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

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2006		**00-00-2006 to 00-00-2006**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Toward the Holodeck: Integrating Graphics, Sound, Character and Story	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of California,Institute for Creative Technologies,13274 Fiji Way,Marina del Rey,CA,90292**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**8**	

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/

Figure 2: ICT Virtual Reality Theater

Boston Dynamics, Inc. (BDI), the producer of PeopleShop™ to put the Haptek™ heads onto the bodies of the PeopleShop™ characters. In addition, we wanted more variety when it came to the character bodies, so we contracted to have BDI create a suite of new character bodies and behaviors. The new character bodies included a Balkan woman (to play the mother), a child, a man holding a TV news camera, civilian characters for the crowd, an Army medic, and an Army sergeant.

The models of the scenes, character heads, and character bodies are stored as files that can be specified and loaded at run-time. This gives us a great deal of flexibility when it comes to creating new scenarios in that we can load any combination of models to fit the situation being created. Currently we only have one 3D town model available to us, but we'll eventually have many different ones. We have several different expressive heads in our repository, and these can be matched with different bodies. Or we can chose to use the standard head that comes with the different characters, which is currently what we do for the lesser roles in the scenario.

Figure 3: Adjusting Audio

3.2 Audio Modeling

In our current scenario the scene begins with the lieutenant driving up to the village in an Army truck known as a Humvee. As the vehicle drives into town and turns a corner, our view out the front windshield and side windows allows us to see the road, buildings, and trees. We perceive the bumps in the road as a jiggle in the scene, and the vehicle appears to change velocity as the gears are shifted. While the visual aspects of the scene give the viewer a sense of being there in that village, the audio system provides a critical dimension to the experience. The distinctive roar of the Humvee's diesel engine, the creaks, rattles, and bumps from the bouncy ride, and the post-ignition knock when the engine shuts off are all synchronized with visual effects. When the lieutenant steps out of the Humvee, one can immediately hear the murmur of a crowd of people speaking in Serbo-Croatian, gathered near the accident site. When the medevac helicopter flies overhead the room literally vibrates with the sound of a Blackhawk helicopter.

To address the problem of matching picture with sound spatially for a large group, a novel multi-channel audio system was developed [2]. It uses three main channels across the front, a pair of wide speakers that recreate reflections from the sides, a pair of height speakers that simulate overhead sound, and three rear surround channels for seamless ambient sound. In addition, two low frequency effects channels are used. The sound signals sent to each channel are processed through a set of psychoacoustically derived filters that simulate directional characteristics of human hearing including the effects of reflections from the pinnae.

The audio effects were pre-rendered and recorded for the simulation and stored as multiple, parallel tracks—up to 64 tracks are used for some effects. The mixing and spatialization of the sounds was all done in advance—the development of techniques to dynamically render spatialized sound remains as an active research topic. The audio effects were designed to match the events in the story and special care is taken to insure that the sound matches the picture. For instance, it is important that when the helicopter enters the picture that it sound like it is coming from the direction where it is located on the screen. But this also means that the timing and location of the sound leading up to the helicopter's appearance must be designed.

At run-time, the audio tracks are triggered by a human scenario controller or by an AI agent (see Figure 4) at the appropriate time in the simulation. A continuous time code is sent to the audio system to synchronize the audio tracks to the simulation.

When the characters speak, a message/action is sent from the character to the audio system to play a particular voice clip. As

mentioned earlier, a voice clip can either be a pre-recorded human voice or a recording of a synthesized voice generated by a text-to-speech system. In the cases where the character has one of the Haptek™ heads with an expressive face, a phoneme file is also triggered that synchronizes the lips of the character to the voice clip.

3.3 Characters

As mentioned previously, there are three classes of agents playing the character roles in the MRE system: scripted, AI, and AI with emotions. The agents with scriptable behaviors come packaged with PeopleShop™. They can be scripted to perform specific actions, such as running along a pre-specified path, and this behavior can be triggered by the scenario controller or by an AI agent. The scripted characters do not perceive anything in the world, nor is their behavior interruptible—their behaviors are generated by playing motion capture sequences.

