Virtual Humans in the Mission Rehearsal Exercise System

Randall W. Hill, Jr., Jonathan Gratch, Stacy Marsella, Jeff Rickel, William Swartout, and David Traum.

1 Introduction

How can simulation be made more compelling and effective as a tool for learning? This is the question that the Institute for Creative Technologies (ICT) set out to answer when it was formed at the University of Southern California in 1999, to serve as a nexus between the simulation and entertainment communities. The ultimate goal of the ICT is to create the Experience Learning System (ELS), which will advance the state of the art in virtual reality immersion through use of high-resolution graphics, immersive audio, virtual humans and story-based scenarios. Once fully realized, ELS will make it possible for participants to enter places in time and space where they can interact with believable characters capable of conversation and action, and where they can observe and participate in events that are accessible only through simulation.

Figure 1: The Mission Rehearsal Exercise System

2. Mission Rehearsal Exercise

The ICT project that most fully embodies the vision of the Experience Learning System is the Mission Rehearsal Exercise (MRE), which is being designed to teach critical decision-making skills to small-unit leaders in the U.S. Army. The situation depicted in Figure 1 illustrates one of the uses of such a training system. A young Army lieutenant has just arrived in a Balkan town, where he was to meet the rest of his platoon before heading for another site on a mission. When he meets the platoon, he finds that one of his soldiers has been involved in an accident with a civilian vehicle. 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. As the lieutenant drives up, the sergeant, who had been bending over the mother and boy, stands up and faces him.

The lieutenant inquires, "Sergeant, what happened here?"

The sergeant replies "There was an accident sir. This woman and her son came from the side street and our driver couldn't see them"

Lt: "Who is hurt?"

Sgt: "The boy and our driver"

LT: "How bad is he hurt?"

Sgt: "The boy or the driver?"

Lt: "The boy"

The sergeant turns to the medic, Tucci, and says "Tucci, how is the boy?"

Looking up, the medic answers, "The boy has critical injuries."

Sgt: "Understood"

Glancing at the lieutenant, the medic continues, “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. 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 in the Army's tactics, 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. Until recently, it has not been possible to simulate human behavior with the degree of fidelity required to support the kind of face-to-face dialogue and interaction that are needed for leadership training. Instead, actors are often hired to play the roles of civilians for non-battlefield scenarios, and role-playing is used to learn about how to deal with interpersonal issues. With the advent of technologies for virtual humans (Gratch, et al, 2002), this type of training is becoming possible.

To create an immersive experience, the Mission Rehearsal Exercise places the participant in an environment that combines the use of ultra-wide-screen graphics and immersive 10.2 audio (10 channels of audio with 2 subwoofer channels, see Kyriakakis, 1998). The human learner interacts with virtual humans by means of spoken dialogue in the context of a structured story, which is designed to achieve specific pedagogical goals. These stories present the participant with a series of dilemmas, and the outcome 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.

The MRE architecture is shown in Figure 2. DIMR provides the core graphics and animation services by integrating commercially available graphics software with the simulation. Static scene elements and special effects are rendered using Multigen-Paradigm’s Vega™ software. The bodies of the virtual humans are animated using the PeopleShop™ Embedded Runtime System (PSERT) from Boston Dynamics, Incorporated (BDI), while the animation software from Haptek, Inc. provides expressive faces and accurate lip synchronization. DIMR also sends messages to audio system to trigger sound events in the environment that correspond to the visual effects.

Virtual humans play the role of the local populace and friendly (or hostile) elements. The characters use expressive faces and gestures to make their interactions more believable A hybrid approach was taken to

