Autonomy, Interaction, and Presence

1 Virtual Environment Technology

Computer simulation of real world objects and processes is certainly not a new endeavor. After all, the very first electronic digital computing engines were employed in the computation of fluid dynamics and ballistic trajectories (Harvard Computation Laboratory, 1985). What is new is the advent of graphics workstations capable of rendering reasonable visual approximations of familiar objects in near-realtime, coupled with commercially available wide field of view (FOV) head-mounted displays (HMD) slaved to the user's head motions (Blanchard, Burgess, Harvill, Lanier, Lasko, Oberman, & Teitel, 1990; Fisher, McGreevy, Humphries, & Robinett, 1986). These rather recent developments promise to make the virtual environment a medium that can engage our attention as can few others, which has generated a great deal of public interest in these systems.

In this article I will present a taxonomy of graphic simulation systems, based on three salient components: *autonomy, interaction,* and *presence.* The resulting *AIP cube* provides a useful qualitative tool for describing, categorizing, comparing, and contrasting virtual environments, as well as more conventional computer animation and graphic simulation systems. Moreover, such a taxonomy can help us to identify application areas as well as avenues of research to pursue.

In addition to the computing platform, graphics engine and associated peripherals, three key components of a virtual environment—or indeed, of any graphic simulation or computer animation system—are (1) a set of computational models of objects and processes to be simulated, (2) some means of modifying the states of these models over the time course of the simulation, and, finally, (3) communication channels that allow the participant to experience the simulated events and processes through one or more sensory modalities. Let us examine each of these in turn.

2 Autonomy

At one extreme, a computational model in computer graphics may be a passive, geometric data structure with no associated procedures. We may apply the usual affine transformations to these models and then render them. At the other extreme are virtual actors capable of reactive planning, and, ultimately, more powerful knowledge-based behaviors. Between these extremes, we can augment models of objects and agents in various ways, for example, with procedures that account for the mechanical properties of rigid and nonrigid objects, yielding what has come to be known as physically based models. *Autonomy* then, is a qualitative measure of the ability of a computational model to act and react to simulated events and stimuli, ranging from 0 for the passive geometric model to 1 for the most sophisticated, physically based virtual agent.

3 Interaction

In this context, interaction means the degree of access to model parameters at runtime (i.e., the ability to define and modify states of a model with immediate response). The range is from 0 for "batch" processing in which no interaction at runtime is possible, to 1 for comprehensive, realtime access to all model parameters.

Most current graphics systems are indeed highly interactive, such that realtime manipulation of rigid objects is often controlled by joysticks, knobs, or a mouse. In other application domains, such as computational fluid

David Zeltzer
Computer Graphics and Animation Group,
The Media Laboratory,
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

dynamics, interaction remains quite limited due to the computational cost of updating the model at each time step.

Note that autonomy implies levels of abstraction, which are crucial to the representation of behaviors. And, as I have discussed in detail elsewhere, interaction paradigms vary as one accesses model parameters at different levels of abstraction (Zeltzer, 1991). Moreover, merely providing direct access to many parameters is not necessarily productive, since it is easy to overwhelm a user with the sheer number of parameters and tasks to be attended. This has been referred to as the ***degrees of freedom problem*** (Turvey, Fitch, & Tuller, 1982), and it means that in a complex work domain, such as a virtual environment with many actors and processes, the input operations must be properly organized so as to reduce functionally the number of degrees of freedom that must be directly controlled (Sheridan, 1987; Vicente & Rasmussen, 1990), regardless of the input modality, whether through the keyboard, conventional graphical input devices, or such hand-motion measurement devices as the DataGlove. The latter, while increasing the sense of "presence" (discussed below), do not necessarily reduce the complexity of the control task; that really depends on understanding the functional relationships among input parameters, and implementing the proper abstraction classes, which is what is represented by the autonomy–interaction plane (Zeltzer, 1991).

