Toward the Holodeck: Integrating Graphics, Sound, Character and Story

W. Swartout,[1] R. Hill,[1] J. Gratch,[1] W.L. Johnson,[2] C. Kyriakakis,[3] C. LaBore,[2] R. Lindheim,[1] S. Marsella,[2]
D. Miraglia,[1] B. Moore,[1] J. Morie,[1] J. Rickel,[2] M. Thiébaux,[2] L. Tuch,[1] R. Whitney[2] and J. Douglas[1]

[1]USC Institute for Creative Technologies, 13274 Fiji Way, Suite 600, Marina del Rey, CA, 90292

[2]USC Information Sciences Institute, 4676 Admiralty Way, Suite 1001, Marina del Rey, CA 90292

[3]USC Integrated Media Systems Center, 3740 McClintock Avenue, EEB 432, Los Angeles, CA, 90089

http://ict.usc.edu, http://www.isi.edu, http://imsc.usc.edu

ABSTRACT

We describe an initial prototype of a holodeck-like environment that we have created for the Mission Rehearsal Exercise Project. The goal of the project is to create an experience learning system where the participants are immersed in an environment where they can encounter the sights, sounds, and circumstances of real-world scenarios. Virtual humans act as characters and coaches in an interactive story with pedagogical goals.

1. INTRODUCTION

A young Army lieutenant drives into a Balkan village expecting to meet up with the rest of his platoon, only to find that there has been an accident (see Figure 1). A young boy lies hurt in the street, while his mother rocks back and forth, moaning, rubbing her arms in anguish, murmuring encouragement to her son in Serbo-Croatian. The lieutenant's platoon sergeant and medic are on the scene.

The lieutenant inquires, “Sergeant, what happened here?”

The sergeant, who had been bending over the mother and boy, stands up and faces the lieutenant. “They just shot out from the side street, sir. The driver couldn't see them coming.”

“How many people are hurt?”

“The boy and one of our drivers.”

“Are the injuries serious?”

Looking up, the medic answers, “The driver's got a cracked rib, but the kid's….” Glancing at the mother, the medic checks himself. “Sir, we've gotta get a medevac in here ASAP.”

The lieutenant faces a dilemma. His platoon already has an urgent mission in another part of town, where an angry crowd surrounds a weapons inspection team. If he continues into town, the boy may die. On the other hand, if he doesn't help the weapons inspection team their safety will be in jeopardy. Should he split his forces or keep them together? If he splits them, how should they be organized? If not, which crisis takes priority? The pressure of the decision grows as a crowd of local civilians begins to form around the accident site. A TV cameraman shows up and begins to film the scene.

This is the sort of dilemma that daily confronts young Army decision-makers in a growing number of peacekeeping and disaster relief missions around the world. The challenge for the Army is to prepare its leaders to make sound decisions in similar situations. Not only must leaders be experts at the Army's tactics, techniques and procedures, but they must also be familiar with the local culture, how to handle intense situations with civilians and crowds and the media, and how to make decisions in a wide range of non-standard (in military terms) situations.

In the post-cold-war era, peacekeeping and other operations similar to the one outlined above are increasingly common. A key aspect of such operations is that close interaction occurs between the military and the local population. Thus, it is necessary that soldiers understand the local culture and how people are likely to react.

Unfortunately, training options are limited. The military does stage exercises using physical mockups of villages with actors playing the part of villagers to give soldiers some experience with such operations. However, these are expensive to produce, and the fact that the actors must be trained and the sets built makes it difficult to adapt to new situations or crises. Computer-based simulators have been used by the military for years for training but these have focussed almost exclusively on vehicles: tanks, humvees, helicopters and the like. Very little exists for the soldier on foot to train him in decision making in difficult circumstances.

Inspired by Star Trek's Holodeck, the goal of the Mission Rehearsal Exercise (MRE) Project at the USC Institute for Creative Technologies (ICT) is to create a virtual reality training environment in which scenarios like the one described above can be played out. Participants are immersed in the sights and sounds of the setting and interact with virtual humans acting as characters in the scenario. At times, these characters may also act as coaches, dispensing advice to the trainee to help him achieve pedagogical goals. The underlying assumption is that people

