Visual Emotive Communication in Lifelike Pedagogical Agents

Stuart G. Towns[1], Patrick J. FitzGerald[2], and James C. Lester[1]

[1] Department of Computer Science
North Carolina State University
Raleigh, NC 27695
{sgtowns, lester}@eos.ncsu.edu
[2] Department of Design and Technology
North Carolina State University
Raleigh, NC 27695
pat_fitzgerald@ncsu.edu
http://multimedia.ncsu.edu/imedia/

Abstract. Lifelike animated agents for knowledge-based learning environments can provide timely, customized advice to support leaners' problem-solving activities. By drawing on a rich repertoire of emotive behaviors to exhibit contextually appropriate facial expressions and emotive gestures, these agents could exploit the visual channel to more effectively communicate with learners. To address these issues, this paper proposes the *emotive-kinesthetic behavior sequencing* framework for dynamically sequencing lifelike pedagogical agents' full-body emotive expression. By exploiting a rich behavior space populated with emotive behaviors and structured by pedagogical speech act categories, a behavior sequencing engine operates in realtime to select and assemble contextually appropriate expressive behaviors. This framework has been implemented in a lifelike pedagogical agent, COSMO, who exhibits full-body emotive behaviors in response to learners' problem-solving activities.

1 Introduction

Recent years have witnessed significant advances in intelligent multimedia interfaces that broaden the bandwidth of communication in knowledge-based learning environments. Moreover, because of the potential benefits of both agent-based technologies and anthropomorphic interfaces, concerted efforts have been undertaken to develop pedagogical agents that can play an important role in learning environment architectures [4, 5, 7, 13, 16] In particular, *animated pedagogical agents* [12, 14] that couple advisory functionalities with a strong lifelike presence offer the promise of providing critical visual feedback, which raises the intriguing possibility of creating learning environments inhabited by pedagogical agents in the form of intelligent lifelike characters.

Engaging lifelike pedagogical agents that are visually expressive could clearly communicate problem-solving advice and simultaneously have a strong motivating effect on learners. If they could draw on a rich repertoire of emotive behaviors to exhibit contextually appropriate facial expressions and expressive gestures, they could exploit the visual channel to advise, encourage, and empathize with learners. However, enabling lifelike pedagogical agents to communicate the *affective* content of problem-solving advice poses serious challenges. Agents' full-body emotive behaviors must support expressive movements and visually complement the problem-solving advice they deliver. Moreover, these behaviors must be planned and coordinated in realtime in response to learners' progress.

To address these issues, this paper proposes the *emotive-kinesthetic behavior sequencing* framework for dynamically sequencing lifelike pedagogical agents' full-body emotive expression. Creating an animated pedagogical agent with this framework consists of a three phase process:

1. **Emotive Pedagogical Agent Behavior Space Design:** Creating a behavior space populated with emotive behaviors with full-body movements including facial expressions with eyes, eyebrows, and mouth and gestures with arms and hands.
2. **Speech Act-Based Behavior Space Structuring:** Constructing a behavior space in which pedagogical speech acts are associated with their emotional intent and their kinesthetic expression.
3. **Full-body Emotive Behavior Sequencing:** Creating an emotive-kinesthetic behavior sequencing engine that operates in conjunction with an explanation system to dynamically plan full-body emotive behaviors in realtime by selecting relevant pedagogical speech acts and then assembling appropriate visual behaviors.

This framework has been used to implement Cosmo (Figure 1), a lifelike pedagogical agent with realtime full-body emotive expression. Cosmo inhabits the Internet Advisor, a learning environment for the domain of Internet packet routing. An impish, antenna-bearing creature who hovers about in a virtual world of routers and networks, he provides advice to learners as they decide how to ship packets through the network to specified destinations. Previous work with the Cosmo project focused on techniques to enable lifelike agents to dynamically create deictic references to particular objects in learning environments [15]. Here, we propose the emotive-kinesthetic behavior sequencing framework and illustrate its use in Cosmo's realtime emotive behavior sequencing as it corrects learners' misconceptions detected in the course of their problem-solving activities.