4 Presence

We are immersed in a very high bandwidth stream of sensory input, organized by our perceiving systems, and out of this "bath" of sensation emerges our sense of being in and of the world. This feeling is also engendered by our ability to affect the world through touch, gesture, voice, etc. The presence axis provides a rough, lumped measure of the number and fidelity of available sensory input and output channels.

A discussion of presence is quite meaningless without specifying the application domain and the task requirements. For example, consider the HMD and the null-stimulus, which we can think of as a structureless, diffuse, evenly illuminated visual field, the so-called *Ganzfeld* (Cornsweet, 1970; Grind, 1986). What visual cues should be provided to the HMD wearer, beyond this null-stimulus, so that he or she feels "present" in some environment? Clearly, the visual field must provide some structured imagery lest the wearer feel immersed in a featureless void. Gibson suggests that the experience of perceiving some volume of space requires at least the presentation of texture gradients that can be interpreted as a surface (Mace, 1977). Alternatively, fixed reference points (Howard, 1987) or a uniformly moving visual field can also give rise to a perception of space (Hochberg, 1986). But is stereo viewing required to feel "present." Should a collimated display be provided? What level of photorealism is appropriate? To answer these and other key questions, we need to specify present ***where*** and for what ***purpose***. In practice, implementation is guided by what has been called ***selective fidelity*** (Johnston, 1987). It is not possible to simulate the physical world in all its detail and complexity, so for a given task we need to identify carefully the sensory cues that must be provided for a human to accomplish the task, and match as closely as possible the human perceptual and motor performance required for the task. Determining the operational parameters inevitably involves many tradeoffs among cost, performance, and efficiency. For visual cues, such parameters include throughput of geometric primitives, visual update rate, and display resolution, all of which are critical numbers for any computer image generator. Unfortunately, in many of these areas, current technology falls far short.

Just as the conventional logical input devices (e.g., locator and valuator) allow us to separate functionality from the operational detail of physical input devices (Foley, van Dam, Feiner, & Hughes, 1990), I distinguish between ***interaction*** (i.e., the degree to which model parameters can be accessed at runtime) and the ***means*** of accessing those parameters. Therefore, an analysis of the kinds of input devices provided belongs also to the measure of presence: are the input devices designed to monitor our movements and speech, or are they constructed to measure the motions we impart to special kinematically constrained assemblies such as buttons (vertical motion), dials (rotary motion), and the mouse (2-D translations)? Put another way, a widely accepted working hypothesis states that using our familiar sensorimotor skills to manipulate virtual objects directly by means of whole-hand input devices, like the VPL DataGlove, contributes to our sense of presence much more than writing programs, twisting knobs, or pushing a mouse to accomplish the same task. Thus the presence dimension provides a measure of the degree to which input and output channels of the machine and the human participant(s) are matched.

5 The AIP Cube

The three axes—*autonomy, interaction,* and *presence*—define a coordinate system we can use as a qualitative measure of virtual environments, computer animation, and graphic simulation systems. Let us examine the positive octant formed by the three axes (see Fig. 1).

