Virtual Humans

William R. Swartout

USC Institute for Creative Technologies
13274 Fiji Way; Marina del Rey, CA 90292
swartout@ict.usc.edu

Extended Abstract

Imagine a simulated world where the characters you interact with are almost human – they converse with you in English, they understand the world they are in, can reason about what to do, and they exhibit emotions. Some of these characters may be your friends, while others will oppose you. Unlike current video games, being successful in this world won't just be a matter of who is quickest on the draw or most adroit at solving puzzles, instead it will be the person who understands the social fabric and cultural context and can use interpersonal skills most effectively. Such a simulation could open up whole new horizons for education, entertainment and simulation. And, given recent advances in AI and graphics, it may not be too far off.

Virtual humans are computer-generated characters that can take the part of humans in a variety of limited contexts. These can include acting as role-players in simulations and training systems (Johnson, Rickel & Lester 2000; Swartout et al. 2005; Traum et al. 2005; Johnson, Vilhjálmsson & Marsella 2005), where they play a variety of parts, such as acting as friendly or hostile forces, or locals in the environment. Other uses for virtual humans include acting as museum guides (Gustafson & Bell 2000), marketing assistants (Cassell, Bickmore et al. 2000) or characters in entertainment systems, where the advent of video games such as *The Sims 2* makes clear the growing interest of the computer game industry in virtual humans (see also (Mateas & Stern 2003)).

The behaviors of virtual humans are not scripted in advance, but instead they use artificial intelligence to understand what is going on in their simulated world and figure out how to respond. They form and modify their beliefs, plans and emotions, and act and communicate using both natural language and non-verbal gestures.

Over the last ten years, virtual human research has made substantial progress, advancing from rather stilted characters with limited interaction capabilities to ones that inter-

act robustly using a much more fluid repertoire for both verbal and non-verbal communication.

In this talk I will give an overview of virtual human research, focusing on the opportunities this area provides and the lessons learned and their larger implications for the field of artificial intelligence. Finally, I will conclude with some thoughts on future directions for virtual human research.

Because virtual humans are intended to mimic a broad range of human behaviors, they must integrate a diverse set of AI technologies, including speech recognition, natural language understanding and generation, dialogue management, non-verbal communication, perception, automated reasoning, and emotion modeling. While integrating such a broad range of technologies can be a daunting task, I believe it has several important benefits for AI research.

First, in our own work on virtual humans at the ICT, by focusing on the overall behavior of the virtual human and trying to make that behavior believable, we have uncovered research areas that have received relatively little attention up to now. One such area is emotion modeling. Arguably, the great preponderance of AI research over the last 50 years has not been concerned with emotions but rather with building systems that behave intelligently in their environment. Yet, we have found that when people see a virtual human, they not only expect it to behave rationally but also to exhibit appropriate emotions. In fact, people sometimes explain a character's behavior on emotional grounds whether or not the character actually uses emotion models. We thus realized that understanding how to build computer models of emotion was essential if we were to avoid unintentionally misleading behaviors from our virtual humans, which led to a major effort in emotion modeling (Gratch & Marsella 2004). Another area where the study of virtual humans has raised new issues is natural language processing. One issue is that the virtual humans are embedded in a social structure and that affects how language is used. Additionally, while most natural language systems assume a one-on-one interaction between a computer and a user, virtual humans may be involved in multiple conversations about multiple topics simultaneously. In this talk I will describe these new areas of re-

search and the approach that we and others have taken to address them.

A second benefit that emerges from the integration of various AI capabilities in the context of a virtual human is that it makes additional knowledge available that can resolve problems that are difficult to resolve in isolation. For example, we have found that some problems in natural language processing, such as resolving the linguistic focus of an ambiguous question can be done much more successfully by taking advantage of information in a character's emotional model. In the talk I will describe this and other synergies we have found that result from integration.

It could be argued that the integration problem posed by virtual humans is daunting. Indeed, if one were to try to build a virtual human that would function in the real world it might well be too difficult: too many possibilities and interactions to consider. But virtual humans are not constructed to function in the real world, instead they operate in a synthetic, simulated world that both people and the virtual humans inhabit. The story or scenario behind that simulation is critical because it establishes a strong context for the experience. This context is important because it limits the range of responses that a person in the simulation is likely to make, which in turn makes it easier to build a virtual human that will function robustly. For example, a person engaged in an emergency medical simulation is unlikely to ask a virtual human about recent baseball scores. In this talk I will elaborate on the relationship between story and virtual humans technology and describe how it can be used not only to limit context but also to deal with impasses in a simulated experience (Mateas & Stern 2003; Swartout, Gratch et al. 2005).

Finally, in this talk I will address some new areas in virtual human research, focusing on two areas in particular:

Perceptive Virtual Humans. Although some progress has been made (Morency, Sidner et al. 2005) current virtual humans have limited abilities to see the gestures, posture or facial expressions of a person interacting with them and limited ability to hear the prosody in his or her voice. This lack of perception makes it much more difficult for virtual humans to interact in a natural way. Gestures cannot be intermixed with speech and since the virtual human cannot see posture, facial expressions or sense the tone of a person's voice it is much more difficult for the virtual human to gauge a person's emotional state. Research directed toward developing the technology that will make virtual humans perceptive will open up new communication channels between people and virtual humans and could significantly improve interaction.

Theory of mind and social reasoning. Current virtual humans have only a limited ability to model their own beliefs, desires and intentions or those of others. This limitation curtails their ability to reason in social situations, negotiate and work as teams. Research in this area could

enhance a virtual human's ability to behave plausibly in the complex social environments opening up new applications for entertainment, training and education.

By focusing on the problem of mimicking human behavior and intelligence and bringing together a variety of AI research threads into an integrated system, virtual humans research is addressing some of the same problems that participants at the Dartmouth Conference described 50 years ago. Benefiting from far greater processing power, much better software architectures and half a century of progress in the various subareas of AI, it's exciting to think about what the next 50 years may bring.

References

Cassell, J., T. Bickmore, et al. 2000. Human conversation as a system framework: Designing embodied conversational agents. *Embodied Conversational Agents*. J. Cassell, J. Sullivan, S. Prevost and E. Churchill. Boston, MIT Press: 29-63.

Gratch, J., & Marsella, S. 2004. A domain independent framework for modeling emotion. Journal of Cognitive Systems Research.

Gustafson, J. & Bell, L., 2000. "Speech Technology on Trial: Experiences from the August System", Natural Language Engineering, volume 6.

Johnson, W.L., Vilhjálmsson, H. H., and Marsella, S. 2005, Serious games for language learning: How much game, how much AI?. In proceedings of AIED 2005.

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.

Mateas, M. and Stern, A. 2003. Facade: An Experiment in Building a Fully-Realized Interactive Drama. In Game Developer's Conference: Game Design Track, San Jose, California, March 2003.

Morency, L-P, Sidner, C., Lee, C.; Darrell, T. "Contextual recognition of Head gestures," International Conference on MultiModal Interacation (ICMI), Oct 2005

Swartout, W., Gratch, J., Hill, R., Hovy, E., Lindheim, R., Marsella, S., Rickel, J., and Traum, D. 2005. Simulation meets Hollywood: Integrating Graphics,Sound, Story and Character for Immersive Simulation, in *Multimodal Intelligent Information Presentation*, Oliviero Stock and Massimo Zancanaro eds. Kluwer.

Traum, D., Swartout, W., Marsella, S. and Gratch, J. 2005. Fight, Flight, or Negotiate: Believable Strategies for Conversing under Crisis presented at the 5th International Working Conference on Intelligent Virtual Agents