Coherent Gestures, Locomotion, and Speech in Life-Like Pedagogical Agents

Stuart G. Towns **Charles B. Callaway**

Jennifer L. Voerman **James C. Lester**

Multimedia Laboratory
Department of Computer Science
North Carolina State University
Raleigh, NC 27695-7534
(919) 515-7534 Fax: (919) 515-7896
{sgtowns,cbcallaw,jlvoerma,lester}@eos.ncsu.edu

ABSTRACT

Life-like animated interface agents for knowledge-based learning environments can provide timely, customized advice to support students' problem solving. Because of their strong visual presence, they hold significant promise for substantially increasing students' enjoyment of their learning experiences. A key problem posed by life-like agents that inhabit artificial worlds is *deictic believability*. In the same manner that humans refer to objects in their environment through judicious combinations of speech, locomotion, and gesture, animated agents should be able to move through their environment, and point to and refer to objects appropriately as they provide problem-solving advice. In this paper we describe a framework for achieving deictic believability in animated agents. A deictic behavior planner exploits a world model and the evolving explanation plan as it selects and coordinates locomotive, gestural, and speech behaviors. The resulting behaviors and utterances are believable, and the references are unambiguous. This approach to spatial deixis has been implemented in a life-like animated agent, Cosmo, who inhabits a learning environment for the domain of Internet packet routing. The product of a large multidisciplinary team of computer scientists, 3D modelers, graphic artists, and animators, Cosmo provides realtime advice to students as they escort packets through a virtual world of interconnected routers.

KEYWORDS: Animated agents, life-like, believability, learning environments, educational applications.

