CHAPTER XX

Cognitive and affective modeling in intelligent virtual humans for training and tutoring applications

Robert Sottilare and John Hart

U.S. Army Research Laboratory
Human Research & Engineering Directorate
Orlando, Florida
{robert.sottilare, john.hartiii}@us.army.mil

ABSTRACT

This chapter reviews current and emerging trends in cognitive and affective (e.g., emotions, motivational) modeling within virtual humans and their application to training and tutoring domains for individuals and small groups. Virtual humans have become commonplace in computer games and other digital entertainment applications, but their use for training and one-to-one tutoring applications is evolving and remains primarily focused on well-defined training/tutoring domains (e.g., procedural tasks and rule-based domains including mathematics and physics). In order to support viable self-regulated learning environments, future training and tutoring systems will require virtual humans with enhanced cognitive and affective capabilities that are adaptive, engaging and motivating in ill-defined domains.

Keywords: virtual humans, training, computer-based tutoring

1. INTRODUCTION

In 2002, Egges, Kshirsagar and Magnenat-Thalmann proposed a goal to create virtual humans that can interact spontaneously using a natural language, emotions and gestures in a manner similar to and even indistinguishable from real humans. Toward this goal, this chapter compares and contrasts the capabilities of virtual humans and reviews their current roles and limitations in entertainment and training. Potential roles and capabilities for future virtual humans in training and computer-based tutoring contexts are also discussed along with recommended design goals and areas for future/continued research.

To engage trainees and gain their confidence as credible actors in training environments, future virtual humans will need to be able to understand the trainee's language and interpret their behaviors to assess the trainee's states (e.g., cognition and affect) and then respond appropriately. They will also need to recognize context in the training environment (e.g., increases or decreases in progress toward training objectives). Using this trainee and training environment information, future virtual humans will be able to adapt their interaction (e.g., direction or support) to influence and even optimize learning. By affiliating with and motivating learners, virtual humans will develop rapport and trust as credible and supportive mentors.

2. COMPARING AND CONTRASTING VIRTUAL HUMANS

There is a great fascination with the creation of artificial humans. It is not a recent idea and can be seen in literature and entertainment throughout many years. In Mary Shelley's *Frankenstein*, Dr. Frankenstein desired to create an artificial human. Modern science fiction introduced us to many artificial humans such as in the Terminator films, and the human clone replicants in *Blade Runner*. There are many other examples of artificial humans developed to perform roles to assist humans or perform dangerous missions as in weapons of war in literature and entertainment. Now, imagine our world today with similar artificial humans. What about an artificial human that could act as a virtual receptionist or assist in training as a tutor or teammate? What if you could deal with a character that was unscripted, had knowledge and could reason about environment, understood and expressed emotion, communicated both verbally and nonverbally, and could play different roles as needed? Recently, renewed interest in artificial humans is making this a reality. This quest to build an artificial human is becoming a reality due to developments in virtual human technologies. It can be seen that there is a role for virtual humans in our world today and into the future.

Just as all men are not created equal, so it is true for virtual humans. Virtual humans can be thought of as "software entities [that] look and act like people and can engage in conversation and collaborative tasks, but they live in simulated environments." (Gratch, et. al., 2002). They can also include different cognitive states to include beliefs, desires, goals, intentions, and attitudes (Rickel and Johnson, 1999; Traum, Swartout, Gratch, and Marsella, 2008). Many virtual humans have been developed over the years, but each one was developed for a specific purpose as in to support training and to provide a new interface for the delivery of information. Forms of virtual humans are being used today as web-based airline reservationists as on Alaska Airlines on *www.alaskaair.com*, building receptionists as in MicroSoft Research's "Situated Interaction" project (Bohus, 2008), museum guides (Swartout, et al., 2010), Army recruiting (Artein, 2009), and training applications (Rickel and Johnson, 1999; Hill, et. al., 2006). In other areas, virtual humans can also be helpful in the medical and social sciences fields that include diagnosis, treatment, and therapy skills.

