Student-Sensitive Multimodal Explanation Generation for 3D Learning Environments*

Brent H. Daniel
Multimedia Laboratory
Department of Computer Science
North Carolina State University
Raleigh, NC 27695-8206

William H. Bares
Center for Advanced Computer Studies
University of Southwestern Louisiana
PO Box 44330
Lafayette, LA 70504-4330

Charles B. Callaway
Multimedia Laboratory
Department of Computer Science
North Carolina State University
Raleigh, NC 27695-8206

James C. Lester
Multimedia Laboratory
Department of Computer Science
North Carolina State University
Raleigh, NC 27695-8206

Abstract

Intelligent multimedia systems hold great promise for knowledge-based learning environments. Because of recent advances in our understanding of how to dynamically generate multimodal explanations and the rapid growth in the performance of 3D graphics technologies, it is becoming feasible to create multimodal explanation generators that operate in realtime. Perhaps most compelling about these developments is the prospect of enabling generators to create explanations that are customized to the ongoing "dialogue" in which they occur. To address these issues, we have developed a student-sensitive multimodal explanation generation framework that exploits a discourse history to automatically create explanations whose content, cinematography, and accompanying natural language utterances are customized to the dialogue context. By these means, they create integrative explanations that actively promote knowledge integration. This framework has been implemented in CINESPEAK, a student-sensitive multimodal explanation generator.

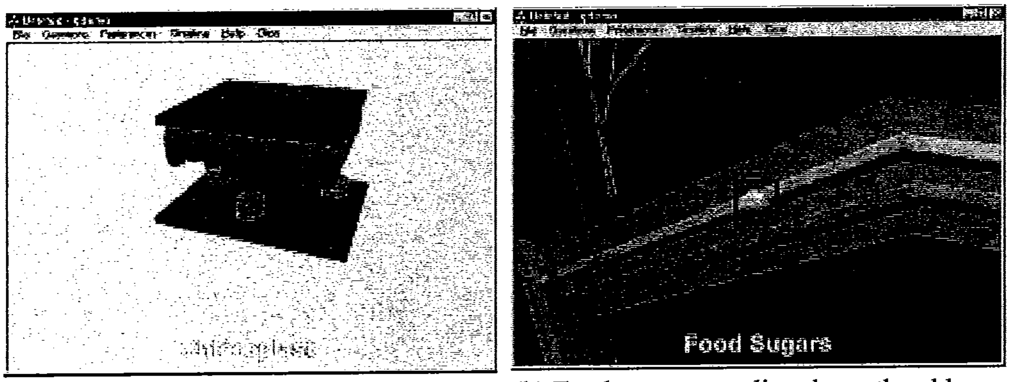
Introduction

Intelligent multimedia systems hold great promise for knowledge-based learning environments. Recent years have witnessed significant strides in techniques for automatically creating 2D graphics (Green *et al.* 1998; McKeown *et al.* 1992; Mittal *et al.* 1995; Wahlster *et al.* 1993), constructing static 3D illustrations (Seligmann & Feiner 1993), and generating 3D animations (Bares & Lester 1997; Butz 1997; Christianson *et al.* 1996; Karp & Feiner 1993; Seligmann & Feiner 1993; Towns, Callaway, & Lester 1998). Because of the central role that communication plays in learning, as graphics capabilities grow in sophistication, knowledge-based learning environments will significantly benefit from full-scale multimodal explanation generation.

Because of recent advances in our understanding of how to dynamically generate multimodal explanations and the rapid growth in the performance of 3D graphics technologies, it is becoming feasible to create multimodal explanation generators that operate in realtime with the compute power available in classrooms. Students will soon be able to pose questions to 3D learning environments depicting complex physical systems such as biological organisms or electro-mechanical devices. These environments will be able to dynamically create explanations by selecting content, constructing rhetorical structures, synthesizing animations, planning cinematic organizations, and generating natural language. As a result, students in the not-too-distant future may be able to interactively explore the physical systems found in science and engineering disciplines in a manner almost unimaginable a decade ago.

Perhaps most compelling about these developments is the prospect of enabling students to have interactions that are customized to the ongoing "dialogue" in which they participate. Multimodal explanation systems for learning environments should be able to dynamically create explanations whose content, rhetorical structure, cinematic organization, and natural language are highly sensitive to the multimodal discourse in which they are constructed. However, previous models of multimodal explanation generation have not focused on taking multimodal discourse context into account.