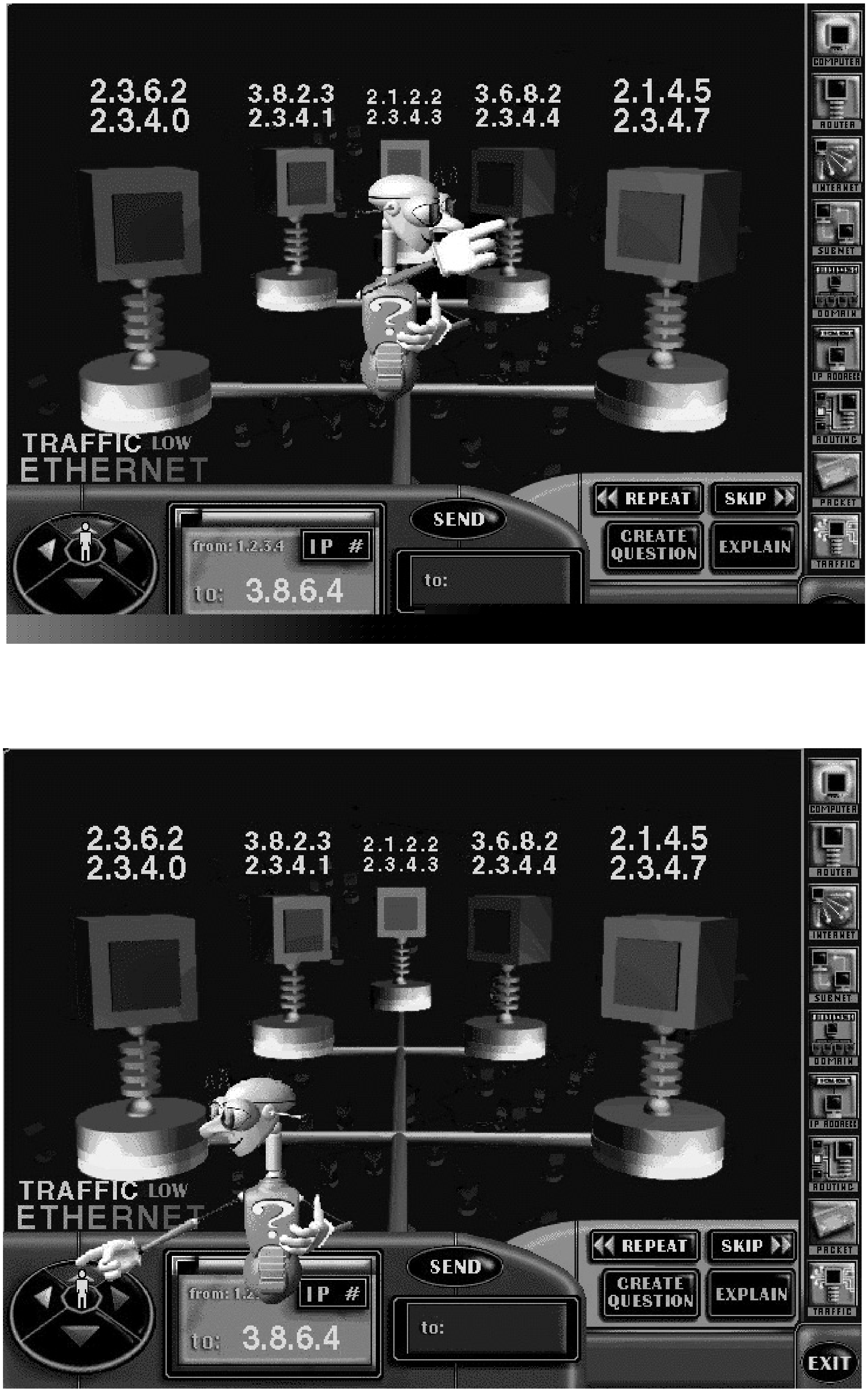
INTRODUCTION

Life-like animated agents offer great promise for knowledge-based learning environments. Because of the immediate and deep affinity that children seem to develop for them, the potential pedagogical benefits these agents provide are perhaps even exceeded by their motivational benefits. By creating the illusion of life, animated agents may significantly increase the time that children seek to spend with educational software, and recent advances in affordable graphics hardware are beginning to make the widespread distribution of real-time animation technology a reality.

Endowing animated agents with believable, life-like qualities has been the subject of a growing body of research [1, 2, 3, 7, 11, 17] and much interesting work has examined the social aspects of human-computer interaction and users' anthropomorphization of software [12, 13]. Animated pedagogical agents [14, 16] constitute an important category of animated agents whose intended use is educational applications. Believability is a key feature of animated pedagogical agents, as demonstrated by a recent large-scale study conducted with an animated pedagogical agent developed in our laboratory. After interacting with this life-like agent in a problem-solving environment, it was determined that students perceived the agent as being very helpful, credible, and entertaining [8].

A key problem posed by life-like agents that inhabit artificial worlds is *deictic believability*. In the same manner that humans refer to objects in their environment through combinations of speech, locomotion, and gesture, animated agents should be able to move through their environment, point to objects, and refer to them appropriately as they provide problem-solving advice. Deictic believability in animated agents requires the design of an agent behavior planner that considers the physical properties of the world inhabited by the agent. The agent must exploit its knowledge of the positions of objects in the world, its relative location with respect to these objects, as well as its prior explanations to create deictic gestures, motions, and utterances that are both natural and unambiguous.

To address these issues, we have developed a spatial deixis framework for achieving deictic believability. Building on our previous work on dynamically sequencing animated pedagogical agents [16] and enhancing pedagogical agent be-



(a)

(b)

Figure 1: Cosmo and the INTERNET ADVISOR World

lievability [10] and on Cassell *et al*'s foundational work on agent deixis [4], a deictic behavior planner exploits a world model and the evolving explanation plan as it selects and coordinates locomotive, gestural, and speech behaviors. The resulting behaviors are believable, and by considering the relative proximity of objects, the references are unambiguous. This approach has been implemented in a life-like animated agent, Cosmo, who inhabits a learning environment for the domain of Internet packet routing.

Cosmo is an impish, antenna-bearing creature who hovers in a virtual world of routers and networks and provides advice to students as they decide how to ship packets through the networks to specified destinations (Figure 1). His appearance, mannerisms, and behavior space of actions and utterances are the combined creation of a large multidisciplinary team of computer scientists, graphic artists, modelers, and animators. In response to students' problem-solving activities and questions, Cosmo interjects explanations that refer to specific routers, subnets and address labels in the environment. By carefully selecting and coordinating speech, gesture, and locomotion, his behavior planner creates deictic references that are natural and unambiguous. Focus group studies with students interacting with Cosmo in the INTERNET ADVISOR learning environment are encouraging.

DEICTIC BELIEVABILITY IN LIFE-LIKE ANIMATED PEDAGOGICAL AGENTS

In the course of communicating with one another, humans constantly employ deictic techniques to create context-specific references. *Spatial deixis*, a much studied phenomenon in linguistics, is used to refer to specific locations and objects in the physical world [6]. Speakers use these techniques to narrow listeners' attention to particular entities. In one popular psycho-social framework for analyzing spatial deixis, the *figure-ground* model [15], the world is categorized into *ground*, which is the common physical environment shared by the speaker and hearer, and the *referent*, which is the aspect of the ground to which the speaker wishes to refer. Through carefully constructed referring expressions and well-chosen gestures, the speaker assists the hearer in focusing on the particular referent of interest.

