Pedagogical Agents Trying on a Caring Mentor Role

Konstantin ZAKHAROV[a], Antonija MITROVIC[a], Lucy JOHNSTON[b]
[a] *Intelligent Computer Tutoring Group,* [b] *Department of Psychology*
University of Canterbury, Christchurch, New Zealand
{konstantin.zakharov, tanja.mitrovic, lucy.johnston}@canterbury.ac.nz

Abstract: We describe the design and evaluation of an affective pedagogical agent persona for Intelligent Tutoring Systems. The goal of our research was to develop an agent embodying a persona of a caring mentor interested in the learner's progress. The agent's behaviour is guided by a set of rules that are triggered by the states of the session history. Four agents were integrated with EER-Tutor for a formative evaluation study. The mentor persona secured strong rapport with the users; the audible narration was seen as a strong feature of the agents.

Introduction

The semantic component of social interaction, most frequently represented as speech, is often underpinned by the affective component, which can be expressed through speech and non-verbal displays such as gestures, posture, facial expression and eye gaze [10]. People have a propensity to transfer the social view of interaction onto their interaction with electronic media. As described by Reeves and Nass [18], people tend to view electronic media in a social way, as if they were other people. However, computers since their early days have been implicitly designed without awareness of the affective communication channel. Computers respond to people as if they too were computers, thus forcing people to adjust to computer protocols and interact on a sub-human level, which contradicts the main assumption guiding Human-Computer Interaction (HCI) research: *"People should not have to change radically to 'fit in with the system' – the system should be designed to match their requirements"* [17].

This inherent lack of affective fit is particularly significant in the area of Intelligent Tutoring Systems (ITS), because research suggests a strong interaction between cognitive and affective processes in human mind. For example, there is a reliable correlation between one's emotional state, memory capacity and motivation [6, 18]. Consequently, a failure to recognise affective processes might impose a serious limitation on interaction types which are fundamentally social in nature, because the affective component is often considered as important as semantic component [3]. Without considering the interaction between the cognitive and affective processes ubiquitous in human interaction, educational systems might never approach their full potential.

HCI does acknowledge the need to avoid negative affective states, such as frustration; common solutions to avoid user frustration include either (a) trying to determine and fix the problem causing the negative feelings, and/or (b) pre-emptively trying to avoid the problem from happening in the first place [7]. However, these approaches have a limited application in ITSs because of the fundamental differences between general HCI and educational protocols. Affective considerations in ITSs are more complex, because learning from a computer is not just about ease of use. It would be unrealistic to expect students to stop making errors during learning. Learning can be frustrating and difficult because it requires exposing learners' errors in thinking and gaps in their knowledge. Theory of Learning from Performance Errors [15] suggests that errors are an inseparable component of learning. Error detection followed by error correction, in fact, is vital for the improvement of future performance.

Recent research suggests using Affective Pedagogical Agents (APAs) in ITSs as a medium for delivering feedback to the users [8, 12]. This research draws on Allport's [1] classic definition of social psychology: *"The scientific investigation of how the thoughts, feelings and behaviours of individuals are influenced by the actual, imagined or implied presence of others"*; in accordance with this statement APAs are known to enhance the social view of interaction with ITSs. This undermines the idea that computers are merely neutral tools and emphasises the importance of the social relationships that can develop between a computer and a learner [14]. A better understanding of these relationships is essential to building smarter tools for learning.

We start by presenting our work on developing an affective persona for agents integrated with EER-Tutor, an ITS for developing Enhanced Entity-Relationship (EER) modelling skills [19, 21]. Section 2 outlines the experiment aimed at assessing learners' perception, expectations and response to the agents, and details the experimental results. Section 3 presents the conclusions and discussion of our research.

