a dream that could be made real only by a "dream machine." The obvious gap between the two conflicting interpretations of our conference title can be bridged by a marriage of the two which suggests "the dreams that audiles have of machines that realize their auditory dreams."

I will focus my discussion on a particular "auditory dream," rather than the machine. I wish to focus on two interrelated application areas that lie somewhat outside the obvious scope of computer music: scientific visualization and virtual reality. I believe that the design of "computer music systems" can be conceived broadly enough to serve goals other than creating musical compositions, and that computer music composition can be conceived of broadly enough to include the creation of the auditory images for new forms of computer technology. Composers have special insights and perspectives on the construction of sound that should inform this work and guide the development of audio for these new technologies.

Computer Music Journal, Vol. 15, No. 4, Winter 1991,

Definitions

Scientific visualization and artificial reality are catch words for a variety of efforts being made to provide more perceptually relevant interfaces to computing machinery. Scientific visualization projects have been concentrated in supercomputer centers. The primary goal of scientific visualization is to provide scientists with improved representations of complex phenomena, whether empirically measured or computationally simulated. This work has been dominated by certain of the tools at hand, most particularly three-dimensional, color computer graphics. Modern computer graphics represents an obvious extension of traditional visual means of scientific data presentation. But when one is forced to grapple with phenomena of many dimensions that are not well related to three-dimensional objects in space and time, the limitations of simple graphic approaches are easily recognized. At present, auditory imagery is essentially absent from scientific visualization.

Virtual reality is a catch word for a collection of human-computer interface technologies that aspire to create perceptual analogs to reality. This development has been largely driven by the demands placed on the human operators of avionic equipment (pilots, air traffic controllers, astronauts, etc.), but it also extends to remote sensing, telerobotics, and telepresence systems. In all of these cases, a human is presented with perceptual images akin to those directly perceived in reality, even though the images represent events far removed and often impossible to experience directly. In a virtual reality the observer's point of view is within the data representation in the same sense that we are within the environment. This is essentially a change in spatial orientation. In a virtual reality one can also interact with the environment and change it—the viewer becomes a participant. There is an integration of visual and auditory senses akin to that experienced in everyday life.

Common Senses

We experience the events of everyday life through all of our senses, and our primary memory and understanding of these events is captured in the information from all our perceptual systems. Even though vision and hearing capture different kinds of information, we are acclimated to their integration and are seldom aware of the extent to which we rely differentially upon either one of them. For example, we may predict the potential collision of two objects by observing their paths with our eyes, but it is the sound of the collision that best reveals how the structure of the objects has been affected by the collision.

Our experiences of events outside of everyday life such as those we might experience through electronic sensing or computer simulation are going to be remembered and understood by scientists or astronauts in a manner that is analogous to the experience of everyday life. The combination of visual and auditory imagery offers a way of presenting and communicating complex events that emulates the richness of everyday experience. This use of auditory information should be consistent with normal experience even if the phenomenon being represented is far outside of normal experience. This is how one would hope to gain a perceptual integration of information that is not possible through visual information alone. Even though in their initial development scientific visualization and artificial reality have been driven by different concerns, their essential similarity is unmistakable: they both profit from encoding representational data into appropriate auditory events.

Sound Events

In everyday life, sound events arise from action, in fact, from the transfer of energy to a sounding object.The auditory system provides us with perceptual characterizations of the energy transfer and of the internal structure of the objects involved. Early in childhood one learns to recognize the occurrence of sound events and to relate them to physical events. Through a lifetime of experience one learns to classify heterogeneous sound events and to identify them.

Some classifications of sound events tend to be categorical. Excitation functions are typically of discrete types such as hitting, scrapping, blowing, vocal glottis, etc. Some classifications of sounding objects are similarly categorical—metal, wood, hollow, solid, vocal tract, etc. These simple categorical distinctions can potentially be exploited in auditory presentations to communicate important distinctions in the data.

Beyond these categorical distinctions, the essential goal is that perceptually continuous auditory attributes are scaled and mapped to data attributes in a way that is meaningful to the observer. Relevant changes in data should insure a change in what is perceived. Changes in what is perceived should signify meaningful changes in the data. The appropriate scaling functions will probably not exist a priori. Psychophysical scaling experiments may be needed in order to create perceptual scaling functions through which collections of auditory stimuli are mapped. This is made feasible only by utilizing a limited number of auditory tokens with well-understood perceptual properties. This suggests that sets of tokens be developed and scaled in advance.

It might seem desirable that steps of equal significance in the data be represented by steps of equal significance in the auditory tokens. Upon reflection, it is more important that a viewer be able to adjust the mapping of the tokens in order to evaluate a variety of sonic viewpoints on the data. This is akin to adjusting the color enhancement so as to highlight certain features of a photograph. A more complex issue arises from the fact that natural acoustic events are not evenly distributed in orthogonal perceptual dimensions: changes in one acoustic attribute may cause simultaneous changes along several perceptual dimensions. Once again, this suggests the design of auditory tokens specifically for the purpose at hand.