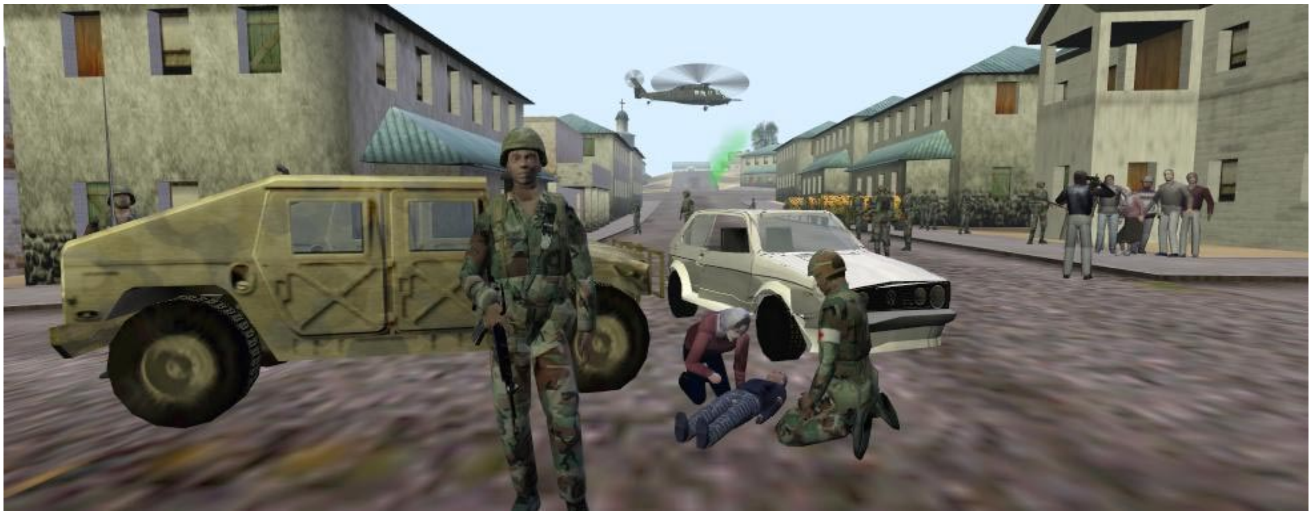

Figure 1: A screen snapshot of the MRE simulation

learn through experiencing a situation and making decisions in the context of a stressful and sometimes confusing environment.

At the ICT, our approach to creating such a simulation is to bring together two communities that have a lot to offer each other: the computer science/technical community that knows how to create the underpinnings for simulation (such as high resolution graphics, immersive sound and AI reasoning) and the entertainment community that knows how to create characters and story lines that people will find compelling. Although the ICT has only been in existence for about a year, we are already starting to see, in projects such as MRE, the synergies that can result from this collaboration.

In line with these ideas, the key features of the MRE system include:

- Interactive stories that engage the learners while achieving pedagogical goals. These stories present the participant with a series of dilemmas. The outcome of the stories depends on the participants' actions. By experiencing these dilemmas in a simulation, participants will be better prepared to make correct decisions when they encounter similar situations in real life.
- Virtual humans play the role of the local populace and friendly (or hostile) elements. We have used a hybrid approach to creating these characters: some are scripted while others use automated reasoning and models of emotion to determine their behavior dynamically. The characters use expressive faces and gestures to make their interactions more believable.
- Ultra-wide-screen graphics and immersive 10.2 audio (10 channels of audio with 2 subwoofer channels) place the participant in a compelling environment.

To create the initial prototype of the MRE system, we integrated a number of commercial components with research on intelligent agents. This paper describes that effort and our initial approach to create a compelling story to test the system.

2. A HYBRID APPROACH

In Hollywood, it is the norm to take a hybrid approach to creating a film. The goal is to create a seamless, compelling presentation, but that doesn't mean that everything has to be created using a single approach. Recognizing that each technique has its own strengths and weaknesses, Hollywood artists select the most appropriate technique for each element of an overall scene and then composite the results together to create a unified whole.

For example, in the movie *Titanic*, the command "All ahead full!" is given as the ship heads out into the open ocean. What follows is a series of shots in the engine room as the ship comes to life under a full head of steam. While it looks as if the shots were taken in the engine room of a very large ship, in fact the scene was created by using live action shots of actors, together with shots of machinery on a ship much smaller than the Titanic that were enlarged to the correct size, and the background of the huge engine room was provided by shots of a model only a few feet high. These were all composited together to create a convincing view of the engine room on the doomed ship.

But the integration must be done skillfully. If the audience notices the "seams" or finds something anomalous in the way things are depicted, the effect will be ruined and the audience will no longer be immersed.

In constructing the MRE system, we found that it was best to take a hybrid approach because the requirements for the various elements of the system varied widely, and no one approach was best for everything.

For example, many of the virtual humans in our scenario had what were essentially "bit" parts with a fairly limited set of movements and behaviors. A soldier moving out to establish security would be one example, a person in a crowd of onlookers would be another. Controlling these characters with a full AI-based reasoner seemed like overkill. The additional capabilities provided by such a controller would not be utilized and the extra runtime and development expense of a sophisticated controller would be wasted.

