Carrying the Role-Playing Metaphor to Interactive Learning Environments

Helmut Prendinger, Mitsuru Ishizuka
Department of Information and Communication Engineering
School of Engineering, University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
+81 3 5841 6347
{helmut,ishizuka}@miv.t.u-tokyo.jp

ABSTRACT
In this paper, we motivate the role-playing metaphor for intelligent educational interfaces in two ways: first, as an enjoyable interaction style between users and animated agent characters, and secondly - taken more literally - we argue that the concept of social role should be considered in the design of agents' mental models. In particular, we introduce social control programs that qualify the agent's expression of its affective state by the social context. We also describe a web-based language learning system that uses animated agent characters as conversational partners in role-playing environments.

Keywords
Social dimension in communication, role playing, affective reasoning and expression, believable animated characters

INTRODUCTION
Recent years show a growing interest in animated characters to enhance learning in computer-based interactive learning environments [7]. Lester and colleagues [8] promote *animated pedagogical agents* for their motivational role in the learning context (the 'persona effect'), in addition to the possibility of increased learning effectiveness. The animated agent approach also allows for new exciting styles of presenting information [1, 6].

Encouraged by those results, we recently started a project with the aim to employ animated characters for the pedagogical task of language conversation training. Specifically, the animated agent approach will be used to improve English conversation skills of native speakers of Japanese. Since we are also engaged in developing presentation agents, we opted to take a more broad view and set up a system that covers different kinds of user-agent communication, which we call *role-playing* interactions [13]. In a typical conversation training situation, the user interacts with one or more agent characters, and plays the role, e.g., of a customer in a virtual interactive coffee shop. When watching a presentation, possibly performed by a presentation team [1], the user may occasionally play the role of a curious critic. The communication between a user and an animated software assistant can also be seen as a form of role-playing interaction.

Our notion of role-playing interaction emphasizes the *social level* of user-agent as well as inter-agent communication. At this level, agents respect interpersonal relationships and modify their behavior according to their social role. In particular, the agent's social role determines its way of emotion expression. Consider a situation where you are angry with your boss and you happen to be a rather aggressive personality. You will presumably not show your emotion, being aware of your social role as an employee. We believe that considering the social dimension in role-playing interactions adds value to intelligent (pedagogical) interfaces for the following reasons:

- It increases the believability of animated agents, which is often captured by emotion and personality only.
- It adds sophistication to the interaction by respecting an important feature of human-human conversation.
- It explains the frequent mismatch between the output of emotional reasoning (the emotional state) and emotional display (emotion expression), as seen in human-human communication.

In our system, social reasoning will be blended with a rather standard theory of reasoning about emotions [12]. We employ Moulin and Rousseau's [10] approach to model and simulate conversations, which provides a rich framework for many aspects of inter-agent communication.

The programmable interface of the Microsoft agent package is used to run our example conversations. This choice put some restrictions from the outset: the characters available for this package have only a limited number of behaviors ('animations'), confining the realization of various emotional displays as well as some features of embodied conversational behavior [2]. On the other hand, the package

comes with a speech recognizer and a text-to-speech engine and allows client-side execution in a web browser.

The rest of this paper is organized as follows. The next section describes a framework for modeling and simulating conversations. In the following section, we first argue that affective reasoning is not sufficient to obtain believable emotion expression. Then, we introduce social control programs as a filter between affective state and emotion expression. After that, we illustrate our approach by a role-playing scenario that uses animated characters. Finally, we briefly discuss and conclude the paper.

A CONCEPTUAL FRAMEWORK FOR SIMULATING CONVERSATIONS

A conversation is typically seen as a cooperative activity where multiple locutor-agents participate and communicate through multiple channels, such as verbal utterances, gestures and facial display. Each agent has its own goals and will try to influence other participants' mental states (e.g., beliefs, goals) and affective states (e.g., emotions).

We distinguish three levels of communication [10]:

- At the *communication level* agents perform activities related to communication maintenance and turn-taking management.
- At the *conceptual level* agents transfer concepts.
- At the *social level* agents manage and respect the social relationships that hold between agents.

Our system integrates all three levels. The communicative level basically implements conversational features of human-human conversation, as proposed by Cassell and colleagues [2]. At the conceptual level, information is passed from one agent to other agents as a (simplified) symbolic representation of the utterance. According to their role in the social context, the social level puts behavioral constraints on agents' actions and emotion expression [9].

