Affective and Social Behavior in Animated Agents

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

Abstract— **This paper introduces a model of interaction between users and animated agents as well as inter-agent interaction that supports basic features of affective conversation and social interaction. As essential requirements for animated agents' capability to engage in and exhibit affective and social communication we motivate reasoning about emotions and emotion expression, personality, and social role awareness. The main contribution of our paper is the incorporation of 'social filter programs' to mental models of animated agents. Those programs may qualify an agent's expression of its emotional state by the social context, thereby enhancing the agent's believability as a conversational partner. Our implemented system is entirely web-based and demonstrates animated agents with affective and social behavior in a virtual coffee shop environment.**

Keywords— **Social dimension in communication, affective reasoning and emotion expression, believability**

I. Introduction and Motivation

Our concept of affective behavior derives from the influential paradigm of *affective computing*, "computing that relates to, arises from, or deliberately influences emotions" (Picard [15]). In line with this work, we assume that emotions are indispensable for effective communication, and thus promote the view that emotions should be integrated to models of human-computer interaction. In particular, we envision that humans interact with animated agents, i.e., cartoon-style characters that may behave in believable and appropriate ways.

Animated agents with believable behavior are used as virtual tutors in interactive learning environments (e.g., Johnson *et al.* [7]), as virtual presenters on the web (e.g., André *et al.* [1], Ishizuka *et al.* [6]), and as virtual actors for entertainment (e.g., Rousseau and Hayes-Roth [17]) and language conversation training (Prendinger and Ishizuka [16]). Although those agents achieve convincing behavior in their respective predefined roles, they might fall short of 'social robustness' when put into a different scenario. E.g., a tutor agent might be believable for a student (in the teacher role) but not in the role of a peer. Humans, on the other hand, are always aware of the roles they play in a certain social setting and typically behave accordingly.

Our goal is to develop a general framework for affective and social communication that covers various forms of user-agent and inter-agent interactions. Specifically, we will propose components of agents' mental models that enable them to process emotions and show affective behavior. The main contribution of our paper is a computational model of affective communication that allows for natural responses and social behavior. Our approach will be used to improve English conversation skills of native speakers of Japanese. Since interactions are set up as role-playing dramas and games, strong requirements are imposed on the agents' affective and social abilities. Social reasoning will be blended with a rather standard theory of reasoning about emotions (O'Rorke and Ortony [13]) and a simple model of personality. We employ Moulin and Rousseau's [11] 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 [8] is used to run our role-playing scenarios. Although these off-the-shelf agent characters are quite restricted in the number of behaviors, the package comes ready with a speech recognizer and text-to-speech engine that allow client-side execution in a web browser.

The rest of the paper is organized as follows. In the next section, the core components of the design of animated agents are introduced. We explain an affective reasoner for reasoning about emotions, a rudimentary model of personality, and a simple model of social role awareness. Also, we briefly describe the theory of emotion expression used. Section 3 introduces so-called 'filter programs' that function as a 'filter' between the agent's affective state and emotion expression. After that, we illustrate our approach by an example run of our web-based system that features animated agents as conversational partners in role-playing environments. In Section 5, we summarize the paper.

II. Mental Models of Animated Agents

Each agent involved in a 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, personality traits, attitudes), goals and plans. In this section, we will focus on the core components underlying socially intelligent reasoning: affective reasoning, i.e., reasoning about emotions, personality, and the social context. Furthermore, we will discuss the output of social reasoning—the agent's emotion expression—which is the expression of its affective state qualified by its personality and the social setting in which the communication takes place.