In the context of a virtual reality similar to the everyday world, sound events would be triggered by the same happenings as in everyday life. In the context of other conceivable virtual realities or in most scientific visualization, other happenings or features of the data could trigger events. The triggers might be threshold crossings, slope reversals, boundary detectors, counter targets, etc. The triggering of events takes on a causal meaning only in the context of data being represented. It is important that communication derives from the pattern of the whole auditory presentation in context—capturing multidimensional attributes in a fashion that is immediate, revealing, and memorable.

Grouping of Events

One of the predominant features of auditory processing is the sequential grouping of events into auditory streams. In presenting auditory events, there are many potential groupings which can be exploited to reveal underlying structures in the data. In everyday experience, the auditory system forms streams with events from a single sound source or related sound sources. This is largely based on the continuity through time of spectral profile maintained by the permanence of the object's mass and structure. Naturally, stream formation is strongly influenced by the choice of sounding objects, but then, every sounding object has a vast potential repertoire of acoustic utterances that depend on the exact excitation function. So, the continuity of composite spectrum is dependent on the continuity of the excitation functions too.

In everyday life, streams of sound events are strongly characterized by their perceived pitch height; there are high streams and low streams. This property is shared and amplified by musical instruments. The use of conventional musical pitch to affect stream formation probably creates a potential for misinterpretation due to the unintended implications of musical grammar. This potential problem can be avoided by utilizing sound sources with pitch height, but not pitch chroma.

Stream formation is also affected by the rate of presentation. The choice of rate can cause streams to merge or segregate. The ultimate result of manipulating all of these organizing factors will be to shape the listener's perception of temporal sequence. Temporal sequence is most clearly grasped and remembered within a single stream. In fact, temporal sequence across streams is often difficult to sort out. The great power of the auditory system to group temporal events contrasts strongly with the visual system. In the context of scientific visualization and virtual reality, it is clearly auditory information that excels in communicating temporal structure.

Spatial Organization

There is no area in which the visual and auditory systems would seem more complementary than in spatial perception. No technological linkage seems more obvious than that of three-dimensional graphics to three-dimensional sound. In fact, the head-mounted visual display systems developed for avionics are ideally suited for use with three-dimensional sound because adjustments for sensed head movement are a very powerful influence on improving auditory localization performance.

There are important contrasts between the two perceptual systems, however. The visual system is more spatially acute than the auditory system, especially in sensing the angle of elevation, and, whereas the visual spatial field is limited to the region in front of the viewer, the auditory spatial field is unbounded and surrounds the listener. Detecting sounds outside the visual field provokes the listener to turn and orient to the event. Sound is not blocked by objects in the same way as light, and so one can hear around obstacles and give attention to unseen events, wherever they are.

Our intrinsic spatial frame of reference is our body and our understanding of spatial dimensionality results from the interaction of our body with the environment. The acoustic and neurological organization of auditory spatial perception suggests an underlying two-dimensional framework (that is not Cartesian). Time and intensity differences between the two ears determine the position of a sound event along a left/right axis; this is referred to as "lateralization" and it underscores the fundamental left/right symmetry of the body. Lateralization is one of the most fundamental dimensions of auditory organization. The acoustics of the torso, head, and pinna provide cues to the location of a lateralized sound source along the circular front/above/back/below dimension, which underscores the asymmetry of the body. These two types of dimensions are referred to as simplex and circumplex, respectively.

The reach of our bodies provides a frame of reference for "near" and "far." Distance perception of near sound events relies on direct acoustic interaction with our bodies. Distance perception of far sound events relies on indirect sound from the environment. One of the most powerful attributes of spatial sound is its ability to position sound events within or outside of the personal space around our bodies. Sound events near the body demand attention.

Indirect sound affects our perception of sound events and environments. The spatial focus of a sound event can be quite high or low depending on the relative intensity (and timing) of the indirect sound. Events with very low spatial focus do not appear to originate from any specific location. Indirect sound also provides us with a sense of environmental context. That environmental context might be open or enclosed, large or small, reverberant or nonreverberant, etc. A region within a virtual reality or scientific phenomena can be imbued with a sense of place by a simulated environment. As in reality, one can pass between environments that have differing characteristics. These environments can be adjacent to each other or one can enclose the other. As in reality, there can exist windows to environments that contain sound events but which we observe from the outside.

The purpose of this discourse on the underlying dimensionality of spatial sound is to reveal the multiplicity and complexity of potential auditory-spatial relationships. While the mapping of three-dimensional graphic events into three-dimensional auditory space is straightforward, there are many other sets of relationships that can be interpreted and exploited within an auditory-spatial framework.

Conclusion: Auditory Visualization as Orchestration

Job Description: selecting acoustic tokens, assigning tokens to events so as to create auditory streams; organizing sound events in auditory space

Job Title: New Technologies Orchestrator

One of the strongest recommendations for the role of computer musicians in the development of sound for these new technologies is the inherent similarity of the tasks described here to musical orchestration. The scientist's data and the computational representation of virtual reality are like musical scores waiting for the orchestrator's guidance before being performed. There are already role models for this interdisciplinary task in the visual artists at supercomputer centers who are now providing the computer scientists with an artistic point of view. Even though the auditory images for scientific visualization and virtual reality may not suggest musical sound per se, they do suggest sound akin to that of everyday life, a synthetic form of *musique concrète*, the sound of unheard scientific phenomena and unvisited virtual worlds—the experience of an auditory dream.