We assume that a conversation is governed by

- a *conversational manager* that is activated when the conversation starts and maintains a model of the conversation, and
- an *environmental manager* that simulates the environment in which the agents are embedded.

For simplicity, we assume that the conversational manager operates on a shared knowledge base that is visible to all agents participating in the conversation. It stores and updates all concepts transferred during the conversation. It also includes descriptions about the agents' roles and their social relationships. The environmental manager simulates the world that agents inhabit and updates its (shared) knowledge base with consequences of their actions.

MENTAL MODELS OF AGENTS

Each agent involved in the conversation is assumed to have its own mental model. A mental model may contain different kinds of entities, including world knowledge (beliefs), affective states (emotions, moods, personality traits), goals and plans. In this paper, we will concentrate on reasoning about affective states and social reasoning.

Reasoning about Emotion vs. Emotion Expression

It is widely accepted that animated agents with emotional behavior are an important contribution to make the interface more accessible and enjoyable for users [8]. Emotional behavior can be conveyed through various channels, such as facial display (expression), speech and body movement. The so-called 'basic emotions' approach [3] distills those emotions that have *distinctive* (facial) expressions associated with them: fear, anger, sadness, happiness, and disgust. Murray and Arnott [11] describe the vocal effects on the five basic emotions found in [3], e.g., if a speaker expresses the emotion 'happiness', her or his speech is typically faster, higher-pitched, and slightly louder.

Although a 'basic emotions' theory allows relating emotion to behavior, it cannot answer the question *why* an agent is in a certain emotional state. However, reasoning about emotions is considered equally important for pedagogical and presentation agents [1,7]. Many systems that reason about emotions (affective reasoners) derive from the influential 'cognitive appraisal for emotions' model of Ortony, Clore, and Collins, also known as the OCC model [4,12,5]. Here, emotions are seen as valenced reactions to events, agents' actions, and objects, qualified by the agents' goals, standards, and preferences. The OCC model groups emotion types according to cognitive eliciting conditions. In total, twenty-two classes of eliciting conditions are identified and labeled by a word or phrase, such as 'joy', 'fears-confirmed', or 'angry-at'. E.g., the emotion types 'joy' and 'angry-at' are described as follows:

Emotion type 'joy': an agent is in the emotional state of 'joy' in situation *S* IF she or he wants that a state-of-affairs *F* holds in *S* AND *F* holds in *S*.

Emotion type 'angry-at': an agent *L1* is angry at another agent *L2* about action *A* in situation *S* IF

- agent *L2* performed action *A* prior to *S*
- AND action *A* causes a state-of-affairs *F* to hold in *S*
- AND agent *L1* wants the opposite of *F*, non-*F*, in *S*
- AND action *A* is blameworthy.

A real-world example for the emotional state 'angry-at' might be the following: You ask your boss to give you some vacation and your boss turns you down. You are now angry at your boss because you cannot make the trip you were looking for (your boss' answer implies the opposite of what you wanted) and you consider the refusal of your boss as blameworthy. How will you react to your boss? Presumably you will nod, showing that you understood your boss' answer, and try to convince your boss that you really need some days off in a calm voice, with a rather neutral facial expression. Your behavior - suppressing the expression of

your emotional state – can be explained in at least two ways. First, you might have *personality traits* that characterize you as friendly and introverted. Second, and probably more important in this scenario, you might be aware of your *social role* as an employee which puts behavioral restrictions on your answer to your boss. Below, we will try to explicate the impact of the social dimension on emotion expression in communication.

Social Control Programs

We borrow the notion of *social control programs* from Gratch [5] who uses them on top of a general purpose planning system. In this system, plan generation and execution are biased by the characteristics of the social context. By contrast, we place social control programs at the interface of the module that reasons about emotion and the module that renders the emotional state to actual behavior. Basically, a social control program consists of a set of rules that encode qualifying conditions for emotion expression. This control program acts as a filter between the agent's affective state and its rendering in a social context, such as a conversation. We consider the agent's personality and the agent's social role as the most important emotion expression qualifying conditions.

