LiveActor: A Virtual Training Environment with Reactive Embodied Agents

Norman Badler
Center for Human Modeling and Simulation
University of Pennsylvania
Philadelphia, PA 19104-6389
Tel: 215-898-5862; Fax: 215-573-5577; badler@seas.upenn.edu

Abstract

The last few years have seen great maturation in using computer graphics to portray 3D embodied virtual agents. Unlike the off-line, animator-intensive methods used for special effects, real-time embodied agents are expected to interact with us "live." They can represent other people in a live VR environment such as autonomous helpers, teammates, or tutors and thus enable novel interactive educational and training applications. We should be able to interact and communicate with them, intentionally or not, through natural modalities such as language, facial expressions, and gesture. *LiveActor* is a virtual training environment where live and virtual people can mutually interact. Various aspects of this system will be discussed, including consistent parameterizations for gesture and facial actions, and the representational basis for character believability, personality, and affect. A Parameterized Action Representation (PAR) allows an agent to act, plan, and reason about its actions or actions of others. Besides embodying the semantics of human action, the PAR is designed for building future behaviors into autonomous agents and controlling animation parameters that portray an agent's internal state.

1 Introduction

Real-time embodied human characters are needed for virtual reality (VR) systems where a live participant must interact with other (virtual) people. The *LiveActor* system is a virtual training environment that facilitates agile development of scenarios, characters, and story, and emphasizes reactive virtual players.

Many VR systems are used for visualization, navigation, or passive participation. A major future application of VR will be for real-time training. The obvious benefits include immersion in a large scale virtual space, physical props and movements commensurate with the real situation, and group interactions between both real and virtual players. The major impediments to achieving VR training have been in effective and reactive virtual players, agile scenario generation, and realistic story variations.

Our research enterprise in human simulation began in the 1970's and continues into the present: e.g., articulated figure modeling and control [7], language-to-animation interfaces [10, 6], expressive gesture control [14], autonomous agents [1], conversational agents [12], and aggregate control and action recognition [2]. Our work has always stressed real-time control across all body components: face, eyes, gestures, body pose, and locomotion.

The graphical portrayal of embodied agents is critical to close encounter training requiring the user to analyze and react to facial actions, body posture, and gesture qualities. Such interactions require virtual human models with realistic contextual behaviors and detailed and controllable parameters. Military simulations often use DI-Guy [8] to portray a connected set of motion captured actions, but DI-Guy lacks the subtle controls needed to express any internal motivations and feelings. Recently we integrated our EMOTE gesture engine [14] onto DI-Guy to enable such expressive controls. Other real-time human models such as Jack [7, 16] and Steve [18], have been used for interactive presentation and training but lacked effective parametric motion controls.

We believe that realism in a synthesized embodied agent will be afforded by careful coordination of its several communication channels [5]. The EMOTE parameters, originally defined for arm and torso movements, have been extended to the face [11]. Because people can often perceive the emotional state of another person by observing their movements, we presume that the body's communication channels can reflect an agent's emotional state. Then we hypothesize that the consistent application of the EMOTE parameters across body channels will result in more genuine (believable) whole body motions. Interestingly, there is evidence that lying and insincerity (or bad acting) may be conveyed through a lack of such consistency [15].

2 Rapid Scenario Authoring

An important part of training is repeatedly practicing a scenario and its variants. Authoring plausible scenarios with training value is a craft: generating stories with realistic variations is a difficult and emerging area of research [19, 13, 21]. From an animation perspective, story variations require the creation of context-sensitive, procedural animation programs (*motion generators*). From an artificial intelligence perspective, stories require planners, behavior models, and a robust system that will dynamically react to the trainee's actions and reactions. Effective stories also require virtual players who are consistent with those that might be encountered in real life. They should have appropriate social and cultural behaviors; some may be allies and some may be adversaries.

We have developed a *Parameterized Action Representation* (PAR) [6, 10]; PAR and the software systems developed around it provide tools for rapid scenario authoring. A primary component of the PAR system is the *Actionary*™. It contains databases of agents, objects, and actions. Parameterized actions are represented in a frame-like structure with fields describing action generation. PARs may be bound to agents. The action hierarchy is derived from motion semantics. An uninstantiated PAR is essentially a definition for an action, containing only default properties; an instantiated PAR is bound to an agent and contains specific information relevant to the agent, objects, and other properties.

PAR allows authoring variable dynamic scenarios for training. The Actionary component of the PAR system is, in essence, a database of actions associated with animation motion generators as well as a natural language verb lexicon. The PAR definitions are context-sensitive and can dynamically react to the current state of the world. For example, the *manner* specification in PAR [14, 3, 11] allows expressive quality changes to movements, allowing actions to be individualized reflecting social, psychological, and physiological aspects of the portrayed player.

Although motion generators such as "reach" or "walk" can, to a degree, adapt to the current state of the world, there is a need for explicit behavior modification based on the internal physiological and psychological state of the agent. One opportunity we will likely exploit is to interface PAR with Performance Moderating Functions [20] defined from empirical human data. Agent state changes are dependent on agent actions and resource loads, and this information is readily obtained from the agent model by monitoring the PARs it executes.