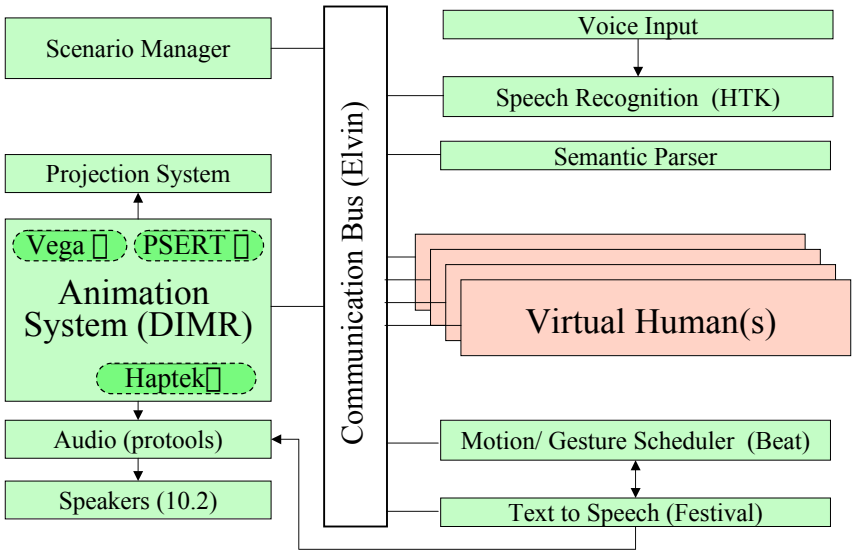

Figure 2: MRE System Architecture

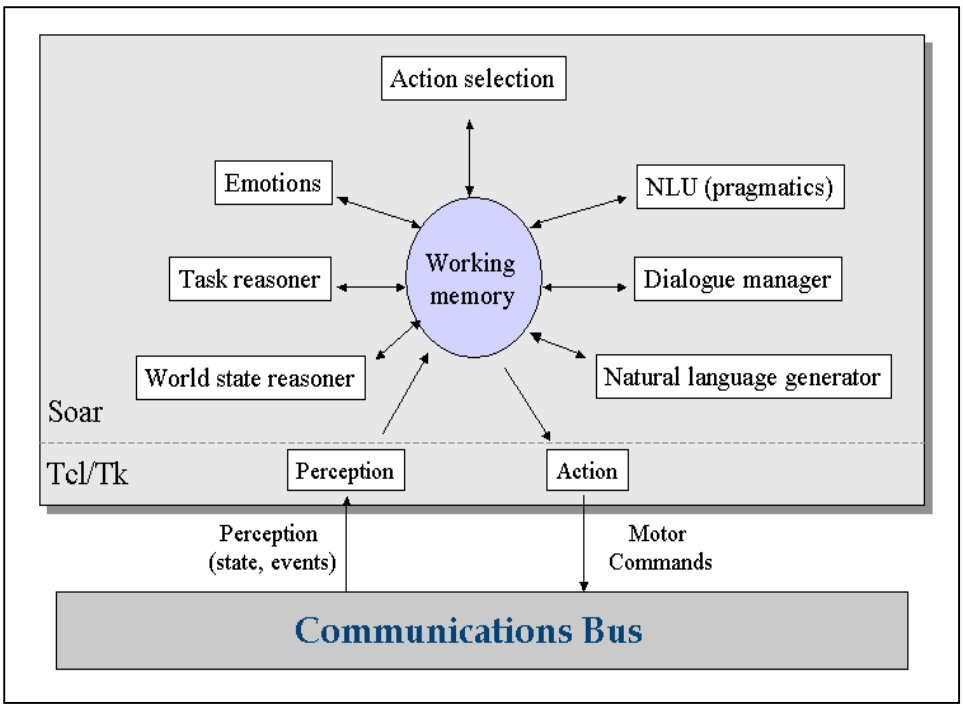

Figure 3: Virtual human architecture

create these characters: some execute pre-scripted reactions to specific stimuli, while others use automated reasoning for perception, dialogue, emotion, and body control.. The remainder of this project overview summarizes the research that has gone into creating virtual humans that fulfill the needs we have identified.

3 Virtual Humans

Our virtual humans build on prior work in the areas of embodied conversational agents (Cassell, et al., 2000b) and animated pedagogical agents (Johnson, et al., 2000), but they integrate a broader set of capabilities than any prior work (see Gratch, et al., 2002 for a recent overview). For the types of training scenarios we are targeting, the virtual humans must integrate three broad influences on their behavior: they must perceive and act in a 3D virtual world, they must engage in face-to-face spoken dialogues with people and other virtual humans in such worlds, and they must exhibit human-like emotions. Classic work on virtual humans in the computer graphics community focused on perception and action in 3D worlds (Badler, et al., 1993), but largely ignored dialogue and emotions. Several systems have carefully modeled the interplay between speech and nonverbal behavior in face-to-face dialogue (Cassell, et al., 1994, 2000a; Pelachaud, et al., 1996) but these virtual humans did not include emotions and could not participate in physical tasks in 3D worlds. Some work has begun to explore the integration of conversational capabilities with emotions (Lester, et al., 2000; Marsella, et al., 2000; Poggi & Pelachaud, 2000), but still does not address physical tasks in 3D worlds. Steve (Rickel & Johnson, 1999a,b, 2000) addressed the issues of integrating face-to-face dialogue with collaboration on physical tasks in a 3D virtual world, but did not include emotions and had limited dialogue capabilities. The tight integration of all these capabilities is the most novel aspect of our current work.