1. Developing Affective Pedagogical Agents

When designing the persona, we had to establish what kind of affective response an agent should provide in order to support the learner's determination in the face of the inevitable stress, anxiety and frustration involved in learning. One simple rule of thumb suggested by Bickmore [3] is to apply what has been found appropriate for human-to-human interaction (HHI) to the design of educational HCI. People use a variety of methods to help manage their emotions, such as interacting with media and/or other people, engaging in sports or work, meditating or praying, using positive thinking, and consuming foods and substances such as alcohol. Klein et al. [9] identify two types of support for emotion regulation. First, passive support is used to manipulate moods without necessarily discussing emotions themselves. Media, activities, food and other substances fall into this category. In contrast, active support occurs when people discuss or otherwise directly address their emotions, as a means of managing them.

Out of possible instructional roles we chose the mentor role, because of positive influence on learning demonstrated in previous research [2]. At the same time, from the affective interaction point of view, passive affective support intuitively seems as adequately congruent with the role of a mentor. Consequently, we tried to create an agent with a persona of a mentor interested in learner's progress. We wanted the agent to acknowledge the leaner's emotions indirectly through its emotional appearance, while trying to keep the learner focused on the task at hand. Thus we designed an agent which expresses solidarity with the user – it will cheer with the users' success, be sympathetic when there are difficulties and keep company to the user in neutral situations. Similar agent's behaviour was earlier adopted in the work of Lester et al. [11].

Human mentors can pick up a lot of clues from non-verbal communication; with varying degrees of depth, the mentor is always aware of the learner's affective state and cognitive state. In combination, these two factors allow a human mentor to choose an appropriate affective response at each step. While in our case, the agent does not have a way of determining users' affective state, prior research shows that in a learning context affective state can be indexed on the basis of cognitive state [4]. Cognitive Theory of Emotions states that the valence of one's emotional reaction depends on the desirability of the situation [16]. Thus we can assume that continuous lack of cognitive progress will be accompanied by a negative affective state, because the user will be dissatisfied with the state of the current task. Conversely, good progress will result in a positive affective state. In our approach, we rely on a simplified emotional model described by a single dimension – affective valence.

EER-Tutor maintains student models; the state of student model can be used to index the student's affective states. In our agents, however, affective logic does not directly rely on the student model. Instead, the logic relies on session history, which includes a wide variety of user actions. The rationale behind this approach is twofold. First, most user actions, such as login, problem selection and so on, may require a response or acknowledgement from the agent; in many cases such responses would not have to carry much affective load, although failing to respond in a similar HHI situation could be interpreted as lack of attention and care. Second, under certain circumstances seemingly neutral actions might indicate changes in the users' affective state and thus should be addressed by the agent. For example, repeated submissions of the same solution to a problem might indicate a negative affective state, such as frustration. In general, we assume that any repeating actions should not go unnoticed.

1.1. Agents' Appearance

EER-Tutor [19, 21] is a web-based ITS whose server carries the application load which includes running the constraint-based modelling engine, applying pedagogical logic, maintaining student models and hosting EER-Tutor curriculum. The client is implemented as a set of AllegroServe[1] dynamic HTML pages and a Java applet providing drawing tools for creating solutions; the interface only accepts user input and delivers feedback. Similarly, the agents' implementation is also split between the server and client. The server carries the agents' affective logic and controls the agents' behaviour, while on the client-side the agent appears to the users as a "talking head" with an upper body, embedded in EER-Tutor's work-space. The agent figures have been designed with the help of PeoplePutty[2] toolkit; the web browser displays the agent with Haptek[3] player plug-in. Haptek's character affective appearance is controlled by a number of parameters called switches. Some switches, for example, control lips and eyebrows positions. Haptek characters communicate with the server through AJAX[4] requests. The communication process revolves around retrieving updated values for the switches along with appropriate narration fragments. Figure 1 shows EER-Tutor workspace with a male agent above the feedback pane. The screenshot shows the state immediately after a solution submission. In this case, the submitted solution was incorrect: the user mistakenly defined the *Chapter_number* attribute as a key attribute, instead of making it a partial key attribute. The erroneous diagram component, the *Chapter_number* key