A. Affective Reasoning

Many systems that reason about emotion, so-called *affective reasoners*, derive from the influential 'cognitive appraisal for emotions' model of Ortony, Clore, and Collins [14], also known as the OCC model (see also O'Rorke and Ortony [13]). Here, emotions (more precisely, emotion *types*) 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, and labels them by a word or phrase, such as *joy*, *happy-for*, or *angry-at*. In total, twenty-two classes of eliciting conditions are identified. The OCC model can be represented by a set of rules and might thus be seen as a model of (emotion) *causation* that allows to reason about emotion [13]. By way of example, we give the definitions of the emotion types *happy-for* and *angry-at*:

- **Emotion type *happy-for*:** agent $L1$ is in the emotional state *happy-for* agent $L2$ in situation S **if**
  - $L1$ experiences joy over a state-of-affairs F presumed to be desirable for $L2$ in S.
- **Emotion type *angry-at*:** agent $L1$ is in the emotional state *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 to hold in S **and**
  - $L1$ considers action A as blameworthy.

Affective reasoners based on the OCC model can be characterized as theories of 'cold' emotions, where agents simply derive an emotional state based on a cognitive reasoning process and in no way 'have' emotions.

B. Personality

Personality traits are typically characterized by patterns of thought, attitude, and behavior that are permanent or at least change very slowly (e.g., Moffat [9]). Consequently, believable agents should be *consistent* in their behavior [17]. For simplicity, we consider only two dimensions of personality, which seem crucial for social interaction. *Extraversion* refers to an agent's tendency to take action: sociable, active, talkative, optimistic. We consider the values: 'outgoing', 'neutral', and 'introverted'. *Agreeableness* refers to an agent's disposition to be sympathetic: good-natured, helpful, forgiving. We consider the values: 'friendly', 'indifferent', and 'unfriendly'.

C. Roles, Conventions, and Social Networks

A significant portion of human conversation takes place in a socio-organizational setting where participating agents have clearly defined *social roles*, such as sales person and customer, or instructor and student (Moulin [10]). *Conventional practices* are guidelines (or restrictions) about socially appropriate behavior in a particular social setting. We can distinguish two kinds of guiding restrictions: *Behavioral constraints* concern responsibilities, rights, duties, prohibitions, and possibilities associated with a social role. *Communicative conventions* function as a regulatory for the agent's choice of verbal expressions in a given social context. We will focus on the choice of verbal and nonverbal behavior (emotion expression), depending on the agent's social role and personality.

Formally, in social or organizational groups roles are ordered according to a *power scale*, which defines the social power of an agent's role over other roles [10]. We consider just two such relations for agents L_i and L_j: L_i's role R_{L_i} is higher ranked in the power scale than that of L_j, R_{L_j}, or L_i's role and L_j's role have the same rank. The power relations are stored as facts of the form

$$\textit{social\_power(Rel,P,}R_{L_i}\textit{,}R_{L_j}\textit{)}$$

where the relation *Rel* is either ">" or "=", and the argument P is a value between 0 and 3. By definition, if the relation is "=", then P is zero. Otherwise, if the value of P is 3, for instance, then the rank of L_i is much higher than the rank of L_j. The *social network* is specified by the social roles and associated power relations.

Following Walker *et al.* [18], we also consider the *social distance* between two agents (used in their framework to determine an appropriate linguistic style). The distance relations have format

$$\textit{social\_distance(D,}L_i\textit{,}L_j\textit{)}$$

where $D \in \{0, 1, 2, 3\}$. If agents know each other very well, the social distance can be set to zero. Note that the social distance can be high even if two agents have the same rank. When agents interact, they do not only exchange information but also establish and maintain social relationships. Hence it is important that agents avoid introducing disharmony into a conversation [10] or threaten other agents' public face [18]. Based on the values of social power $P = p(L_i, L_j)$ and social distance $D = d(L_j, L_i)$, the agent L_i computes the *threat* θ from L_j of expressing a certain emotion by using the following simple equation

$$\theta = p(L_i, L_j) + d(L_j, L_i)$$

and storing the resulting value in facts of the form "*threat(T,*L_i*,*L_j*)*" (see also [18]). High values for θ lead to the agent's suppression of the expression of its emotional state, whereas low values allow the agent to show its emotional state. Additionally, we explicitly condition responses on whether an agent L_i respects the conventions associated with its social role towards the role of L_j, by stating (for simplicity, agents are identified with their roles) "*respects_conventional_practices(*L_i*,*L_j*)*". If conventional practices are not respected, the following fact is added: "*violates_conventional_practices(*L_i*,*L_j*)*".