The virtual humans (Figures 2 & 3) are implemented in Soar, a general architecture for building intelligent agents (Newell, 1990) and build on the Steve system (Rickel and Johnson, 1999a,b, 2000). As such, their behavior is not scripted; rather, it is driven by a set of general, domain-independent capabilities discussed below. The virtual humans perceive events in the scenario, 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 Communications Bus shown in Figures 2 and 3.

Steve consists of a set of general, domain independent capabilities operating over a declarative representation of domain tasks. We can apply Steve to a new domain simply by giving him declarative knowledge of the virtual world—its objects, their relevant simulator state variables, and their spatial properties—and the tasks that he can perform in that world. We give task knowledge to Steve using a relatively standard hierarchical plan representation. Each task model 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 each step's role in the task; each link specifies that one step achieves a particular goal that is a precondition for a second step (or for the task's termination). Steve's 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 his student and teammates.

a. Perception and Spatial Awareness

One obstacle to creating believable interaction with a virtual human is omniscience. When human participants realize that a character in a computer game or a virtual world can see through walls or instantly knows events that have occurred well outside its perceptual range, they lose the illusion of reality and become frustrated that the virtual humans have an unfair advantage. Many current applications of virtual humans finesse the issue of how to model perception and spatial reasoning. We have developed a model of perceptual resolution based on psychological theories of human perception (Hill, 1999, 2000). Hill's model predicts the level of detail at which an agent will perceive objects and their properties in the virtual world. He applied his model to synthetic helicopter pilots in simulated military exercises. Complementary research by Sonu Chopra-Khullar and Norman Badler provides a model of visual attention for virtual humans (Chopra-Khullar and Badler, 2001). Their work, which is also based on psychological research, specifies the types of visual attention required for several basic tasks (such as locomotion, object manipulation, or visual search), as well as the mechanisms for dividing attention among multiple tasks.

We have begun to put these principles into practice, beginning by making the virtual humans' models of perception more realistic. We implemented a model that simulates many of the limitations of human perception, both visual and aural, and limited the virtual human's simulated visual perception to 190 horizontal degrees and 90 vertical degrees. The level of detail perceived about objects is high, medium, or low, depending on where the object is in the virtual human's field of view and whether attention is being focused on it. The virtual human perceives dynamic objects, under the control of the simulator, by filtering updates that the system periodically broadcasts.

While omniscience was one problem that had to be addressed, a second obstacle to believable behavior was blindness. The simulator does not represent static objects, such as buildings and trees, in the same way that it represents dynamic objects, which meant that the virtual humans could not perceive them. To address this issue we developed a method for perceiving the locations of static objects by using the scene graph representation employed by Vega™ and an edge-detection algorithm to determine the locations of these objects. But knowing where the edges of the static objects were was not enough—the virtual humans also needed to be able to make sense of the space so that they could move freely without colliding with walls and other objects. To accomplish used a method called the Absolute Space Representation (Yeap and Jeffries, 1999) and extended it to enable our virtual humans to incrementally encode perceptions of the environment in a cognitive map (Hill, Han, and van Lent, 2002). The cognitive map contains the surfaces of static objects, the locations of exits, labeled as known and hypothesized, and maintains a memory of spaces that have only been partially explored. We have begun experimenting with using the cognitive map for way-finding and other spatial tasks.