2 Pedagogical Agents

As a result of developments in *believable* intelligent agents [2], the intelligent tutoring systems community is now presented with opportunities for exploring new technologies for *pedagogical agents* and the roles they can play in communication. Work to date on pedagogical agents is still in its infancy, but progress is being made on two fronts. First, research has begun on pedagogical agents that can

Fig. 1: Cosmo and the INTERNET ADVISOR learning environment

facilitate the construction of component-based tutoring system architectures and communication between their modules [13, 16], provide multiple context-sensitive pedagogical strategies [7], reason about multiple agents in learning environments [5], and act as co-learners [4]. Second, projects have begun to investigate techniques by which animated pedagogical agents can behave in a lifelike manner to communicate effectively with learners both visually and verbally [1, 12, 14] It is this second category, lifelike animated pedagogical agents, that is the focus of the work described here.

Lifelike pedagogical agents hold much promise because they could play a central communicative role in learning environments. Through an engaging persona, a lifelike pedagogical agent could simultaneously provide students with contextualized problem-solving advice and create learning experiences that offer high visual appeal. Perhaps as a result of the inherent psychosocial nature of learner-agent interactions and of humans' tendency to anthropomorphize software, recent evidence suggests that an ITS with a lifelike character can be pedagogically effective [9] while at the same time can have a strong motivating effect on learners [8].

> **Situated Emotive Communication:** Lifelike agents' emotive behaviors must support the goal of facilitating students' learning. The behaviors must therefore be situated, i.e. agents should exhibit their behaviors directly in support of problem solving.

Although work is underway on two projects on lifelike pedagogical agents, neither has focused on runtime inference techniques for providing visual feedback via the exhibition of continuous full-body emotive behaviors. The STEVE (Soar Training Expert for Virtual Environments) project has produced a full complement of animated pedagogical agent technologies for teaching procedural knowledge. Although the STEVE agent can create on-the-fly demonstrations and explanations of complex devices and its creators are beginning to examine more complex animations [11], its focus to date has been on the realtime generation of behaviors using a non-emotive agent. The DESIGN-A-PLANT project [14] has produced effective animated pedagogical agent technologies that are the creation of a multidisciplinary team of ITS researchers and animators, but it also has not yet addressed realtime inference about the creation of full-body emotive behaviors.

3 Emotive-Kinesthetic Behavior Sequencing Framework

To enable a lifelike pedagogical agent to play an active role in facilitating learners' progress, its behavior sequencing engine must be driven by learners' problem-solving activities. As learners solve problems, an explanation system monitors their actions in the learning environment (Figure 2). When they reach an impasse, as indicated by extended periods of inactivity or sub-optimal problem-solving actions, the explanation system is invoked to construct an explanation plan that will address potential misconceptions. By examining the problem state, a curriculum information network, and a user model, the explanation system determines the sequence of pedagogical speech acts that can clearly repair the misconception and passes the types of the speech acts to the emotive-kinesthetic behavior sequencing engine. By assessing the speech act categories and then identifying full-body emotive behaviors that the agent can perform to communicate the affective impact appropriate for those speech act categories, the behavior sequencing engine selects relevant behaviors and binds them to the verbal utterances determined by the explanation system. The behaviors and utterances are then performed by the agent in the environment and control is returned to the learner who continues her problem-solving activities.

3.1 Emotive-Kinesthetic Behavior Space Design

To exhibit full-body emotive behaviors, a pedagogical agent's behavior sequencing engine must draw on a large repertoire of behaviors that span a broad emotional spectrum. For many domains, tasks, and target learner populations, agents that are fully expressive are highly desirable. To this end, the first phase in creating a lifelike pedagogical agent is to design an *emotive-kinesthetic behavior space* that is populated with physical behaviors that the agent can perform when called upon to do so. Because of the aesthetics involved, an agent's behaviors are perhaps best designed by a team that includes character animators. Creating a behavior space entails setting forth precise visual and audio specifications that describe in great detail the agent's actions and utterances, rendering the actions, and creating the audio clips.

