The More the Merrier: Multi-Party Negotiation with Virtual Humans

Patrick Kenny, Arno Hartholt, Jonathan Gratch, David Traum, Stacy Marsella, Bill Swartout

University of Southern California's Institute for Creative Technologies
13274 Fiji Way
Marina del Rey, CA 90292
United States of America
{kenny, hartholt, gratch, traum, marsella, swartout}@ict.usc.edu

Abstract

The goal of the Virtual Humans Project at the University of Southern California's Institute for Creative Technologies is to enrich virtual training environments with virtual humans – autonomous agents that support face-to-face interaction with trainees in a variety of roles – through bringing together many different areas of research including speech recognition, natural language understanding, dialogue management, cognitive modeling, emotion modeling, non-verbal behavior and speech and knowledge management. The demo at AAAI will focus on our work using virtual humans to train negotiation skills. Conference attendees will negotiate with a virtual human doctor and elder to try to move a clinic out of harm's way in single and multi-party negotiation scenarios using the latest iteration of our Virtual Humans framework. The user will use natural speech to talk to the embodied agents, who will respond in accordance with their internal task model and state. The characters will carry out a multi-party dialogue with verbal and non-verbal behavior. A video of a single-party version of the scenario was shown at AAAI-06. This new interactive demo introduces several new features, including multi-party negotiation, dynamically generated non-verbal behavior and a central ontology.

Introduction

Existing virtual worlds, such as military simulations and computer games, often incorporate virtual humans with varying degrees of intelligence. However, these characters' ability to interact with human users is usually very limited. Typically, users can shoot at them and they can shoot back. The social interactions are usually scripted and offer human users no ability to carry on a dialogue. In contrast, we envision virtual humans that co-habit virtual worlds with people and support face-to-face dialogues situated in those worlds, serving as guides, mentors and teammates.

The goal of the Virtual Humans Project at the University of Southern California's Institute of Creative Technologies is to create a state of the art, realistic and immersive, interactive training environment with rich virtual humans. Our latest scenario builds upon the technology used for the Mission Rehearsal Exercise (MRE) (Rickel at al., 2001) and Stability And Support Operations – Simulation and Training (SASO-ST) (Swartout et al., 2006). The latter has been extended to include two virtual humans: a Spanish doctor, head of a Non-Government Organization clinic, and an Iraqi village elder. Set in a small Iraqi town plagued by violence, the human trainee takes on the role of an Army captain with orders to move the clinic to a safer location. It is up to the user to negotiate this goal with both the doctor and the elder in order to reach agreement between all stakeholders. This scenario introduces various new features, including multi-party negotiation, dynamically generated non-verbal behavior and a central ontology.

Multi-Party Negotiation

The SASO - Extended Negotiation scenario features two virtual humans with rich cognitive, emotional and conversational skills that are fully capable of a wide range of verbal and non-verbal behaviors. Both agents pursue their own agenda and need both the human trainee as well as each other to reach their goals. Agents can share elements of their respective task model, but having different concerns and world views ensures both conflicts and dependencies between the characters. One of the doctor's demands for moving the clinic, for instance, is that the new facility will be renovated by the elder. The elder in turn demands money from the captain before he is willing to commit to this action, a necessity that catches the doctor unaware.

The Dialogue Manager has been extended to support full multi-party dialogue. Each agent has its own way of speaking and they react to both each other as well as the trainee equally in terms of turn taking and grounding.

In order to enforce consistency and increase productivity, the Virtual Humans Project uses a central ontology. The ontology serves as the lynchpin for the system and contains the task model of the agents as well as all the language assets used for natural language understanding.