To keep things simple, we consider only two dimensions of personality, which seem crucial for social interaction [1]. Extraversion refers to an agent's tendency to take action (values: 'outgoing', 'neutral', 'introverted'). Agreeableness refers to an agent's disposition to be sympathetic (values: 'friendly', 'indifferent', 'unfriendly').

Social roles are ordered according to a power scale, which defines the social power of an agent's role over other roles, and imposes certain conventional practices on the agents' behavior [9]. Our initial theory contains just three such relations: 'higher(L1,L2)', 'lower(L1,L2)', and 'equal'(L1,L2) express that L1's role is higher (lower) ranked in the power scale than that of L2, and L1's role and L2's role have the same rank, respectively. This is of course a very simple view of a social network but it already allows us to explain various phenomena in actual conversations. If the conversational partner has more social power, emotion expression is typically 'neutralized'.

Emotion expression 'neutral': an agent *L1* has the emotion expression 'neutral' towards agent *L2* IF

- agent *L2* is higher than *L1* on the power scale
- AND (*L1* is angry-at *L2* OR *L1* feels reproach towards *L2*).

Here, the agent's emotion expression is assumed to be independent of its personality traits. The second condition in this rule describes the output of the affective reasoner, e.g., the emotional state that one agent is angry at the other. If an agent communicates with an agent whose role is equal or lower, personality traits come into effect.

Emotion expression 'happiness': an agent *L1* has the emotion expression 'happiness' towards agent *L2* IF

- agent *L2* is lower than OR equal to *L1*
- AND *L1* has personality traits 'outgoing', 'unfriendly'
- AND *L1* is gloating about something regarding *L2*.

Our current implementation contains about twenty such rules. A more complete set of rules is in the process of development. In particular, we try to define rules that describe the effects of violations of conventional practices.

ROLE-PLAYING IN AN INTERACTIVE LEARNING ENVIRONMENT

Our interactive learning environment for English conversation training for Japanese speakers assumes that users (language students) would enjoy to get involved in a role-play with animated characters, and thereby overcome their uneasiness to converse in a foreign language. We already implemented two scenarios. In our *interactive theater*, the user may take the role of Rosencrantz, the companion of Guildenstern in Tom Stoppard's famous play. Inspired by [13], our *interactive drama* offers the role of a customer in a virtual coffee shop (see Fig. 1).

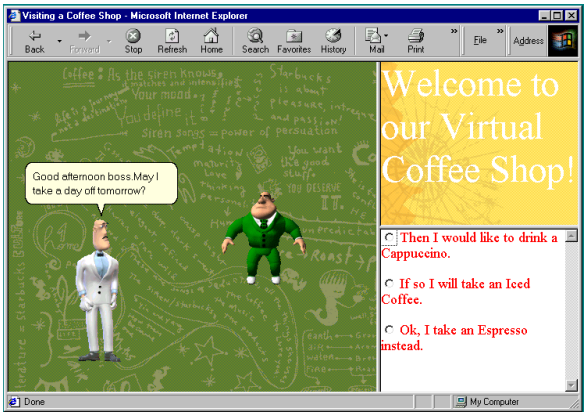

Fig. 1: Screenshot of the coffee shop scenario.

The Microsoft Agent package provides controls to embed animated characters into a web page based JavaScript interface, and includes a voice recognizer and a text-to-speech engine. The user can promote the development of the conversation by uttering one of a set of predefined sentences. The character will respond by synthetic speech, facial display, and gestures. The parameters for speech output are set in accordance with the vocal effects associated with the five basic emotions [11]. Of course, the facial display of characters is limited to the predefined 'animations' from the Agent package (e.g., 'pleased', 'sad'). To some extent, we also implemented conversational behavior [2]. E.g., the animations 'confused' (confused look, lifting shoulders) and 'don't-recognize' (put hand to ear) are used if the user's utterance is not recognized. For multi-character conversation, we implemented cues that regulate the conversational process, such as quick nods, and initiative demand behavior (e.g., offer turn, asking for turn).

The following is an annotated trace of a run of our conversation system. Here, the user interacts with a waiter-style character (extrovert, unfriendly) as a customer, who himself interacts with a boss-style character (neutral, friendly) as an employee.