[1] http://allegroserve.sourceforge.net – AllegroServe web application server
[2] http://www.haptek.com/products/peopleputty/ — Haptek's PeoplePutty SDK
[3] http://www.haptek.com/ — PeoplePutty is a product of Haptek
[4] http://www.adaptivepath.com/publications/essays/archives/000385.php

attribute, is shown in red. In general, the error messages presented by the agent refer the user to the errors in the diagram; in our example the first error message says: *A weak entity type must have a partial key. The highlighted entity type has no partial key*. This approach to engaging the user in active information processing is supported by Mayer's first principle of multi-media design, *Multimedia effect for transfer*, which states that providing the learner with a narration results in better learning when it is supplemented by corresponding visual aids [13].

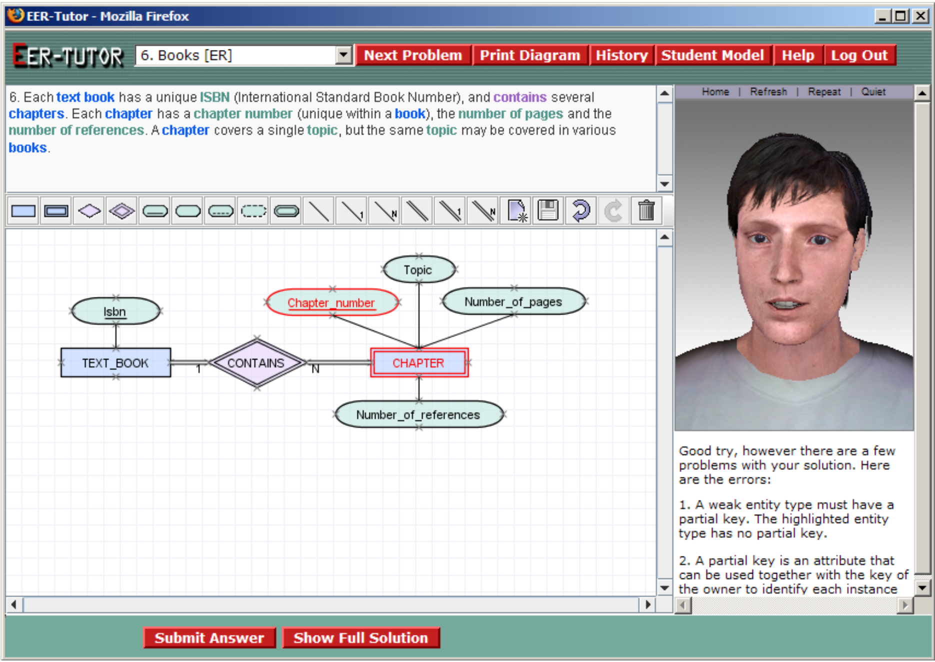

Figure 1. The view of EER-Tutor's work-space with the male agent

Figure 2. The agents designed with the PeoplePutty toolkit.

Along with feedback in audible form, the user is also presented with the most recent feedback messages in the textual form. Even though the sixth principle of multimedia design, *Redundancy effect for transfer*, states that such redundancy negatively affects learning [13], we consider that the EER-Tutor context justifies such redundancy, because the complexity of the domain knowledge and abundance of information may be difficult to process without textual representation. For example, when the user makes several mistakes, the corresponding messages may help the user to stay focused on the task instead of wasting efforts on remembering multiple feedback messages.

Haptek figures rely on Microsoft's Speech Application Programming Interface (SAPI), SAPI4 and SAPI5-compatible Text-to-Speech (TTS) engines to display realistic mouth movements while producing verbal narrations. In order to enhance the agents' realism, we obtained two reportedly higher quality Cepstral[5] voices – one male and one female, instead of using the default TTS voices supplied with Windows XP. Figure 2 shows the four agents we created – two male and two female characters. We chose a generic youthful appearance in order to remain consistency with agent's role of a mentor' since the mentoring group is traditionally represented by younger people.