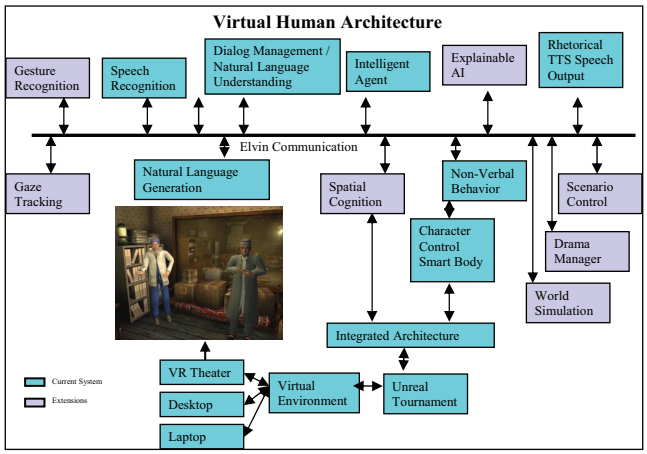
Demonstration

The AAAI-07 demonstration will consist of a single party and a multi-party scenario. The single-party scenario was shown as a video in a senior member talk at AAAI-06 (Swartout, W., 2006). The multi-party scenario is an extension of our single-party negotiation work (Swartout et al., 2006). The multi-party version of the demo will be a highly interactive engagement with a doctor and a village elder, further complicating the issues involved in moving the clinic as both have somewhat different goals. The job of the user will be to talk to both in order to work out a common solution. The demonstration will show off the underlying technology of the virtual humans.

General Architecture

The virtual human system is based on an architecture developed at the USC Institute for Creative Technologies (Gratch et al., 2002, Swartout et al., 2006). The architecture supports a wide range of virtual humans from simple question/answering to more complex ones that contain cognitive and emotional models. The system is a set of distributed components that communicate with message passing. The system works as follows: the trainee speaks into a speech recognition engine that converts this speech to text frames. These are then sent to the agent architecture that chooses a response based on its goals, emotional state and internal knowledge. The response is sent to a non-verbal behavior module that applies rules to create the appropriate gesture behaviors. These are then sent to a procedural animation system called SmartBody (Kallmann, 2005) that synchronizes verbal and non-verbal behavior along with the generated voice for the character. The major components in the system are:

- *Speech recognition*: digitizes the trainee's speech and produces a string of words as output.
- *Natural Language Understanding*: parses the word string produced by speech recognition and forms an internal semantic representation.
- *Intelligent Agent*: reasons about plans and generates actions. Complex agents are created using the Soar Cognitive architecture. The agents contain task models, a dialogue manager and a model of emotions.
- *Non-verbal behavior generation*: takes the response output string and applies a set of rules to select gestures, postures and gazes for the virtual characters.
- *SmartBody*: synchronizes speech output with gestures and other non-verbal behavior to perform character control over the characters in the virtual environment.
- *Unreal Tournament game engine*: is the underlying graphics engine used for the Virtual Environment.
- *Rhetorical*: performs speech synthesis from text generated by the agent. Or a set of pre-recorded voice strings can be used as speech.



Discussion / Future Work

The SASO System has been extensively tested with over 50 subjects of varying levels of skills and the results were positive. We plan to create more densely populated virtual worlds with high fidelity virtual characters that have more social and cultural knowledge and develop tools that allow authors to quickly create scenarios, knowledge and environments for training and education.

We wish to acknowledge Ed Fast, Andrew Marshall and the whole SASO team. This work was sponsored by the U.S. Army Research, Development, and Engineering Command, and the content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

References

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

Kallmann, M., Marsella, S. 2005. Hierarchical motion controllers for real-time autonomous virtual humans. *Proceedings of the 5th International working conference on Intelligent Virtual Agents (IVA),* 243–265, Kos, Greece.

Rickel, J., Gratch, J., Hill, R., Marsella, S., Swartout, W., 2001. Steve Goes to Bosnia: Towards a New Generation of Virtual Humans for Interactive Experiences. *In AAAI Spring Symposium on Artificial Intelligence and Interactive Entertainment*, Stanford University, CA

Swartout, W., Gratch, J., Hill, R., Hovy, E., Marsella, S., Rickel J., Traum, D., 2006. Toward Virtual Humans, *AI Magazine,* v.27(1)

Swartout, W. 2006, Virtual Humans, *Twenty-First National Conference on Artificial Intelligence (AAAI-06). (Senior paper)* Boston, MA