The ability to handle *spatial deixis* effectively is especially critical for animated pedagogical agents that inhabit virtual worlds. To provide problem-solving advice to students who are interacting with objects in the world, the agent must be able to refer to objects in the world to clearly explain their function and to assist students in performing their tasks. Deictic mechanisms for animated pedagogical agents should satisfy two criteria:

- *Lack of Ambiguity:* In a learning environment, an animated agent's clarity of expression is of the utmost importance. To effectively communicate advice and explanations to students, the agent must be able to create deictic references that are unambiguous. Avoiding ambiguity is critical for problem solving in virtual environments, where an ambiguous deictic reference can cause mistakes and foster misconceptions. Ambiguity is particularly challenging in virtual environments housing a multitude of objects, especially when many of the objects are similar.
- *Immersivity:* It has been demonstrated that agents that are believable [2] hold much appeal for children [8]. Hence, though it is reasonably simple to create unambiguous references to objects in a scene through object highlighting, our goal is to create life-like agents that make references in much the same manner that humans do. Just as humans are immersed in their environment and gesture and move within it, e.g., by walking across a scene to a cluster of objects and pointing to one of them, to achieve believability, agents should behave accordingly.

The lack-of-ambiguity requirement implies that deictic planning mechanisms must make use of an expressive representation of the world. While unambiguous deictic references can be made by object highlighting or by employing a relatively stationary agent with a long pointer, e.g., [1], the immersivity requirement suggests that agents should artfully combine speech, gesture, and locomotion.

A Life-like Animated Agent Testbed

To investigate life-like animated agents for learning environments, we have developed a life-like agent and the testbed learning environment he inhabits.[1] The environment and agent were designed to foster evaluation of mechanisms for animation behavior sequencing of life-like characters, human-agent conversational interaction, and realtime problem-solving assistance. For example, environmental features that force spatial deixis issues to the forefront are (1) a world populated by a multitude of objects, many of which are similar, (2) an agent that provides advice and explanations that must refer to these objects, and (3) a problem-solving task that requires students to make decisions based on factors present in the environment. The **Internet Advisor** (Figure 1) provides such a "deictic laboratory." It consists of a virtual world with many routers and networks and inhabited by Cosmo, a helpful antenna-bearing creature with a hint of a British accent. The INTERNET ADVISOR serves as an excellent testbed for exercising spatial deixis because each subnet has a variety of routers attached to it and the agent must refer unambiguously to them as it advises the students about their problem solving.

Students interact with Cosmo as they learn about 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. Students' journeys are complete when they have successfully navigated the network and delivered their packet to the proper destination.

[1] In addition to the authors, the INTERNET ADVISOR was created by 9 graphic artists (environment designers, 3D modelers, and animators), as well as a musician, a voice actor, and several programmers.

