Virtual Reality, Art, and Entertainment *

Joseph Bates
School of Computer Science
and College of Fine Arts
Carnegie Mellon University

June 1991

Abstract

Most existing research on virtual reality concerns issues close to the interface, primarily how to present an underlying simulated world in a convincing fashion. However, for virtual reality to achieve its promise as a rich and popular artistic form, as have the novel, cinema, and television, we believe it will be necessary to explore well beyond the interface, to those issues of content and style that have made traditional media so powerful.

We present a case for the importance of this research, then outline several topics we believe are central to the inquiry: developing computational theories for cognitive/emotional agents, presentation style, and drama.

1 The Fantasy Machine

A writer of popular science articles visits a laboratory to learn about the new "virtual reality" research. The writer is ushered into the demo room and warned to move slowly until adjusted to the system. The goggles slip on, then the dataglove, and a switch is thrown. As a world appears, the writer's head turns, and slowly the light of understanding dawns.

Three months later the article appears. "I opened my eyes and had been transported in time and space. I could look around and almost feel things with my gloved hand, as if I were living in a different world. Hollywood watch out, soon VR will let us enter *into* our fantasies. We'll be able to go anywhere and do anything imaginable."

The public is beginning to understand that virtual reality portends a new medium, new entertainment, a new and very powerful type of art. It has a

*Accepted to *Presence, The Journal of Teleoperators and Virtual Environments*, MIT Press, 1991. This is a revised version of "Deep Structure for Virtual Reality", Carnegie Mellon School of Computer Science technical report CMU-CS-91-133.

potential that beckons to the average person, as well as to the scientists developing the technology. The vision of this writer, to let us "go anywhere and do anything" is an important one. Indeed, the wish to build such a fantasy machine is the fundamental driving force for many VR researchers.

Achieving this wish requires us to examine it closely. As engineers we need to know what sorts of considerations are involved in constructing such a machine. Do the goggles and glove bring us most of the way to success? Some of the way?

Let us consider several highly developed, successful media for presenting fantasy worlds: novels, television, and film. Some people are interested in worlds solely involving physical space, objects, and natural forces. Films about these topics are usually documentaries with narration, with an occasional art film produced. However almost always, television, movies, and novels include at least three other key elements.

First, there are living creatures, usually human, and usually embodying some intelligence and emotion. These let the viewer see the world as a place of life, purpose, and feeling. Second, there is long term structure to the events portrayed, which is to say that some kind of story is told. The story gives intensity and meaning to the world. Finally, the world is presented in an effective, emotionally powerful style. Cinematic and narrative technique are highly developed examples of the art of presenting worlds.

Each of these aspects of traditional media play a great role in allowing the author to affect the viewer emotionally and intellectually. Lacking any one, the author will have special difficulty conveying a rich, significant world. Thus, when our writer says we will soon "go anywhere and do anything", these key elements of traditional media are certainly implicit in the vision.

Let us now consider the primary research directions in our very young field of virtual reality. Most of the well known work concerns the interface hardware and software. Developing the goggles and glove, and getting them to work together to present any kind of world at all, has been a battle. Much existing work, such as that at NASA and UNC (Fisher et al., 1986; Brooks, 1988), has been substantially concerned with solving this problem.

Animating the structure and behavior of physical worlds is also an active research area directly relevant to VR. Work on physically based modeling for animation, such as that reported in (Badler et al., 1990; Witkin and Welch, 1990; McKenna et al., 1990) and elsewhere, may make it much easier to portray and simulate familiar physical worlds, including the physical bodies and low level behaviors of agents.

Yet looking at what makes traditional media so powerful for viewers, and thus what may make virtual reality a genuinely powerful and popular artistic form, we are struck by the absence of research into the VR analogs of our three key areas: computational models of cognitive/emotional agents, long term dramatic structure, and presentation style. While an insightful few have suggested the importance of these areas (especially see Brenda Laurel's work (Laurel, 1986; Laurel, 1991)), to our knowledge the AAAI-90 Workshop on Interactive Fiction and Synthetic Realities (Bates, 1990b) is the only place that workers in the relevant fields have come together to seriously discuss broad research toward inhabited, dramatic virtual worlds.

In the early days of our field's efforts, it is natural that almost all work has focused on the hardware interface and system software. Indeed some virtual worlds undoubtedly will exist solely to provide an apparently real physical space for humans to inhabit. The notion of "cyberspace", presented so forcefully in the works of William Gibson, is of this character. So too is the use of virtual reality as a visualization technology, in medicine, science, architecture, and elsewhere.