1.2. Agents' Behaviour

EER-Tutor is driven entirely by user-generated interface events, such as submissions of solutions. Every interface event results in a corresponding request type being sent to the server. When the server processes the request and the client receives the response, the agent control script requests the server for the updates of the agents' emotional display and verbal feedback to be returned in response to the users' most recent action. At this stage the agent's behaviour rules, described later in this Section, are matched against the session history. On completion of this process, the server sends the response defined by the selected rule. In this way, even though the Haptek character is completely unaware of the users' actions and student model, to the user it appears that the agent is actively observing their actions.

The agent control script also runs an uninterrupted loop (repeating at the 1.5s intervals) continuously querying the server for affective appearance and narration updates even in the absence of interface events. While user actions may colour the agent's affective appearance, the agent is capable of emotional self-regulation mimicking human behaviour. The agent's affective state tends to gradually return to neutral, so if the agent momentarily may appear very upset or happy, after a few minutes the agent inevitably "settles down" even in the absence of affect-provoking actions of the user.

The agent's persona is controlled by a set of rules, each of which corresponds to a unique state of the session history and produces a certain reaction from the agent; every rule consists of a pattern (defined in Allegro Prolog[6]) to match a certain history state, a number of equivalent alternative verbal responses and the affective state update command. We have defined just under 20 rules, which make the agent respond to a variety of situations. The rules can be roughly divided into three categories: affectively-neutral rules, rules causing current affective valence to nudge slightly towards the positive end, and rules resulting in a move towards the negative end. The values of positive/negative affect changes are defined individually in each rule. When dominated by positive affect, the agent smiles; negative affect makes the agent appear sad. The following are examples of situations, with descriptions of corresponding responses and affective changes in the agent's appearance:

---

[5] http://www.cepstral.com/ — Cepstral Text-to-Speech engines.
[6] http://www.franz.com/products/prolog/ — Allegro Prolog – integrated extension for Common Lisp.

First-time login – agent introduces itself and welcomes the user with an extended message; the agent's affective state is changed so that the agent smiles.

Selection of a new agent – agent introduces itself and smiles.

Submission with a few errors – agent starts with a friendly introductory phrase and reads the errors; the agent's affective state is nudged slightly towards the negative side. If this attempt happens to come in a series of unsuccessful attempts, the agent will have an unhappy/concerned look as the result of its affective valence being dampened by the repeated triggering of this rule. However, if this is a single unsuccessful attempt, the change in the agent's appearance will be subtle.

Submission with a single error – the agent starts an encouraging phrase and presents the error; affective valence is nudged slightly to the positive direction.

Correct solution – the agent congratulates the user with solving the problem in one attempt and while happily smiling suggests that the user try another problem.

There are also rules defining behaviours for repeated submissions of identical solutions, repeated abandoning of problems, logging out etc. Precedence of session states is implicitly encoded in the order of the rules; during the rule matching process, the first match results in the corresponding action on behalf of the agent. Both male and female agents are guided by the same set of rules.

2. The Experiment

In July 2006, we recruited 20 volunteers (16 male and 4 female) from the third and fourth year students at the University of Canterbury. All participants were familiar with the EER model. The disparity between the users' levels of EER expertise was not an important factor, because the study was aimed at collection of qualitative data only. The experiment was carried out as a series of individual sessions. At the start of a session, a participant was provided with a verbal description of the task. Participants were expected to spend 45 to 60 minutes solving problems. The participants were asked to choose an agent before they were able to navigate to the workspace. During the session, the participants were free to choose a different agent at any time. At the end, the participants were required to fill out a questionnaire, for assessing their experience of learning with the agent and their perception of the agent's persona.