Differences in virtual humans start with how they are controlled. Basically, virtual humans are controlled by either a human via a keyboard and mouse, joystick, and other interfaces such cameras and motion sensors. Human or user controlled characters are popular in computer games (ie., Call of Duty, Battlefield, Halo), massively multiplayer online games (ie., World of Warcraft, Sims Online), and virtual worlds (ie., Second Life, Active Worlds). The user controls the actions of the character in these environments. There environments do allow for some characters to be controlled via the computer and they are known as "bots". The computer controlled characters have become more than just "bots" with the development of computational algorithms in natural language, emotions, and behaviors. These developments have allowed virtual humans to perceived and respond to and within the environment without human intervention (Johnsen, Beck, & Lok, 2010; Kenny, et al., 2007; Swartout, 2010).

Johnsen (2008) describes four categories that affect the human-computer interaction when dealing with virtual humans. Those categories are the virtual human, the simulation system, the user environment, and the user. In recent years, there has been a large body of research focused on the virtual human development in the areas of anthropomorphism (Dehn & van Mulken, 2000; Yee, Bailenson, &, Rickersten, 2007); appearance (Garau, et al., 2003; Bailenson, et al., 2005; MacDorman, Coram, Ho, & Patel, 2010); and behaviors (Garau, et al., 2003; Bailenson, et al., 2005; Gratch, et al., 2007). The simulation system includes the human to computer interface. There are a number a methods for interacting with virtual humans that include speech and natural language (Traum, et al., 2007; Johnsen & Lok, 2008); text and natural language (Rizzo, et al., 2010, Sproull, et al., 1996); and menu systems (Hill, et al., 2006; Lester, Stone, & Stelling, 1999).

2.1 Roles of virtual humans in games and digital entertainment applications

As one participates in a game, virtual world, or virtual training environment, it can be analogous to an interactive drama where the user experiences the story first hand (Bates, 1992; Mateas & Stern, 2002). Virtual humans are then considered the actors within the virtual environment. In screenwriting and other entertainment, actors typically take on primary or supporting roles within the story (Stout, 2011). Primary roles include both the protagonist and antagonist. One could develop a single user interactive play model for the virtual environment where the user assumes the role of the protagonist. In this model, the antagonist and other supporting roles are played by virtual humans. Using the interactive play model, a more appropriate description of the interaction might be an improvisational play where the actors are free to an act or react to other actors within the limits of the story or script (E. H. LeMasters, personal communications, Sept. 1, 2011). Supporting roles for virtual humans include that of a team mate, adversary, instructor, tutor, or mentor.

In developing the improvisational play model, the roles a virtual human can play can be divided into either roles as an actor or as a supporting role. The supporting role provides realism for the scenario and may cause the actors to act or react to the supporting members. An example may be a crowd of people in the street, where they may become angry based on the action or decision of the primary actors. The actor roles in the improvisational play help to "drive" the story by taking actions based on the current state of the scenario. A virtual human could participate as a team member or ally, an adversary, or as a mentor or instructional role.

The training domain makes use of virtual humans in instructional roles. Frenchette (2008) describes some of these roles as supplanting agents, scaffolding agents, demonstrating agents, modeling agents, coaching agents, and testing agents. These types of agents provide different forms of instructional support to the training environment. Steve, a half-bodied human-like agent, was an early implementation of an instructional agent (Johnson, Rickel, & Lester, 2000). Steve shared the virtual environment with the student and aided them in learning how to operate shipboard equipment. In the environment, Steve would first start with a demonstration of how to perform an activity opposed to just explaining the activity. The student could ask questions of Steve, and even ask to complete an activity he had started. Steven monitored the student's performance and provided assistance when needed. Another agent was Herman the Bug (Lester, et al., 1997). Herman helped students understand biological processes for plants and provided advice in response to students' problem solving activities. Lester, et al. (1997) proposed that virtual humans can play a critical role in the motivation of students through the way the agent interacts with the student.