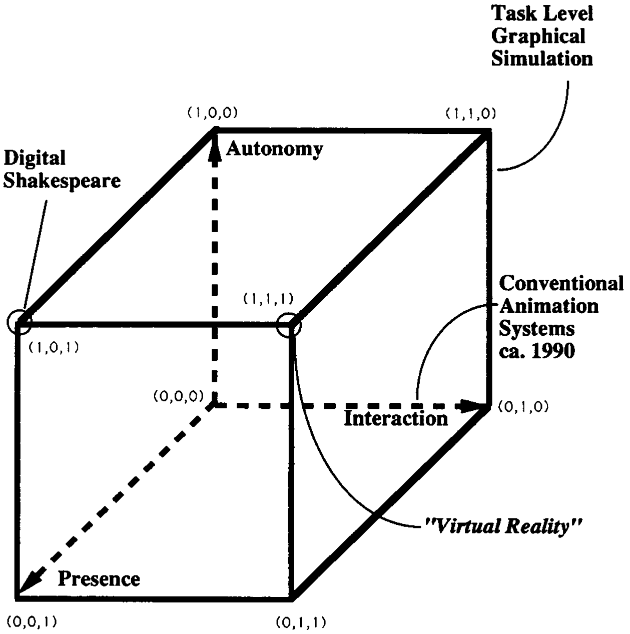

Figure 1. *The AIP-cube.*

At the origin (0,0,0) we have essentially the situation as obtained in the early 1960s—models with no autonomy, and systems with no interaction and no presence (i.e., batch processing of simple graphic models, with the results portrayed on a bed plotter or perhaps output to a film recorder).

In contrast, the corner (1,1,1) is our "grail": fully autonomous agents and objects that act and react according to the state of the simulation, and that are equally responsive to the actions of the human participant(s). In addition, sensory simulation provided to the participant(s) in the virtual environment is indistinguishable from what would be expected in a physical setting. That is to say, this node represents a hypothetical future system in which one could feel a cool breeze on the back of the neck, or experience what it is like to pull 9 Gs in a high-performance jet aircraft—truly a "virtual reality." Is such a system achievable? Probably not without some form of direct stimulation of nervous tissue, about which I will not speculate. Is such a system desirable? Certainly not for all applications. For example, if one were teleoperating a real or virtual robot performing what for a human would be a physically demanding task, one would quickly be overcome with fatigue if all the forces experienced by the robot were reflected back to the human participant. Moreover, if the robot were in some hostile environment—say, a nuclear reactor on fire, or a battlefield—one certainly would want to restrict the presence cues!

The point (0,1,0) represents a system where all model parameters are subject to user control at runtime. Conventional commercial animation systems approach such a point, providing dozens if not hundreds of parameters that can be modified directly by the user, generally through the use of conventional input devices (e.g., tablet and mouse). For the most part, users of such systems must operate on passive two- and three-dimensional geometric abstractions, which forces them to construct synthetic events in all their detail. The sheer number of details that must be specified can be overwhelming, which makes the control task very difficult, and also requires users to be quite expert with design and animation techniques.

Currently available commercial virtual environment systems begin to approach the point (0,1,1), since, in general, there is a nontrivial amount of interaction with object models in the system and an interesting (and often compelling) level of presence is implemented through the use of a binocular HMD with wide, head-slaved FOV, and glove-like hand motion measurement devices. For example, in the current "RB2" system offered by VPL, participants can interact with a graphic simulation of a robot manipulator, by "grabbing" the end-effector with the DataGlove and moving the arm (i.e., updating the position and orientation of the robot gripper) (Blanchard et al., 1990). Autonomy is low, however, since the computational models in these systems are for the most part rather passive geometric data structures with few attached procedures.

Continuing around the base of the cube, the point (0,0,1) represents a noninteractive system with sophisticated sensory stimulation devices, which is based on passive, geometric models that are neither autonomous nor modifiable in real time. For example, a precomputed conventional animation sequence viewed with HMD and tactual feedback would approach this corner of the AIP cube. The "Star Tours" attraction at Disneyland is another example of such a system.

The point (1,0,1) represents an interesting combination in which autonomy is high (i.e., the simulation contains virtual actors and physically based models), and where presence is high, but there is little or no interaction with characters. I think such a system could be, and likely will be, quite successful commercially. Imagine being able to put on an HMD and stroll among virtual actors performing, say, a play by Shakespeare. High autonomy means that the virtual actors could respond to, say, natural language scripts describing the performance, a topic of much recent research (Badler & Webber, 1991; Ridsdale, Hewitt & Calvert, 1986), and which would go far toward making such "virtual theater" pieces economical to produce. One would be able to view the action from various viewpoints, and perhaps even rewind or fast forward the piece. However, the virtual actors would be entirely oblivious to the viewer's virtual presence.