If one views a VR system as producing surface level phenomena via the interface and associated software, then the organization and content of software well behind the interface constitute a "deep structure" for the virtual world. This is the arrangement of code and data that produces the essential meaning of the interactive experience. For the reasons suggested above, we believe that for VR to join the novel, cinema, and television as a broadly successful artistic medium, the technology must provide a suitably rich deep structure, particularly containing computational theories for agents, presentation, and drama. This structure is largely missing from existing VR work and is an area whose development we wish to encourage, along with work on interface and physical modeling, to help VR achieve its potential of letting us "go anywhere and do anything".

2 An Approach to the Problem

Having suggested that there are applications of virtual reality that require a certain kind of deep structure, let us examine how it might be provided. We will now be focused solely on those worlds that demand this particular approach.

A virtual world must respond appropriately to users without assistance from its original creators. This means that agent behavior, dramatic content, and presentation style must vary according to explicit artistic models built into the world by its creators. Thus, the primary task in building a world is to transfer artistic knowledge, of both general and specific kinds, into the machine.

The study of knowledge representation, processing, and transfer is one of the central topics of artificial intelligence research. We believe that much existing AI technology, for example in cognitive modeling, agent architectures, story understanding, natural language processing, and adversary search, can help provide a framework for the deep structure we seek. Note that this is primarily a problem of applying known AI results for the benefit of VR. We do not propose to gain leverage on the fundamental problems of AI by examining them in the context of VR.

Much of this existing artificial intelligence research, aimed at building intelligent systems for the real world, has been criticized for working only in small, narrowly defined micro-worlds. But virtual worlds are precisely such micro-worlds, so this is not an objection as long as the virtual worlds are rich enough to hold artistic interest. The question then is whether the best of the created worlds are of sufficient quality to warrant the term "art", and whether existing technology is rich enough to support continued artistic innovation within the medium. This question seems to us answerable only by experiment.

The Oz project at Carnegie Mellon is attempting to perform this experiment. We are a group of faculty, staff, students, and visitors from the School of Computer Science, the College of Fine Arts, and elsewhere. Our goal is to bring together appropriate existing technologies, especially AI technologies, and on this framework to help artists build dramatic simulated worlds that include simulated people. Oz thus is an attempt to provide a deep structure for virtual reality, and to explore the potential of virtual worlds built on such foundations.

3 Three Central Concerns

To help convey our view of what is needed in this research area, we will sketch briefly the approach taken in Oz to the construction of computational theories of agents, presentation, and drama. The intent here is to illuminate the areas of investigation, rather than to present specific research results.

3.1 Cognitive/Emotional Agents

One of the keys to an effective virtual world is for the user to be able to "suspend disbelief". That is, the user must be able to imagine that the world portrayed is real, without being jarred out of this belief by the world's behavior.

This requirement can be turned to our advantage if we take a non-traditional view of the problem of building intelligent agents. Instead of demanding that our agents be especially active and smart, we require only that they not be clearly stupid or unreal. An agent that keeps quiet may appear wise, while an agent that oversteps its abilities will probably destroy the suspension of disbelief. Thus, we propose to take advantage of the "Eliza effect" (Weizenbaum, 1966), in which people see subtlety, understanding, and emotion in an agent as long as the agent does not actively destroy the illusion.

In order to foster this illusion of reality, we believe our agents must have broad, though perhaps shallow, capabilities. To this end, we are attempting to produce an agent architecture that includes goals and goal directed behavior, emotional state and its effects on behavior, some natural language abilities (especially pragmatics based language generation), and some memory and inference abilities. Each of these capacities can be as shallow as is necessary to allow us to build a broadly capable, integrated agent.

This approach is unlike most AI efforts to build agents, which generally result in deep but narrow agents. Rather than attempting to drive any particular aspect of agent architecture to new highs, we are trying to integrate existing technology into a single unusually broad architecture. This is a little studied area, and we are hopeful that moderate effort in integration may yield good results, such as believable hints of thought and emotion in our limited micro-world domains.

Some work in this direction was discussed at the AAAI Spring Symposium on Integrated Intelligent Architectures (Vere and Bickmore, 1990; Bates et al., 1991a). More detail on our particular architecture, called Tok, is provided in several technical reports (Kantrowitz, 1990; Loyall and Bates, 1991; Bates et al., 1991b).

3.2 Presentation

It is clear that presentation style plays a central role in traditional media. In film, the feeling of the underlying world is greatly influenced by the lighting, camera angles, focus, editing, and so on. Music plays a key role in foreshadowing and emphasizing the mood of the visuals.

The first films were little more than visual records of traditional theater and naturally visible scenes. As time passed and film makers invented new ways of showing, the cinematic language of presentation became richer. Early audiences, accustomed to a certain kind of reality, might well not understand aspects of modern film technique.