Another type of percept that was not initially modeled was sound—our virtual humans could not "hear" the explosions, truck and helicopter sounds that were being broadcast to the human participants. As result, they did not react in any way to aural cues and loud sounds. To model aural perception we estimate the sound pressure levels of objects in the environment and determining their individual and cumulative effects on each listener based on the distances and directions of the sources. This enables the agents perceive aural events involving objects not in the visual field of view. For example, when the sergeant hears that a vehicle is approaching from behind, he can choose to look over his shoulder to see who is coming. When he observes that it is his lieutenant, he can get up and turn around in order to enable efficient face to face

conversation. Another effect of modeling aural perception is that some sound events can mask others. A helicopter flying overhead can make it impossible to hear someone speaking in normal tones a few feet away. The noise could then prompt the virtual speaker to shout and could also prompt the addressee to cup his ear to indicate that he cannot hear.

b. Natural Language Dialogue

The example in section 2 shows some of the demands on dialogue ability for virtual humans. They must be aware of their environment and perceptual considerations. They must be able to relate utterances to a task and emotional model. They must use physical behaviors as well as verbal behaviors to communicate. Moreover, in the MRE scenario, they must be able to participate in multi-party conversations, such as when the sergeant brings the medic into the dialogue. In many ways, our natural language processing components and architecture mirror fairly traditional dialogue systems. There is a speech recognizer, semantic parser, dialogue manager, NL generator, and speech synthesizer. However, the challenges of the MRE project, including integration within an immersive story environment as well as with the other virtual human components required innovations in most areas. The Speech recognizer was built using the Hidden Markov Model Toolkit (http://htk.eng.cam.ac.uk/) currently employing a limited domain finite-state language model with a several hundred word vocabulary and several thousand phrase generative capacity, and with locally trained acoustic models (Wang & Narayanan, 2002). Speech recognition messages are interpreted using a hybrid semantic parser, consisting of finite state and statistical parsers, each producing a best-guess at semantic information from the input word stream. In cases in which imperfect input is given, the parser will pass on possibly incomplete or partially incorrect representations. See (Traum, 2003) for more details about the semantic representation. Dialogue Processing and Generation are part of the integrated Virtual Human module, implemented in Soar. The Soar module for each agent receives the output of the speech recognizer and semantic parser. This information is then matched against the agent's internal representation of the context, including the actions and states in the task model, current expectations, and focus to determine a set of candidate interpretations. Some of these interpretations may be underspecified, due to impoverished input, or over-specified in cases of incorrect input (either an out of domain utterance by the user, or an error in the speech recognizer or semantic parser). The dialogue component of the Soar agent also produces a set of dialogue act interpretations of the utterance. Some of these are traditional speech acts (e.g., assert, request, info-request) with content being the semantic interpretation, while others represent other levels of action that have been performed, such as turn-taking, grounding, and negotiation. See (Traum & Rickel, 2002) for details on the levels of dialogue acts.

Dialogue management follows the approach of the TRINDI Project (Larsson & Traum, 2000), and specifically the EDIS system (Matheson, Poesio, & Traum, 2000). Dialogue acts are used to update an *Information State* that is also used as context for other aspects of agent reasoning. More on the aspects of information state can be found in (Traum & Rickel, 2002). Decisions of how to act in dialogue are tightly coupled with other action selection decisions in the agent. The agent can choose to speak, choose to listen, choose to act related to a task, etc. Aspects of the information state provide motivations to speak, including answering questions, negotiating with respect to a request or order, giving feedback of understanding (acknowledgements, repairs, and repair requests), and making suggestions and issuing orders, when appropriate according to the task model (Traum et al., 2003b). Once a decision is made to speak, there are several phases involved in the language production process, including content selection, sentence planning, and realization (Fleischman & Hovy, 2002; Traum, et al., 2003a). The final utterance is then augmented with communicative gestures and sent to the synthesizer and rendering modules to produce the speech. The speech synthesizer uses Festival and Festvox, with locally developed unit-selection limited-domain voices to provide the emotional expressiveness needed to maintain immersiveness (Johnson et al., 2002).

c. Emotion

A key aspect of leadership is recognizing and managing interpersonal emotion. Thus, virtual humans must exhibit the emotional behaviors and cognitive biases one would expect in a high-stake encounter such as the Bosnia scenario. Our work on modeling emotion is motivated by cognitive appraisal theory, a psychological theory that emphasizes the relationship between emotion and cognition (Lazarus, 1991; Scherer, 1984). Following this view, two processes underlie emotional behavior. *Appraisal* characterizes the organism's goals/desires vis-à-vis the environment. For example, if the injured boy's mother believes