The corner (1,0,0) represents a graphic simulation with physically based agents and virtual actors, viewed on a conventional display device with no interaction. Portions of the animation *Grinning Evil Death* are prototypical of such a piece (McKenna, Atherton, & Sabiston, 1990).

Finally, the corner (1,1,0) represents something close to the current virtual environment system we have implemented at the Computer Graphics and Animation Group of the MIT Media Lab, which supports a number of autonomous models, and which also supports comprehensive and varied access to model parameters (Zeltzer, Pieper, & Sturman, 1989). However, the level of presence is not high, since "natural" input is limited to kinematic, whole-hand input, and we currently do not use an HMD.

6 Conclusion

The AIP cube with its three axes—autonomy, interaction, and presence—provides a conceptual tool for organizing our understanding of current virtual environment technology. In this scheme, for example, "virtual reality" is an appropriate label for the unattainable node in which the value of all three components is unity. It is not clear how to quantify rigorously these components, however, and much work remains, for example, in understanding how to measure selective fidelity.

In the autonomy domain, work on physically based modeling of rigid and nonrigid bodies continues in many labs, including work on anthropometric (Lee, Wei, Zhao, & Badler, 1990) and physically based jointed figure motion (Wilhelms, Moore & Skinner, 1988) and reactive planners (Zeltzer & Johnson, 1991). In terms of interaction, the problem is to understand how to organize and functionally abstract complex control spaces to make them more amenable to safe and effective human operation. In terms of presence, we need to improve our understanding of human perception for several reasons. First, a better understanding of human sensory mechanisms will make it possible for us to design and implement effective devices for enhancing presence (e.g., tactual feedback). Other perceptual phenomena, such as binaural hearing, may be stimulated artificially, but at great computational expense (Wenzel & Foster, 1990), so that algorithm and hardware development is necessary to realize economical, realtime devices. Second, there are presence cues that are simply not practical to synthesize, and we therefore need to know how to suggest sensations or substitute other kinds of cues when possible or appropriate. For example, applying the linear and angular accelerations experienced by the pilot of a high-performance jet aircraft is simply not possible to achieve in a ground-based flight simulator (Rolfe & Staples, 1986). However, a sufficient understanding of the sensory processes involved makes it possible to provide a variety of cues that sum to a fairly convincing experience of being in a moving cockpit. Finally, it is important to develop a taxonomy of tasks in terms of sensory input: for a given task, which sensory cues are necessary, and which are dispensable but improve performance? Are there sensory cues that do not affect performance per se, but that enhance the aesthetics of the operations or the workplace? Are there sensory cues that interfere with performance and that should be avoided?

As we proceed with research and development of virtual environment systems, we should not lose sight of the fact that each time we open a book, or go to a movie, or simply close our eyes and daydream, we enter a "virtual reality." For millennia, artists, musicians, writers, and storytellers have sought to involve our senses and imaginations in worlds that have no physical basis. Our new electronic tools, therefore, rather than offering entirely new experiences, in many ways merely transform media with which we are already familiar. In our pursuit of autonomous computational models, more effective and manageable interaction paradigms, and increased presence, we will doubtless learn as much or more about ourselves as we will about the new technologies we develop and explore.

Acknowledgments

This work was supported in part by NHK (Japan Broadcasting Corp.), National Science Foundation Grant IRI-8712772, and an equipment grant from Hewlett-Packard, Inc.

References

Badler, N. I., & Webber, B. L. (1991). Animation from instructions. In N. Badler, B. Barsky, & D. Zeltzer (Eds.), *Making them move: Mechanics, control and animation of articulated figures* (pp. 51–93). San Mateo, CA: Morgan Kaufmann.

Blanchard, C., Burgess, S., Harvill, Y., Lanier, J., Lasko, A., Oberman, M., & Teitel, M. (1990). Reality built for two: A virtual reality tool. *Proceedings of the 1990 Symposium on Interaction 3D Graphics,* March 25–28, Snowbird, UT, 35–36.

Cornsweet, T. N. (1970). *Visual perception* (p. 371). San Diego: Harcourt, Brace Jovanovich.