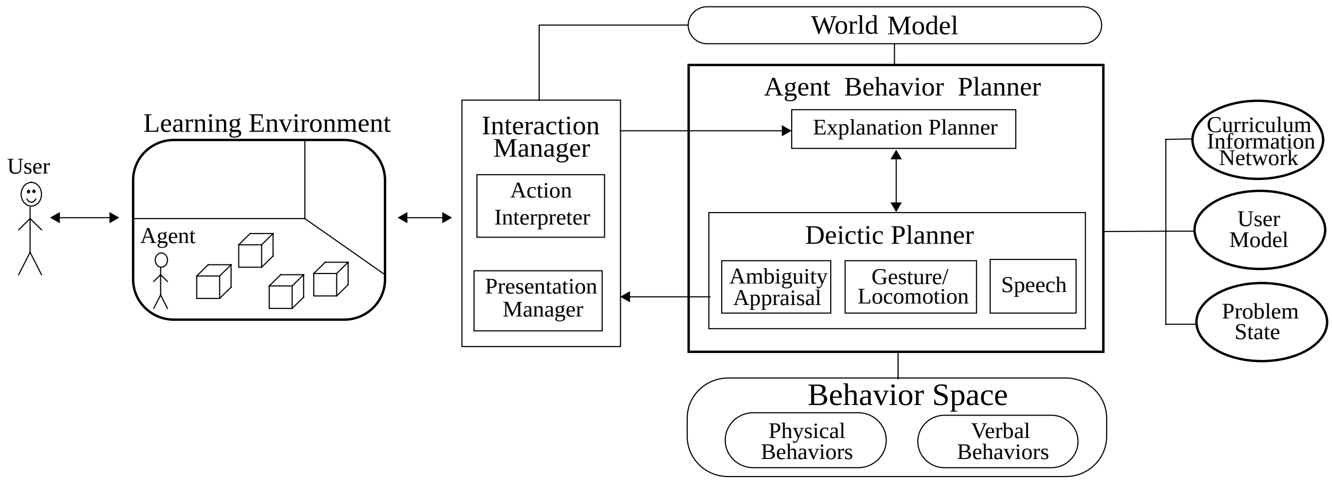


Figure 2: Life-like Animated Pedagogical Agent Behavior Planning Architecture

PLANNING DEICTIC GESTURE, LOCOMOTION, AND SPEECH

As students solve problems in the learning environment, the animated agent provides advice to assist them. In the course of observing a student 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, or (2) when a student commits an error. When the action interpreter detects a situation in which the agent should provide advice, it invokes the agent behavior planner (Figure 2). The agent behavior planner consists of an explanation planner and a deictic planner. The explanation planner determines the content and structure of explanations by examining a curriculum information network, a user model, the current problem state, and the student's 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 student are punctated by customized explanations delivered by the agent.

Deictic planning comes into play when the behavior planner determines that an explanation must refer to an object in the environment. For each utterance that makes a reference to an environmental object, the explanation planner invokes the deictic system and supplies it with the intended referent R. The deictic system operates in the following three phases to plan the agent's gestures, locomotion, and speech:

1. **Ambiguity Appraisal:** The deictic system first assesses the situation by determining whether a reference to R may be ambiguous. By examining the evolving *explanation plan*, which contains a record of the objects the agent has refered to in utterances spoken so far in the current explanation sequence, the deictic planner evaluates R's initial potential for ambiguity. This assessment will contribute to gesture, locomotion, and speech planning decisions.

2. **Gesture and Locomotion Planning:** To determine the agent's physical actions, the deictic system uses both a *world model* representing the relative positions of the objects in the scene, as well as the previously made ambiguity assessment to plan the agent's pointing gestures and movement. By considering the proximity of objects in the world, the deictic system computes whether the agent should point to R, and if so, whether it should move to R.

3. **Utterance Planning:** To determine what the agent should say to refer to R, the deictic system considers focus information, the ambiguity assessment, and the world model. Utterance planning pays particular attention to the relative locations of the referent and the agent, taking into account its planned locomotion from the previous phase. The result of utterance planning is a referring expression consisting of the appropriate proximal/non–proximal demonstratives and pronouns.

The deictic system passes the agent's behaviors and utterances back to the agent behavior planner which integrates them into the final behavior specifications. These typically include a variety of non-deictic actions and utterances as well. The behavior planner then passes the specifications to the presentation engine, which extracts the relevant behaviors from the behavior space, cues them up, and orchestrates the agent's actions and speech in the environment.

Ambiguity Appraisal

For each utterance in the evolving explanation plan that makes a reference to an object in the environment, the explanation planner invokes the deictic system. Deictic decisions depend critically on an accurate assessment of the discourse context in which the reference will be communicated. To correctly plan the agent's gestures, movements, and utterances, the deictic system determines whether the situation has the potential for ambiguity within the current explanation.[2]

Because focus indicates the prominence of the referent at the current juncture in the explanation, the deictic system uses focus as the primary predictor of ambiguity: potentially

[2]This initial phase of ambiguity assessment considers only discourse issues; spatial considerations are handled in the following two phases.

ambiguous situations can be combatted by combinations of gesture and locomotion.

A referent R has the potential for ambiguity if it is currently not in focus or it is in focus but is one of multiple objects in focus. To determine if the referent is in focus, the deictic system examines the evolving explanation plan and inspects it for previous deictic references to R. Suppose the explanation planner is currently planning utterance U_i. It examines utterances U_{i-1} and U_{i-2} for preceding deictic references to R. There are three cases:

- *Novel Reference:* If the explanation planner locates no decitic reference to R in U_{i-1} or U_{i-2},then R is ambiguous, and is therefore deserving of greater deictic emphasis. For example, if a student interacting with the INTERNET ADVISOR chooses to send a packet to a particular router which does not lie along the optimal path to the packet's destination, Cosmo interrupts the student and makes an initial reference to that router. He should therefore provide a proper introduction to the referent.
- *Unique Focus:* If the explanation planner locates a reference to R in U_{i-1} and U_{i-2} but not to other entities, then R has already been introduced and the potential for ambiguity is less. For example, when Cosmo's explanation consists of multiple utterances about a particular router, a reference to that router will be in unique focus. Consequently, the need for special deictic treatment is less.
- *Multiple Foci:* If the explanation planner locates a reference to R but also to other entities in U_{i-1} and U_{i-2}, then the situation is potentially ambiguous. For example, if Cosmo points to one router and subsequently points to another which the student has just selected, but he now needs to refer to the first router again for purposes of comparison, multiple referents are in focus and he should therefore take precautions against making an ambiguous reference.

The result of this determination is recorded for use in the following two phases.

Gesture and Locomotion Planning

When potential ambiguities arise, endowing the agent with the ability to point and move to objects to which it will be referring enables it to improve its clarity of reference. The deictic system plans two types of physical behaviors: gestures and locomotion. In each case, it first determines whether a behavior of that type is warranted. If so, it then computes the behavior. To determine whether the agent should exhibit a pointing gesture to physically designate the referent within the environment, the behavior planner inspects the conclusion of the ambiguity computation in the previous phase. If the referent was deemed ambiguous or potentially ambiguous, the system will plan a pointing gesture for the agent.

In addition to pointing, the agent can also move from one location to another to clarify a deictic reference which otherwise might be ambiguous. If the referent has been determined to be unambiguous, i.e., it is in a unique focus, the agent will remain stationary.[3] In contrast, if the referent is ambiguous, i.e., if it is a novel reference, the deictic system instructs the agent to move towards the object specified by the referent as the agent points at it. For example, if Cosmo is discussing a router which has not been previously mentioned in the last two utterances, he will move to that router as he points to it. If the referent is potentially ambiguous, i.e., it is a reference to one of the concurrently active foci, then the Deictic Planer must decide if locomotion is needed. If no locomotion is needed, the agent will point at R without moving towards it. In contrast, if any of the following three conditions hold, the agent will move towards R as it points:

- *Multiple Proximal Foci:* If the object specified by R is near another object that is also in focus, the agent will move to the object specified by R. For example, if two nearby routers are being compared, Cosmo will move to the router to which he is referring to ensure that his reference is clear.
- *Multiple Proximal Similarity:* Associated with each object is an ontological category. If the object specified by R is near other objects of the same category, the agent will move to the object specified by R. For example, if Cosmo were referring to a router and there were several routers nearby, he would move to the intended router.
- *Diminutiveness:* If the object specified by R is unusually small, the agent will move to the object specified by R. Small objects are labeled as such in the world model. For example, many interface control buttons are relatively small compared to objects in the environment. If Cosmo needs to make a clear reference to one of them, he will move toward that button.

Once a sequence of gestures and possible locomotions are computed, the sequence is returned to the *interaction manager* which determines the direction the agent needs to point. To enable the agent to correctly point to the object specified by the referent, the interaction manager first consults the world model. It obtains the location of the agent (L_A) and the referent (L_R) in the environment. It then determines relative orientation of the vector from (L_A) to (L_R). For example, Cosmo might be hovering in the upper right-hand corner of the environment and need to point to a router in the lower left-hand corner. In this case, he will point down and to the left towards the router. The interaction manager must also decide whether or not the agent really needs to move based on his current location. If the deictic system determines that locomotion is called for, the interaction manager must first determine if the agent is already near the object, which would obviate the need for moving towards it. Nearness of two objects is computed by measuring the distance between them and ascertaining whether it is less than a *proximity bound*. If the distance between the agent and the intended object is less than the proximity bound, then there is no need for the agent to move because it can already clearly point to the object, and so it will remain in its current position. If locomotion is appropriate, the system computes a direct motion path from the agent's current location to the object specified by R. Finally, so that the pointing will be performed en route, it computes a rate of locomotion so that the duration of the pointing action will complete when the agent reaches the object.

[3] More precisely, the agent will not perform a locomotive behavior. In fact, for purposes of believability, the agent is always in motion, performing actions such as an "anti-gravity" bobbing motion.