This is of course a very simple view of a social network but, as shown below, it already allows us to explain various phenomena in actual conversations.

D. Emotion Expression

Emotional behavior can be conveyed through various channels, such as facial display (expression), speech and body movement. The so-called 'basic emotions' approach (Ekman [5]) distills those emotions that have distinctive (facial) expressions associated with them and seem to be universal: *fear*, *anger*, *sadness*, *happiness*, *disgust*, and *surprise*. In order to avoid confusion and since there is only a

Emotion expression *neutral*: an agent $L1$ displays emotion expression *neutral* towards agent $L2$ **if**

the threat for $L1$ from $L2$ is high, $\theta > 3$
and $L1$ respects conventional practices towards $L2$
and [$L1$ is *angry-at* $L2$
or $L1$ is in state *distress*
or $L1$ is in state *fear*].

Emotion expression *happiness*: an agent $L1$ displays *happiness* towards agent $L2$ **if**

the threat for $L1$ from $L2$ is low, $\theta \leq 1$
and $L1$ respects conventional practices towards $L2$
and $L1$ has unfriendly personality
and $L1$ is *gloating* over $L2$.

Fig. 1. Partial definitions for expressions *neutral* and *happiness*.

Emotion expression *anger*: an agent $L1$ displays *anger* towards agent $L2$ **if**

the threat for $L1$ from $L2$ is low, $\theta \leq 1$
and $L1$ respects conventional practices towards $L2$
and $L1$ has unfriendly personality
and $L1$ is *angry-at* $L2$

Emotion expression *anger*: an agent $L1$ displays *anger* towards agent $L2$ **if**

the threat for $L1$ from $L2$ is high, $\theta > 3$
and $L1$ violates conventional practices towards $L2$
and $L1$ has unfriendly personality
and $L1$ is *angry-at* $L2$.

Fig. 2. Partial definition for expression *anger*.

limited number of comprehensive 'emotion words', we use *slanted* when referring to basic emotions instead of *italics* for emotional states.

The 'basic emotions' approach provides a useful list of emotions as an emotion family inventory for animated agents, as it explicitly relates emotion to behavior. Besides facial expressions, Ekman also suggests other signals, such as speech and body movement to express emotions. Most importantly, Murray and Arnott [12]) describe the vocal effects of five of Ekman's basic emotions. E.g., if a speaker expresses *happiness*, then his or her speech is typically faster (or slower) and higher-pitched, whereas a speaker's *sadness* is usually reflected in slow and lower-pitched speech.

III. Social Filter Programs

Basically, a social filter program consists of a set of rules that encode qualifying conditions for emotion expression (see Figures 1 and 2). The 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. By way of example, let us explain the *angry-at* emotion type. Assume that a student is angry at his or her instructor because s/he is treated in an unfair way, which s/he considers as blameworthy. How will s/he react to his or her instructor? Presumably s/he will nod, showing that s/he understood the instructor's point of view, and try to argue that the instructor's interpretation is wrong, in a calm voice with a rather neutral facial expression.

The student's behavior—*suppressing* the expression of his or her emotional state—can be explained in at least two ways. *First*, s/he might have personality traits that characterize him or her as very friendly. *Second*, and probably more important in this scenario, s/he might be aware of his or her social role as a student which puts behavioral restrictions on his or her answer to the instructor.

A. Social Filter Rules

In the following, we will give some examples of social filter rules. We assume that emotion expression (e.g., facial display or linguistic style) is determined by personal experience, background knowledge, and cultural norms (Walker *et al.* [18]), as well as the 'organizational culture' (Moulin [10]). Consequently, human agents determine the values of the social variables 'social power' and 'social distance'. Our rules are consistent with Brown and Levinson's theory of social interaction, as reported in [18].