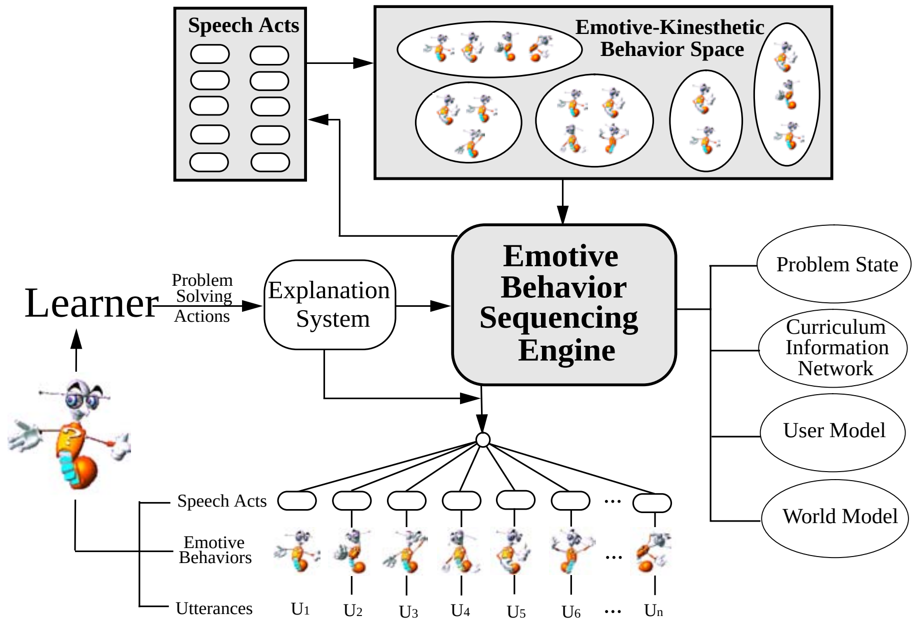

Fig. 2: The Emotive-Kinesthetic Behavior Sequencing Architecture

Stylized Emotive Behaviors. It is important to draw the critical distinction between two approaches to animated character realization, life-quality vs. stylized [3]. In the *life-quality* approach, character designers and animators follow a strict adherence to the laws of physics. Characters' musculature and kinesthetics are defined entirely by the physical principles that govern the structure and movement of human (and animal) bodies. For example, when a character become excited, it raises its eyebrows and its eyes widen. In contrast, in the *stylized* approach, a consistency is obeyed yet the laws of physics (and frequently of human anatomy and physiology) are broken at every turn. When a character animated with the stylized approach becomes excited, e.g., as in the animated films of Tex Avery [3], it may express this emotion in an exaggerated fashion by rising from the ground, inducing significant changes to the musculature of the face, and bulging out its eyes.

Expressive Range. To be maximally entertaining, animated characters must be able to express many different kinds of emotion. As different social situations arise, they must be able to convey emotions such as happiness, elation, sadness, fear, envy, shame, and gloating. In a similar fashion, because lifelike pedagogical agents should be able to communicate with a broad range of speech acts, they should be able to visually support these speech acts with an equally broad range of emotive behaviors. For example, they should be able to exhibit body language that expresses joy and excitement when learners do well, inquisitiveness for uncertain situations (such as when rhetorical questions are posed), and disappointment when problem-solving progress is less than optimal.

3.2 Behavior Space Structuring with Pedagogical Speech Acts

An agent's behaviors will be dictated by design decisions in the previous phase, which to a significant extent determine its personality characteristics. Critically, however its runtime emotive behaviors must be somehow modulated to a large degree by ongoing problem-solving events driven by the learner activities. Although, in principle, behavior spaces could be structured along any number of dimensions such as degree of exaggeration of movement or by type of anatomical components involved in movements, experience with the implemented agent suggests that the most effective means for imposing a structure is based on *speech acts* and is inspired by foundational work on affective reasoning [6]. By creating emotive annotations that connect pedagogical speech acts to relevant physical behaviors, behavior spaces can be augmented with representational structures that enable an emotive behavior sequencing engine to identify relevant behaviors at runtime. To illustrate, the Cosmo agent deals with cause and effect, background, assistance, rhetorical links, and congratulatory acts as follows:

