UTEP's AGENT Architecture

Iván Gris, David Novick, Diego A. Rivera, Mario Gutiérrez

The University of Texas at El Paso

1 Main Research Themes

Two related goals of IVA research involve increasing the believability and perceived trustworthiness of agents and increasing the user's sense of engagement. In our research, we use human-human interaction as a model for agent behaviors that build rapport between humans and agents. Our paralinguistic model of rapport (Novick & Gris, in press) comprises a sense of emotional connection, a sense of mutual understanding, and a sense of physical connection. Our research, which focuses primarily on the sense of physical connection, seeks to increase the naturalness of non-verbal interaction to correspondingly increase human-IVA rapport (Tickle-Degnan & Rosenthal, 1987; Huang et al., 2011). For these studies, we have developed two agents, with variable non-verbal behaviors, in immersive games designed to maintain users' engagement.

1.1 Familiarity Agent

This agent is part of a longer-term project to provide IVAs with behaviors that enable them to build and maintain rapport with their human partners. We focus on paralinguistic behaviors, especially nonverbal behaviors, and their role in communicating rapport. The Our IVA guides its players through a speech-controlled game called "Escape from the Castle of the Vampire King (see Figure 1), through which we measure the familiarity between humans and agents across two interaction sessions. We studied whether increasing amplitude of nonverbal paralinguistic behaviors leads to an increased perception of physical connectedness between humans and ECAs (Novick & Gris, 2013; Gris, Novick, Gutierrez & Rivera, in press).

1.2 Two-Way Virtual Rapport Agent

This agent is a work in progress in which we make use of full-body gesture recognition to create a sense of physical connection between users and agents. To achieve this, we guide users through a jungle-island survival scenario (see Figure 2). We focus on collaborative physical activities that take place in the virtual environment but require the user's physical action, such as lighting a fire or spearing a fish.

Figure 1. Interaction with the "Escape from the Caastle of the Vampire King" agent.

Figure 2. Jungle survival agent.

2 Current Architectures and Standards

We developed our agents independently of SAIBA, BML, and FML, in part because we wanted to understand agent development from the ground up and in part because our research requires some less-common features, such as recognition of full body gestures, blending animations, and agent portability and reuse.

Our implementation uses Unity 4, a Microsoft Kinect, and the Windows Speech SDK, interfaced and networked with each other and synchronized to handle the agent's complex behavior. Figure 3 presents a high-level outline of the system's components and their relationships. We use the Unity 4 game engine to display our agents and Unity's Mecanim system to create an extensive array of animations.

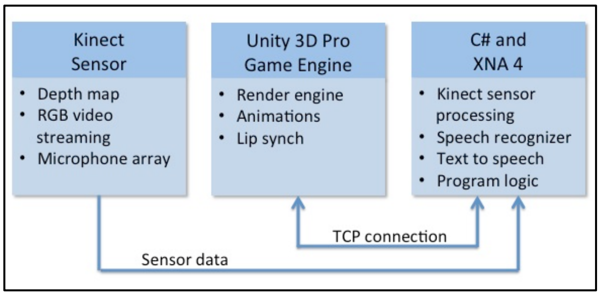


Figure 3. High-level architecture of the software-hardware interfaces and feature handling.

Animations are played by a state graph that follows user-specified parameters of when an animation should start, end, or blend with another animation. Multiple animations can be blended to obtain a completely different animation in real time and give the user the impression that the agent never moves in exactly the same way twice. Animations are divided layers that can control different parts of the body, so multiple animations can be played at the same time and affect different limbs of the agent, for example playing a blinking and talking animation on the face only, an explaining animation on the arms, and a walking animation from the hips down. The animations are played when the system decodes a message sent by the dialog tree that has the information about the specific animation to be played, the length of the dialog that the agent will say, and the position where the agent and the player should be.

To describe interactive scenes, we developed a dialog interpreter that parses an XML document and links dialog states through conditionals. The file includes the responses anticipated by the systems at relevant parts of the scene. After the interpreter compiles the file, it builds a dialog tree that contains the relationships of the dialogs segments through the storyline.

3 Future of Architectures and Standards for IVAs

Current generations of users are accustomed to hyper-realistic videogame characters and movie assets. A key aspect of these advanced-gen characters is the fluid and realistic way in which they move, which is usually handmade through motion capture. IVAs are a step behind, as their movement is generated on the fly. We would like to see a standard that enables research groups to exchange animation sets or subsets (animation of specific joints) that can be blended at a later time to create unlimited specialized movements.'