2.2 Roles of virtual humans in training and tutoring domains

This section examines the roles of virtual humans in sample training and tutoring environments. There are several working examples of embodied conversational agents in both training and tutoring contexts (Rickel & Johnson, 1999; Traum, et al, 2007; Person, Graesser, Kreuz, Pomeroy & the Tutoring Research Group, 2001). In the training domain, virtual humans are used to take the place of human role-players and thereby reduce the labor required to support training exercises. In the tutoring domain, virtual humans are the interface for providing feedback (e.g., direction, support, information) to the trainee.

Examples of the use of virtual humans in the training domain include virtual patients used for both training clinical therapists (Kenny, Parsons, Gratch, Leuski and Rizzo, 2007) and medical doctors performing military sick call (Kenny, Parsons and Garrity, 2010). In the tutoring domain, both AutoTutor (Graesser, Chipman, Haynes and Olney, 2005) and the Cognitive Tutor (Ogan, Aleven, Kim & Jones, 2010) use virtual human interaction to guide instruction and provide feedback to the trainee.

2.3 Cognitive and affective modeling of virtual humans in training and tutoring domains

"Emotion pulls the levers of our lives, whether it be by the song in our heart, or the curiosity that drives our scientific inquiry." (Picard, 1995). In order to fulfill the goal of being indistinguishable if one is interacting with a virtual human or human, emotion is needed. Emotion is important to the social interaction. The Institute for Creative Technologies of the University of Southern California has been researching and developing virtual human technologies over the last decade and has found that an emotional model can have an impact on the virtual human's cognitive processing. The emotional model can provide insight to assist in natural language processing such as the disambiguation of ambiguous references. The model can also inform the decision making process (Swartout, 2010). de Melo, Carnevale, and Gratch (2010) developed a virtual human to play games of prisoner's dilemma and found indications that emotions in the virtual human players had influence on the human participant's decision making.As one considers the development of a computational model for a virtual human, one quick finds that it involves a multi-disciplinary approach. The theory emotions are rooted in the science of psychology. As one begins to use facets of artificial intelligence, the development of a computational model can unveil assumptions and hidden complexities with the theory of the emotional model (Marsella, Gratch, and Petta, 2010). Some of the more popular models are based on appraisal theory, which is based on an individual's judgment concerning the relationship between events and the individual's beliefs (Lazarus, 1991). Dimensional theories characterize emotion not as discrete events but a points in a continuous multi-dimensional space. Other models include anatomic theories which attempts to reconstruct neural links and process that control emotions; rationale approaches that take a more abstract approach looking at the purpose of the adaptive function of emotion; and communicative approaches in which one views the emotional processes as a communicative system (Marsella, Gratch, and Petta, 2010).

3. DESIGN GOALS FOR EFFECTIVE VIRTUAL HUMANS IN TRAINING AND TUTORING

This section addresses constraints, learning-related design goals and desired attributes for virtual humans to be effective training and tutoring tools. Constraints focus on barriers to real-time interaction with virtual humans while learning-related design goals highlight the potential influence of virtual humans on engagement and motivation during training/tutoring sessions. In conclusion, we address virtual human perception as a gateway to providing adaptable training and tutoring.

3.1 Constraints in training and tutoring domains

As in all human-technology interactions, there is an expectation of real-time (or near real-time) natural language and gesture interaction to maintain engagement.

The process of determining what the trainee says/does, what a particular phrase/gesture means and then selecting an appropriate response takes time. It would be frustrating for trainees to wait several seconds for virtual human responses. This delay can be compounded when we are thinking about applying virtual human technology in: training domains with a large volume of phrases; and in mobile and/or other distributed training/tutoring applications. Until understanding natural language technologies become more efficient, the authors recommend that application of virtual human technologies in very specific and localized contexts. More specific training applications will limit the size of the corpus that must be understood by the virtual human and reduce the search time. Localizing virtual humans to run in the same geographic location as the trainee and training application will minimize delays due to data transport on wide-area networks. Next, we will explore learning-related design goals for virtual humans to enable them to support training and tutoring.