To address these problems, we propose the *multimodal discourse generation* framework for automatically creating customized explanations whose animation and natural language are tailored to the discourse



(a) Shot of the chloroplast (b) Food sugars traveling down the phloem

Figure 1: The CINESPEAK multimodal generator in the PLANTWORLD learning environment

in which they occur. In particular, this framework is designed to produce *integrative multimodal explanations* whose content and form are optimized for enabling students to integrate new knowledge into their current knowledge of the domain. By exploiting a discourse history of recently generated multimodal explanations and employing integrative multimodal explanation planning strategies, the framework dynamically produces extended animations whose content, cinematography, and accompanying natural language utterances are customized to the context in which they are generated. It has been implemented in CINESPEAK, a student-sensitive multimodal explanation system consisting of a media-independent explanation planner, a visuo-linguistic mediator, a 3D animation planner, and a realtime natural language generator with a speech synthesizer. CINESPEAK has been used in conjunction with PLANTWORLD (Figure 1), a prototype 3D learning environment for the domain of botany, to generate realtime multimodal explanations.

Integrative Multimodal Explanations

The context in which multimodal explanation generators for learning environments will operate consists of a complex matrix of knowledge activated by student-system exchanges. Multimodal explanation systems should therefore exploit this context to assist students in incorporating new knowledge into their existing set of beliefs about the domain. In their landmark work on reading comprehension, Adams and Bruce set forth rhetorical requirements for authors who wish to clearly communicate with their readers (Adams & Bruce 1982).

> The task of constructing an effective linguistic message consists of (1) correctly guessing what sorts of related knowledge the intended readers already have, (2) producing expressions that will evoke appropriate subsets of that knowledge, and (3) presenting those expressions in a way that will induce readers to interrelate the evoked knowledge into a structure that most nearly captures the meaning intended. (Adams & Bruce 1982)

Analogously, multimodal explanation systems should construct *integrative* explanations whose content, rhetorical structure, cinematography, and natural language promote knowledge integration. They should follow what have been termed *epistemological gradients*, which are "relations between the to-be-communicated and whatever guesses are available about the hearer's knowledge state" (Suthers 1993). Computationalizing such integrative desiderata entails providing solutions to each of the following problems:

- *Integrative Content and Structure:* The knowledge selected for presentation must not be determined in a vacuum but rather must take into account the discourse in which it will be presented (Moore 1995).
- *Integrative Animation Synthesis:* The visual elements of animations must be selected so that students can make visual connections between the imagery associated with currently understood concepts and new ones. Optimizing for knowledge integration can be significantly aided by showing new objects/processes in the physical context of familiar ones and by transitioning from known to new concepts.
- *Integrative Natural Language Generation:* Each of the traditional tasks of generating natural language, e.g., sentence planning and surface generation, must be performed in such a manner that the conceptual relations between familiar concepts and new concepts are communicated verbally.

Planning Customized Explanations

Generating integrative multimodal explanations hinges on dynamically performing each of the tasks above in a manner that takes into account the student's evolving

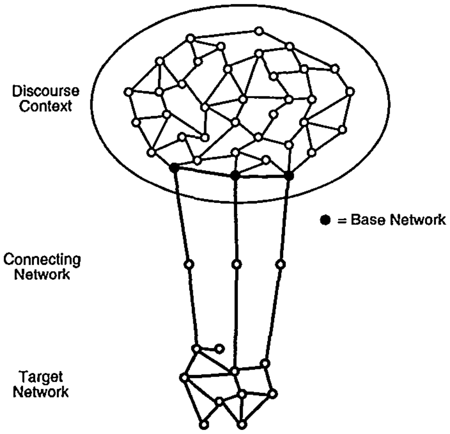


Figure 2: Integrative Explanation

interests and knowledge of the domain. To do so, explanation systems must be able to (1) recognize opportunities in the discourse for generating integrative explanations and (2) construct explanations' content and organization (including rhetorical, visual, and linguistic structure) in such a manner that they provide multiple links in multiple modalities from familiar concepts to new concepts. We formalize this notion of multimodal integrative explanations in a graph-theoretic manner. This framework defines three networks of relevant concepts and relations for integrative multimodal explanation generation (Figure 2). The *target* network is the set of concepts and relations that an explanation generator seeks to communicate to respond to the student's question. The *base* network is the set of concepts and relations that model what the student already understands and are relevant in some way to the target. The *connecting* network is the set of concepts and relations that relate the target to the base.