Fisher, S. S., McGreevy, M., Humphries, J., & Robinett, W. (1986). Virtual environment display system. *Proceedings of the 1986 ACM Workshop on Interactive Graphics,* October 23–24, Chapel Hill, NC, 77–87.

Foley, J. D., van Dam, A., Feiner, S. K., & Hughes, J. F. (1990). *Computer graphics: Principles and practice* (2nd ed.). Reading, MA: Addison-Wesley.

Grind, W. A. van de. (1986). Vision and the graphical simulation of spatial structure. *Proceedings of the 1986 ACM Workshop on Interactive Graphics,* October 23–24, Chapel Hill, NC, 197–235.

Harvard Computation Laboratory. (1985). *Proceedings of a symposium on large-scale calculating machinery, January 1947.* Cambridge, MA: The MIT Press. The Charles Babbage Institute Reprint Series for the History of Computing, Vol. 7.

Hochberg, J. (1986). Representation of motion and space in video and cinematic displays. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 1, pp. 22-1–22-64). New York: Wiley.

Howard, I. P. (1987). Spatial vision within egocentric and exocentric frames of reference. *Proceedings NASA Conference on Spatial Displays and Spatial Instruments,* August 31–September 4, NASA CP 10032.

Johnston, R. S. (1987). *The SIMNET Visual System,* Proc. 9th ITEC Conf., Washington, D.C., Nov. 30–Dec. 2, 264–273.

Lee, P., Wei, S., Zhao, J., & Badler, N. (1990). Strength guided motion. *Computer Graphics, 24*(4), 253–262.

Mace, W. M. (1977). James J. Gibson's strategy for perceiving: Ask not what's inside your head, but what your head's inside of. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 43–65) Hillsdale, NJ: Erlbaum.

McKenna, M., Atherton, D., & Sabiston, B. (1990). *Grinning evil death.* Cambridge, MA: MIT Media Lab, Computer Animation.

Ridsdale, G., Hewitt, S., & Calvert, T. W. (1986). The interactive specification of human animation. *Proceedings of Graphics Interface 86,* Vancouver, Canada, May 26–30, 121–130.

Rolfe, J. M., & Staples, K. J. (Eds.) (1986). *Flight simulation.* Cambridge: Cambridge University Press.

Sheridan, T. (1987). Supervisory control. In G. Salvendy (Ed.), *Handbook of human factors* (pp. 1243–1268). New York: Wiley.

Turvey, M. T., Fitch, H. L., & Tuller, B. (1982). The problems of degrees of freedom and context-conditioned variability. In J.A.S. Kelso (Ed.), *Human motor behavior* (pp. 239–252). Hillsdale, NJ: Erlbaum.

Vicente, K. J., & Rasmussen, J. (1990). Mediating "direct perception" in complex work domains. *Ecological Psychology, 2*(3), 207–249.

Wenzel, E. M., & Foster, S. H. (1990). Realtime digital synthesis of virtual acoustic environments. *Proceedings of the 1990 Symposium on Interactive 3D Graphics,* March 25–28, Snowbird, UT, 139–140.

Wilhelms, J., Moore, M., & Skinner, R. (1988). Dynamic animation: Interaction and control. *The Visual Computer, 4*(6), 283–295.

Zeltzer, D. (1991). Task level graphical simulation: Abstraction, representation and control. In N. Badler, B. Barsky, & D. Zeltzer (Eds.), *Making them move: Mechanics, control and animation of articulated figures* (pp. 3–33). San Mateo, CA: Morgan Kaufmann.

Zeltzer, D., & Johnson, M. (1991). Motor planning: Specifying and controlling the behavior of autonomous animated agents. *Journal of Visualization and Computer Animation, 2*(2), 74–80.

Zeltzer, D., Pieper, S., & Sturman, D. (1989). An integrated graphical simulation platform. *Proceedings Graphics Interface '89,* June 19–23, London, Ontario, 266–274.