- *Congratulatory:* When a learner experiences success, a `congratulatory` speech act triggers an `admiration` emotive intent that will be expressed with behaviors such as applause, which depending on the complexity of the problem will be either restrained or exaggerated. The desired effect is to encourage the learner.
- *Causal:* When a learner requires problem-solving advice, a `causal` speech act posed rhetorically triggers an `interrogative` emotive intent that will be expressed with behaviors such as head scratching or shrugging. The desired effect is to underscore questioning.
- *Deleterious effect:* When a learner experiences problem-solving difficulties or when the agent needs to pose a rhetorical question with unfortunate consequences, `disappointment` is triggered which will be expressed with facial characteristics and body language that indicate sadness. The desired effect is to build empathy.
- *Background* and *Assistance:* In the course of delivering advice, `background` or `assistance` speech acts trigger `inquisitive` intent that will be expressed with "thoughtful" restrained manipulators such as finger drumming or hand waving. The desired effect is to emphasize active cognitive processing on the part of the agent.

To create a fully operational lifelike agent, the behavior space includes auxiliary structuring to accommodate important emotive but non-speech-oriented behaviors such as dramatic entries and exits from the learning environment. Moreover, sometimes the agent must connect behaviors generated by two separate speech acts. To achieve these `rhetorical link` behaviors, it employs subtle "micro-movements" such as blinks or slight head nods.

3.3 Dynamic Emotive Behavior Sequencing

To dynamically orchestrate full-body emotive behaviors that achieve situated emotive communication, complement problem-solving advice, and exhibit realtime visual continuity, the emotive behavior sequencing engine selects and assembles behaviors in realtime. By exploiting the pedagogical speech act structuring, the sequencing engine navigates coherent paths through the emotive behavior space to weave the small local behaviors into continuous global behaviors. Given a communicative goal G (such as explaining a particular misconception that arose during problem solving), a simple overlay user model, a curriculum information network, and the current problem state, it employs the following algorithm to select and assemble emotive behaviors in realtime:

1. **Determine the pedagogical speech acts $A_1 \dots A_n$ to achieve G.** When the explanation system is invoked, it employs the by-now-classic techniques from the explanation community to determine a set of relevant speech acts.[1] For each speech A_i, perform steps (2)–(5).
2. **Identify a family of emotive behaviors F_i to exhibit when performing A_i.** Using the emotive annotations in the behavior speech act structure, index into the behavior space to determine a relevant family of emotive behaviors F_i.
3. **Select an emotive behavior B_i that belongs to F_i.** Either by using additional contextual knowledge, e.g., the level of complexity of the current problem, or simply randomly when all elements of F_i are relevant, select an element of F_i.
4. **Select a verbal utterance U_i that is appropriate for performing A_i.** Using a audio library of voice clips that is analogous to physical behaviors, extract a relevant voice clip. The voice clip may be either *connective* (e.g., "but" or "and"), *phrasal* (e.g., "this subnet is fast,"), or *sentential* (i.e., a full sentence).
5. **Coordinate the exhibition of B_i with the speaking of U_i.** Couple B_i with U_i on the evolving presentation timeline.
6. **Establish visual continuity between $B_1 \dots B_n$.** Examine the final frame of each B_i, compare it with the initial frame of each B_{i+1}, and if they differ, introduce transition frames between them.

The resulting behaviors are then presented directly in the learning environment and control is immediately returned to the learner. The net effect of the sequencing engine's activities is the learner's perception that an expressive lifelike character is carefully observing their problem-solving activities and behaving in a visually compelling manner.

4 An Implemented Full-Body Emotive Pedagogical Agent

The emotive-kinesthetic behavior sequencing framework has been implemented in Cosmo, a lifelike (stylized) pedagogical agent that inhabits the Internet Advisor learning environment. Cosmo and the Internet Advisor environment are implemented in C++ using the Microsoft Game Software Developer's

[1] For example, the particular class of explanations focused on in the current implementation were inspired by McCoy's seminal work on discourse schemata for correcting misconceptions [10].