Identifying Target and Base Networks

The integrative framework identifies the target and base networks in different ways. Employing traditional content determination techniques, e.g., (Moore 1995; Suthers 1993), target networks are selected based on the the student's question type and the concept of interest. To illustrate, when requested to explain a process, the explanation system inspects the knowledge base to obtain the `location`, principle `actors`, and `sub-events` of the specified process. In contrast to target networks, which are determined by "context-insensitive" strategies, base networks must be cognizant of the ongoing discourse and students' interests. This knowledge is provided by two knowledge sources:

- **Multimodal Discourse History:** The explanation system will ultimately construct explanation trees whose leaves are annotated with media realization specifications (animation and/or natural language) for each proposition to be communicated. To track the ongoing explanation session, it maintains the sequence of trees for the past n explanations it has generated. This provides a rich, structured representation of (1) concepts the student has expressed interest in, (2) the communication strategies that have been employed to explain these concepts, and (3) the order in which the explanation session has unfolded.
- **Conceptual Centrality Specification:** The explanation system also maintains a *centrality concept* that indicates which concept the student would most benefit from new knowledge being "integrated into." For example, a high school student studying a curricular unit on food production in the domain of botanical anatomy and physiology might wish to have new information related to photosynthesis.

Creating Connecting Networks

For a given concept in the target network, the explanation system must obtain the concepts in the connecting and base networks that are appropriate for the student's current knowledge state. Typically, for a concept in the target network, there are a plethora of connections to other concepts. Moreover, many subsets of concepts in a knowledge base may be known to a given student at a particular juncture in a learning session. Hence, there may be many connections between concepts in the target network and concepts that are familiar to the student. Presenting arbitrary connections between new concepts and familiar concepts would result in confusing explanations that meander aimlessly. To combat this, the explanation system finds the connecting network by searching for the following categories of connections between (1) concepts T in the target network and *historical* concepts H in the base network or (2) concepts T in the target network and central concepts C in the base network.

- **Base Super-Structure**
  - *Triggering Connection:* Some H or C is a superstructure of some concept T.
  - *Integrative Visual/verbal Specifications:* ((Show H/C), (Say H/C super-structural-relationship T)).
  - *Sequencing:* Integrative → Target
  - *Example:* To generate an explanation of xylem, the system will note that the tree, a superstructure, has recently been discussed. It will therefore show the tree and say that the xylem is part of the tree.
- **Base Sub-Structure**
  - *Triggering Connection:* Some H or C is a substructure of some concept T.
  - *Integrative Visual/verbal Specifications:* ((Show T), (Say H/C sub-structural-relationship T)).
  - *Sequencing:* Integrative → Target
  - *Example:* To generate an explanation of a leaf, the system will note that the stomata, a part of the the leaf, is of central interest to the student. It

will therefore show the leaf and say that the leaf contains the stomata.