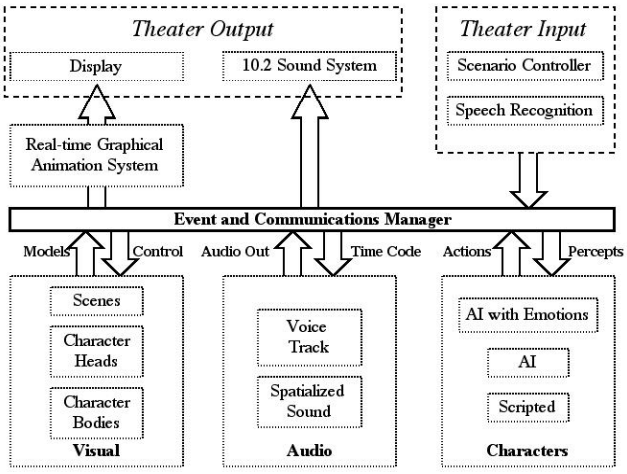

Figure 4: MRE Architecture

The AI agents, which include the sergeant, medic, and mother in the scenario, are implemented in Soar, a general architecture for building intelligent agents [8]. They are built on top of the Steve pedagogical agent [4,5,6]. As such, their behavior is not scripted; rather, it is driven by a set of general, domain-independent capabilities operating over a declarative representation of domain tasks. Task knowledge is given to the AI agents using a relatively standard hierarchical plan representation. Each task consists of a set of steps (each a primitive action or another task), a set of ordering constraints on those steps, and a set of causal links. The causal links describe the role of each step in the task; each one specifies that one step achieves a particular goal that is a precondition for a second step (or for termination of the task). The AI agents' general capabilities use such knowledge to construct a plan for completing a task from any given state of the world, revise the plan when the world changes unexpectedly, and maintain a collaborative dialogue with the human participant and teammates about the task.

For the mother character in the scenario we used an AI agent with emotions. Our model uses Gratch's [1] approach to generate emotions based on an appraisal of the agent's plans, and it uses the approach of Marsella et al. [3] to express those emotions. These emotions in turn affect the agent's subsequent decisions.

The AI characters perceive events in the simulation, reason about the tasks they are performing, and they control the bodies and faces of the PeopleShop™ characters to which they have been assigned. They send and receive messages to one another, to the character bodies, and to the audio system via the Event and Communications Manager shown in Figure 4.

3.4 Human Interface to System

The student speaks to the virtual humans on the screen via a speech recognition system, and the characters respond by shifting their gaze toward the center of the room and uttering the most appropriate statement in return. The characters at this point only recognize key words and phrases and do not have a deep understanding of the human's utterance[2]. We have plans to integrate research on natural language understanding with this work in the coming year. We will also integrate head tracking in the coming year.

4. CREATING CONTENT: STORY-BASED APPROACH

For years, Hollywood has known that an effective story is critical for creating a compelling experience that will grip the audience and cause them to suspend disbelief. But what makes a story effective? It's more than just a simple list of events. A successful story structures the interrelationships between characters and events so that emotions build and ebb throughout the story. It's hard to achieve a peak abruptly—a peak event usually only works (and is believable) if the events leading up to it have built toward it.

Consider for a moment the Disney animated film *Bambi*. Watching the film, almost everyone in the audience is deeply saddened when Bambi's mother is shot by a hunter. Yet why should we care? After all, in the real world many deer are shot during hunting season, and it's not a cause for great grief. Furthermore, the film doesn't even show a real deer being shot; it's just a cartoon character. The reason we care, and the reason the film works, is because of all the events that come earlier in the film, where we see the relationship between Bambi and his mother depicted in an anthropomorphic fashion.

So what does story have to do with interactive animated agents? Quite a lot, if we want people to be engaged by the agents we create. Even though an agent might behave in a very realistic fashion, it won't be very interesting unless there is a compelling underlying story, just as a videotape of an average person's average day would be realistic but boring.

But how can we structure stories for interactive animated agents? Several problems present themselves. Traditional stories are structured in a linear fashion — one event follows another in a fixed sequence. This allows an author to structure events so that they build toward a climactic event. At the same time we want

[2] Currently the speech recognition system works in the room near where the computer is located, but it has not been connected yet to the virtual reality theater. This is because the necessary serial connections do not yet exist in the theater.

the human participant in the simulation to feel that he has free will—he can do whatever he wants. But if he has complete free will, how can we create tension or emotional builds if he fails to pick the "correct" sequence. If there is too much flexibility, events may not build toward a climax and the story may not be compelling. For training systems like MRE, an additional complication is that if the trainee can do anything he wants in the environment, he may fail to experience any of the lessons or dilemmas that will instruct him.

Additionally, much of the current work on agents that operate in simulated environments is successful because these agents operate in rather limited and focussed task domains, such as starting an air compressor or finding a leaky pipe. However, a mission rehearsal exercise is potentially very broad. How can we hope to cover the range of possible situations that might arise?

