Simulation Meets Hollywood: Integrating Graphics, Sound, Story and Character for Immersive Training

William R. Swartout
USC/Institute for Creative Technologies
swartout@ict.usc.edu

Extended Abstract

keywords: Natural Language Generation, Conversational Spoken Dialogue, Multimodal Communication, Life-like Characters

How can training simulations be made more effective? An important insight in answering that question is to recognize that effective training depends both on the technology that is used to present the material and the content of the material itself. The Institute for Creative Technologies (ICT) was created at the University of Southern California with the goal of bringing together researchers in simulation technology to collaborate with people from the entertainment industry. The idea was that if those who understood how to create high resolution graphics, immersive sound, and believable virtual humans worked together with those who understood how to create compelling stories and characters a synergy would emerge that would allow them to create much more compelling simulation experiences.

Although the ICT has only been in existence for a short time, we are already beginning to see some of the results of this collaboration. These are reflected both in the kinds of projects that the ICT takes on and the approach that we take to implementing systems. While most military simulations involve simulating a vehicle such as a tank, an airplane or a helicopter, ICT's simulations put trainees into a human-oriented simulation, where they interact with real and virtual (computer-generated) humans. While scenarios in most military simulations tend to proceed in a straightforward fashion, our scenarios engage the trainee with plot twists, turns and surprises, much like one might find in a good Hollywood script. In constructing our simulations we have used a hybrid approach, mixing different techniques and technologies to produce the best overall effect. In that way, we are following Hollywood film production techniques where what appears as a single seamless scene in film may actually be the result of integrating a large number of disparate elements produced using filmed live action, computer generated imagery, and models.

One of the ICT's projects that illustrates these ideas well is the Mission Rehearsal Exercise (MRE) project. Since the end of the cold war, the kinds of operations that the US military is involved with has expanded greatly. The need for peacekeeping and nation-building operations has grown, and humanitarian efforts such as disaster relief are common. One of the hallmarks of these operations is that they frequently involve close interactions between the military and the local civilian populace. To function effectively and avoid misunderstandings that could have unintended consequences, it is important that soldiers understand the customs, norms, habits and taboos of the local population and they need to be exposed to the thorny dilemmas and decisions that may await them.

Figure 1: Mission Rehearsal Exercise System

The Mission Rehearsal Exercise system, shown in Figure 1, is designed to provide that kind of experience in simulation, before trainees encounter it in reality. Presented on a 30 foot by 8 foot curved screen, the MRE system places the trainee in a location. The trainee interacts with life-sized virtual humans that can play the role of local civilians, friendly forces and hostile forces. A 10.2 sound system (10 channels of audio, 2 subwoofer channels) enhances the immersive effect.

The scenario we are currently using is situated in a small town in Bosnia. It opens with a lieutenant (the trainee) in his humvee. Over the radio, he gets orders to proceed to a rendezvous point to meet up with his soldiers to plan a mission to assist in quelling a civil disturbance. When he arrives at the rendezvous point, he discovers a surprise. One of his platoon's humvees has been involved in an accident with a civilian car. There's a small boy on the ground with serious injuries, a frantic mother, and a crowd is starting to form. A TV camera crew shows up and starts taping. What should the lieutenant do? Should he stop and render aid? Or should he continue on with his mission? Depending on decisions he makes, different outcomes will occur.

The initial version of the Mission Rehearsal Exercise system was first shown in September, 2000. That initial version showed the vision for the project, but it was extensively scripted and supported only limited interactivity. Since then, the MRE project has been actively engaged in research to improve the MRE system and make it more interactive. Below I outline some of the issues we have confronted. These will be discussed in the talk.

Natural Language

To keep the simulation as realistic as possible, we feel that it is critical that participants interact with the virtual characters in the simulation using natural language. We need to deal with the whole range of natural language processing, ranging from speech recognition, to natural language understanding, to generation, to speech synthesis. Specifically, some of the issues we have confronted include:

- **Speech recognition.** Today, speech recognition works reasonably well using

commercial products such as IBM's Via Voice, assuming that one uses a good quality microphone in a quiet environment. But we are intentionally making the MRE environment noisy by introducing sound effects such as helicopter fly overs, troops running by, and so forth. To be able to operate in that environment, we have trained speech recognizers to operate in noisy environments.

- **Natural language understanding/Dialogue management.** A simulated environment presents some unique challenges and opportunities for NLU and dialogue management. NLU must be able to handle errors that may be introduced by the speech recognition component, and engage in clarification dialogues when an utterance is partially understood. However, the fact that the dialogue occurs in the context of an overarching story structure may actually facilitate understanding. The story structure provides a very strong context that limits the range of things that a trainee may reasonably be expected to do. Given that, it is possible to approach NLU as both a bottom-up process (as is normally done) *and* as a topdown process, so that NLU becomes more of a diagnosis process: given an utterance, figure out which one of a limited number of alternatives it could be. This approach can also provide help for clarification dialogues.
- **Natural language generation.** The fact that there are often multiple participants involved in or overhearing a dialogue can affect processing. For example, some responses must be carefully crafted to avoid alarming some participants or giving them too much information, while still getting the point across to others.
- **Speech synthesis.** Most text-to-speech systems do not sound natural. To address this, we have taken a hybrid approach. For virtual humans with small parts in the scenario, we have used actors to pre-record a library of lines they may say. This allows them to sound very natural, and the limited range of utterances is not a problem given their small roles. For characters with larger roles, we have used text-to-speech systems based on a concatenative approach like AT&T's Next Gen http://www.research.att.com/projects/tts/ or we have created our own voices using the Festival framework.
- **Gestures.** To make our animated characters appear lifelike, it is critical that we integrate gestures with speech. This integration has a fundamental effect on the whole communication architecture. The MIT Media Lab's BEAT has been a useful tool in this integration.

In this talk, I will discuss these issues and the approaches we have taken to resolving them.

References

Swartout, W., Hill, R., Gratch, J., Johnson, W.L., Kyriakakis, C., Labore, K., Lindheim, R., Marsella, S., Miraglia, D., Moore, B., Morie, J., Rickel, J., Thiebaux, M., Tuch, L., Whitney, R. Toward the Holodeck: Integrating Graphics, Sound, Character and Story, in Proceedings of 5th International Conference on Autonomous Agents, 2001

J. Rickel, J. Gratch, R. Hill, S. Marsella, and W. Swartout. Steve Goes to Bosnia: Towards a New Generation of Virtual Humans for Interactive Experiences. In AAAI Spring Symposium on Artificial Intelligence and Interactive Entertainment, Stanford University, CA, March 2001

Lindheim, R. and Swartout, W. Forging a new simulation technology at the ICT in IEEE Computer, Jan 2001