Achieving Affective Impact: Visual Emotive Communication in Lifelike Pedagogical Agents

James C. Lester, Stuart G. Towns, Patrick J. FitzGerald *Multimedia Laboratory, Department of Computer Science, North Carolina State University, Raleigh, NC 27695.*
lester@csc.ncsu.edu, www.multimedia.ncsu.edu/imedia

Abstract. Lifelike animated agents for knowledge-based learning environments can provide timely, customized advice to support learners' 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 article 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.

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 (Dillenbourg *et al.*, 1997; Eliot & Woolf, 1996; Frasson, 1997; Ritter, 1997; Chan and Chan, 1997). In particular, *animated pedagogical agents* (Lester *et al.*, 1999a; Rickel & Johnson, 1999; Stone & Lester, 1996) 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 a pedagogical agent in the form of an intelligent lifelike character.

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. In short, to create the illusion of life typified by well

crafted animated characters, animated pedagogical agents must be able to communicate through both visual and aural channels.

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 the 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 *agents* (Lester *et al.*, 1999b). 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.

This article is structured as follows. Section 2 outlines the communicative functionalities that animated pedagogical agents should provide to learners. Section 3 describes the emotive-kinesthetic behavior sequencing framework, including methods for designing emotive-kinesthetic behavior spaces, for structuring these spaces with pedagogical speech acts, and the algorithm for dynamically sequencing emotive behaviors in realtime. Section 4 presents an implemented animated pedagogical agent, COSMO, that employs the emotive-kinesthetic behavior sequencing framework, illustrates its operation in a problem-solving episode, and describes an informal focus group study with COSMO. The article concludes with a discussion of directions for future work.

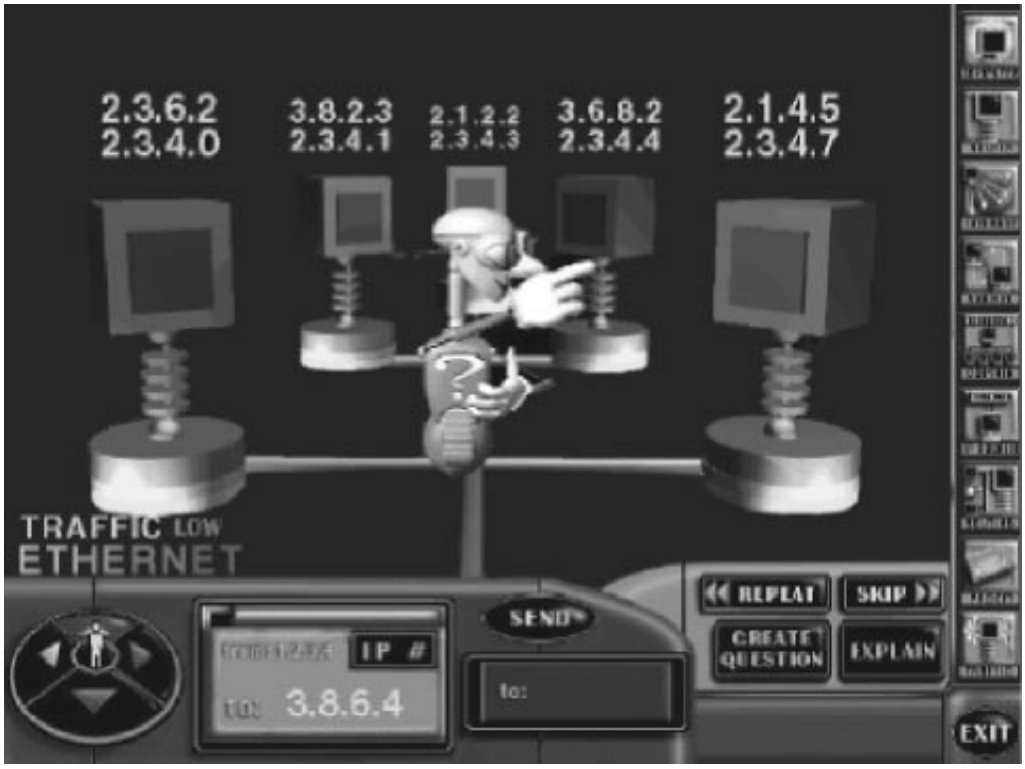

Figure 1. COSMO and the INTERNET ADVISOR learning environment

PEDAGOGICAL AGENTS

Although knowledge-based graphical simulations (Hollan *et al.*, 1987) are virtually *de rigueur* in contemporary learning environments, it is only in recent years, as a result of rapid advances in multimedia technologies, that full-scale intelligent multimedia interfaces have become standard components through which tutoring systems can provide clear visual feedback to learners. A particularly promising line of work underway outside of the intelligent tutoring systems community is that of lifelike animated intelligent agents. Because of these agents' compelling visual presence and their high degree of interactivity, there has been a surge of interest in *believable* intelligent characters (André & Rist, 1996; Bates, 1994; Blumberg & Galyean, 1995; Granieri *et al.*, 1995; Kurlander & Ling, 1995), including the runtime incorporation of gesture and facial expression in communication (Cassell, 1999; Pelechaud *et al.*, 1996).