If the conversational partner has more social power and/or the social distance is large (i.e., θ is high), the expression of 'negative' emotions is typically suppressed, resulting in *neutral* emotion expression (see Fig. 1, top). The first and second condition of the (top) rule concern the social context, the third condition accounts for the output of the affective reasoner, the emotional state. Here, the agent's emotion expression is assumed to be independent of its personality traits. If an agent communicates with an agent that imposes low threat, personality traits come into effect. If an agent's personality traits can be characterized as unfriendly, the agent may show its emotional state without violating conventional practices. An agent might even express happiness about something which—the agent believes—distresses another agent (see Fig. 1, bottom), or show its anger (see Fig. 2, top). Moreover, if the agent violates conventional practices, it might show its emotional state even if the social and/or power distance is high (see Fig. 2, bottom).

B. Social Feedback

Social filter programs certainly support high social accuracy of agents' responses. However, we currently have no principled way to introduce *social feedback* mechanisms to our model, i.e., values for social power and distance are kept fixed. However, when humans establish and maintain social relationships, e.g., the value of their social distance changes. This phenomenon is not reflected in agents' mental models. Agents have no memory about the social implications of their interactions with a human or other agents. We can imagine that if a user repeatedly interacts with an agent, the agent should not only remember the user but also show more familiarity. Another interesting phenomenon of human-human communication is the *reciprocal feedback loop* where, e.g., one agent's use of polite linguistic style results in another agent adapting its linguistic style and adjusting its value of social distance. Of course, this phenomenon should only occur over a longer course of social interaction.

IV. Role-Playing with Socially Intelligent Animated Agents

Our interactive environment for English conversation training for Japanese speakers is based on the idea that users (language students) would enjoy getting involved in a role-play with animated, socially intelligent agents, and thereby overcome their uneasiness to converse in a foreign language. Although Japanese speakers have a good knowledge of English grammar and vocabulary, they mostly lack opportunities to actually listen to and speak English. Inspired by the Virtual Theater project (Rousseau and Hayes-Roth [17]), we implemented an interactive drama that offers the role of a customer in a virtual coffee shop.

A. Implementation

The programmable interface of the Microsoft Agent package [8] is used to run a virtual coffee shop session in a web browser (Internet Explorer 5). This choice put some serious restrictions from the outset: the characters available for this package have only a limited number of behaviors ('animations'), confining the realization of various emotions as well as some features of embodied conversational behavior (Cassell [3]). However, our goal is believablity on the level of socially adequate emotion expression rather than lifelikeness (in the sense of realistic behavior). 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. Prolog programs implement all reasoning related to conversation and environmental management and agents' mental models (affective and social reasoning). We use Jinni 2000 to communicate between Prolog code and the Java objects that control the agents through JavaScript code. Tarau's Jinni 2000 (Java Inference engine and Networked Interactor) can be used as a "scripting tool for gluing together knowledge processing components and Java objects in distributed applications" (BinNet Corp. [2]).

In a role-playing session, the user can promote the development of the conversation by uttering one of a set of predefined sentences that are displayed on the screen (see Fig. 3). Unlike the setup of Hayes-Roth's Virtual Theater project [17], the user does not have an avatar in the play. Animated characters will respond by synthetic speech, facial display, and gestures. However, there is no time-critical management of verbal and non-verbal behavior (as in Cassell [4]). Depending on the output of the reasoning process

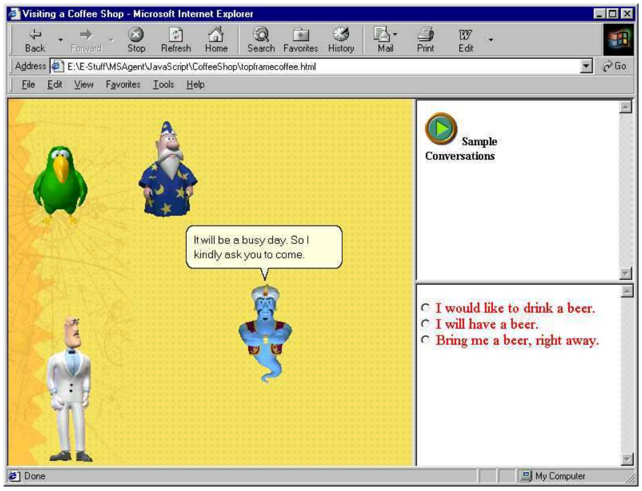

Fig. 3. Interaction involving user, James as waiter agent, Genie as manager agent, and two customer agents.