3 Adversaries and Allies

Having virtual players perform actions in varying contexts and with individualized movements is necessary but not sufficient for the creation of virtual training environments. These embodied agents must also appear to have recognizable, realistic social and cultural behaviors. This must be true of both virtual teammates and adversaries.

PAR contains parameters for the specification of agent traits. Many software systems have been created for the generation of emotions based on the OCC model [17]. Personality models have also been applied to agents [22]. While incorporating these models into embodied players would enrich the simulation, more is needed to create effective training environments. Teammates and adversaries need to be realistically distinguishable. Genuine social and cultural behaviors are also necessary. It is also necessary to include situations relevant to the training purpose. This could be done by manipulating the plot or by creating characters truly reflective of those who would be found in the situation depicted in the simulation. We are extending an OCC emotion model and a personality model with role and culture variations [1, 4] where agent instances can be constructed through graphical interfaces rather than writing program code.

4 Reading the Trainee's Interactions

Training participants react to situations differently depending on their motivations. A virtual player that can sense a situation and choose actions appropriate either to an adversary or ally role can produce a dramatic training experience. Thus it is critical that some aspects of trainee actions be sensed to influence virtual player action selection.

While there are important VR issues such as navigation and omni-directional treadmills that focus on the trainee experience, we focus on one neglected area: providing non-verbal cues from the trainee back to the virtual players. Rather than looking at recognition of a limited gesture (manual sign) library, we extract the motion qualities of trainee movements. We have shown that simple feature detectors and trained neural nets can reliably identify the EMOTE parameters in a 3D motion captured human performance; even with a single 2D (camera) view, most of the EMOTE parameters can be readily extracted [23]. Since the EMOTE parameters carry the expressive content of a movement, this information supplies the virtual player with insight into the psychological and physiological state of the trainee. For example, the virtual player could detect *indirect*, *sustained*, and *heavy* qualities (interpreted as fatigue) in the trainee's movements and use that to strike for advantage. Or the virtual player could sense *direct*, *sustained*, and *light* qualities (alert caution) in the trainee and thus choose its actions more warily. A trainee's lack of genuine engagement in the training exercise would make him more vulnerable: exactly one of the effects desired.

5 Using Language for Future Actions and After Action Reviews

Language interfaces to agent actions are mediated through PAR. We have investigated the use of language to import possible future contextually-triggered behaviors into agents for a checkpoint monitoring scenario [10]. The language processing creates and binds PARs to agents with explicit (but parameterized) preconditions. In our (Python) implementation, this new rule is immediately available to the agent though it may not be invoked until a future time when the preconditions are satisfied. We used this capability to insert appropriate reactive behaviors into supporting player agents, e.g., to take cover behind a barrier when they draw their weapons and to warn the other players if the car driver grabs a gun.

Language can also be used for compressing activity reports. During a training session, the trainees will act and perceive events that shape the overall experience. To date, the primary means for studying this simulation for After Action Review (AAR) have been through recorded graphics and repositioned synthetic cameras. PAR can be used to create a textural description generated from the actions of live and virtual players [9], as well as groups [2]. By design, a PAR can be graphically animated but also translated directly into a sentence. The actions of virtual agents are based on PAR executions which are themselves triggered by internal rules, goals, and motivations as well as by external events. The gestures of live agents can be captured by standard motion capture techniques, abstracted into PARs, and matched against lexical semantics of action concepts; we have applied this as well to aggregate actions with changing group memberships.

6 Summary

The LiveActor facility is under construction in the Center for Human Modeling and Simulation at the University of Pennsylvania. LiveActor is designed to develop the next stages of interactivity between real and virtual agents in close interpersonal engagements. It will lead us into strong couplings between language and animation, sensitize users to nuances of facial and gesture generation, allow their own gesture performances to be known to the virtual agents, and provide a unique application testbed for virtual training for interpersonal situations. The Parameterized Action Representation yields cognitive and behavioral models of general utility. Real-time capture of motion qualities from live gesture performance will inform virtual agents of the psychological and physiological state of the participant, and will allow a level of virtual agent responsiveness essential for VR interactions. The LiveActor software enables interactive training for military as well as other emergency personnel who must now be particularly adept and observant in the subtleties of human action, intent, and opportunity.

7 Acknowledgements

This work includes contributions from my colleagues Martha Palmer, Aravind Joshi, and Dimitris Metaxas, research staff Jan Allbeck, and students Charles Adams, Koji Ashida, Matt Beitler, Rama Bindiganavale, Aaron Bloomfield, MeeRan Byun, Bjoern Hartmann, Karin Kipper, Seung-Joo Lee, Sooha Lee, Ying Liu, William Schuler, Liwei Zhao, and Elena Zoubanova. Special thanks to Catherine Pelachaud and Eric Petajan for 3D face models, and Fred Azar for acting. The support of sponsors: NSF, ONR, NASA, USAF/AFRL, USC-ICT-ARO STRICOM, EDS, Ascension Technologies, and Alias-Wavefront is greatly appreciated.