One can imagine a case being made in the early days of cinema that such a photographically accurate medium as film could not possibly embody any subjectivity. This argument, that film technique made no conceptual sense, of course would have turned out to be wrong.

One can imagine the same kind of evolution in the art of virtual reality. While VR appears to be a first-person, "realistic", form, we believe that a language of presentation will develop over time. Someone accustomed to having normal control over their bodies and perceptions might find the partial lack of control implied by such a development confusing and perhaps unpleasant. However, we suspect that a language of presentation for VR will allow world builders to enrich the experience for participants. Worlds avoiding such style may come to appear dull and emotionally flat.

The Oz group has made some effort to understand presentation in the context of narrative. We are exploring language generation under stylistic constraints, such as the work by Hovy (Hovy, 1988), and generation architectures that integrate text planning and surface structure generation. Our initial work is described in (Smith and Bates, 1989; Kantrowitz, 1990), but this work remains very preliminary.

Language based presentation research has relevance to virtual reality, but there is clearly a need to study and automate visual style. We see the work of Witkin and Kass on space-time constraints as an excellent example (Witkin and

Kass, 1988). By adjusting simple control parameters of a physical model, they were able mechanically to reproduce certain famous aspects of Disney style animation. We hope animation researchers will continue to find interest in automating artistic styles used in the visual presentation of simulated worlds.

Film, of course, goes beyond the visual, and it seems clear that music, speech, and other sound will be similarly important to the broad success of virtual reality. A particularly relevant, and to our knowledge unstudied, research topic would be generating sound as a by-product of physically based animation.

A final variation of the theme of presentation is the possibility of interactive radio drama. This is a relatively low bandwidth, potentially high impact way of presenting a virtual world. We know of no existing work in developing computational theories of presentation for radio drama.

3.3 Drama

Some worlds are best modeled as free running simulations. The physical setting and the characters are enough to express the creator's vision. However, at least some of the time, the creator will want to impose longer term structure on the evolution of the world, and this in the presence of the free-willed and active user. It is the job of a computational theory of interactive drama (which we refer to here as the "director") to make this possible.

If the user had no influence on the world, the creator could use the writer's traditional method of carefully planning the exact sequence of story events. However, the user is not a peripheral element. Indeed usually the whole purpose of the world is to give the active user a certain kind of experience. Thus, the creator must convey to the director enough general information about stories and enough specific information about the intended kind of experience to allow the director to achieve the creator's goals. This must be done in such a way as to leave the user with an undiminished feeling of free-will.

How can we view the relationship between the director and the user? One approach is to see them in a kind of two-player game, such as chess. The director and user are taking turns, the user acting as a free agent in the world, the director looking down from above and very gently pushing the elements of the world in various ways. The director is constantly trying to maximize the chances of a pleasing overall experience, no matter what the user does along the way. If we elaborate on the chess analogy, the pushes of the director and the actions of the user are the moves of the game. The director wins if the complete history of the world is consistent with the creator's aesthetic goals, thereby (presumably) pleasing the user.

Computer chess programs based on adversary search have gradually come close to the capabilities of the world's best human players. The two-player game model for interactive drama may be amenable to the same kinds of search technology, and might yield similarly impressive results.

Formalizing the drama game for adversary search requires us to make precise the moves and the evaluation function. The first has been studied in the form of plot units by traditional narrative theorists and by some computer scientists (Lehnert, 1981; Lebowitz, 1985; Dyer, 1983). The second requires a theory of the nature of good stories, which has been discussed also to some extent (Laurel, 1986; Lebowitz, 1985; Dyer, 1983).

One of the problems that arises in this model is the large number of fine-grained actions that may occur during an interactive experience, possibly on the order of thousands. In addition, there is a wide variety of possible such actions at each moment. Together, these produce a deep search with high branching factor, a computationally difficult task.

The difficulty of search in the space of primitive actions suggests that we might search instead in a space of abstractions, such as the plot unit space. However, abstract adversary search is a technique that is not well understood. Perhaps in compensation, a possible advantage here over typical two-player games is that the user is not particularly trying to make the director win or lose, but is instead moving through the world intent on personal goals. Thus, the director is really not facing an adversary, and a shallow search may suffice.

Besides the computational complexity of the problem, a natural concern about any theory of drama is the extent to which a director can influence the world without attracting the attention of the user. To study this issue, we have performed several simulations of this model for virtual worlds, placing a user into a real physical environment with live actors and with a live director hidden away from view (Kelso and Bates, 1991).