> *Customer*: I would like to drink a beer. [User may also choose to drink other beverages]
>
> *Waiter* (to customer): This is a coffee shop. Get out of here! [Considers it as blameworthy to be asked for alcohol and shows his anger. We assumed equal social power of waiter and customer (which is arguable, of course)]
>
> *Waiter* (to boss): Good afternoon, boss. May I take a day off tomorrow?
>
> *Boss*: It will be a busy day. So I kindly ask you to come. [Uses polite linguistic style in accordance with his personality traits]
>
> *Waiter*: Sorry, I forgot about this. You are perfectly right. I will be here. [Considers it as blameworthy to be denied a vacation and is angry. However, he is aware of his lower social role and therefore does not show his anger. Instead, he shows neutral emotion expression]

Since the characters do not understand English, the exact wording has to be prepared for each personality/social role pair. However, this adds considerable 'social accuracy' to the conversation, and increases the characters' believability.

We started experimenting with Jinni [14], a programming tool that allows us to glue together Java applets (animating the characters), and the Prolog programs that implement all reasoning related to conversation management and agents' affective and social reasoning.

CONCLUSION

In this paper, we propose role-playing with animated characters as an enjoyable interaction style for language learning students. This idea is also present in work on tutoring [7] and information presentation [1,6]. The novel aspect of our work is that we explicate the social role of agents involved in the conversation, which allows enhanced believability of animated characters beyond reasoning about emotion and personality. The social dimension might also become an issue in future, truly conversational interfaces where (animated) agents play the roles of broker or security agents with varying rights, duties, and decision power.

We presented work in progress. One of our near-term goals is to test our language learning environments. Besides role-playing in interactive theater and drama, we will set up an interactive game environment for conversation training. Another goal is to refine our social control programs (and the notion of social role), which only allow for very restricted forms of social networks, and hence lack most of the sophistication encountered in actual conversations.

REFERENCES

1. André, E. and Rist, T. Presenting through performing: On the use of multiple lifelike characters in knowledge-based presentation systems. *Proceedings of IUI-2000*.
2. Cassell, J., Bickmore, T., Vilhjálmsson, H., and Yan, H. More than just a pretty face: Affordances of Embodiment. *Proceedings of IUI-2000*.
3. Ekman, P. An argument for basic emotions. *Cognition and Emotion 6*, 3-4, 1992, 169-200.
4. Elliott, C. *The Affective Reasoner: A process model of emotions in a multi-agent system.* PhD Thesis, Institute for the Learning Sciences, Northwestern University, 1992.
5. Gratch, J. Socially situated planning. Submitted to *AAAI Fall Symposium on Socially Intelligent Agents.*
6. Ishizuka, M., Tsutsui, T., Saeyor, S., Dohi, H., Zong, Y., and Prendinger, H. MPML: A multimodal presentation markup language with character control functions. *Proceedings of Agents'2000 Workshop on Achieving Human-like Behavior in Interactive Animated Agents*, 2000, 50-54.
7. Johnson, W.L., Rickel, J.W., and Lester, J.C. Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education*. To appear.
8. Lester, J.C., Converse, S.A., Stone, B.A., and Kahler, S.E. Animated pedagogical agents and problem-solving effectiveness: A large-scale empirical evaluation. *Artificial Intelligence in Education*, IOS Press: Amsterdam, 1999, 23-30.
9. Moulin, B. The social dimension of interactions in multiagent systems. In W. Wobcke, M. Pagnucco, C. Zhang (Eds.), Agents and Multi-Agent Systems, Formalisms, Methodologies and Applications, Springer Verlag, 1998, 109-122.
10. Moulin, B. and Rousseau, D. An approach for modelling and simulating conversations. In D. Vanderveken and S. Kubo (Eds.), *Essays in Speech Act Theory*, John Benjamins. To appear.
11. Murray, I.R and Arnott, J.L. Implementation and testing of a system for producing emotion-by-rule in synthetic speech. *Speech Communication 16*, 1995, 369-390.
12. O'Rorke, P. and Ortony, A. Explaining Emotions. *Cognitive Science 18*, 1994, 283-323.
13. Rousseau, D. and Hayes-Roth, B. Improvisional synthetic actors with flexible personalities. Technical Report KSL97-10, Knowledge Systems Laboratory, Stanford University, 1997.
14. Tarau, P. Jinni: Intelligent mobile agent programming at the intersection of Java and Prolog. *Proceedings of PAAM'99*, 109-123.