References

[1] J. Allbeck and N. Badler. Embodied autonomous agents. In Kay Stanney, editor, *Handbook of Virtual Environments*, pages 313–332. Lawrence Erlbaum Associates, 2002.

[2] J. Allbeck, K. Kipper, C. Adams, W. Schuler, E. Zoubanova, N. Badler, M. Palmer, and A. Joshi. ACUMEN: Amplifying Control and Understanding of Multiple ENtities. In *Autonomous Agents and Multi-Agent Systems*, 2002.

[3] K. Ashida, S.-J. Lee, J. Allbeck, H. Sun, N. Badler, and D. Metaxas. Pedestrians: Creating agent behaviors through statistical analysis of observation

data. In *Proc. Computer Animation*. IEEE Computer Society, 2001.

[4] N. Badler and J. Allbeck. Towards behavioral consistency in animated agents. In N. Magnenat-Thalmann and D. Thalmann, editors, *Deformable Avatars*, pages 191–205. Kluwer Academic Publishers, New York, NY, 2001.

[5] N. Badler, J. Allbeck, L. Zhao, and M. Byun. Representing and parameterizing agent behaviors. In *Proc. Computer Animation*. IEEE Computer Society, 2002.

[6] N. Badler, R. Bindiganavale, J. Allbeck, W. Schuler, L. Zhao, and M. Palmer. A parameterized action representation for virtual human agents. In J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors, *Embodied Conversational Agents*. MIT Press, Cambridge, MA, 2000.

[7] N. Badler, C. Phillips, and B. Webber. *Simulating Humans: Computer Graphics Animation and Control*. Oxford University Press, New York, NY, 1993.

[8] DI-Guy, 1996. Boston Dynamics Inc., Cambridge, MA.

[9] R. Bindiganavale. *Building parameterized action representations from observation*. PhD thesis, CIS, University of Pennsylvania, 2000.

[10] R. Bindiganavale, W. Schuler, J. Allbeck, N. Badler, A. Joshi, and M. Palmer. Dynamically altering agent behaviors using natural language instructions. In *Autonomous Agents*, pages 293–300, New York, June 2000. ACM Press.

[11] M. Byun and N. Badler. FacEMOTE: Qualitative parametric modifiers for facial animations. In *Symposium on Computer Animation*, pages 65–71. ACM Press, July 2002.

[12] J. Cassell, C. Pelachaud, N. Badler, M. Steedman, B. Achorn, W. Becket, B. Douville, S. Prevost, and M. Stone. Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In *Computer Graphics, Annual Conf. Series*, pages 413–420. ACM, 1994.

[13] M. Cavazza, F. Charles, and S. Mead. AI-based animation for interactive storytelling. In *Proc. Computer Animation '01*, Seoul, Korea, 2001. IEEE Press.

[14] D. Chi, M. Costa, L. Zhao, and N. Badler. The EMOTE model for Effort and Shape. In *ACM SIGGRAPH Computer Graphics*, pages 173–182, New York, July 2000. ACM Press.

[15] W. Lamb. *Posture and Gesture: An Introduction to the Study of Physical Behavior*. Duckworth & Co., London, 1965.

[16] T. Noma, L. Zhao, and N. Badler. Design of a virtual human presenter. In *IEEE Joural of Computer Graphics and Applications*, volume 20(4), pages 79–85, 2000.

[17] A. Ortony, G.L. Clore, and A. Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, 1988.

[18] J. Rickel and W. L. Johnson. Task-oriented collaboration with embodied agents in virtual worlds. In J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors, *Embodied Conversational Agents*. MIT Press, Cambridge, MA, 2000.

[19] N. M. Sgouros, G. Papakonstantinou, and P. Tsanakas. A framework for plot control in interactive story systems. In *Proc. AAAI'96*, Portland, 1996. AAAI Press.

[20] B. Silverman, R. Might, R. Dubois, H. Shin, and M. Johns. Toward a human behavior modeling anthology for developing synthetic agents. In *10th Conf. On Computer Generated Forces and Behavioral Representation*. SISO, May 2001.

[21] W. Swartout, R. Hill, J. Gratch, W. L. Johnson, C. Kyriakakis, C. LaBore, R. Lindheim, S. Marsella, D. Miraglia, B. Moore, J. Morie, J. Rickel, M. Thiebaux, L. Tuch, R. Whitney, and J. Douglas. Toward the holodeck: Integrating graphics, sound, character and story. In *Proc. Autonomous Agents '01*, pages 409–416, Montreal, 2001.

[22] J.S. Wiggins. *The Five-Factor Model of Personality: Theoretical Perspectives*. The Guilford Press, New York, 1996.

[23] L. Zhao, S. Lu, N. Badler, and D. Metaxas. Structure and style: Motion acquisition from video. In *Third International Workshop on Intelligent Virtual Agents*, September 2001. Madrid, Spain.