Kit (SDK). Cosmo's behaviors run at 15 frames/second with 16 bits/pixel color on a Pentium Pro 200 Mhz PC with 64 MB of RAM. Cosmo, as well as the routers and subnets in the virtual Internet world, were modeled and rendered in 3D on SGIs with Alias/Wavefront. The resulting bitmaps were subsequently post-edited with Photoshop and AfterEffects on Macintoshes and transferred to PCs where users interact with them in a $2\frac{1}{2}$D environment. Cosmo has a head with movable antennae and expressive blinking eyes, arms with bendable elbows, hands with a large number of independent joints, and a body with an accordion-like torso. This allows him to perform a variety of behaviors including pointing, blinking, clapping, and raising and bending his antennae. His verbal behaviors (recorded by a voice actor) include 200 utterances ranging in duration from 1–20 seconds.

Students interact with COSMO as they navigate through a series of subnets to learn about network routing mechanisms. Given a packet to escort through the Internet, they direct it through several networks connected by routers. At each subnet, the learner uses factors such as address resolution, subnet type, and traffic congestion to decide whether they should send their packet to a router on the current subnet, or should view adjacent subnets. Learners' journeys are complete when they have successfully navigated the network and have delivered their packet to the proper destination.

Suppose a student has just routed her packet to a fiber optic subnet with low traffic. She surveys the connected subnets and selects a router which she believes will advance it one step closer to the packet's intended destination. Although she has chosen a reasonable subnet, it is suboptimal because of non-matching addresses, which will slow her packet's progress. The emotive behavior planner chooses speech acts and then relevant emotive behaviors as follows. First, it uses its deictic behavior planner [15] to refer to the onscreen subnet information and direct COSMO to say, "You chose the fastest subnet." It then directs him to perform a micro-movement, exhibiting a blink, as he says, "Also, ..." Following that, he refers to the traffic information ("...it has low traffic.") Then, the emotive behavior planner selects a congratulatory and enthusiastic applauding behavior to indicate the learner made a great choice on those factors. As COSMO then poses a rhetorical question ("But more importantly, if we sent the packet here, what will happen?"), the emotive behavior sequencer directs him to scratch his head inquisitively. Next, because the packet will not arrive at its intended destination (a deleterious effect) since the learner chose a poor address, the behavior sequencer directs him to behaved disappointedly by using a sad facial expression, slumping body language, and dropping his hands as he says, "If that were the case, we see it doesn't arrive at the right place." After another remark, he employs a restrained manipulator in the form of finger tapping as he explains the role of addresses in internet routing. Finally, he points out a better choice, which the learner enacts.

The emotive-kinesthetic behavior sequencing framework has been "stress tested" in a very informal focus group study in which 10 students interacted with Cosmo for approximately half an hour each[2]. The subjects of the study (7 men and 3 women with ages ranging from 14 to 54) expressed genuine delight in interacting with Cosmo. Their typical reaction was that he was fun, engaging, interesting, and full of charisma. Although a few subjects voiced the opinion that Cosmo was overly dramatic, almost all exhibited particularly strong positive responses when Cosmo performed the exaggerated congratulatory behaviors. In short, they seemed to find him very entertaining and his advice very helpful.

5 Conclusions and Future Work

Because of their strong lifelike presence, animated pedagogical agents offer significant potential for playing the dual role of providing clear problem-solving advice and keeping learners highly motivated. By endowing them with the ability to exhibit full-body emotive behaviors to achieve situated emotive communication, to complement problem-solving advice, and to exhibit realtime visual continuity, an emotive behavior sequencing engine can select and assemble expressive behaviors in realtime. In the emotive-kinesthetic behavior sequencing framework for dynamically planning lifelike pedagogical agents' full-body emotive expression, a behavior sequencing engine can navigate a behavior space populated with a large repertoire of emotive behaviors. By exploiting the structure provided by pedagogical speech act categories, it can weave small expressive behaviors into larger visually continuous ones that are then exhibited by the agent in response to learners' problem-solving activities.

This work represents a small step towards the larger goal of creating fully interactive and fully expressive lifelike pedagogical agents. To make significant progress in this direction, it will be important to develop a comprehensive theory of pedagogical speech acts and leverage increasingly sophisticated computational models of affective reasoning. We will be pursuing these lines of investigation in our future work.