Deictic Referring Expression Planning

To effectively communicate the intended reference, the deictic system must combine gesture, locomotion, and speech. Having completed gesture and locomotion planning, the deictic planner turns to speech. To determine an appropriate referring expression for the agent to speak as it performs the deictic gestures and locomotion, the deictic system first examines the results of the ambiguity appraisal. If it was determined that R is in a unique focus, there is no potential for ambiguity because R has already been introduced and no other entities are competing for the student's attention. It is therefore introduced with a simple referring expression using techniques similar to outlined in [5], e.g., "the router" or "it" will be pronominalized.

In contrast, if R is ambiguous or potentially ambiguous, i.e. R is a novel reference or is one of multiple foci, the deictic planner makes three assessments: (1) it determines the demonstrative category called for by the current situation; (2) it examines the ontological type of R and the other active foci; and (3) it considers the number of R. First, it categorizes the situation into one of two deictic families:

- *Proximal Demonstratives:* If the deictic planner determined that the agent must move to R or that it would have moved to R if it were not already near R, then employ a proximal demonstrative such as "this" or "these."
- *Non-Proximal Demonstratives:* If the deictic planner determined that R was not nearby but that the agent did not need to move to R, then employ a non-proximal demonstrative such as "that" or "those."

Second, after it has determined which of the demonstrative categories to use, the deictic planner narrows its selection further by considering the ontological type of R and the previous two utterances in the evolving explanation plan. If R belongs to the same ontological type as the other entities which are in focus, then the deictic planner selects the phrase, "This one . . . ." For example, suppose the system has determined that a proximal demonstrative should be used and that the preceding utterance referred to one router, e.g., "This router has more traffic." To refer to a second router in the current utterance, rather than saying, "This router has less traffic," it will say, "This one has less traffic."

Finally, it uses the number of R to make the final lexical choice. If R is singular, it uses "this" for proximal demonstratives and "that" for non-proximals. If R is plural, it uses "these" and "those." The final referring expression, as well as the selected gestures and locomotive actions, are then passed to the behavior planner, which integrates them into the other speech and behaviors and passes them to the interaction manager's presentation engine in the order in which the agent should exhibit them. The agent's combined behaviors are then sequenced in realtime, providing a convincing illusion of a sentient being delivering advice that integrates life-like locomotion, gesture, and speech.

AN IMPLEMENTED LIFE-LIKE ANIMATED PEDAGOGICAL AGENT

Cosmo (Figure 1) is a realtime implementation of a life-like animated agent that inhabits the INTERNET ADVISOR learning environment.[4] 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 $2\frac{1}{2}$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 200 utterances ranging in duration from 1–20 seconds.

Life-like Behavior Sequencing in Cosmo

Cosmo's behavior planner operates according to the coherence-based behavior sequencing framework for animated pedagogical agents [16].[5] Applying this framework to create an agent entails constructing a behavior space, imposing a coherence structure on it, and developing a behavior sequencing engine that dynamically selects and assembles behaviors. 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 user). 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.

When invoked, the explanation planner first consults the knowledge sources noted above to select a sequence of communicative acts. These acts include

- State-Correct: Provides information about the factors of the student's selection which were correct, e.g., showing which fields of an address match.
- Cause: Poses a question to the student with regards to what would happen if her choice were implemented.
- Effect: Answers the question just posed by the agent in the Cause utterance with regards to the current problem, e.g., telling the student that the packet will move through the network slowly.
- Rationale: Gives detail on why the student's choice was incorrect, e.g. showing the student that the current subnet is highly congested.
- Give-Background: Provides foundational information about entities in the domain, e.g., an explanation of a router's function.

[4] 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 64 MB of RAM.

[5] This framework was originally developed for the "Herman the Bug" agent who inhabits the DESIGN-A-PLANT design-centered learning environment [9].

- Assistance: Gives a hint, e.g. to try a subnet with lower traffic.

To facilitate the student's acquisition of problem-solving skills, the explanation planner frequently supplies relevant causality knowledge when the student requests advice, as well as justifications for its suggestions.

Cosmo remains onscreen at all times throughout problem-solving sessions to accompany students on their journeys through the Internet. He remains alive by blinking, performing his "anti-gravity" bobbing action, and occasionally leaning on one of the routers. When the student fails to take an action for an extended period of time or commits an error by making an incorrect or sub-optimal routing decision, the action interpreter invokes the explanation planner, which in turn invokes the deictic planner. Together, the explanation planner and deictic planner specify a series of utterances and actions that Cosmo should perform. These are passed to the interaction manager, which extracts the specified movements, gestures, and utterances from the behavior space and enables Cosmo to exhibit them. Each individual frame of every selected behavior is positioned and displayed in the environment according to the motion path specifications constructed by the interaction manager. The net effect of realtime behavior planning is the dynamic creation of a helpful, somewhat whimsical onscreen persona with a convincing presence.