Our preliminary results suggest that the director's influence easily can be intrusive, but that substantial influence can be subtly expressed by modifying the actions of agents. We found that people engaged in an experience easily accepted odd behavior in the actors, even when that behavior was judged as irrational or manipulative by third parties given access to the whole picture. In retrospect, this particular method for wielding power in subtle ways may be already well known in society.

The Oz group is actively studying several computational models of drama, including the adversary search model (Bates, 1990a). We are excited by the possibility of effective computer performance in this unusual domain.

4 Conclusion

Virtual reality developed out of the technical community, from a vision of what was technically possible and from the requirements of certain technically demanding applications. Of course, some of the creators had visions of applications far beyond the needs of their funding sources, but generally the community has explored virtual reality as a human-computer interface technology.

We see this focus on interface as something like studying celluloid instead of cinema, paper instead of novels, cathode ray tubes instead of television, hardware instead of software. While the interface is important, we must also look beyond it to the underlying structure of the worlds we want to model. Traditional arts and media provide excellent examples worthy of our inspection. Since VR is new, it is natural that we haven't explored the whole problem yet, but we must indeed go forth and carry out a broad exploration if we want virtual reality to achieve its promise of letting us "go anywhere and do anything".

5 Acknowledgments

We gratefully acknowledge the support of Fujitsu Laboratories, Ltd. for our research.

References

Badler, N., Webber, B., Esakov, J., and Kalita, J. (1990). *Making Them Move: Mechanics, Control, and Animation of Articulated Figures*. Morgan-Kaufmann.

Bates, J. (1990a). Computational drama in Oz. In *Working Notes of the AAAI-90 Workshop on Interactive Fiction and Synthetic Realities*, Boston, MA.

Bates, J., editor (1990b). *Working Notes of the AAAI-90 Workshop on Interactive Fiction and Synthetic Realities*, Boston, MA. American Association of Artificial Intelligence.

Bates, J., Loyall, A. B., and Reilly, W. S. (1991a). Broad agents. In *Proceedings of AAAI Spring Symposium on Integrated Intelligent Architectures*, Stanford, CA. Available as a special issues of SIGART News, Summer 1991.

Bates, J., Loyall, A. B., Reilly, W. S., and Kantrowitz, M. (1991b). Tok: A broad architecture for intelligent, emotional agents. Technical report, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.

Brooks, Jr., F. (1988). Grasping reality through illusion – interactive graphics serving science. In *CHI'88 Proceedings*, pages 1–11.

Dyer, M. (1983). *In-Depth Understanding*. The MIT Press, Cambridge, MA.

Fisher, S., McGreevy, M., Humphries, J., and Robinett, W. (1986). Virtual environment display system. In *ACM 1986 Workshop on 3D Interactive Graphics*, pages 77–87, Chapel Hill, NC.

Hovy, E. (1988). *Generating Natural Language under Pragmatic Constraints.* Lawrence Erlbaum Associates, Hillsdale, NJ.

Kantrowitz, M. (1990). Glinda: Natural language text generation in the Oz interactive fiction project. Technical Report CMU-CS-90-158, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.

Kelso, M. and Bates, J. (1991). Live simulation of a model for dramatic virtual worlds. Working Paper, Oz Project, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.

Laurel, B. (1986). *Toward the Design of a Computer-Based Interactive Fantasy System.* PhD thesis, Drama Department, Ohio State University.

Laurel, B. (1991). *Computers as Theater.* Addison-Wesley Publishing Company, Reading, MA.

Lebowitz, M. (1985). Story-telling as planning and learning. *Poetics*, 14:483–502.

Lehnert, W. (1981). Plot units and narrative summarization. *Cognitive Science*, 5:293–332.

Loyall, A. B. and Bates, J. (1991). Hap: A reactive, adaptive architecture for agents. Technical Report CMU-CS-91-147, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.

McKenna, J., Pieper, S., and Zelter, D. (1990). Control of a virtual actor: The roach. In *Proceedings of the 1990 Symposium on Interactive 3D Graphics*, Snowbird, Utah.

Smith, S. and Bates, J. (1989). Toward a theory of narrative for interactive fiction. Technical Report CMU-CS-89-121, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.

Vere, S. and Bickmore, T. (1990). A basic agent. *Computational Intelligence*, 6:41–60.

Weizenbaum, J. (1966). Eliza. *Communications of the ACM*, 9:36–45.

Witkin, A. and Kass, M. (1988). Spacetime constraints. *Computer Graphics*, 22:159–168.

Witkin, A. and Welch, W. (1990). Fast animation and control of nonrigid structures. In *Proceedings of SIGRAPH '90.*