Acknowledgements

Thanks to Dorje Bellbrook, Tim Buie, Charles Callaway, Mike Cuales, Jim Dautremont, Amanda Davis, Rob Gray, Mary Hoffman, Alex Levy, Will Murray, Roberta Osborne, and Jennifer Voerman of the North Carolina State University IntelliMedia Initiative for their work on the behavior sequencing engine implementation and the 3D modeling, animation, sound, and environment design for the Internet Advisor. Support for this work was provided by the following organizations: the National Science Foundation under grants CDA-9720395 (Learning and Intelligent Systems Initiative) and IRI-9701503 (CAREER Award

[2]The study simultaneously studied two aspects of the agent's behavior sequencing: its ability to generate unambiguous deictic references, which also yielded encouraging results, and the overall effects of his emotive behaviors.

Program); the North Carolina State University IntelliMedia Initiative; Novell, Inc.; and equipment donations from Apple and IBM.

References

1. Elisabeth André and Thomas Rist. Coping with temporal constraints in multimedia presentation planning. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, pages 142–147, 1996.
2. Joseph Bates. The role of emotion in believable agents. *Communications of the ACM*, 37(7):122–125, 1994.
3. Shamus Culhane. *Animation from Script to Screen*. St. Martin's Press, New York, 1988.
4. P. Dillenbourg, P. Jermann, D. Schneider, D. Traum, and C. Buiu. The design of MOO agents: Implications from an empirical cscw study. In *Proceedings of Eighth World Conference on Artificial Intelligence in Education*, pages 15–22, 1997.
5. Christopher R. Eliot and Beverly P. Woolf. A simulation-based tutor that reasons about multiple agents. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, pages 409–415, 1996.
6. Clark Elliott. *The Affective Reasoner: A Process Model of Emotions in a Multi-agent System*. PhD thesis, Northwestern University, 1992.
7. Claude Frasson, Thierry Mengelle, and Esma Aimeur. Using pedagogical agents in a multi-strategic intelligent tutoring system. In *Proceedings of the AI-ED '97 Workshop on Pedagogical Agents*, pages 40–47, 1997.
8. James C. Lester, Sharolyn A. Converse, Susan E. Kahler, S. Todd Barlow, Brian A. Stone, and Ravinder Bhogal. The persona effect: Affective impact of animated pedagogical agents. In *Proceedings of CHI'97 (Human Factors in Computing Systems)*, pages 359–366, Atlanta, 1997.
9. James C. Lester, Sharolyn A. Converse, Brian A. Stone, Susan E. Kahler, and S. Todd Barlow. Animated pedagogical agents and problem-solving effectiveness: A large-scale empirical evaluation. In *Proceedings of Eighth World Conference on Artificial Intelligence in Education*, pages 23–30, Kobe, Japan, 1997.
10. Kathleen F. McCoy. Generating context-sensitive responses to object-related misconceptions. *Artificial Intelligence*, 41:157–195, 1989 1990.
11. Jeff Rickel. Personal communication, January 1998.
12. Jeff Rickel and Lewis Johnson. Integrating pedagogical capabilities in a virtual environment agent. In *Proceedings of the First International Conference on Autonomous Agents*, pages 30–38, 1997.
13. Steven Ritter. Communication, cooperation, and competition among multiple tutor agents. In *Proceedings of Eighth World Conference on Artificial Intelligence in Education*, pages 31–38, 1997.
14. Brian A. Stone and James C. Lester. Dynamically sequencing an animated pedagogical agent. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, pages 424–431, Portland, Oregon, 1996.
15. Stuart Towns, Charles Callaway, Jennifer Voerman, and James Lester. Coherent gestures, locomotion, and speech in life-like pedagogical agents. In *Proceedings of the Fourth International Conference on Intelligent User Interfaces*, pages 13–20, San Francisco, 1998.
16. Wen-Cheng Wang and Tak-Wai Chan. Experience of designing an agent-oriented programming language for developing social learning systems. In *Proceedings of Eighth World Conference on Artificial Intelligence in Education*, pages 7–14, 1997.