the soldiers are abandoning him, this is appraised as a threat to her goals and leads to distress or anger. *Coping* translates this assessment into strategic acts that maintain positive or overturn negative appraisals. Following Lazarus (1991), there are two classes of coping strategies. *Problem-focused* strategies act on the world and include acting, planning, or certain forms of speech acts (request, order, etc). For example, the mother might implore the troops to remain on the scene. *Emotion-focused* strategies act internally to change beliefs or desires. For example, the mother might convince herself that the troops are going for help, despite perceptual evidence. Coping and appraisal interact and unfold over time. For example, the mother agent may "feel" distress over the accident (appraisal), which motivates the agent blame the accident on the troops (emotion-focused coping), which transforms the agent's distress into anger (re-appraisal).

In re-casting this theory as a computational model, we have tied appraisals and coping to the agent's task-related information in Soar's working memory including the current state of the world, as derived by the perception module, expectations of future acts derived from the task model, and the consequences of past events, stored in a causal history. This facilitates reasoning about blame and indirect consequences of action. For example, a threat to a sub-goal (such as the landing zone being secure) might be distressing, not because the sub-goal is intrinsically important, but because it facilitates a larger goal (such as evacuating the boy). These data structures provides a uniform representation of past and future actions (this action caused an effect which I can use to achieve that goal) and facilitate reasoning about different agents' perspectives (I think this outcome is good but I believe you think it is bad).

Our approach to appraisal assesses the agent-environment relationship via features of this explicit task representation (Gratch, 2000). Speaking loosely, we treat appraisal as a set of feature detectors that map features of this representation into appraisal variables that characterize the consequences of an event from the agent's perspective. These variables include the desirability of those consequences, the likelihood of them occurring, who deserves credit or blame and a measure of the agent's ability to alter those consequences. The result of this feature detection is one or more data structures, called appraisal frames, which characterize the agent's emotional reactions to an event. Thus, the belief that another agent has caused an undesirable outcome leads to distress and possibly anger.

Our computational model of coping—as described in (Marsella & Gratch, 2002)—similarly exploits the task representation to uncover what features led to the appraised emotion, and what potential there may be for altering these features. In essence, coping is the inverse of appraisal. To discharge a strong emotion about some situation, one obvious strategy is to change one or more of the factors that contributed to the emotion. Coping operates on the same representations as the appraisals, the agent's beliefs, goals and plans, but in reverse, seeking to make a change, directly or indirectly, that has the desired impact on appraisal. Coping could impact the agent's beliefs about the situation, such as the importance of a threatened goal, the likelihood of the threat, responsibility for an action, etc. Further, the agent might form intentions to change external factors, for example, by performing some action that removes the threat. Indeed, our coping strategies can involve a combination of such approaches, mirroring how coping processes are understood to operate in human behavior whereby people may employ a mix of problem-focused coping and emotion-focused coping to deal with stress.

A final aspect of the model concerns cognitive focus of attention. At any point in time, the virtual humans have many potential emotions corresponding to multiple appraised features of the task representation. Only appraisal frames that are in cognitive focus lead to expressed emotion or coping behavior. Appraisals are brought into focus if their constituent data elements are referenced by a mental operator currently under consideration. These mental operators consist of the atomic Soar operators that implement the various agent functions (e.g. updating a belief, forming an intention, understanding a speech act, repairing a plan, etc.). For example, if the trainee asks the virtual sergeant "what happened," the operator that resolves the referent of this question must access several data structures and the associated appraisals are brought into focus. Via Soar's decision cycle, multiple operators are proposed in parallel and one is selected for execution. For example, if the agent is considering two possible courses of action (help-the-boy vs. help-the-inspection-team), operators are proposed in parallel suggesting the adoption of each. Any appraisal frames associated with each of these actions are brought into focus. This focus mechanism allows expressed emotion to more closely mirror the agent's mental processes and is important for modeling certain emotion-focused coping strategies that are related to mental focus of attention (such as rumination and avoidance).