As a result of these developments, the ITS 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 a variety of pedagogical agents that can facilitate the construction of component-based tutoring system architectures and communication between their modules (Chan and Chan, 1997), provide multiple context-sensitive pedagogical strategies (Frasson, 1997), reason about multiple agents in learning environments (Eliot & Woolf, 1996), provide assistance to trainers in virtual worlds (Maraella & Johnson, 1998), and act as co-learners (Dillenbourg *et al.*, 1997). 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 (André & Rist, 1996; Johnson *et al.*, 1998; Paiva & Machado, 1998; Rickel & Johnson, 1997; Stone & Lester,

1996). It is this second category, lifelike animated pedagogical agents, that is the focus of the work described here.

Creating lifelike pedagogical agents that are endowed with facilities for exhibiting learner-appropriate emotive behaviors potentially provides four important educational benefits (Elliott *et al.*, 1999). First, a pedagogical agent that appears to care about a learner's progress may convey to the learner that it and she are "in things together" and may encourage the learner to care more about her own progress. Second, an emotive pedagogical agent that is in some way sensitive to the learner's progress may intervene when she becomes frustrated and before she begins to lose interest. Third, an emotive pedagogical agent may convey enthusiasm for the subject matter at hand and may foster similar levels of enthusiasm in the learner. Finally, a pedagogical agent with a rich and interesting personality may simply make learning more fun. A learner that enjoys interacting with a pedagogical agent may have a more positive perception of the overall learning experience and may consequently opt to spend more time in the learning environment.

In short, lifelike pedagogical agents seem to 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 (Reeves & Nass, 1998), recent evidence suggests that ITSs with lifelike characters can be pedagogically effective (Lester *et al.*, 1997b), while at the same time having a strong motivating effect on learners (Lester *et al.*, 1997a). For example, the latter study, which was conducted with one hundred middle school students, demonstrated that well-designed pedagogical agents are perceived as being very helpful, credible, and entertaining. It is even becoming apparent that particular features, e.g., personal characteristics, of lifelike agents, can have an important impact on learners' acceptance of them (Hietala & Niemirepo, 1998).

In the same manner that human-human communication is characterized by multi-modal interaction utilizing both the visual and aural channels, agent-human communication can be achieved in a similar fashion. As master animators have discovered repeatedly over the past century, the quality, overall clarity, and dramatic impact of communication can be increased through the creation of emotive movement that underscores the affective content of the message to be communicated:

> **Situated Emotive Communication**: By carefully orchestrating facial expression, full-body behaviors, arm movements, and hand gestures, animated pedagogical agents could visually augment verbal problem-solving advice, give encouragement, convey empathy, and perhaps increase motivation.

Although work has been underway for several years on two large-scale projects involving lifelike pedagogical agents, STEVE and DESIGN-A-PLANT, 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 (Rickel, 1998), its focus to date has been on the realtime generation of behaviors using a visually simple agent, originally based on the JACK model (Granieri *et al.*, 1995). The DESIGN-A-PLANT project (Stone & Lester, 1996) has produced effective animated pedagogical agent technologies that are the creation of a multidisciplinary team of ITS researchers and animators. However, research on its behavior sequencing mechanisms has not addressed realtime inference about the creation of full-body emotive behaviors. Finally, initial forays have begun on emotion generation in pedagogical environments (Abou-Jaoude & Frasson, 1998) and reasoning about learners' emotions (de Vicente & Pain, 1998), indicating the potential richness offered by affective learner-system interactions.

Animated pedagogical agents can be introduced into learning environments with a variety of forms and functions. In this work, we make the following three simplifying assumptions about the role and form of the agent. First, it assumes that only one agent inhabits the learning environment and this agent serves as a "coach." Second, it assumes that a full-body agent is used. While emotions can be communicated solely with facial expressions, employing a body including arms enables the agent to gesture emotively. Third, it assumes that an explanation system is used to drive the content and organization of the agent's advice. While the explanation system's decisions may be informed by a student model or plan recognition system—in fact the implemented explanation system uses a simple overlay student model (Carr & Goldstein, 1977)—the emotive behavior sequencing framework described here only requires that the explanation system *somehow* provides the content and organization of the advice that will be presented.

