Full-Body Gesture Interaction with Improvisational Narrative Agents

Andreya Piplica, Chris DeLeon, and Brian Magerko

Digital Media, Georgia Institute of Technology
apiplica@gmail.com, chris@deleonic.com, magerko@gatech.edu

Abstract. Interactive narrative research strives to allow humans and intelligent agents to *co-create* narratives in real-time as equal contributors. While intelligent agents can interact with humans through expressive embodied representations, typical interactive narrative interfaces provide humans no way to reciprocate with embodied communication of their own. This disparity in interaction capabilities has been informally discussed as the *human puppet problem*. Improvisational theatre (improv) provides a real-world analogue to embodied co-creative interactive narrative experiences. This paper presents an overview of a system for combining improvisational acting with full-body motions to support human-AI co-creation of interactive narratives. The human begins an improvised narrative with an AI improviser while an intelligent avatar mediates interaction and gives the human an embodied presence in the scene.

Keywords: improvisational or dramatic interaction, conversational and storytelling agents, postures and gestures

1 Overview

This paper presents a system for combining improv acting with full-body motions to support the co-creation of interactive narratives with an AI agent and a human interactor. Our system translates a human interactor's movements into actions that an AI improviser can respond to within the context of an improvised story. We have constructed a framework for human interaction with an AI improviser for beginning an improvised narrative with interaction mediated through an intelligent user avatar [1]. This approach uses human gestural input to contribute part, but not all, of an intelligent avatar's behavior. While joint human-AI agents have been employed in interactive narrative [2, 3], this is the first system to do so with full-body gestures. Our framework is derived from the interactions involved in the improv games *Three Line Scene*, which focuses actors in establishing the initial elements of a story within three lines (e.g. what characters are in the scene, where they are, and what they are doing together), and *Moving Bodies*, which allows audience members to control the bodies of human improvisers while the improvisers interpret those inputs and perform discourse acts.

We have built a system based on our socio-cognitive studies of improvisational actors [4] where a human interactor and an AI can improvise a pantomimed three-line scene. An intelligent avatar uses motion data from a Microsoft Kinect sensor to represent the interactor's motions in the same virtual space as the AI improviser. This gives the human interactor an embodied presence in the scene and shows how the Kinect senses their motion. This feedback can help the user understand the avatar's interpretations and adjust their movements to accommodate the sensor's limitations. The avatar reasons about the human's intentions for the scene and creates its own mental model. This mental model is used to inform the avatar about potential discourse acts to select. While the intelligent avatar, inspired by *Moving Bodies*, does provide a means of joint AI / human control of a character, our current implementation omits dialogue in favor of specifically studying gestural interaction.

The user stands before a large screen – like a television or projector screen – where the AI improviser and the intelligent avatar are displayed on a stage. The user faces the screen while standing approximately four to ten feet away from a Microsoft Kinect below the screen (i.e. within the Kinect's sensor range). The intelligent avatar and the AI improviser are shown as two-dimensional animated characters on a virtual stage. These simplified visual characters map 3D motion data from the Kinect directly onto the characters' 2D animations. The system does not currently support animated facial expressions, though such animations may be supported in future iterations.

The user performs a motion to begin a three-line scene, such as putting one fist on top of the other and moving their hands from side to side. The Kinect sends the sensed motion data to the intelligent avatar and the AI improviser, who each interpret the motion as an action [1]. The avatar displays the user's motion while it reasons about what character and joint activity the user may be portraying, which in turn can inform select of discourse acts. The AI improviser then reasons about the user's character and joint activity, as well as its own character. The AI improviser and the intelligent avatar draw on the same reasoning processes to understand how the human interactor contributes to the scene. They utilize background knowledge about a specific domain to make inferences about the platform. They incorporate the human's motions into their reasoning based on joint position data from the Microsoft Kinect sensor. Additionally, the AI improviser reasons about how to contribute presentations to the scene with its own motions.

The inputted motion data from the Microsoft Kinect consists of 3D coordinates for joint and limb positions, which we convert to 2D and use to animate the on-screen characters. The 2D coordinates are evaluated as "signals," which are hand-authored sets of joint angles and positions. Signals can be *simple* (joint angles and positions at one time) or *temporal* (joint angles and positions varying across time). A neutral stance is defined as standing still with feet together and hands at one's sides. Actions are defined as sets of positive and negative signals. If the Kinect data satisfies all positive signals for an action, it becomes a candidate. However, if the data triggers any negative signals for that action, it is removed as a candidate. When the user's turn ends, the intelligent agent and AI improviser select an action from the candidate interpretations based on their mental model of the scene [1]. After the AI improviser interprets the user's motion and reasons about how it contributes to the scene, the AI improviser must select an action to present and a motion to present it with based on its own perception of the ambiguous knowledge communicated in the scene.

2 References

1. O'Neill, B., Piplica, A., Fuller, D., and Magerko, B. A Knowledge-Based Framework for the Collaborative Improvisation of Scene Introductions. In the Proceedings of the 4th International Conference on Interactive Digital Storytelling, Vancouver, Canada (2011)
2. Hayes-Roth, B., Sincoff, E., Brownston, L., Huard, R., Lent, B.: Directed Improvisation. Technical Report KSL-94-61. Stanford University, Palo Alto, CA (1994)
3. Brisson, A., Dias, J., Paiva, A.: From Chinese Shadows to Interactive Shadows: Building a storytelling application with autonomous shadows. In: Proc. Workshop on Agent-Based Systems for Human Learning and Entertainment (ABSHLE), AAMAS 2007. ACM Press, New York, NY, USA (2007)
4. Magerko, B., Manzoul, W., Riedl, M., Baumer, A., Fuller, D., Luther, K., and Pearce, C. An Empirical Study of Cognition and Theatrical Improvisation. In the Proceedings of ACM Conference on Creativity and Cognition, Berkeley, CA (2009)