Presenting Realtime Advice: An Example

To illustrate how the agent behavior planner produces deictic gestures, motion, and verbal advice as it provides problem-solving assistance in realtime, consider the following situation in an INTERNET ADVISOR learning session. 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. She has made multiple mistakes on address resolution already, and so the explanation is somewhat detailed. The behavior planner selects behaviors with the following communicative acts:

- State-Correct: The action interpreter determines that the agent should interject advice and invokes the deictic planner. Because no referents are in focus, Cosmo moves towards and points at the onscreen subnet information and says, "You chose the fastest subnet."
- State-Correct: Cosmo then tells the student that the choice of a low traffic subnet was also a good one. Because the amount of traffic is not in focus, Cosmo moves to the onscreen congestion information and points to it. However, because he had mentioned the subnet in the previous utterance, he refers to the subnet as "it" and says, "Also, it has low traffic. Fabulous!"
- Cause: Cosmo wants the student to rethink her choice, so he scratches his head and poses the question, "But more importantly, if we sent the packet here, what will happen?"
- Effect: Cosmo tells the student of the ill-effect of chosing that router, and saying dejectedly, "If that were the case, we see it doesn't arrive at the right place."
- Rationale: To give a reason why the packet won't arrive at the right place, he adds, "This computer has no parts of the address matching." This utterance is coupled with a pointing action to the chosen computer since it was not in focus.
- Background: To emphasize the previous point, the explanation planner choses a background utterance. "Addresses are used by networked computers to tell each other apart."
- Assistance: Finally, Cosmo assists the student by making a suggestion on what to do next. He knows where the correct computer is located, and therefore he points to that location (since it was not in focus) and states, "We want to chose a router on the subnet with the best matching address."

EVALUATION

To gauge the effectiveness of the spatial deixis framework for deictic believability in animated pedagogical agents, a focus group study was conducted with students interacting with Cosmo and the INTERNET ADVISOR learning environment. The exploratory study was designed to investigate (1) how well the spatial deixis approach produces explanations that are clear and helpful and (2) how an agent-based approach to deixis in learning environments compares with a non-agent-based approach. The subjects of the study were 7 men and 3 women. To obtain a broad spectrum of responses, subjects with a broad range of ages (14–54) were chosen. On average, each subject interacted with the INTERNET ADVISOR for 30 minutes.

Subjects interacted with two different versions of the INTERNET ADVISOR. In the first version, as students solved Internet routing problems, the agent's behavior planner selected and coordinated locomotive, gestural, and speech behaviors according to the method described above. In the second version, students solved the same type of internet routing problems, but rather than interacting with an agent, they interacted with an "agent-less" version of the system. In the agent-less version, a voice spoke the same advice as in the agent-based version, but deictic gesture and locomotion were absent. To compensate for these missing functionalities, the system flashed a blinking red line under the referent in the environment. To avoid an ordering bias, approximately half the subjects experienced the agent-based system first, and the other half experienced the agent-less system first.

Results of the study suggest that the spatial deixis framework produces clear explanations. Most participants understood the agents advice most of the time. Although some wished it were less dramatic and suggested alternative organizations for the advice, its clarity was positively received. Subject unanimously preferred the agent-based version over the agent-less version. Some suggested that a combination of agent gestures with the blinking indicators might be more effective than either in isolation. In general, especially given the age of the subjects, the agent's motivating role was surprisingly strong.

CONCLUSION

Because of their strong life-like presence, animated pedagogical agents can capture students' imaginations and play a critical motivational role in keeping them deeply engaged in a learning environment's activities. Believability is a key feature of life-like pedagogical agents, and deictic believability is an important characteristic for animated agents that inhabit artificial worlds. To dynamically sequence animated pedagogical agents in a manner that promotes deictic believability, we have designed the spatial deixis framework for planning gesture, locomotion, and speech. This framework has been used to implement Cosmo, a life-like animated agent for a testbed learning environment.

