Multimodal Interaction with a Virtual Character in Interactive Storytelling

(Extended Abstract)

Nikolaus Bee, Johannes Wagner
and Elisabeth André
Augsburg University
Institute of Computer Science
86135 Augsburg, Germany
{bee, wagner,
andre}@informatik.uni-augsburg.de

Fred Charles, David Pizzi
and Marc Cavazza
University of Teesside, School of Computing
Middlesbrough TS1 3BA, United Kingdom
{f.charles, d.pizzi,
m.o.cavazza}@tees.ac.uk

ABSTRACT

A number of interactive storytelling (IS) systems offer the user the possibility to input natural language input which determines how a story progresses. We introduce a framework for real-time signal processing (SSI) to analyze the users' state which then can influence the feelings of the story characters and their actions in the story. SSI is not only used to enable more natural character responses that are sensitive to the user's state at runtime. In addition SSI offers the possibility to collect a large variety of synchronized user data which can be used to analyze the user's experience in offline mode. The underlying narrative in which the approach was tested is based on a classical XIX$^{\text{th}}$ century psychological novel: Madame Bovary, by Flaubert.

Categories and Subject Descriptors

H.1 [**User/Machine Systems**]: Human factors

General Terms

Measurement, Experimentation, Human Factors

Keywords

eye gaze, interactive storytelling, virtual agent

1. ANALYSIS OF CONVERSATIONAL AND SOCIAL BEHAVIORS

A frequent metaphor for interactive storytelling is that of the Holodeck [4, 8], the science-fiction ultimate entertainment system, where narratives take the form of virtual reality world in which the user is totally immersed, interacting with other characters and the environment in a way which drives the evolution of the narrative. As a character in the narrative, the user communicates with virtual characters much like an actor communicates with other actors. This requirement introduces a novel context for multimodal communication as well as several technical challenges. Acting involves attitudes and body gestures that are highly significant for both dramatic presentation and communication. Apart from our earlier work [5] where we developed a story character that responds to the user's emotive tone, there is, however, hardly any conversational interface to interactive storytelling that emphasizes the socio-emotive aspects of interaction and integrates sophisticated technologies to recognize the user's emotive state. Furthermore, hardly any attempt has been made to study the role of eye gaze in interactive storytelling.

Unlike earlier systems [6], our focus is not on the analysis of the semantics, but on the socio-emotional aspects of such a conversation. By expressing feelings and showing interest, the user may influence the story. To analyze the user's affective and attentive behaviors when interacting with Emma Bovary, we use the SSI tool [9], short for Smart Sensor Integration, a framework for multimodal signal processing in real-time. SSI suits our needs in two ways. Firstly, it offers the possibility to collect synchronized sensor data from users interacting with the system. In order to get realistic data we still simulate the system response at this point. Afterwards we analyze the recordings and based on the observations we use SSI to implement a pipeline that extracts the observed user behavior in real-time. This information is then used to automate the system response. The information provided by SSI ranges from raw sensor data, such as eye coordinates or skin conductivity level, over low level features, such as voice pitch or heart rate, to high level description, such as the level of interest or emotional states.

To gather different kind of information from the user, we decided to use a various set of sensors, namely video camera, eye-tracker, microphone and two physiological sensors skin conductance (SC) and blood volume pulse (BVP). From each sensor type we extract different cues, which are assembled to derive the user state.

2. VIRTUAL CHARACTER

Emma Bovary, a virtual character created at University of Teesside, is able to show a huge variety of facial expression implementing the FACS. Emma (see Fig. 1), was enhanced to use the Facial Action Coding System (FACS) to synthesize a huge set of different facial expressions. The ac-

Cite as: Multimodal Interaction with a Virtual Character in Interactive Storytelling (Extended Abstract), N. Bee, J. Wagner, E. André, F. Charles, D. Pizzi and M. Cavazza, *Proc. of 9th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, van der Hoek, Kaminka, Lespérance, Luck and Sen (eds.), May, 10–14, 2010, Toronto, Canada, pp. 1535-1536

tion units were designed using morph targets. The system includes a tool to control the single action units [3]. The system interfaces the Microsoft Speech API to synchronize the audio output with the lip movements. This allows us to use any text-to-speech that supports SAPI 5. As the quality of common TTS systems may not be satisfactory, we integrated a module to synchronize prerecorded audio speech files with the lip movements of the virtual character. This allows us to use highly emotional sentences or affect bursts to be spoken through a virtual character.