THE EMOTIVE-KINESTHETIC BEHAVIOR 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 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 selecting full-body emotive behaviors that the agent can perform to communicate the affective impact appropriate for those speech act categories, the behavior sequencing engine identifies 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.

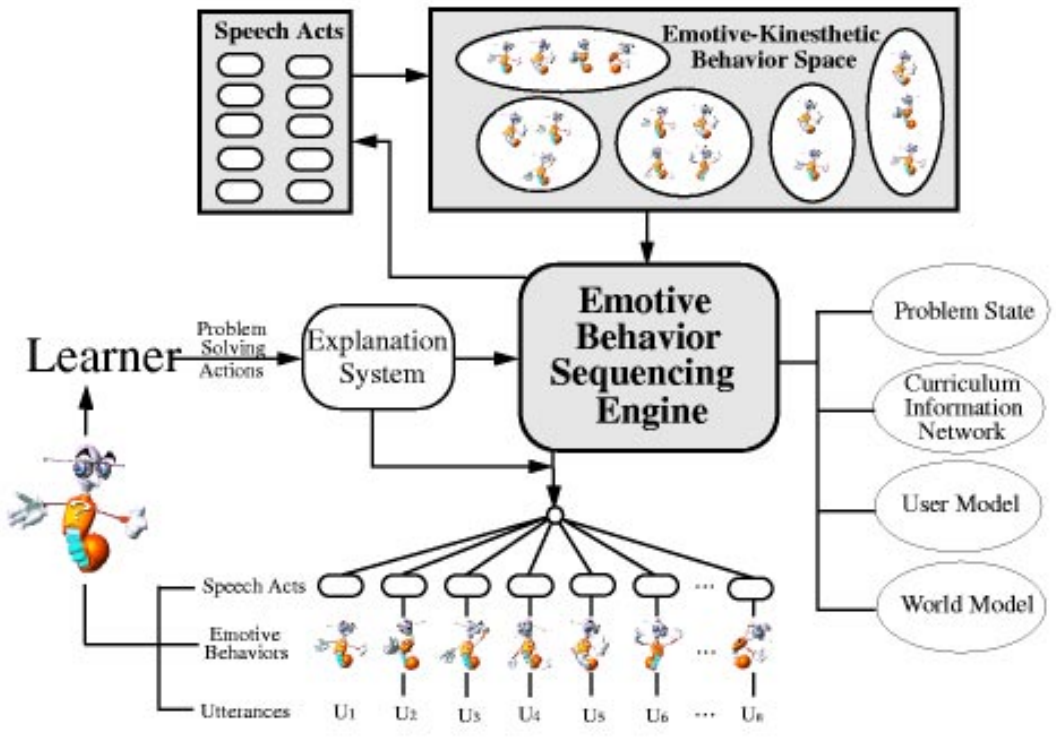

Figure 2. The Emotive-Kinesthetic Behavior Sequencing Architecture

The techniques for designing emotive-kinesthetic behavior spaces, structuring them with pedagogical speech act categories, and the computational mechanisms that drive the emotive behavior sequencing engine are described below.

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 narrative utterances.[1] By exploiting the character behavior canon of the animated film (Culhane, 1988) (which itself drew on movement in theater) and then adapting it to the specific demands posed by learning environments, we can extract general emotive animation techniques that artists in this medium have developed over the past hundred years.

[1] An important technical decision in creating an emotive behavior space is the decision of whether the agent's utterances will be created by a voice actor or via natural language generation (NLG) coupled with speech synthesis. Although NLG plays a central role in the authors' research programme, e.g., (Lester & Porter, 1997), because of the current quality of speech synthesizers, it was determined that the COSMO agent's behavior space should be populated with utterances created by a professional voice actor. As speech synthesis improves, the authors believe that NLG for emotive pedagogical agents will become an increasingly important research issue.

Stylized Emotive Behaviors

It is important to draw a critical distinction between two approaches to animated character realization, life-quality vs. stylized (Culhane, 1988). 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, although a consistency is obeyed, the laws of physics (and frequently the laws 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 (Culhane, 1988), 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. Not all stylized animation features such exaggerated emotive overstatement—for learning environments, a more restrained approach is called for—but its ability to communicate with dramatic visual cues can be put to good use in the realtime animation of pedagogical agents. For example, when a learner solves a complex problem in the INTERNET ADVISOR environment, the COSMO agent smiles broadly and uses his entire body to applaud the learner's success.