3.2 Learning-related design goals

In assessing the design goals of virtual humans as tutors, we examined validated studies of "expert" human tutors and came across a set of studies (Lepper & Chabay, 1988; Lepper, Aspinwall, Mumme & Chabay, 1990; Lepper, Woolverton, Mumme & Gurtner, 1993; Lepper, Drake & O'Donnell-Johnson, 1997) that provided an analysis of factors contributing to successful tutoring outcomes and resulting in the INSPIRE model of tutoring success (Lepper, Drake & O'Donnell-Johnson, 1997). INSPIRE is an acronym of the attributes of a successful human tutor: intelligent, nurturing, Socratic, progressive, indirect, reflective and encouraging.

If we expect to use virtual humans as coaches, mentors and tutors, we should have design goals that include the capability for them to influence learning by influencing engagement and motivation. Below we posit how virtual humans might be designed to allow them to manage trainee engagement and motivation.

3.2.1 Virtual human influence on trainee engagement

In implementing the INSPIRE model within virtual humans to influence trainee engagement, key attributes to consider in the virtual human design are intelligence, Socratic interaction, reflection and indirect feedback. Virtual humans in the role of the tutor must be knowledgeable of the subject matter and pedagogy to demonstrate credibility and maintain the trust/engagement of the trainee. The virtual human must be able to ask leading questions and provide hints instead of directions/answers that could result in trainees losing interest in the subject matter.

Klein & Baxter (2006) assert that advanced problem solving on the part of trainees requires the recognition of flaws in their existing mental models in a process called the Cognitive Transformation Theory (CTT). CTT links learning objectives to the person's current mental models and promotes reflective processes for shedding flawed mental models for less flawed models. Virtual humans that are designed to support discovery and reflective processes per the INSPIRE model and CTT are anticipated to be more effective tutors/training partners.

3.2.2 Virtual human influence on trainee motivation

In implementing the INSPIRE model within virtual humans to influence trainee motivation, key attributes to consider in the virtual human design are the capabilities to nurture and encourage. Nurturing virtual humans should be able to develop rapport with trainees which indicates a persistent trainee model that is maintained by the virtual human (or in its cognitive architecture or in a tutoring architecture) and used to demonstrate a history of "shared experiences" and mutual trust. Lepper & Malone (1987) and Malone & Lepper (1987) noted five complementary sources of motivation for learning under the heading of encouragement. Virtual humans should be able to encourage trainees and affect their motivation by: enhancing the trainee's feelings of competence and mastery; challenging the trainee to accomplish more; piquing the trainee's sense of curiosity; providing trainees with a sense of control over the learning process; and by providing context and relevance for training content.

3.3 Desired virtual human attributes for training and tutoring

In addition to the influence on trainee engagement and motivation, it is desirable for virtual humans to be able to perceive their environment including the trainee and the training environment (e.g., virtual world, game or other simulation). Perception is a prerequisite for the virtual human to interpret the trainee's state (e.g., cognition and affect) and then use this trainee state data along with training context to formulate optimal training strategies (e.g., feedback, direction, questions, support).

Adapting training strategies based on individual differences (traits), changes in state, context and progress allows for tailored training that is optimized for the individual trainee. A tutoring architecture will provide the capability for virtual humans to recognize the state of the trainee and the training environment. Computer-based tutoring systems have concept maps to describe the trainee state and training context. Markov decision processes might be used to weight and optimize the reward resulting from movement from one state to another. For example, the tutor might determine that a trainee's learning outcome might be optimized by selecting a reflective training strategy over a directive strategy based on the trainee's current motivational level, engagement level, progress and competency. To optimize this decision, it is essential for the tutoring architecture to be able to sense the behaviors, interactions and even the trainee's physiology to determine the trainee's cognitive and affective states.