Accordingly, we decided to adopt a hybrid approach to the control of our virtual humans:

- *Scripted.* Characters with a limited range of behaviors were scripted in advance. Their behaviors were triggered by events in the simulation, such as a command given by one of the characters, or an event in the environment.
- *AI-based.* Characters with a broad range of behaviors, and those that interact directly with the trainee and thus need to be able to deal with unanticipated situations and interactions used an AI reasoner [4,5,6] to control their behavior.
- *AI-based with emotion model.* The most sophisticated character in our simulation, the mother, used an AI reasoner coupled with a model of emotion [1,3] to determine behavior.

Another area where a hybrid approach was required was speech. To create a realistic training scenario, the virtual humans need to be able to interact with the trainee in natural language. Two approaches to generating speech seemed possible:

- One approach, which is frequently used in computer games, is to have an actor pre-record a number of possible utterances. While the simulation is running, the computer selects the most appropriate pre-recorded response and plays it back.
- The second approach is to use a text-to-speech (TTS) speech synthesis system and dynamically create the speech output that is needed.

Both approaches have strengths and weaknesses. Using a text-to-speech system allows for great flexibility: one doesn't have to anticipate in advance all the possible speech fragments that might be required. But a significant weakness of this approach is that even the best TTS system lacks the emotional range that is possible with the human voice, and most current systems sound very artificial. For a part like the mother of the injured boy in our scenario we felt that a TTS system would lack the emotion needed to be convincing, and we were concerned that an unnatural sounding voice would destroy believability. On the other hand, while pre-recorded voice sounds just as good as the actor recorded, it is clearly limited by the range of responses recorded in advance.

To deal with these problems, we realized that neither approach to speech would work best in all cases. After considering the various parts involved, we decided to use pre-recorded voices for the minor parts, and those like the mother's that require significant emotional range. We decided that the sergeant should use a TTS system because he serves as the main "interface" between the simulation and the trainee, and as we evolve the system he will need to deal with the greatest range of inputs (some of which may be unanticipated). Thus, TTS is appropriate for his part because we need great flexibility in *what* he can say, but because his role is not an emotional one there is less concern about *how* he says it.

For speech one concern remained. We argued that integration must be done skillfully if a hybrid approach is used—if the seams are apparent the effect won't work. We were concerned that if we used TTS for the sergeant, it would be very jarring to have an artificial sounding voice mixed in with natural voices. After an extensive search of research in TTS systems, we selected the AT&T Next-Gen TTS system[1] because it produces a very natural sounding voice that integrated well with the pre-recorded voices.

In addition to allowing us to make the best match between task requirements and technology capabilities, taking a hybrid approach allowed us to create a complete MRE scenario so that we could assess the impact of the system as a whole without having to solve all the sub-problems in the most general way. This also allows us to identify the areas where further research will have the most impact. Over time we expect to replace our simple solutions with more sophisticated ones as technology progresses. The hybrid approach allows us to do this in an incremental fashion.

3. ARCHITECTURE

Now that we have asserted the utility of using a hybrid approach for developing the MRE prototype, we will describe the architecture in more detail. Four aspects of the system are considered here—visual modeling, audio modeling, the modeling of characters, and the interface between the participants and the virtual characters.

3.1 Visual Modeling

At its foundation, the MRE prototype is built on a visual simulation of the characters and environment in a story-based scenario. The system is hosted in the theater setting shown in Figure 2. The participant (e.g., a lieutenant) stands before a large curved screen—8.75 feet tall and 31.3 feet unwrapped—where images are rendered by three separate projectors and blended together to form a 150 degree field of view. An SGI™ Onyx Reality Monster™ with sixteen processors and four graphics pipes provides the computational resources to drive the system. It takes three graphics pipes to drive the projectors, and the fourth pipe is used for the control console.

A 3D model of a Balkan village was developed to fit the types of scenarios we had in mind. Texture mapped surfaces were applied to the buildings, vehicles, and characters to give them a more authentic look and feel. The Real-time Graphical Animation System shown in Figure 4 performs the real-time rendering of the visual scenes and characters. We use an integration of two commercial products, Vega™ and PeopleShop™, to provide this capability to the MRE system. Vega™ renders the environment and the special effects. The environment includes the buildings, roads, trees, vehicles, and so on, while the special effects include explosions and the dynamic motion of objects like cars and helicopters. The PeopleShop™ Embedded Runtime System (PSERT) is integrated with Vega™ and provides the animation of the characters' bodies.

To make the experience more compelling, we felt that it was important to give some of the characters expressive faces. Since the PeopleShop™ characters did not have this ability, we found a product made by Haptek™ called VirtualFriend™, which has realistic looking faces that can change expression and synchronize the character's lips with speech. We contracted with

[1] http://www.research.att.com/projects/tts/

moved. The fact that our early prototype had such a strong effect on people surprised us greatly.