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. However, because their role is primarily to facilitate positive learning experiences, only a critical subset of the full range of emotive expression is useful for pedagogical agents. 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. The COSMO agent, for instance, can scratch his head in wonderment when he poses a rhetorical question.

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's activities. Consequently, after the behavior space has been populated with expressive behaviors, it must then be structured to assist the sequencing engine in selecting and assembling behaviors that are appropriate for the agent's communicative goals. 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*. While it could be indexed by a

full theory of speech acts, our research to date leverages a highly specialized collection of speech acts that occur in pedagogical dialogue with great frequency.

Given the primacy of the speech act in this approach, the question then arises about the connection between rhetorical goals on the one hand and physical behaviors on the other. This linkage is supplied by emotive categories inspired by foundational research on affective reasoning. Work on the Affective Reasoner (AR) (Elliott, 1992) uses Ortony's computational model of emotion to design agents' that can respond emotionally. In the AR framework, agents are given unique pseudo-personalities modeled as both an elaborate set of *appraisal frames* representing their individual goals (with respect to events that arise), *principles* (with respect to perceived intentional actions of agents), *preferences* (with respect to objects), *moods* (temporary changes to the appraisal mechanism), and as a set of about 440 differentially activated *channels* for the expression of emotions (Elliott, 1992; Elliott & Ortony, 1992). Situations that arise in the agents' world may map to twenty-six different emotion types (e.g., *pride*, as approving of one's own intentional action), twenty-two of which were originally theoretically specified by Ortony and his colleagues (Ortony *et al.*, 1988). Qualities, and intensity, of emotion instances in each category are partially determined by some subset of roughly twenty-two different *emotion intensity variables* (Elliott & Siegle, 1993). To communicate with users, Elliott's implementation of the AR framework uses line-drawn facial expressions, which are morphed in real time.

The emotive-kinesthetic behavior sequencing framework exploits the fundamental intuition behind the AR, namely, that the emotive states and communication are intimately interrelated. It creates emotive annotations that connect pedagogical speech acts to relevant physical behaviors. Computationally, this is accomplished by employing a model of communication that places pedagogical speech acts in a *one-to-one* mapping to emotive states: each speech act type points to the behavior type that expresses it. 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 is performed in which the agent communicates 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.

The one-to-one mapping is used to enact a three-fold adaptation of the AR framework. First, while the AR is intended to be generic, the emotive-kinesthetic behavior framework is designed specifically to support problem-solving advisory communication. Second, while the AR framework is enormously complex, the emotive-kinesthetic framework employs only the speech acts and only the emotive intentions that arise frequently in tutorial situations.[2] Third, while work on computational models of social linguistics indicates that the combination of speech and gesture in human-human communication is enormously complex (Cassell *et al.*, 1994), the one-to-one mapping approach turns out in practice to be a reasonable starting point for realtime emotive behavior sequencing.

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 into and exits from the learning environment. Moreover, sometimes the agent must connect two behaviors induced by multiple utterances that are generated by two speech acts. To achieve these `rhetorical link` behaviors, it employs subtle "micro-movements" such as slight head nods or blinking.

Dynamic Emotive Behavior Sequencing

As students solve problems in the learning environment, the pedagogical agent provides advice to assist them. In the course of observing a learner attempt different solutions, the agent explains concepts and gives hints. It provides advice in two situations: (1) when a student pauses for an extended period of time, which may signal a problem-solving impasse, and (2) when a learner proposes a solution that is either incorrect or sub-optimal. When it has been determined that the agent should provide advice, the emotive behavior sequencing engine is invoked. First, an explanation planner determines the content and structure of explanations by examining a curriculum information network, a simple overlay user model (Carr & Goldstein, 1977), the current problem state, and the learner's proposed solution. It constructs a sequence of explanatory behaviors and explanations (typically 6-10 utterances) which will collectively constitute the advice that will be delivered. In this way, problem-solving actions performed by the learner are punctuated by customized explanations delivered by the agent.

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 $A_1 \ldots A_n$ used to achieve *G*.** When the explanation system is invoked, employ a top-down goal decomposition

[2] An extensive discussion of adapting the Affective Reasoning framework to emotive models of tutoring may be found in (Elliott *et al.*, 1999).

planner to determine a set of relevant speech acts. For each speech act 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 structuring, 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 from the library of utterances 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.