The first-choice preference was given to female agents (14 vs. 6) irrespective of the participants' sex. Among male participants only 38% chose a male agent, while all female participants chose a female agent at the start of the session. On the basis of questionnaire responses, it appears that male agents won in their effectiveness and overall impression on the participants. The six participants who chose a male agent reported they enjoyed EER-Tutor more, giving it a rating of 4.1 ($\sigma = 0.4$) out of 5, compared to the rating of 3.7 ($\sigma = 0.7$) by the participants who chose the female agent. The average learning success (rated on the scale of 1 to 5) reported by participants with male agents is higher then the female agents' group: 3.5 ($\sigma = 0.5$) vs. 3.1 ($\sigma = 0.7$). This learning estimate difference is consistent with the actual learning outcomes, measured by the number of constraints learned during the session: 3.3 ($\sigma = 4.8$) for male agents vs. 2.1 ($\sigma = 2.3$) for female agents. However, the small number of participants does not allow us to treat these results as being statistically reliable.

We attribute the apparent association between higher ratings and better learning outcomes for the male agents to the difference in the quality of male and female Cepstral TTS voices used. The male TTS voice sounds more human-like then the female voice; even though the participants were not specifically asked to rate the quality of TTS voices, 50% of the participants who used female agents stated that they were frustrated by or disliked that voice. No such comments were made about the male voices. Our interpretation of the voice quality effect is supported by the Cognitive Load Theory [20], which states that processing unnatural-sounding or machine-like voices imposes higher cognitive demands on learners than natural voices do.

The general response to the agents was positive – 75% rated the agents as a useful feature. At the same time, half of the participants who thought the agent's presence was unnecessary rated audible narration as useful. Overall, the participants were enthusiastic about narration – 50% stated that narration was the most helpful feature, because it made it possible for them to keep their eyes on the diagram and begin correcting errors while listening to the narration. Participants commented that this helped save time and enabled them solve problems faster.

Both male and female agents' responses and appearance were rated as adequate by around 90% of both male and female participants. Participants' feedback indicates that the agents' persona was appreciated for not being "in-your-face". Some users commented that they liked the "feeling of company" and "human presence" created by the agents. Some users made comments suggesting that the agents should be more active and versatile in their behaviour; others commented on being distracted by the agents' movements, such as blinking and breathing, in the background.

The study highlighted the need for the agents' behaviour to appear more natural. Many comments suggested enabling the user to have greater control over the agents' behaviour and voice properties. For example, some users wanted the agents to be more dynamic, even "have them fly over the workspace" and "make emotional changes more apparent", while others wanted the agents to be less obtrusive; some participants said they would like to be able to control the speed of the narration to make it faster or slower. Some users (around 25%) did not pay attention to the agents' affective appearance, but a small proportion (around 15%) commented on the agents "flipping out" and getting "too emotional" too soon. One participant commented that he or she felt "emotionally blackmailed and manipulated by the agent".

3. Conclusions and Future Work

We described an approach to modelling affective pedagogical agent persona and reported evaluation results. The persona was designed as a set of rules corresponding to the interaction states which require responses from a pedagogical, affective or social point of view. Our approach is characterised by simplicity, flexibility and scalability. The persona can be developed incrementally by adding new rules; the larger the number of rules, the more versatile and interactive the persona is. With this approach it is possible to develop different sets of rules defining different types of personas.

We created a mentor-like persona, mimicking a somewhat informal HHI interaction with passive emotional support. Four animated characters were designed aimed at soliciting perception of the agents' mentor persona. The results indicate a good outlook on the viability of the suggested approach and we will enhance our implementation in future work. Our agents' personae secured positive ratings during the evaluation; however, the lack of scale in the study and the confounding factor of difference between the quality of male and female TTS voices does not allow us to make conclusions about participants' preferences on the agent's sex and its effect on learning. We have learned that the quality of agents' voice is critical for maintaining the agents' rapport with the users and maintaining the flow of the learning process.