related to an agent's mental model, JavaScript functions for the agent's speech and animation are invoked and interpreted in the browser. The parameters for speech output are set in accordance with the vocal effects associated with the basic emotions (Ekman [5]). 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 (Cassell [3]). E.g., the animations 'confused' (lifting shoulders) and 'don't-recognize' (put hand to ear) are used if the user's utterance is not recognized.

B. Illustration

We will illustrate our system by showing an example run. The user takes the role of a (friendly) customer who interacts with an unfriendly waiter agent James, who himself interacts with a friendly manager agent as an employee. Fig. 4 shows the annotated trace from our conversation system. Since our agents do not understand English, a library is used to associate the user's utterance with an 'effect', e.g., that the regulations of the coffee shop are violated, and an evaluation of its linguistic style, such as polite, rude or neutral. Moreover, as an emotion (expression) recognition module is not part of our system, we set *neutral* as the default value for user input (but see the work of Picard [15]). Similarly, a library is employed to generate the syntactic form of the agent's response. In accordance with the beliefs in the waiter agent's mental model, the rules for affective reasoning, and the social filter program, the agent expresses its anger in situation $s1$. Here we assume that the customer's role is higher ranked than the waiter's role, the social power P has the value 1, and the social distance between waiter and customer is very high, D has value 3. Hence we obtain

$$threat(4,waiter,customer)$$

However, the waiter agent violates conventional practices and shows its anger. On the other hand, in situation $s5$,

[s0] Customer: *I would like to drink a beer.* [User can also choose other beverages, and for each, he or she may select the linguistic style (polite, neutral, rude).]

[s1] James (to customer): *No way, this is a coffee shop.* [The waiter considers it as blameworthy to be asked for alcohol and shows his anger. He violates conventional practices towards the customer, as the social distance between waiter and customer is high.]

[s2] [The manager of the coffee shop appears.]

[s3] James (to manager): *Good afternoon, boss. May I take a day off tomorrow?* [Performs welcome gesture. Following conventional practices towards the manager, the waiter uses polite linguistic style.]

[s4] Manager: *It will be a busy day. So I kindly ask you to come.* [Uses polite linguistic style in accordance with his personality traits.]

[s5] Waiter: *Ok, I will be there.* [Considers it as blameworthy to be denied a vacation and is angry. However, he is aware of the threat from his boss and thus does not show his anger. Instead, he shows neutral emotion expression.]

Fig. 4. Conversation involving friendly customer, unfriendly waiter, and his friendly manager.

the waiter agent suppresses the expression of its emotional state, as the agent respects conventional practices towards the manager and the threat from the manager is high. The manager's role is higher than the waiter's role, the social power value is, say $P = 2$, and the social distance is high too, e.g., $D = 2$.

Initial experiments showed that users can identify the social roles played by the animated agents, which are easily detected in the coffee shop environment. Users could also recognize when conventional practices have been violated. However, answers varied when asked for what they think went wrong in case of violation: including, that the agent is in a bad mood, or does not like the boss (or customer agent). When we run the experiment with agents that only reason about emotion and personality (i.e., without social role awareness), users would generally not consider those agents as 'unbelievable', but they expected to get hints regarding the motivation for the violation of conventional practices, and appropriate reactions from the other agents. Our (preliminary) findings reinforce the belief that an agent should show an overall consistency in its affective behavior in order to come across as believable and that social role awareness facilitates consistency.