5. **Coordinate the exhibition of B_i with the speaking of U_i.** Couple B_i with U_i on the evolving timeline schedule.

6. **Establish visual continuity between $B_1 \ldots B_n$.** 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.

First, the behavior sequencing engine must determine the content and organization of the problem-solving advice to be communicated (Step 1). To do so, it performs a function that is analogous to that performed by discourse planners of natural language generation systems (Cawsey, 1992; Hovy, 1993; Lester & Porter, 1997; Mittal, 1993; Moore, 1995; Suthers, 1991). Natural language generators typically consist of a discourse planner that determines the content and structure of multi-sentential texts and a realization system that plans the surface structure of the resulting prose. Analogously, given a communicative goal, the emotive behavior sequencing engine uses the by-now-classic techniques of goal decomposition planning to determine the content and structure of the agent's explanations. For example, the particular class of explanations focused on in the current agent implementation were inspired by McCoy's seminal work on discourse schemata for correcting misconceptions (McCoy, 1989-90). The sequencing engine typically first points out the strong points (if any) of the learner's proposed solution, then compares and contrasts it with the properties that an ideal solution would exhibit. The leaves of the resulting hierarchical plan are instantiated speech acts that will achieve the initial top-level communicative goal.

For each speech act A_i identified by the sequencing engine above, it performs the following actions. First, during Step 2, it identifies a family of emotive behaviors F_i that can be exhibited while the agent is performing A_i. It accomplishes this by employing pedagogical speech act indices that have been used to index the agent's physical behavior space. For example, a `congratulatory` speech act created during top-down planning will cause the sequencing engine to identify the `admiration` emotive behavior family.

Next, during Step 3, it selects one of the physical behaviors in F_i. By design, all of the behaviors have the same emotive intent, so they are all legitimate candidates. However, because a key aspect of agent believability is exhibiting a variety of behaviors, the behavior space was constructed so as to enable the agent to perform a broad range of facial expression and gestures. Hence, the sequencing engine selects from a collection of behaviors, any of which will effectively communicate the relevant emotive content. For example, in the current implementation of the COSMO agent, the behavior sequencing engine makes this decision pseudo-randomly with elimination, i.e., it randomly selects from among the behaviors in F_i that have not already been marked as having been performed. After all behaviors in a given F_i have been performed, they are unmarked, and the process repeats. Empirical evidence suggests that this pseudo-random element contributes significantly to believability.

During the final three steps the behavior sequencing engine determines the narrative utterances to accompany the physical behaviors and assembles the specifications on an evolving timeline. In Step 4, it selects the narrative utterances U_i, which are of three types: *connective* (e.g., "but" or "and"), *phrasal*, e.g., "this subnet is fast" or *sentential*, i.e., a full sentence. Because each instantiated speech act specifies the verbal content to be communicated, narrative utterance selection is straightforward. In Step 5, it lays out the physical behaviors and verbal utterances in tandem on a timeline. Because the emotive physical behaviors were determined by the same computational mechanism that determined the utterances, the sequencing engine can couple their exhibition to achieve a coherent overall behavior.

Finally, in Step 6, it ensures that the visual continuity is achieved by introducing appropriate transition frames. To do so, for each of the visual behaviors selected above, it inspects the first and final frames. If adjacent behaviors are not visually identical, it splices in visual transition behaviors and installs them, properly sequenced into the timeline. As it delivers advice, sometimes the agent must refer to objects in the environment through judicious combination of gesture, locomotion, and speech. It employs a *deictic behavior planner* (Lester *et al.*, 1999b) to make these decisions. In addition, for purposes of believability, the agent is always in subtle but constant motion, even when it is not delivering advice. COSMO, for example, typically performs "anti-gravity bobbing" and blinking behaviors as learners solve problems.

The sequencing engine passes all behaviors and utterances to the learning environment, which cues them up and orchestrates the agent's actions and speech in realtime. 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. The resulting behaviors are then exhibited by the agent in the learning environment and control is immediately returned to the learner who continues her problem-solving activities.

AN IMPLEMENTED 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^{++} and employ the Microsoft Game Software Developer's Kit (SDK). COSMO's

behaviors run at 15 frames/second with 16 bits/pixel color on a Pentium Pro 200 Mhz PC with 128 MB of RAM. He 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. His speech was supplied by a voice actor. 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 $2^1/_2$D environment. COSMO can perform a variety of behaviors including locomotion, pointing, blinking, leaning, clapping, and raising and bending his antennae. His verbal behaviors include 240 utterances ranging in duration from 1-20 seconds.

