Social Perception and Steering for Online Avatars

Claudio Pedica[1,2] and Hannes Vilhjálmsson[1]

[1] Center for Analisys and Design of Intelligent Agents, Reykjavik University, Iceland
[2] Camerino University, Italy

Abstract. This paper presents work on a new platform for producing realistic group conversation dynamics in shared virtual environments. An avatar, representing users, should perceive the surrounding social environment just as humans would, and use the perceptual information for driving low level reactive behaviors. Unconscious reactions serve as evidence of life, and can also signal social availability and spatial awareness to others. These behaviors get lost when avatar locomotion requires explicit user control. For automating such behaviors we propose a steering layer in the avatars that manages a set of prioritized behaviors executed at different frequencies, which can be activated or deactivated and combined together. This approach gives us enough flexibility to model the group dynamics of social interactions as a set of social norms that activate relevant steering behaviors. A basic set of behaviors is described for conversations, some of which generate a social force field that makes the formation of conversation groups fluidly adapt to external and internal noise, through avatar repositioning and reorientations. The resulting social group behavior appears relatively robust, but perhaps more importantly, it starts to bring a new sense of relevance and continuity to the virtual bodies that often get separated from the ongoing conversation in the chat window.

1 Introduction

Massively Multiplayer Online Role Playing Games (MMORPGs) are a rapidly growing form of mass entertainment delivered over the Internet in the form of live game worlds that persist and evolve over time. Players connect to these worlds using client software that renders the world from their perspective as they move about and meet fellow players. The community is the cornerstone of these games. Therefore, any effort spent on supporting communication and social interaction between players has to be considered valuable for the application. When games wish to use avatars to represent players in environments where they can virtually meet face-to-face, all the animated behaviors that normally support and exhibit social interaction become important. Since players cannot be tasked with micro-management of behavior, the avatars themselves have to exhibit a certain level of social intelligence [1] [2]. The purpose of this avatar AI is in fact twofold: to give players helpful cues about the social situation and to ensure that the whole scene appears believable and consistent with the level of game world realism.

This paper presents ongoing work, which is a collaborative effort between a major MMORPG developer and a research center in the field of embodied conversational agents that specializes in multi-modal behavior generation based on social and cognitive models. The work presented here is one piece of the project which deals with modeling

H. Prendinger, J. Lester, and M. Ishizuka (Eds.): IVA 2008, LNAI 5208, pp. 104–116, 2008.

the relatively low-level motion dynamics in conversational interaction, which despite several ongoing efforts, is largely an unsolved issue.

Many approaches propose interesting solutions for generating the stream of actions that an agent, or in our case, an automated avatar needs to perform in order to believably simulate the positional and orientational movements of a human engaged in a conversation. Each action usually triggers some motor function directly at locomotion level in order to animate the agent. The sequence of actions usually needs to pass through an intermediate layer in order to achieve the desired fluidity of movements and reactions. This extra level between action generation and locomotion, is responsible for smoothing the agent's overall behavior by applying steering forces directly to the underling physical model. Therefore a steering layer provides a suitable solution for filling the gap between two consecutive actions, generating a net continuous fluid behavior. This approach is particular well suited for modeling unconscious reactions and motion dynamics in conversational interactions.

A conversation defines a positional and orientational relationship of its participants. This arrangement has been described by Kendon [3] as an instance of an F-formation system. Moreover, since a conversation is not a fully rigid formation, external events from the environment or individual behaviour of single participants may produce fluctuations inside this system that lead to compensational rearrangement that avatars can automate without requiring input from their human users. In fact, this is the kind of reactive behaviour which is ill suited for explicit control [1].

2 Related Work

2.1 Automating Avatar Control

In most commercial avatar-based systems, the nonverbal expression of communicative intent and social behavior relies on explicit user input. For example, in both Second Life and World of Warcraft, user can make their avatars emote by entering special emote commands into the chat window. This approach is fine for deliberate acts, but as was argued in [4], requiring the users to think about how to coordinate their virtual body every time they communicate or enter a social situation places on them the burden of too much micromanagement. When people walk through a room full of people, they are not used to thinking explicitly about their leg movements, body orientation, gaze direction, posture or gesture, because these are things that typically happen spontaneously without much conscious effort [3]. Some of these behaviors are continuous and would require very frequent input from the user to maintain, which may be difficult, especially when the user is engaged in other input activities such as typing a chat message. In the same way that avatars automatically animate walk cycles so that users won't have to worry about where to place their virtual feet, avatars should also provide the basic nonverbal foundation for socialization.

Automating the generation of communicative nonverbal behaviors in avatars was first proposed in BodyChat where avatars were not just waiting for their own users to issue behavior commands, but were also reacting to events in the online world according to preprogrammed rules based on a model of human face-to-face behavior [4]. The focus was on gaze cues associated with establishing, maintaining and ending conversations.