V. Conclusion

In this paper, we propose to integrate social reasoning to mental models of animated agents, in addition to an affective reasoning component. The novel aspect of our work is that we explicate the social role of agents and associated constraints on emotion expression, which allows for enhanced believability of animated characters beyond reasoning about emotion and personality. We believe that considering the social dimension in animated agents approaches adds value for the following reasons:

- *Believability.* It may increase the illusion of life, which is often captured by emotion and personality only.
- *Social Communication.* By respecting an important feature of human conversation, it adds 'social robustness' to agent-human and inter-agent communication.
- *Explanatory power.* It explains the frequent mismatch between the output of emotional reasoning (the emotional state) and emotion expression.

We have described a web-based interactive drama scenario featuring animated agents as an entertaining testbed to experiment with new capabilities of agents.

References

[1] Elisabeth André, Thomas Rist, Susanne van Mulken, Martin Klesen, and Stephan Baldes. The automated design of believable dialogue for animated presentation teams. In J. Cassell, S. Prevost, J. Sullivan, and E. Churchill, editors, *Embodied Conversational Agents*, pages 220–255. The MIT Press, 2000.

[2] BinNet Corp. *Jinni 2000: A high performance Java based Prolog for agent scripting, client-server and internet programming*, 2000.

[3] Justine Cassell. More than just another pretty face: Embodied conversational interface agents. *Communications of the ACM*, 2000. To appear.

[4] Justine Cassell. Nudge nudge wink wink: Elements of face-to-face conversation for embodied conversational agents. In J. Cassell, S. Prevost, J. Sullivan, and E. Churchill, editors, *Embodied Conversational Agents*. The MIT Press, 2000.

[5] Paul Ekman. An argument for basic emotions. *Cognition and Emotion*, 6(3–4):169–200, 1992.

[6] Mitsuru Ishizuka, Takayuki Tsutsui, Santi Saeyor, Hiroshi Dohi, Yuan Zong, and Helmut Prendinger. MPML: A multimodal presentation markup language with character control functions. In *Proceedings Agents'2000 Workshop on Achieving Human-like Behavior in Interactive Animated Agents*, pages 50–54, 2000.

[7] W. Lewis Johnson, Jeff W. Rickel, and James C. Lester. Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education*, 2000.

[8] Microsoft. *Developing for Microsoft Agent*. Microsoft Press, 1998.

[9] Dave Moffat. Personality parameters and programs. In R. Trappl and P. Petta, editors, *Creating Personalities for Synthetic Actors*, pages 120–165. Springer, 1997.

[10] Bernard Moulin. The social dimension of interactions in multiagent systems. In W. Wobcke, M. Pagnucco, and C. Zhang, editors, *Agents and Multi-Agent Systems, Formalisms, Methodologies and Applications*, pages 109–122. Springer Verlag, 1998.

[11] Bernard Moulin and Daniel Rousseau. An approach for modeling and simulating conversations. In D. Vanderveken and S. Kubo, editors, *Essays in Speech Act Theory*. John Benjamins, 2000.

[12] Iain R. Murray and John L. Arnott. Implementation and testing of a system for producing emotion-by-rule in synthetic speech. *Speech Communication*, 16:369–390, 1995.

[13] Paul O'Rorke and Andrew Ortony. Explaining emotions. *Cognitive Science*, 18:283–323, 1994.

[14] A. Ortony, G. L. Clore, and A. Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, 1988.

[15] Rosalind W. Picard. *Affective Computing*. The MIT Press, 1997.

[16] Helmut Prendinger and Mitsuru Ishizuka. Carrying the role-playing metaphor to interactive learning environments. In *Proceedings International Conference on Intelligent User Interfaces (IUI-2001)*, 2001. Accepted as Poster/Demo. To appear.

[17] Daniel Rousseau and Barbara Hayes-Roth. A social-psychological model for synthetic actors. In *Proceedings 2nd International Conference on Autonomous Agents (Agents'98)*, pages 165–172, 1998.

[18] Marilyn A. Walker, Janet E. Cahn, and Stephen J. Whittacker. Improvising linguistic style: Social and affective bases for agent personality. In *Proceedings First International Conference on Autonomous Agents (Agents'97)*, 1997.