Another key feature of IVAs is the real-time perception of a user's full-body gestures. For most IVAs, this ability is limited, partly because the necessary setup is more elaborate, requiring more than a computer and a screen, and partly because the field lacks standards for gesture recognition. Now that it is possible to track skeleton joint movement, gesture recognition can be more than a set of image-analysis algorithms with limited tracking capabilities, complicated setups, and long render times. We hope that a standard can be set to capture gestures or poses based on skeleton data, so that studied gestures can easily be ported into the recognizers and applied across agents.

More broadly, we see a need to leverage standards and shared tools through more effective organization of the research community. Newcomers to the field would be well served by having a single place from which to obtain standards and tools rather than having to visit our research groups' individual sites.

4 Suggestions for Discussion

Our approach for IVA design is modular. Our agents recognize full-body gesture recognition and produce nonverbal behaviors using multiple technologies that we later synchronize through a network. We are interested in how to develop these modular technologies, how can we share them across research groups, and how we can address difficult interaction problems with them. Specific questions include:

- How can shared resources (e.g., BML, SABIA) support extension to novel modes of interaction?
- Given the popularity and accessibility of the latest gaming technologies, how can we make use of them to develop quick, inexpensive and portable IVA prototypes?
- With the increased IVA functionality and interactive capabilities, what requirements and standards should we take into consideration when developing agents that interact not only with humans but with other agents at the same time?
- Can we define a clear separation of IVAs' architectural components between their domain and the agent's features (e.g., gesture recognition, speech recognition, virtual environments, gesture and pose handling, AI elements)? How can we use this to create IVAs, and specific features that can be shared across research groups and disciplines?

References

Gris, I. Adaptive virtual rapport for embodied conversational agents. In *Proceedings of the 15th ACM on International conference on multimodal interaction*. ACM, 2013.

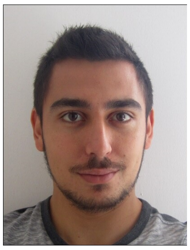
Huang, L., Morency, L.-P., and Gratch, J. Virtual rapport 2.0. In Intelligent Virtual Agents, *Lecture Notes in Computer Science*, Vol. 6895, chapter 8, pages 68-79. Springer, Berlin, Heidelberg, 2011.

Novick, D., and Gris, I. Building rapport between human and ECA: a pilot study. In *Proceedings of HCI International 2014*, in press.

Gris I., Novick, D., Rivera D.A., Gutierrez, M. Recorded speech, virtual environments, and the effectiveness of embodied conversational agents. In *Proceedings of Intelligent Virtual Agents, 2014*, in press.

Tickle-Degnen, L., and Rosenthal, R. Group rapport and nonverbal behavior. In *Group processes and intergroup relations*, pages 113-136. Sage, Newbury Park, CA, 1987.

Biographical Sketches



Ivan Gris is a Ph.D. student at the University of Texas at El Paso working under the supervision of Dr. David Novick. He works in the Interactive Systems Group as part of the Advanced aGent ENgagement Team as the project manager for developing full body embodied conversational agents and immersive, interactive environments. When he is not working on his dissertation, he is working on the research and development for two start-up companies he created. One of these companies is a tech start-up that uses embodied conversational agents, animatronics, scene development, and visual and special effects to develop a unique role-playing themed experience.

Diego A. Rivera is an undergraduate student majoring in Computer Science. He works with the Advanced aGent ENgagement Team as a lead animator. He works integrating and developing the systems that handle the animations for the agents using Unity 3D. In addition he is the main programmer of the dialog interpreter and several other tools and pipelines that control the agent's behavior. Recently Diego obtained an internship on the Institute for Creative Technologies at the University of South-California where he is working under the supervision of Dr. Chad Lane as an integration programmer for Virtual Human and Intelligent Tutoring Research.

Mario Gutierrez is a Master's student at the University of Texas at El Paso. Mario is the lead programmer for the Advanced aGent Engagement Team. He is currently working on his Master's project, which consists of implementing a memory knowledge base for an ECA. He aims to develop agents with memory of previous interactions and dialogs with the users, and then recall them at the appropriate time and in a correct maner.

David Novick is the Mike Loya Distinguished Chair and Professor of Computer Science at the University of Texas at El Paso. He earned his J.D. at Harvard University in 1977 and his Ph.D. in Computer and Information Science at the University of Oregon in 1988. His research has included work on turn-taking and gaze in artificial agents. He built UTEP's Advanced aGent ENgagement Team, whose research focuses on interactive systems and, especially, building rapport in multimodal conversation.