Our gaze model was extended with further parameters as our virtual agent is capable to react to the user's current gaze using an eye tracker. The maximal and minimal duration of mutual gaze can now be set as well. Furthermore, we may indicate the maximal duration the virtual agent gazes around. Finally, we may specify how long the virtual agent waits until the user responds with mutual gaze [2]. We modeled three different gaze modes for our agent. In the *non-interactive normal* mode, the character looks for about 2 s (between 1 and 3 s) at the user before she averts her gaze again. The agent's gaze model in the *interactive* mode is parameterized as in the non-interactive condition, but the agent notices whether the user is looking at her or not and responds accordingly, for example, tries to establish mutual gaze or looks away when the user starts staring. In the *non-interactive staring* mode the agent's gaze model is parameterized in such a way that the agent seems to stare at the user.

System for Interaction. We set up SSI to provide the Horde3D GameEngine [1] with the eye gaze of the users to detect where the user is looking at in the dynamic 3D scene. In this vein, we are able to detect whether the user looks at the virtual agent, the left eye or the right eye or something else in the virtual scene. Further, it allows us to detect the focused object in the 3D world in real-time. This was necessary for the eye gaze based interaction on a level of mutual gaze and to see if the user is looking at the virtual character's eyes, face or away.

Setting. The user is placed in front of a table on which the eye tracker was placed. The eye tracker with an incline of 23° is placed 80 cm above ground and 140 cm away from the projection surface. The user is seated 60 - 80 cm in front of the eye tracker. In total the user is about 2 m away from the virtual agent, which is within the *social space* according to [7]. The projection surface sizes 120 × 90 cm, which displays the virtual agent in life-size (see Fig. 1). To offer an enriched scene where the user has the choice to look away from the virtual agent, Emma was placed in the dining room of her house, which includes chairs and tables (see Fig. 1).

3. CONCLUSION

In this paper, we presented a framework for real-time signal processing that has been integrated into an existing storytelling system to enable richer interactions with the characters. Unlike earlier work, our focus is not on the analysis of the semantics of user utterances. Rather, we are interested in interactions between humans and characters that were mainly driven by the user's emotive and attentive state. Even though we recorded a large variety of signals of users (eye gaze, mimics, bio signals, speech) interacting with Emma using SSI, our current interest focused on eye gaze. The analysis of the remaining channels will be topic of our future research.

Figure 1: Set-up for the interaction with Emma.

Acknowledgments

This work has been funded in part by the European Commission under the grant agreement IRIS (FP7-ICT-231824).

4. REFERENCES

[1] Augsburg University. Horde3D GameEngine. `http://mm-werkstatt.informatik.uni-augsburg.de/projects/GameEngine/`.

[2] N. Bee, E. André, and S. Tober. Breaking the ice in human-agent communication: Eye-gaze based initiation of contact with an embodied conversational agent. In *9th International Conference on Intelligent Virtual Agents (IVA)*, pages 229–242. Springer, 2009.

[3] N. Bee, B. Falk, and E. André. Simplified facial animation control utilizing novel input devices: A comparative study. In *International Conference on Intelligent User Interfaces (IUI '09)*, pages 197–206, 2009.

[4] M. Cavazza, J. L. Lugrin, D. Pizzi, and F. Charles. Madame bovary on the holodeck: immersive interactive storytelling. In *MULTIMEDIA '07: Proc. of the 15th international conference on Multimedia*, pages 651–660. ACM, 2007.

[5] M. Cavazza, D. Pizzi, F. Charles, T. Vogt, and E. André. Emotional input for character-based interactive storytelling. In *AAMAS '09: Proc. of The 8th International Conference on Autonomous Agents and Multiagent Systems*, pages 313–320, 2009.

[6] S. Dow, M. Mehta, E. Harmon, B. MacIntyre, and M. Mateas. Presence and engagement in an interactive drama. In *CHI '07: Proc. of the SIGCHI conference on Human factors in computing systems*, pages 1475–1484, New York, NY, USA, 2007. ACM.

[7] E. T. Hall. A system for notation of proxemic behavior. *American Anthropologist*, 65:1003–1026, 1963.

[8] J. H. Murray. *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. The MIT Press, 1998.

[9] J. Wagner, E. André, and F. Jung. Smart sensor integration: A framework for multimodal emotion recognition in real-time. In *Affective Computing and Intelligent Interaction (ACII 2009)*, 2009.