4. DISCUSSION

Finally, our discussion leads us to understanding of general research domains for modular and integrated virtual humans, their capabilities and limitations and recommendations for future research. In general, virtual humans are difficult to author and modify, they have limited understanding of their environment, limited natural language understanding and prescriptive cognitive and affective responses. Virtual humans have almost no ability to retain and use knowledge gained during a training/tutoring scenario in subsequent scenarios. So, how can we improve the adaptability of virtual humans to be more like their human counterparts?

A major objective is to enable more adaptive interaction between virtual human agents, their live human counterparts, and their environment the following research domains are on the critical path for enhanced virtual humans: natural language understanding and generation, sensing technologies, cognitive modeling, affective modeling and value modeling. The ability for virtual humans to perceive and interpret voice and gestures of multiple live trainees is limited. The ability for virtual humans to identify spoken words/phrases is improving, but the corpus is generally limited to a small number of phrases from a single user that can be recognized and interpreted for a rationale response. Research is needed to improve the capability to recognize, interpret and respond in near real-time to multi-sided conversations that include human and virtual human participants and larger topical domains to support social interaction.

Research is needed to enhance a virtual human's capability to recognize and interpret its environment including the behaviors of human participants. To support a tutoring context, virtual humans must be able to recognize changes in the cognitive state (e.g., engagement) and affective state (e.g., frustration, boredom) of trainees. Research is needed to improve the capabilities of unobtrusive physiological and behavioral sensing technologies to this end.

The modeling of how virtual humans perceive and judge their environment is tied to their cognitive and affective modeling. These models are in turn closely tied to value modeling and affect the decision-making processes of virtual humans. Incorporating value models that include ethics and personality preferences would have a significant impact on the variability of actions taken by virtual humans and would move them from prescriptive beings to true decision makers.

Finally, another major objective is to facilitate the development and authoring of new virtual human characters. Interface standards and a framework for virtual humans would go a long way toward making virtual humans easier to construct and move them from the purview of computer scientists to application domain experts (e.g., training and tutoring developers) who could use virtual human tools and standards to create virtual humans specific to their application needs.

REFERENCES

Cassell, J., Bickmore, T., Campbell, L. and Vilhjalmsson, H. (2000). Human Conversation as a System Framework: Designing Embodied Conversational Agents. In Cassell, J., Sullivan, J. Prevost, S. and Churchill, E. eds. Embodied Conversational Agents. MIT Press, Cambridge, MA.

de Melo, C. M., Carnevale, P., and Gratch, J. (2010). The Influence of Emotions in Emobdied Agents on Human Decision-Making. In J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud, and A. Safonova (Eds.) Intelligent Virtual Agents. 10^{th} International

Conference, IVA 2010. Springer.

Egges, A., Kshirsagar, S. and Magnenat-Thalmann N. Imparting individuality to virtual humans. In First International Workshop on Virtual Reality Rehabilitation (Mental Health, Neurological, Physical, Vocational) (November 2002).

Graesser, A.C., Chipman, P., Haynes, B.C. and Olney, A. (2005). AutoTutor: An Intelligent Tutoring System With Mixed-Initiative Dialogue. IEEE Transactions on Education, VOL. 48, NO. 4, November 2005.

Johnsen, K. (2008). Design and Validation of a Virtual Human System for Interpersonal Skills Education. (Doctoral dissertation, University of Florida, 2008) Dissertation Abstracts International, DAI-B 69/10. (UMI No. 3334476).

Johnsen, K., Beck, D., and Lok, B. (2010). The impact of a mixed reality display configuration on user behavior with a virtual human. In J. Allbeck et. al. (Eds.), Intelligent Virtual Agents 2010. (pp. 42-48). Springer-Verlag Berlin Heidelberg 2010.

Johnson, W. L., Rickel, J. W., & Lester, J. C. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. International Journal of Artificial intelligence in education, 11(1), 47–78.