Whereas there has been prior work in computational models of appraisal, there has been little prior work in modeling the myriad ways that people cope with emotions. And yet coping behavior is a key aspect of human behavior. People employ a rich set of coping strategies and different individuals tend to adopt stable and characteristic "coping styles" that are correlated with personality type. Our work is building a library of these strategies and uses personality-inspired preference rules to model consistent differences in style across different agents. For example, our virtual humans may take preemptive action to circumvent a stressful factor, they may choose to shift blame to another agent or they may behaviorally disengage from attempts to achieve a goal that is being thwarted or threatened.

d. Physical Behavior

The virtual humans continually perceive and internally process the events surrounding them, understanding utterances, updating their beliefs, formulating and revising plans, generating emotional appraisals, and choosing actions. Our goal is to manifest the rich dynamics of this cognitive and emotional inner state through each character's external behavior using the same verbal and nonverbal cues that people use to understand one another. The key challenge is the range of behaviors that must be seamlessly integrated: each character's body movements must reflect its awareness of events in the virtual world, its physical actions, the myriad of nonverbal signals that accompany speech during social interactions (e.g., gaze shifts, head movements, and gestures), and its emotional reactions.

Since gaze indicates a character's focus of attention, it is a key element in any model of outward behavior, and must be closely synchronized to the character's inner thoughts. Prior work on gaze in virtual humans has considered either task-related gaze (Chopra-Khullar & Badler, 2001) or social gaze (Cassell et al., 1994) but has not produced an integrated model of the two. Our gaze model is driven by our cognitive model, which interleaves task-related behaviors, social behaviors, and attention capture. Task-related behaviors (e.g., checking the status of a goal or monitoring for an expected effect or action) trigger a corresponding gaze shift, as does attention capture (e.g., hearing a new sound in the environment). Gaze during social interactions is driven by the dialogue state and the state of the virtual human's own processing, including gaze at an interlocutor who is speaking, gaze aversion during utterance planning (to claim or hold the turn), gaze at an addressee when speaking, and gaze when expecting someone to speak. This tight integration of gaze behaviors to our underlying cognitive model ensures that the outward attention of the virtual humans is synchronized with their inner thoughts.

Body movements are also critical for conveying emotional changes, including facial expressions, gestures, posture, gaze and head movements (Marsella, Gratch, & Rickel, 2001). In humans, these behaviors are signals and as such they can be used to intentionally inform or deceive, but they can also unintentionally reveal information about the individual's internal emotional state. Thus a person's behavior may express anger because they feel it or because they want others to think they feel it or for both reasons. Prior work on emotional expression in virtual humans has focused on either the intentional emotional expression or as a window on internal emotional state (Neal Reilly, 1996). Our work attempts to integrate these aspects by tying expressive behavior to coping behavior. As noted earlier, emotional changes in the virtual human unfold as a consequence of Soar operators updating the task representation. These operators provide a focus for emotional processes, invoking coping strategies to address the resulting emotions, that in turn leads to expressive behaviors. This focus on operators both centers emotional expression on the agent's current internal cognitive processing but also allows coping to alter the relation of the expression to those internal cognitive processes. For example, when the lieutenant asks "What happened here?" the sergeant agent might openly express true appraisal-based feelings of guilt and concern, for example through facial expressions, gestures, posture, gaze and head movements. However, if the sergeant is coping by shifting responsibility for the accident to the mother, the initial expression of guilt will be quickly suppressed. Instead, the sergeant will express anger at the mother, nonverbally through an angry facial expression, a shaking of his head and a dismissive wave of his hand toward the mother, as well as verbally by saying "they rammed into us", all in a more calculated attempt to persuade the lieutenant.

Finally, a wide range of body movements are typically closely linked to speech, movements that emphasize, augment and even supplant components of the spoken linguistic information. Consistent with

this close relation, this nonverbal behavior, which can include hand-arm gestures, head movements and postural shifts, is typically synchronized in time with the speech. Realizing this synchronization faces the challenge that we do not have an incremental model of speech production. Such a model would allow us to tie nonverbal behaviors to speech production operations much like the gaze and coping behaviors are tied to cognitive operations. Rather, our approach is to plan the utterance out and annotate it with nonverbal behavior. The annotated utterance is then passed to a text-to-speech generation system that schedules both the verbal and nonverbal behavior, using the BEAT system (Cassell, et al., 2001). This approach is similar to the work of Cassell et al. (1994). Our work differs in the structure passed to the gesture annotation process, in order to capture the myriad ways that the nonverbal behavior can relate to the spoken dialog and the internal state of the virtual human. Specifically, while both systems pass the syntactic, semantic and pragmatic structure of the utterance, we additionally pass the emotional appraisal and coping information associated with the components of the utterance. The gesture annotation process uses this information to annotate the utterance with gestures, head movements, eyebrow lifts and eyeflashes.