This study also reveals the breadth of user preferences when it comes to interacting with affective pedagogical agents and suggests that finding an ideal persona to "click" with all users is unrealistic. Some learners see the agent as a distraction, while others wish the agent to be more interactive and entertaining. Needless to say, affective agents need to respect human individuality in the style of interaction. Even though the study shows the agents were a well-received addition to EER-Tutor, the participants' comments indicate the need for making the agents more flexible. Audible narration was welcomed by most participants irrespective of their expectations of the agents' persona.

In future work, we intend to extend our system to identify students' affective states via real-time facial feature tracking and physiological sensors. Incorporating data from the sensory channel and changing the agents' persona rules will give the agent a finer level of affective awareness.

References

[1] G. W. Allport, *Attitudes*, in C. A. Murchison, ed., *A handbook of Social Psychology*, 1935, pp. 798-794.

[2] A. L. Baylor, Y. Kim, *Pedagogical Agent Design: The Impact of Agent Realism, Gender, Ethnicity, and Instructional Role*, in J. C. Lester, R. M. Vicari and F. Paraguaçu, eds., *Intelligent Tutoring Systems*, 2004, pp. 592-603.

[3] T. W. Bickmore, *Unspoken Rules of Spoken Interaction*, CACM, 47, 2004, 38-44.

[4] C. Conati, H. Maclaren, *Evaluating a Probabilistic Model of Student Affect*, *Intelligent Tutoring Systems*, 2004, pp. 55-66.

[5] C. Gama, *Metacognition in Interactive Learning Environments: The Reflection Assistant Model*, *Intelligent Tutoring Systems*, 2004, pp. 668-677.

[6] D. Goleman, *Emotional Intelligence*, Bantam Books, New York, 1995.

[7] J. Gratch et al., *Creating Interactive Virtual Humans: Some Assembly Required*, IEEE Intelligent Systems, 17, 2002, 54-63.

[8] W. L. Johnson, J. W. Rickel, J. C. Lester, *Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments*, Artificial Intelligence in Education, 11, 2000, 47-78.

[9] J. Klein, Y. Moon, R. W. Picard, *This Computer Responds to User Frustration: Theory, Design, and Results*, Interacting with Computers, 14, 2002, 119-140.

[10] H. Klingzing, T. Tisher, *Expressive nonverbal behaviors; A review of research on training with consequent recommendations for teacher education*, in J. Raths and L. Katz, eds., *Advances in Teacher Education*, Ablex Publishing, 1986.

[11] J. Lester et al., *Deictic and emotive communication in animated pedagogical agents*, Embodied Conversational Agents, 2000, pp. 123-154.

[12] J. Lester, S. Towns, P. Fitzgerald, *Achieving Affective Impact: Visual Emotive Communication in Life-like Pedagogical Agents*, Artificial Intelligence in Education, 10, 1999, 278-291.

[13] R. E. Mayer, *Multimedia Learning*, Cambridge University Press, 2001.

[14] P. Mishra, M. Nicholson, S. Wojcikiewicz, *Does My Word Processor Have a Personality? Topffer's Law and Educational Technology*, Adolescent and Adult Literacy, 44, 2001, 634-641.

[15] S. Ohlsson, *Learning From Performance Errors*, Psychological Review, 103, 1996, 241-262.

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

[17] J. Preece et al., *Human-Computer Interaction*, Addison-Wesley, New York, 1994.

[18] B. Reeves, C. Nass, *The Media Equation.* Cambridge Publications, 1996.

[19] P. Suraweera, A. Mitrovic, *An Intelligent Tutoring System for Entity Relationship Modelling*, Artificial Intelligence in Education, 14, 2004, 375-417.

[20] J. Sweller, J. van Merrienboer, F. Paas, *Cognitive Architecture and Instructional Design*, Educational Psychology Review, 10, 1998, 251-296.

[21] K. Zakharov, A. Mitrovic, S. Ohlsson, *Feedback Micro-engineering in EER-Tutor*, in J. Breuker, ed., *Artificial Intelligence in Education*, IOS Press, 2005, pp. 718-725.