We were well aware that there were imperfections and technical limitations in all of the elements we used: story, graphics, speech, immersive sound, AI reasoning and emotional modeling. But when all of them were brought together, the audience was willing to suspend disbelief, overlook the imperfections, and the overall effect was greater than the sum of its parts.

7. ACKNOWLEDGEMENTS

The authors wish to thank the other members of the MRE team for their contributions. Marc Raibert, Adam Crane and Whitney Crane of Boston Dynamics developed the animated bodies and and Marlon Veal of Haptek developed the expressive faces for the virtual humans. Michael Murguia of New Pencil and Erika Sass developed the 3D graphic models. Brigadier General Pat O'Neal (Ret.) and Elke Hutto provided military knowledge to structure the scenario. Finally we wish to thank Michael Andrews, Deputy Assistant Secretary of the US Army for Research and Technology, James Blake, Program Manager at STRICOM, and Michael Macedonia, Chief Scientist at STRICOM, for the vision that started this program. The work described in this paper was supported in part by the U.S. Army Research Office under contract #DAAD19-99-C-0046. The content of this article does not necessarily reflect the position or the policy of the US Government.

8. REFERENCES

[1] Gratch, J. Émile: Marshalling Passions in Training and Education. Proceedings of the Fourth International Conference on Autonomous Agents, ACM Press, 2000, 325-332.

[2] Kyriakakis, C. Fundamental and Technological Limitations of Immersive Audio Systems. Proceedings of the IEEE, 86(5), 941-951, 1998.

[3] Marsella, S.C., Johnson, W.L. and LaBore, C. Interactive Pedagogical Drama. Proceedings of the Fourth International Conference on Autonomous Agents, ACM Press, 2000, 301-308.

[4] Rickel, J. and Johnson, W.L. Animated Agents for Procedural Training in Virtual Reality: Perception, Cognition, and Motor Control. Applied Artificial Intelligence, 13:343-382, 1999.

[5] Rickel, J. and Johnson, W.L. Virtual Humans for Team Training in Virtual Reality. Proceedings of the Ninth International Conference on Artificial Intelligence in Education, 578-585, 1999, IOS Press.

[6] Rickel, J. and Johnson, W.L. Task-Oriented Collaboration with Embodied Agents in Virtual Worlds. In Embodied Conversational Agents, edited by J. Cassell, J. Sullivan, S. Prevost and E. Churchill. MIT Press: Boston, 2000.

[7] Johnson, W.L., Rickel, J. and Lester, J. Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments. International Journal of Artificial Intelligence in Education, 11:47-78, 2000.

[8] Newell, A. Unified Theories of Cognition. Cambridge, MA: Harvard University Press, 1990.

[9] Cassell, J. and. Thorisson, K. The Power of a Nod and a Glance: Envelope vs. Emotional Feedback in Animated Conversational Agents. Applied Artificial Intelligence, 13:519-538, 1999.

[10] Cassell, J., Bickmore, T., Campbell, L., Chang, K., Vilhjalmsson, H. and Yan, H. Requirements for an Architecture for Embodied Conversational Characters. Proceedings of Computer Animation and Simulation. Springer-Verlag: Berlin, 1999, 109-120.

[11] Cassell, J., Sullivan, J., Prevost, S., and Churchill, E., editors. Embodied Conversational Agents, MIT Press: Boston, 2000.

[12] Bindiganavale, R., Schuler, W., Allbeck, J.M., Badler, N.I., Joshi, A.K., and Palmer, M. Dynamically Altering Agent Behaviors Using Natural Language Instructions. Proceedings of the Fourth International Conference on Autonomous Agents, ACM Press, 2000, 293-300.

[13] Kelso, M.T., Weyhrauch, P. and Bates, J. Dramatic Presence. In Presence: Journal of Teleoperators and Virtual Environments 2(1), 1993, MIT Press.

[14] Weyhrauch, P. Guiding Interactive Drama. Ph.D. Thesis. Tech Report CMU-CS-97-109. Carnegie Mellon University.

[15] Sgouros, N.M. Dynamic Generation, Management and Resolution of Interactive Plots. Artificial Intelligence, 107:29-62, 1999.

[16] Galyean, T. Narrative Guidance of Interactivity. Ph.D. Thesis, Media Arts and Sciences, MIT, June 1995.

[17] Mateas, M. and Stern, A. Towards Integrating Plot and Character for Interactive drama. In Socially Intelligent Agents: The Human in the Loop. Papers from the 2000 AAAI Fall Symposium, Technical Report FS-00-04, AAAI Press, 113-118.