In our initial version of the MRE system, there are only a few options, so the trainee cannot get far off track. However, because we anticipate expanding the range of options greatly in the next few months, we have been exploring a hybrid approach to story creation, called StoryNet (see Figure 5), that accommodates unstructured interactive "freeplay" with agents as well as structured sequences of events that can be used to create vignettes that engage participants. The key to our approach is to structure the story into *nodes*—confined freeplay areas in which the human participant can exercise initiative and *links*—linear sequences of events that connect the nodes and over which the participant has little control.

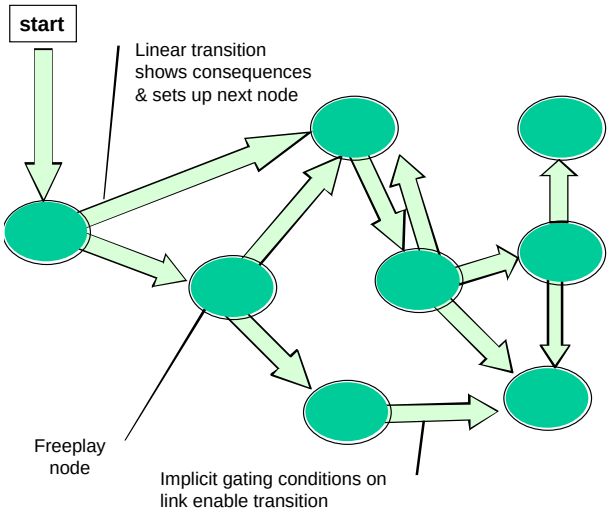

Figure 5: StoryNet

In this approach, the nodes in the network correspond to either well-defined tasks that the participant is intended to learn (for example, in the mission rehearsal scenario, tasks such as securing the accident site or arranging for a medevac) or decision points (such as whether or not to attend to the injured boy). Within a node, the participant can take control and initiate actions and make choices. The limited nature of the task limits what can be done by the participant and limits the scope of behaviors that the system must be able to handle.

The links between nodes are sequences of events in which the participant is passive. These sequences are used to set up the next situation (much like the opening "movie" in many video games) or show the consequences of successful (or unsuccessful) completion of the previous task. Loops are possible in this structure and could be used in situations where a task must be done repeatedly until it is done correctly.

Traversing a link may cause a sequence of events to occur in the simulation. For example, a single link traversal might involve the events of leaving a room, walking down the street and meeting a stranger. Such a link could be used to set up the task (represented by a node) of "dealing with strangers." Depending on how well the participant did, other links would lead out of that node and show him the consequences of successful or unsuccessful completion of the task.

Each link has a set of conditions that must be met before the link may be traversed. These conditions are implicit in the sense that the participant is not aware of what they are. Traditional branching story structures make choices and branch points obvious to the reader (e.g. "Turn to page 23 if you think Bobby finds his dog…"). We feel this destroys some of the immersion of the story, hence the choice points and the entire underlying structure of the story should be hidden from the human participant.

Figure 6 shows how a portion of our initial MRE scenario could be structured using the StoryNet approach. Events appear as boxes on the links, while choice points and tasks appear in the ovals. Gating conditions are shown on the links. If there is only one way to exit a node the gating condition is that any tasks to be performed in the node must be complete.

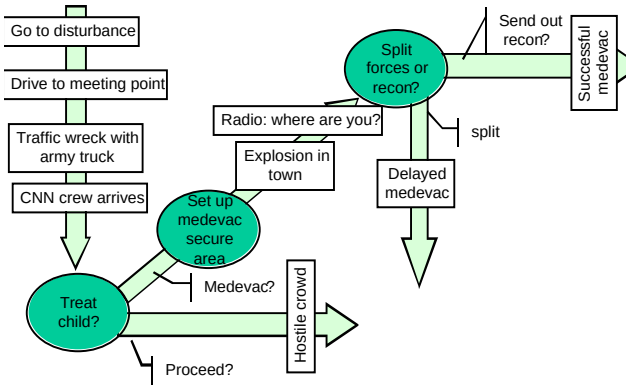

Figure 6: Partial MRE Scenario in StoryNet

In the future, we will be adding a Director agent to the MRE simulator. The Director is an agent that has overall control of the simulation and insures that the human participant has the experiences necessary to achieve the pedagogical goals we have for him. The conditions on some of the links could be under the control of the Director. This would provide a way for moving a participant out of a situation if it seemed that the pedagogical goals were not being met or things were proceeding too slowly. For example, if a student was struggling with a task and making little progress, it might be best to move on to something else and mark this situation as something to address during after-action

review. This mechanism would allow the Director to take control and minimize unproductive time in the simulation.

