

Automatic Generation of Conversational Behavior for Multiple Embodied Virtual Characters: The Rules and Models behind Our System

Werner Breitfuss¹, Helmut Prendinger², and Mitsuru Ishizuka¹

¹ Graduate School of Information Science and Technology, University of Tokyo
7-3-1 Hongo, Bunkyo-ku
Tokyo, Japan
werner@mi.ci.i.u-tokyo.ac.jp

² National Institute of Informatics 2-1-2 Hitotsubashi, Chiyoda-ku Tokyo, Japan
helmut@nii.ac.jp

Abstract. In this paper we presented the rules and algorithms we use to automatically generate non-verbal behavior like gestures and gaze for two embodied virtual agents. They allow us to transform a dialogue in text format into an agent behavior script enriched by eye gaze and conversational gesture behavior. The agents' gaze behavior is informed by theories of human face-to-face gaze behavior. Gestures are generated based on the analysis of linguistic and contextual information of the input text. Since all behaviors are generated automatically, our system offers content creators a convenient method to compose multimodal presentations, a task that would otherwise be very cumbersome and time consuming.

1 Introduction and Motivation

Combining synthetic speech and human-like conversational behavior like gaze and gestures for virtual characters is a challenging and tedious task for human animators. As virtual characters are used in an increasing number of applications, such as computer games, online chats or virtual worlds like Second Life, the need for automatic behavior generation becomes more pressing. Thus, there have been some attempts to generate non-verbal behavior for embodied agents automatically by some researchers ([3],[4]). A drawback of most current systems and tools, however, is that they consider only one agent, a salient feature of our system is that we generate the behavior not only for the speaker agent, but also for the listener agent, who might use backchannel behavior in response to the speaker agent. A previous version of this system has been presented in [1] in this paper we focus mainly on novel the rules that are used to plan and generate the behavior and are were refined after conducting a study which was described in [2].

2 Behavior Generation

Behavior generation in our system operates on the utterance level, for which rules are defined. The input we use thus consists of a text line spoken by one of our two agents.

Based on contextual and linguistic information of the text, the behavior for the speaking and the listening agent is suggested. (For further information on the input, see [1]). Three levels of rules are applied. The first set of rules traverses a tree like structure and suggests certain behavior based on the nodes in the tree. In the next level, we identify words and phrases that might be connected to a specific gesture, using WordNet. In the third level we iterate and recheck the suggested gestures and gaze patterns and align them to each other. One advantage of this approach is that the rule sets can be easily extended, and other non-verbal and verbal behavior could be included easily, such as emotion expression and verbal back channeling.

<pre>GAZE FOR each THEMA node in the tree IF at the beginning of the utterance Or 85% of the time Look at speaker FOR each RHEMA node in the tree IF at the end of the utterance Or 95% of the time Look at speaker</pre>	<pre>GESTURE FOR each OBJECT node in the tree IF contains NEW node Or 70% of the time Single beat SPECIFIC IF word == (big, huge, wide) IF word form == (superlative) set p = 2 Show vertical size (p)</pre>
---	--

Fig. 1. Different Generation Rules

After this iteration, which is the last step of our behavior module, the generated behavior trees are forwarded to our output module that actually produces the final script in form of an MPML3D File ([5]). that can be displayed on the local computer or in the Virtual World of Second Life

References

1. Breitfuss, W., Prendinger, H., Ishizuka, M.: Automated Generation of Non-verbal Behavior for Virtual Embodied Characters. In: Proc. of the Int'l Conf. on Multimodal Interfaces (ICMI 2007), pp. 319–322. ACM Press, New York (2007)
2. Breitfuss, W., Prendinger, H., Ishizuka, M.: Automatic Generation of Gaze and Gestures for Dialogues between Embodied Conversational Agents: System Description and Study on Gaze Behavior. In: Proc. of the AISB 2008 Symposium on Multimodal Output Generation (MOG 2008), pp. 18–25 (2008)
3. Cassell, J., Vilhjálmsón, H., Bickmore, T.: BEAT: the Behavior Expression Animation Toolkit. In: Proceedings of SIGGRAPH 2001, pp. 477–486 (2001)
4. Kipp, M.: Creativity meets automation: Combining nonverbal action authoring with rules and machine learning. In: Gratch, J., Young, M., Aylett, R.S., Ballin, D., Olivier, P. (eds.) IVA 2006. LNCS (LNAI), vol. 4133, pp. 230–242. Springer, Heidelberg (2006)
5. Nischt, M., Prendinger, H., André, E., Ishizuka, M.: MPML3D: a reactive framework for the Multimodal Presentation Markup Language. In: Gratch, J., Young, M., Aylett, R.S., Ballin, D., Olivier, P. (eds.) IVA 2006. LNCS (LNAI), vol. 4133, pp. 218–229. Springer, Heidelberg (2006)