University of Pennsylvania
ScholarlyCommons

Center for Human Modeling and Simulation

Department of Computer & Information Science

October 1994

Automatically Generating Conversational Behaviors in Animated Agents

Justine Cassell
University of Pennsylvania

Catherine Pelachaud
University of Pennsylvania

Norman I. Badler
University of Pennsylvania, badler@seas.upenn.edu

Mark Steedman
University of Pennsylvania, steedman@seas.upenn.edu

Follow this and additional works at: http://repository.upenn.edu/hms

Automatically Generating Conversational Behaviors in Animated Agents

Abstract

In the creation of synthetic computer characters, the creators shouldn't have to create or control every move of their life like human agents: for example, during the progress of a search or planning system, responding to knowledge based queries, or portraying autonomous agents during real-time virtual environment simulations. For these *automated* characters we must *generate* behavior on the basis of rules abstracted from the study of human behavior.

Comments

Copyright Massachusetts Institute of Technology. Presented at *Lifelike Computer Characters*, October 1994.

Automatically Generating Conversational Behaviors in Animated Agents

Justine Cassell   Catherine Pelachaud   Norman Badler   Mark Steedman

University of Pennsylvania*†

November 9, 1994

In the creation of synthetic computer characters, the creators shouldn't have to create or control every move of their lifelike human agents: for example, during the progress of a search or planning system, responding to knowledge base queries, or portraying autonomous agents during real-time virtual environment simulations. For these *automated* characters we must *generate* behavior on the basis of rules abstracted from the study of human behavior.

The behavior that we concentrate on in this project is conversation (that is, an interactive dialogue between two agents). Conversation includes spoken language (words and contextually appropriate intonation marking topic and focus), but it also includes facial movements (lip shapes, expressions, gaze, head movement), and hand gestures (points, beats, and movements representing the topic of accompanying speech). Without all of these verbal and non-verbal behaviors, one cannot have realistic, lifelike, autonomous agents. To this end, our system *automatically* animates conversations between multiple human-like agents with appropriate and synchronized speech, intonation, facial expressions, and hand gestures.

The system is composed of a *Dialogue Generation* program which allows gesture and conversational intonation to be generated along with speech. The output of the dialogue generation program is speech annotated with descriptions of appropriate intonation and gesture, which are then sent on to an intonation synthesis module, facial expression specification module, and gesture and facial synthesis modules. The *Intonation Synthesis* model generates actual intonational tunes as a function of the information structure of the discourse. The *Facial Expression Specification* module generates head and eye movements as a function of dialogic categories such as `planning what to say, feedback to speaker's contribution`. The *Gesture and Facial Movement Synthesis* module has two parts;

1 - A *Synchronization module*: Interaction between agents and synchronization of gaze, hand and head movements to the dialogue for each agent are accomplished using Parallel Transition Networks (PaT-Nets), which allow coordination rules to be encoded as simultaneously executing finite state automata.

2- A *Movement Specification module*: It selects and generates nods, gaze direction, handshapes, wrist and arm motion.

The conversation below is an example of the discourse output from the dialogue generation program. Following it is a description of some of the nonverbal and intonational behaviors generated with the speech.

The dialogue is unnaturally repetitive and explicit in its goals because the dialogue generation program that produced it has none of the conversational inferences that allow humans to follow leaps of reasoning.

Gilbert:	Do you have a blank check?
George:	Yes, I have a blank check.
Gilbert:	Do you have an account for the check?
George:	Yes, I have an account for the check.
Gilbert:	Does the account contain at least fifty dollars?
George:	Yes, the account contains eighty dollars.
Gilbert:	Get the check made out to you for fifty dollars and then I can withdraw fifty dollars for you.
George:	All right, let's get the check made out to me for fifty dollars.

When Gilbert asks a question, his voice rises. When George replies to a question, his voice falls. When Gilbert asks George whether he has a blank check, he stresses the word "check". When he asks George whether he has an account for the check, he stresses the word "account".

Every time Gilbert replies affirmatively ("yes"), or turns the floor over to Gilbert, he nods his head, and raises his eyebrows. George and Gilbert look at each other when Gilbert asks a question, but at the end of each question, Gilbert looks up slightly. During the brief pause at the end of affirmative statements the speaker blinks.

In saying the word "check", Gilbert sketches the outlines of a check in the air between him and his listener. In saying "account", Gilbert forms a kind of box in front of him with his hands: a metaphorical representation of a bank account in which one keeps money. When he says the phrase "withdraw fifty dollars," Gilbert draws his hand towards his chest.

Although the two agents do not visually *perceive* each other's gestures, speech, etc., their actions are nevertheless determined by the evolving conversation. The sequence, and hence their motions, is not pre-determined. Thus if new information becomes available, then all the communication acts will adjust – and be animated – accordingly. It is this expressive flexibility and response to novel situations that make these automated characters lifelike.

References

[1] J. Cassell, C. Pelachaud, N.I. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost, C. Seah, and M. Stone. Animated conversation: Rule based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In *SIGGRAPH '94*, 1994.