The key advantages for the StoryNet approach are:

- StoryNet allows us to mix linear and non-linear story elements together in a simulation experience. If this is done skillfully, the participant should not be aware of the transitions.
- StoryNet gives us a way of decomposing very large tasks (like a peacekeeping operation) where complete freeplay would be impractical (due to the large number of possible interactions) into manageable subtasks where we can allow freeplay but still have boundaries that delimit the range of possibilities.
- Finally, StoryNet provides a way for us to embed control points for a Director agent into the story structure. This allows the Director to move the story along when things are stuck at an impasse, but by embedding control in the story structure, it also means that the Director will not have unlimited control over the environment — using links that are pre-embedded in the story structure helps assure that the Director's control remains consistent with the overall structure of the story.

5. RELATED WORK

Our work builds on prior research on animated pedagogical agents [7]. However, our emphasis on more realistic sound and graphics, and a more engaging story line, makes our system more immersive than prior systems. Our work also builds on prior research on embodied conversational agents [11]. The work of Cassell and her colleagues [9,10] is particularly relevant, as they have built several virtual humans that can carry on multimodal dialogues. However, their virtual humans do not collaborate with users in a scenario with a story line; they focus on conversation for the purpose of information exchange. Bindiganavale et al. [12] have also done work on training systems with virtual humans for the military. One of the differences from that work is that the human user can only give natural language instructions to the agents; he can't participate in the scenario directly.

There is also related work on modeling and controlling interactive stories [3,13,14,15,16,17]. For example, Kelso et al. [13] propose a Plot Graph model whereby the major moments or scenes of the story are linked together by a "must precede" relation to form a partial order. A scene can only happen if all preceding scenes occur. Weyhrauch [14] structures the space of all possible stories in terms of possible permutations of plot points or important moments and a preference function on which orderings of those points it prefers. When an important plot transition is recognized then the manager does a forward projection of all possible extensions of the current story, chooses one that it prefers and proceeds to try to manipulate the world so that story occurs. Marsella et al's IPD system [3] uses an abstract model of ideal paths through the story. When significant deviations are detected, the characters are cajoled to get back onto an ideal path. Sgouros [15] uses an approach to controlling interactive plots based on an Aristotelian model of the dramatic significance of character behavior and interaction. Based on this conceptualization, the plot manager controlled the story at a fine grain by controlling what all the non-player-controlled characters do and delimiting what the single user-controlled player can do. All these approaches provide a very useful source of ideas for developing MRE. However, in contrast to these other efforts, we are attempting to create rich immersive environments with a high degree of relatively unhindered multimodal interaction between the immersed user and a wide range of autonomous characters. In addition, we need to factor in both pedagogical and dramatic goals. It is these features which have lead us to explore an approach to story management that employs extensive "free play zones" with well-structured, potentially looping, transitions between them.

6. RESULTS AND LESSONS LEARNED

In developing the MRE system, we had a number of surprises along the way, some good, some bad. Our first surprise was simply the magnitude of the integration effort. Since we were integrating a number of state-of-the-art systems that had not been integrated before we expected the usual headaches one encounters when trying to integrate systems that run in different software environments and on different hardware platforms. We anticipated the need to put together a multi-faceted technical team to handle these problems. But those of us with technical backgrounds were surprised to discover that it was necessary to put together a team of similar size to deal with content development, and that similar integration issues awaited us on the content side.

To develop the content for the MRE system, we needed a writer to develop the scenario, an art director to specify the overall look for the environment, actors to serve as models for the virtual humans and to record the pre-recorded speech, and artists to create the animated characters, the buildings and the environment for the simulation. In many cases there were dependencies between the development of content and the progression of the technology. For example, the emotion modeling software could not be completely tested until we had expressive motions for the virtual humans. Similarly, the technology imposed constraints on what could be in the scenario. For example, at one point we wanted a group of soldiers to get into a humvee and drive off. To show the animated characters actually entering the humvee turned out to be technically difficult, so the scenario was revised so that the action occurred out of view.

We used a fairly casual approach to software development issues such as version control. Often such an approach works well for research projects. However, due to the scope and size of the integration issues we encountered, our casual approach was not sufficient. On several occasions out-of-sync versions led to apparent errors that we spent quite a bit of time tracking down. The lesson we learned was that we need to take a much more systematic approach to both software and content development in the future.

But the biggest surprise (and the most pleasant) was the impact that the integrated system had on the audience. We have shown the system to several hundred people, including Army personnel and the general public. Army officers leaving the theater have remarked, “I've been there. This is almost too real.” People who had been in similar situations in real life were emotionally

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.