COSMO's behavior sequencing engine operates according to the framework outline above. Given a request to explain a concept or to provide a hint, the behavior planner selects the explanatory content by examining the curriculum information network (a partially ordered structure of topics and skills) and the user model (a representation of the individual problem-solving skills previously demonstrated by the learner). Explanatory content is determined in large part by the current problem state, which includes both the logical state of the problem and the student's proposed solution. Problems in the INTERNET ADVISOR are defined by factors such as the current packet's destination address, subnet type, IP numbers for the computers and routers on the current subnet, and network congestion.

Learners interact with COSMO as they study network routing mechanisms by navigating through a series of subnets. Given a packet to escort through the Internet, they direct it through networks of connected routers. At each subnet, they may send their packet to a specified router and view adjacent subnets. By making decisions about factors such as address resolution and traffic congestion, they learn the fundamentals of network topology and routing mechanisms. Helpful, encouraging, and with a bit of attitude, COSMO explains how computers are connected, how routing is performed, what types of networks have particular physical characteristics, how address schemes work, and how traffic considerations come into play. Learners' journeys are complete when they have successfully navigated the network and 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 sub-optimal because of non-matching addresses, which will slow her packet's progress. Working in conjunction with the deictic behavior planner, the emotive behavior planner chooses pedagogical speech acts and the relevant emotive behaviors as follows.

- ***State-Correct(Subnet-Type)***: The learning environment determines that the agent should interject advice and invokes the sequencing engine. As a result of the deictic behavior planner's directives, COSMO moves towards and points at the onscreen subnet information and says, "You chose the fastest subnet."

- ***State-Correct(Traffic)***: COSMO then tells the student that the choice of a low traffic subnet was also a good one. The gesture focus history indicates that, while the type of subnet has already been the subject of a deictic reference, the traffic information

has not. COSMO therefore moves to the onscreen congestion information and points to it. However, the utterance focus history indicates that he has mentioned the subnet in a recent utterance, he pronominalizes the subnet as "it" and says, "Also, it has low traffic."

- ***Congratulatory()***: Responding to a *congratulatory* speech act, the sequencing engine selects an *admiration* emotive intent which is realized with an enthusiastic applauding behavior as COSMO exclaims, "Fabulous!"

- ***Causal()***: The sequencing engine's planner selects a *causal* speech act, which causes the *interrogative* emotive behavior family to be selected. These include actions such as head scratching and shrugging, for which the desired effects are to emphasize a questioning attitude. Hence, because COSMO wants the student to rethink her choice, he scratches his head and poses the question, "But more importantly, if we sent the packet here, what will happen?"

- ***Deleterious-Effect(Address-Resolution)***: After the *causal* act, the sequencing engine's planner now selects a *deleterious-effect* speech act, which causes it to index into the *disappointment* behavior family. It includes behaviors that indicate sadness, which is intended to build empathy with the learner. COSMO therefore informs the learner of the ill-effect of choosing that router as he takes on a sad facial expression, slumping body language, and dropping his hands, and says, "If that were the case, we see it doesn't arrive at the right place."

- ***Rationale(Address-Resolution)***: To explain the reason why the packet won't arrive at the correct destination, COSMO adds, "This computer has no parts of the address matching," as he moves and gestures to the problematic computer.

- ***Background(Address-Resolution)***: The sequencing engine has selected a *background* speech act. Because all *background* and *assistance* speech acts cause the sequencing engine to index into the *inquisitive* behavior family, it obtains one of several "thoughtful" restrained manipulators such as hand waving. In this case, it selects a form of finger tapping which he performs as he explains, "Addresses are used by networked computers to tell each other apart."

- ***Assistance(Address-Resolution)***: Finally, COSMO assists the learner by making a suggestion about the next course of action to take. Because she has committed several mistakes on address resolution problems, COSMO provides advice about correcting her decision by pointing to the location of the optimal computer and stating, "This router has two parts of the address matching."

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. The subjects of the study were 7 men and 3 women with ages ranging from 14 to 54. All of the subjects expressed genuine delight in interacting with COSMO. Their typical reaction was that he was fun, engaging, interesting, and full of charisma. Taking into account the important caveat that the study was very limited, the findings are nonetheless informative. Although some subjects voiced the opinion that COSMO was overly dramatic, almost all exhibited

particularly strong positive responses when he performed the congratulatory behaviors. In short, they seemed to find him very entertaining and his advice very helpful.