4. Summary

As the previous section indicates, virtual humans require a number of sophisticated abilities in order to produce human-like behavior in a complex world. In each of these areas our work has generally taken a broader view of the issue than most other work. An important result is not just the range of functioning of the individual components, but the way they work together to provide human-like behavior. Emotion reasoning is not done in isolation, but with respect to cognitive function, perception, and dialogue interaction. The focusing mechanism is used not just for emotional reasoning, but also in providing dialogue responses to what might otherwise be interpreted as very broad and open questions. Perception not only updates belief, but motivates action such as moving into contact, or the voice level to use in dialogue. The Soar architecture, greatly facilitates this interactivity, since it makes facilitates both parallel and serial processing and allows easy access from one component to the data structures used by others.

The Mission Rehearsal Exercise project takes a step toward realizing the ultimate goal of creating an immersive environment where people can hold extended dialogues with virtual characters in a 3D world. While the goal of MRE is to help military leaders develop the ability to make decisions under stress, it is easy to see the applicability of the underlying technologies to a host of other domains in education, training, and entertainment. This paper summarizes some of the progress that has been made toward populating such a world with embodied conversational agents.

References

Badler, N. I., Phillips, C. B., & Webber, B. L. (1993). *Simulating Humans*. New York: Oxford University Press.

Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., et al. (1994). "Animated Conversation: Rule-Based Generation of Facial Expression, Gesture and Spoken Intonation for Multiple Conversational Agents,". Proceedings of the SIGGRAPH, ACM Press, Reading, MA.

Cassell, J., Bickmore, T., Campbell, L., Vilhjálmsson, H., & Yan, H. (2000a). "Human conversation as a system framework: Designing embodied conversational agents," In J. Cassell, J. Sullivan, S. Prevost & E. Churchill (Eds.), *Embodied Con-versational Agents* (pp. 29-63). Boston: MIT Press.

Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. (Eds.). (2000b). *Embodied Confersational Agents*. Cambridge, MA: MIT Press.

Cassell, J., Vilhjálmsson, H., & Bickmore, T. (2001). *BEAT: The Behavior Expressive Animation Toolkit.* Paper presented at the SIGGRAPH, Los Angeles, CA.

Chopra-Khullar, S., & Badler, N. (2001). Where to Look? (Automating) Attending Behaviors of Virtual Human Characters. *Autonomous Agents and Multi-Agent Systems, 4*(1-2), 9-23.

Fleischman, M., & Hovy, E. (2002). *Emotional variation in speech-based natural language generation.* Paper presented at the International Natural Language Generation Conference, Arden House, NY.

Gratch, J. (2000). *Émile: marshalling passions in thraining and education.* Paper presented at the Fourth International Conference on Intelligent Agents, Barcelona, Spain.

Gratch, J., Rickel, J., André, E., Badler, N., Cassell, J., and Petajan, E. (2002), "Creating Interactive Virtual Humans: Some Assembly Required," *IEEE Intelligent Systems*, July/August 2002, pp. 54-63.

Hill, R., Han, C., van Lent, M. (2002). "Perceptually Driven Cognitive Mapping in a Virtual Urban Environment," *AI Magazine*, Fall 2002.

Hill, R. (1999). "Modeling Perceptual Attention in Virtual Humans," *Proceedings of the 8th Conference on Computer Generated Forces and Behavioral Representation*, SISO, Orlando, Fla., pp. 563–573.

Hill, R. (2000). "Perceptual Attention in Virtual Humans: Toward Realistic and Believable Gaze Behaviors," *Proceedings of the AAAI Fall Symposium on Simulating Human Agents*, AAAI Press, Menlo Park, Calif., 2000, pp. 46–52.

Johnson, W. L., Narayanan, S., Whitney, R., Das, R., Bulut, M., & LaBore, C. (2002). *Limited Domain Synthesis of Expressive Military Speech for Animated Characters.* Paper presented at the 7th International Conference on Spoken Language Processing, Denver, CO.

Johnson, W. L., Rickel, J., & Lester, J. C. (2000). Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments. International Journal of AI in Education, 11, 47-78.