In this framework, an agent behavior planner evaluates potential ambiguities and exploits a world model representing position and orientation in the virtual world to plan the agent's deictic actions and utterances. To do so, it first examines the evolving discourse plan to ascertain the focus status of the referent. It then inspects the world model to determine the referent's proximity to similar objects and to the agent itself. In this manner, it determines whether to move and point to the referent and what type of demonstrative utterance the agent should use to indicate it. Finally, the behavior planner integrates the gestures, locomotion, and speech into a communicative acts specification that produces a sequence of utterances and actions that are unambiguous and believable. The net result of the behavior planning is a life-like character who provides clear problem-solving advice in realtime with a strong visual presence.

This work represents a promising first step towards the goal of creating enchanting life-like characters for learning environments. Perhaps the greatest challenges ahead lie in creating a full-scale learning environment that extends believability (deictic and other aspects) to accommodate rich conversational contexts and 3D learning environments. Although the INTERNET ADVISOR is currently a prototype, plans for fielding a full-scale 3D learning environment are currently underway.

ACKNOWLEDGEMENTS

Thanks to Dorje Bellbrook, Tim Buie, Mike Cuales, Jim Dautremont, Amanda Davis, Mary Hoffman, Alex Levy, Will Murray, and Roberta Osborne of the North Carolina State University IntelliMedia Initiative for their work on 3D modeling and animation for the INTERNET ADVISOR. Special thanks to Patrick FitzGerald of North Carolina State University's School of Design for leading this effort. Support for this work is provided by a grant from the National Science Foundation (Faculty Early Career Development Award IRI-9701503), the North Carolina State University IntelliMedia Initiative, and an industrial gift from Novell.

REFERENCES

1. E. André and T. Rist. Coping with temporal constraints in multimedia presentation planning. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, pages 142–147, 1996.
2. J. Bates. The role of emotion in believable agents. *Communications of the ACM*, 37(7):122–125, 1994.
3. B. Blumberg and T. Galyean. Multi-level direction of autonomous creatures for real-time virtual environments. In *Computer Graphics Proceedings*, pages 47–54, 1995.
4. J. Cassell, M. Steedman, N. Badler, C. Pelchaud, M. Stone, B. Douville, S. Prvost, and B. Achorn. Modelling the interaction between speech and gesture. Technical report, University of Pennsylvania, February 1994.
5. R. Dale. *Generating Referring Expressions*. MIT Press, 1992.
6. R. J. Jarvella and W. Klein, editors. *Speech, Place, and Action: Studies in Deixes and Related Topics*. John Wiley & Sons, 1982.
7. D. Kurlander and D. T. Ling. Planning-based control of interface animation. In *Proceedings of CHI '95*, pages 472–479, 1995.
8. J. C. Lester, S. A. Converse, S. E. Kahler, S. T. Barlow, B. A. Stone, and R. Bhogal. The persona effect: Affective impact of animated pedagogical agents. In *Proceedings of CHI'97 (Human Factors in Computing Systems)*, pages 359–366, 1997.
9. J. C. Lester, P. J. FitzGerald, and B. A. Stone. The pedagogical design studio: Exploiting artifact-based task models for constructivist learning. In *Proceedings of the Third International Conference on Intelligent User Interfaces*, pages 155–162, 1997.
10. J. C. Lester and B. A. Stone. Increasing believability in animated pedagogical agents. In *Proceedings of the First International Conference on Autonomous Agents*, pages 16–21, 1997.
11. P. Maes, T. Darrell, B. Blumberg, and A. Pentland. The ALIVE system: Full-body interaction with autonomous agents. In *Proceedings of the Computer Animation '95 Conference*, 1995.
12. C. Nass, Y. Moon, B. J. Fogg, B. Reeves, and D. C. Dryer. Can computer personalities be human personalities. *International Journal of Human-Computer Studies*, 43:223–239, 1995.
13. C. Nass, J. Steuer, L. Henriksen, and H. Reeder. Anthropomorphism, agency and ethopoeia: Computers as social actors. In *Proceedings of the International CHI Conference*, 1993.
14. J. Rickel and L. Johnson. Integrating pedagogical capabilities in a virtual environment agent. In *Proceedings of the First International Conference on Autonomous Agents*, pages 30–38, 1997.
15. L. D. Roberts. *How Reference Works: Explanatory Models for Indexicals, Descriptions, and Opacity*. State University of New York Press, 1993.
16. B. A. Stone and J. C. Lester. Dynamically sequencing an animated pedagogical agent. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, pages 424–431, 1996.
17. M. A. Walker, J. E. Cahn, and S. J. Whittaker. Improvising linguistic style: Social and affective bases of agent personality. In *Proceedings of the First International Conference on Autonomous Agents*, 1997.