It is also important to note the limitations of the framework. First, because the sequencing engine does not employ a natural language generation system, it's flexibility is necessarily limited by the narrative utterances of the behavior space. As the quality of speech produced by synthesizers improves, generation will undoubtedly come to the forefront of research on lifelike pedagogical agents. Second, the subjects' perception that COSMO is overly dramatic is a by-product of his initial design by the animation team. In creating pedagogical agents, it is critical to take into account the target learner audience, and an important feature of this is the personality characteristics of the users themselves (Isbister & Nass, 1998). Third, in interacting with COSMO, it quickly becomes clear that his emotions tend to come and go very quickly. While this is certainly in keeping with the stylized approach to character animation, it could become a distraction over time. Further theoretical work needs to be done to create sequencing engines that smooth out emotive transitions and provide mechanisms for the attenuation of emotive expression.

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, the behavior sequencing engine navigates a behavior space populated with a large repertoire of full-body 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 addressing the limitations of the framework noted above and 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. Thanks also to Charles Callaway and Bradford Mott for comments on an

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

REFERENCES

Abou-Jaoude, S. and Frasson, C. (1998). Emotion computing in competitive learning environments. In *Working Notes of the ITS '98 Workshop on Pedagogical Agents*, pages 33-39, San Antonio, Texas.

André, E. and Rist, T. (1996). Coping with temporal constraints in multimedia presentation planning. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, pages 142-147.

Bates, J. (1994). The role of emotion in believable agents. *Communications of the ACM*, 37(7):122-125.

Blumberg, B. and Galyean, T. (1995). Multi-level direction of autonomous creatures for real-time virtual environments. In *Computer Graphics Proceedings*, pages 47-54.

Carr, B. and Goldstein, I.P. (1977). Overlays: A theory of modeling for computer aided instruction. Technical Report AI Memo 406, Massachusetts Institute of Technology, Artificial Intelligence Laboratory.

Cassell, J. (1999). Embodied conversation: Integrating face and gesture into automatic spoken dialogue systems. In Luperfoy, S., editor, *Automatic Spoken Dialogue Systems*. MIT Press. Forthcoming.

Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S., and Stone, M. (1994). Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In *SIGGRAPH'94*.

Cawsey, A. (1992). *Explanation and Interaction: The Computer Generation of Explanatory Dialogues*. MIT Press.

Chan, W.W., and Chan, T. (1997). Experience of Designing and Agent-Oriented Programming Language for Developing Social Learning Systems. In *Proceedings of Eighth World Conference on Artificial Intelligence in Education*, pages 7-14, Kobe, Japan.

Culhane, S. (1988). *Animation from Script to Screen*. St. Martin's Press, New York.

de Vicente, A. and Pain, H. (1998). Motivation diagnosis in intelligent tutoring systems. In *Proceedings of the Fourth International Conference on Intelligent Tutoring Systems*, pages 86-95, San Antonio, Texas.

Dillenbourg, P., Jermann, P., Schneider, D., Traum, D., and Buiu, C. (1997). 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.

Eliot, C.R. and Woolf, B.P. (1996). A simulation-based tutor that reasons about multiple agents. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, pages 409-415.

Elliott, C. (1992). *The Affective Reasoner: A Process Model of Emotions in a Multi-agent System*. PhD thesis, Northwestern University.

Elliott, C. and Ortony, A. (1992). Point of view: Reasoning about the concerns of others. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*, pages 809-814, Bloomington, IN.

Elliott, C., Rickel, J., and Lester, J. (1999). Lifelike pedagogical agents and affective computing: An exploratory synthesis. In Wooldridge, M. and Veloso, M., editors, *Artificial Intelligence Today*, pages 195-212. Springer-Verlag, Berlin.

Elliott, C. and Siegle, G. (1993). Variables influencing the intensity of simulated affective states. In *AAAI Spring Symposium on Reasoning about Mental States: Formal Theories and Applications*, pages 58-67, Palo Alto, CA.

Frasson, C., Mengelle, T., and Aimeur, E. (1997). Using pedagogical agents in a multi-strategic intelligent tutoring system. In *Proceedings of the AI-ED '97 Workshop on Pedagogical Agents*, pages 40-47.

Granieri, J. P., Becket, W., Reich, B.D., Crabtree, J., and Badler, N.I. (1995). Behavioral control for real-time simulated human agents. In *Proceedings of the 1995 Symposium on Interactive 3D Graphics*, pages 173-180.

Hietala, P. and Niemirepo, T. (1998). The competence of learning companion agents. *International Journal of Artificial Intelligence in Education*, 9(3-4):178-192.