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

Larsson, S., & Traum, D. (2000). Information state and dialogue management in the TRINDI Dialogue Move Engine Toolkit. Natural Language Engineering, 6, 323-340.

Lazarus, R. (1991). *Emotion and Adaptation*. NY: Oxford University Press.

Lester, J. C., Towns, S. G., Callaway, C. B., Voerman, J. L., & FitzGerald, P. J. (2000). Deictic and Emotive Communication in Animated Pedagogical Agents. In J. Cassell, S. Prevost, J. Sullivan & E. Churchill (Eds.), *Embodied Conversational Agents* (pp. 123-154). Cambridge: MIT Press.

Marsella, S., & Gratch, J. (2002). "A Step Toward Irrationality: Usign Emotion to Change Belief,". *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems*, Bologna, Italy.

Marsella, S., Gratch, J., & Rickel, J. (2001). "The Effect of Affect: Modeling the Impact of Emotional State on the Behavior of Interactive Virtual Humans*,"* Proceedings of the Agents 2001 Workshop on Representing, Annotating, and Evaluating Non-Verbal and Verbal Communicative Acts to Achieve Contextual Embodied Agents, Montreal, Canada.

Marsella, S., Johnson, W. L., & LaBore, C. (2000). *Interactive Pedagogical Drama.* Paper presented at the Fourth International Conference on Autonomous Agents, Montreal, Canada.

Matheson, C., Poesio, M., & Traum, D. (2000). "Modeling Grounding and Discourse Obligations Using Update Rules," *Proceedings of the First Conference of the North American Chapter of the Association for Computational Linguistics*.

Neal Reilly, W. S. (1996). *Believable Social and Emotional Agents* (Ph.D Thesis No. CMU-CS-96-138). Pittsburgh, PA: Carnegie Mellon University.

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Pelachaud, C., Badler, N. I., & Steedman, M. (1996). "Generating Facial Expressions for Speech," *Cognitive Science, 20*(1).

Poggi, I., & Pelachaud, C. (2000). Emotional Meaning and Expression in Performative Faces. In A. Paiva (Ed.), *Affective Interactions: Towards a New Generation of Computer Interfaces*. Berlin: Springer-Verlag.

Rickel, J., & Johnson, W. L. (1999a). "Animated Agents for Procedural Training in Virtual Reality: Perception, Cognition, and Motor Control," *Applied Artificial Intelligence*, 13, 343-382.

Rickel, J., & Johnson, W. L. (1999b). "Virtual Humans for Team Training in Virtual Reality," *Proceedings of the Ninth International Conference on Artificial Intelligence in Education.*

Rickel, J., & Johnson, W. L. (2000). "Task-Oriented Collaboration with Embodied Agents in Virtual Worlds," In J. Cassell, J. Sullivan, S. Prevost & E. Churchill (Eds.), *Embodied Conversational Agents*. Boston: MIT Press.

Scherer, K. (1984). "On the nature and function of emotion: A component process approach," In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 293-317).

Traum, D. (2003). "Semantics and Pragmatics of Questions and Answers for Dialogue Agents," Paper presented at the *Fifth International Workshop on Computational Semantics*, Tilburg.

Traum, D., & Rickel, J. (2002). *Embodied Agents for Multi-party Dialogue in Immersive Virtual Worlds.* Paper presented at the First International Conference on Autonomous Agents and Multi-agent Systems, Bologna, Italy.

Traum, D., Fleischman, M., & Hovy, E. (2003a). "NL Generation for Virtual Humans in a Complex Social Environment," In *Proceedings of he AAAI Spring Symposium on Natural Language Generation in Spoken and Written Dialogue*, pp. 151-158, 2003.

Traum, D., Rickel, J., Gratch, J. and Marsella, S. (2003b). Negotiation Over Tasks in Hybrid Human-Agent Teams for Simulation-Based Training. In *Proceedings of the Second International Joint Conference on Autonomous Agents and Multi-Agent Systems, 2003*. ACM Press.

Wang, D., & Narayanan, S. (2002). *A confidence-score based unsupervised MAP adaptation for speech recognition.* Paper presented at the Proceedings of 36th Asilomar Conference on Signals, Systems and Computers.

Yeap, W. and Jefferies, M. (1999). "Computing a Representation of the Local Environment," *Artificial Intelligence*, vol. 107, no. 2, Feb. 1999, pp. 265–301.