Kenny, P., Parsons, T., Gratch, J., Leuski, A., and Rizzo, A. (2007). Virtual Patients for Clinical Therapist Skills Training. In C. Pelachaud et al. (Eds.), Intelligent Virtual Agents, pp. 197-210. Springer-Verlag Berlin Heidelberg 2007.

Kenny, P., Parsons, T., Pataki, C., Pato, M., St-george, C., Sugar, J. and Rizzo, A.A. (2008). Virtual Justina: A PTSD Virtual Patient for Clinical Classroom Training. Review Literature And Arts Of The Americas, 6(1), 113-118.

Kenny, P., Parsons, T. and Garrity, P. Virtual Patients for Virtual Sick Call Medical Training. In Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2010 (November 2010).

Klein, G. and Baxter, H. (2006). Cognitive Transformation Theory: Contrasting Cognitive and Behavioral Learning. Presented at the Interservice/Industry Training Systems and Education Conference, Orlando, Florida, December 2006.

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

Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., & Bhogal, R. S. (1997). The persona effect: affective impact of animated pedagogical agents. Proceedings of the SIGCHI conference on Human factors in computing systems (p. 359–366).

Lester, J. C., Towns, S. G., & Fitzgerald, P. J. (1999). Achieving affective impact: Visual emotive communication in lifelike pedagogical agents. International Journal of Artificial Intelligence in Education, 10(3-4), 278–291.

Marsella, S., Gratch, J. and Petta, P. (2010). Computational Models of Emotion. In in Scherer, K.R., Bänziger, T., & Roesch, E. (Eds.) A blueprint for a affective computing: A sourcebook and manual. Oxford: Oxford University Press.

Ogan, A., Aleven, V., Kim, J., & Jones, C. (2010). Intercultural negotiation with virtual humans: The effect of social goals on gameplay and learning. In V. Aleven, J. Kay, & J. Mostow (Eds.), Proceedings of the 10th international conference on intelligent tutoring systems, ITS 2010 (Vol. 1, pp. 174-83). Berlin: Springer.

Person, N. K., Graesser, A. C., Kreuz, R. J., Pomeroy, V., & the Tutoring Research Group (2001). Simulating human tutor dialogue moves in AutoTutor. International Journal of Artificial Intelligence in Education, 12, 23-39.

Picard, R. W. (1995). Affective Computing. (Technical Report No. 321). Cambridge, MA: M.I.T Media Laboratory Perceptual Computing Section. Retrieved from http://affect.media.mit.edu/pdfs/95.picard.pdf on February 26, 2012.

Rickel, J. and Johnson, W.L.: Virtual humans for team training in virtual reality. In: Proceedings of the Ninth International Conference on Artificial Intelligence in Education, IOS Press (1999) 578-585.

Rizzo, A. a, Lange, B., Buckwalter, J. G., Forbell, E., Kim, J., Sagae, K., Williams, J., et al. (2010). An intelligent virtual human system for providing healthcare information and support. Proceeding of 8th International Conference on Disability, Virtual Reality & Associated Technologies (Vol. 163, pp. 503-9). Vina del Mar/Valparaiso, Chile. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21335847

Swartout, W. (2010). Lessons Learned from Virtual Humans. AI Magazine. Spring 2010.

Swartout, W., et al. (2010). Ada and Grace: Toward Realistic and Engaging Museum Guides. In J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud, and A. Safonova (Eds.) Intelligent Virtual Agents. 10th International Conference, IVA 2010. Springer.

Traum, D., Roque, A., Leuski, A., Georgiou, P., Gerten, J., Martinovski, B., Narayanan, S., Robinson, S., & Vaswani, A. (2007). Hassan: A Virtual Human for Tactical Questioning. Proceedings of the 8th SIGdial Workshop on Discourse and Dialoque, pages 71-74 (Antwerp, Belgium, September 2007).