Hollan, J.D., Hutchins, E.L., and Weitzman, L.M. (1987). STEAMER: An interactive, inspectable, simulation-based training system. In Kearsley, G., editor, *Artificial Intelligence and Instruction: Applications and Methods*, pages 113-134. Addison-Wesley, Reading, MA.

Hovy, E.H. (1993). Automated discourse generation using discourse structure relations. *Artificial Intelligence*, 63:341-385.

Isbister, K. and Nass, C. (1998). Personality in conversational characters: Building better digital interaction partners using knowledge about human personality preferences and perceptions. In *Workshop on Embodied Conversational Characters*, pages 103-111, Tahoe City, California.

Johnson, W.L., Shaw, E., and Ganeshan, R. (1998). Pedagogical agents on the web. In *Working Notes of the ITS '98 Workshop on Pedagogical Agents*, pages 2-7, San Antonio, Texas.

Kurlander, D. and Ling, D.T. (1995). Planning-based control of interface animation. In *Proceedings of CHI '95*, pages 472-479.

Lester, J.C., Converse, S.A., Kahler, S.E., Barlow, S.T., Stone, B.A., and Bhogal, R. (1997a). The persona effect: Affective impact of animated pedagogical agents. In *Proceedings of CHI '97 (Human Factors in Computing Systems)*, pages 359-366, Atlanta.

Lester, J.C., Converse, S.A., Stone, B.A., Kahler, S.E., and Barlow, S.T. (1997b). 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.

Lester, J.C. and Porter, B.W. (1997). Developing and empirically evaluating robust explanation generators: The KNIGHT experiments. *Computational Linguistics*, 23(1):65-101.

Lester, J., Stone, B., and Stelling, G. (1999a). Lifelike pedagogical agents for mixed-initiative problem solving in constructivist learning environments. *User Modeling and User-Adapted Interaction*, 9(1-2):1-44.

Lester, J., Voerman, J., Towns, S., and Callaway, C. (1999b). Deictic believability: Coordinating gesture, locomotion, and speech in lifelike pedagogical agents. *Applied Artificial Intelligence*, 13(4-5):383-414.

Marsella, S. and Johnson, W.L. (1998). An instructor's assistant for team-training in dynamic multi-agent virtual worlds. In *Proceedings of the Fourth International Conference on Intelligent Tutoring Systems*, pages 464-473, San Antonio, Texas.

McCoy, K. (1989 1990). Generating context-sensitive responses to object-related misconceptions. *Artificial Intelligence*, 41:157-195.

Mengelle, T., Léan, C. D., and Frasson, C. (1998). Teaching and learning with intelligent agents: Actors. In *Proceedings of the Fourth International Conference on Intelligent Tutoring Systems*, pages 284-293, San Antonio, Texas.

Mittal, V. (1993). *Generating Natural Language Descriptions with Integrated Text and Examples*. PhD thesis, University of Southern California.

Moore, J. D. (1995). *Participating in Explanatory Dialogues*. MIT Press.

Ortony, A., Clore, G.L., and Collins, A. (1988). *The Cognitive Structure of Emotion*. Cambridge University Press.

Paiva, A. and Machado, I. (1998). Vincent, an autonomous pedagogical agent for on-the-job training. In *Proceedings of the Fourth International Conference on Intelligent Tutoring Systems*, pages 584-593, San Antonio, Texas.

Pelachaud, C., Badler, N.I., and Steedman, M. (1996). Generating facial expressions for speech. *Cognitive Science*, 20(1):1-46.

Reeves, B. and Nass, C. (1998). *The Media Equation: How People Treat Computers, Television and New Media Like Real People and Places*. CSLI, New York.

Rickel, J. (1998). Personal communication.

Rickel, J. and Johnson, L. (1997). Integrating pedagogical capabilities in a virtual environment agent. In *Proceedings of the First International Conference on Autonomous Agents*, pages 30-38.

Rickel, J. and Johnson, W. L. (1999). Animated agents for procedural training in virtual reality: Perception, cognition, and motor control. *Applied Artificial Intelligence*, 13(4-5):343-382.

Ritter, S. (1997). Communication, cooperation, and competition among multiple tutor agents. In *Proceedings of the Eighth World Conference on Artificial Intelligence in Education*, pages 31-38.

Stone, B.A. and Lester, J.C. (1996). Dynamically sequencing an animated pedagogical agent. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, pages 424-431, Portland, Oregon.

Suthers, D. (1991). A task-appropriate hybrid architecture for explanation. *Computational Intelligence*, 7(4):315-333.

Towns, S., Callaway, C., Voerman, J., and Lester, J. (1998). 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.

Wang, W.-C. and Chan, T.-W. (1997). 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.