- **Base Categorical**
  - *Triggering Connection:* Either some H has the same superstructure as T, or some C has the same superstructure as T.
  - *Integrative Visual/verbal Specifications:* ((Show super-structure H/C T), (Say H/C categorical-relationship T)).
  - *Sequencing:* Integrative → Target
  - *Example:* To generate an explanation of a chloroplast, the system will note that the stomata, which is also a part of the leaf, has recently been discussed. It therefore shows the leaf and then says, "Like the stomata, the chloroplasts are located in the leaf."
- **Recent Pre-Temporal**
  - *Triggering Connection:* H and T are involved in the same process and the role of T in the process occurs immediately before the role of H in the process.
  - *Integrative Visual/verbal Specifications:* ((Show (Role-Of T wrt P)) (Say Temporal-Relation (Role-Of T wrt P) (Role-Of H wrt P))), where P is the process involved in the roles of H and T.
  - *Sequencing:* Target → Integrative
  - *Example:* To generate an explanation of chloroplasts, if the system has recently explained the role of phloem, it will first explain the basics of the role chloroplasts perform and then, to preserve appropriate temporal ordering, turn to the integrative knowledge. The latter will be communicated by showing the role of the chloroplasts and verbally explaining that the role of phloem (food sugar distribution) occurs directly after the role of the chloroplast (food sugar production) by generating the sentence, "After the chloroplasts produce food sugars, the food sugars are distributed by the phloem."
- **Recent Post-Temporal**
  - *Triggering Connection:* H and T are involved in the same process and the role of T in the process occurs immediately after the role of H in the process.
  - *Integrative Visual/verbal Specifications:* ((Show (Role-Of T wrt P)) (Say Temporal-Relation (Role-Of T wrt P) (Role-Of H wrt P))), where P is the process involved in the roles of H and T.
  - *Sequencing:* Integrative → Target
  - *Example:* To generate an explanation of xylem, if the system has recently explained the role of the chloroplast, it will show the role of the xylem and verbally explain that the role of the xylem (water and mineral transport) occurs directly before the role of the chloroplast (food sugar synthesis) by generating the sentence, "Before the chloroplast can produce food sugars, water and minerals must enter the chloroplast from the xylem."
- **Recurring Target**
  - *Triggering Connection:* H and T are identical (H was discussed initially in low detail.)
  - *Integrative Visual/verbal Specifications:* ((Show T) (Say T :higher-detail))
  - *Sequencing:* Increased detail.
  - *Example:* If the chloroplast was recently discussed and the student asks about it again, the system will generate a more detailed explanation that describes more specific layers in the process structure. In this case, it will include animated behaviors and utterances communicating the details of glucose production while relating them to the previous explanation, including narration such as, "As you saw earlier, chloroplasts produce food sugars for the plant."
- **Central Super-Process**
  - *Triggering Connection:* C is a super-process of T, i.e., T is a "step" of C.
  - *Integrative Visual/verbal Specifications:* ((Show (location C)) (Say Super-Process-Relationship C T))
  - *Sequencing:* Integrative → Target
  - *Example:* If the student (or the learning environment) has noted that photosynthesis is currently of central interest to the student and the student asks about the light reactions, the system will show the location of photosynthesis (the leaf) and say that the light reactions are a step of photosynthesis.
- **Central Sub-Process**
  - *Triggering Connection:* C is a sub-process of T, i.e., C is one of the "steps" of T.
  - *Integrative Visual/verbal Specifications:* ((Show (location T)) (Say Sub-Process-Relationship C T))
  - *Sequencing:* Integrative → Target
  - *Example:* If the learning environment asserts that in the current unit on the Calvin Cycle, new information should be explained in relation to the dark reactions, when the student asks about photosynthesis, the explanation system shows the location of photosynthesis (the leaf) while it explains that the dark reactions are a step of photosynthesis.

Student-Sensitive Multimodal Explanation Architecture

In the multimodal discourse generation framework (Figure 3), explanation construction begins when a student poses a question about a particular concept in the domain. For example, while perusing a 3D visualization of the leaf, the student might become curious about the chloroplast and ask about its physiological role. When

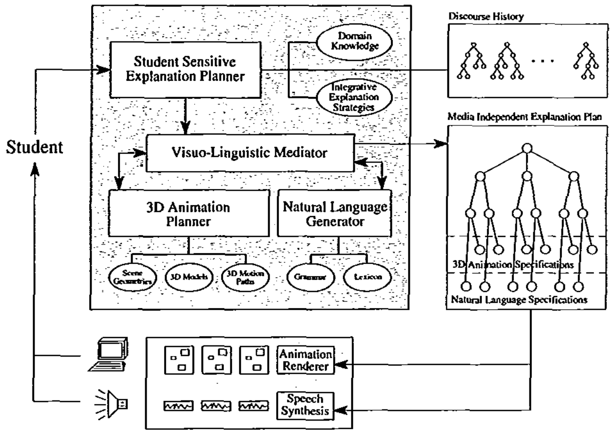


Figure 3: The student-sensitive multimodal explanation architecture

she poses a question, the student-sensitive explanation generator operates in three interleaved phases to construct an integrative multimodal explanation:

1. **Integrative Explanation Planning:** By inspecting the discourse context, which is represented by the discourse history and the current centrality concept, it selects the student-sensitive content, i.e., the propositions in the target, base, and connecting networks, and rhetorically organizes them.
2. **Integrative Animation Planning:** To visually communicate the selected integrative content, the animation planner directs the student's attention along Suthers' "epistemological gradients" by (a) selecting relevant 3D models, (b) selecting relevant 3D motion paths, and (c) issuing camera directives to the virtual camera.
3. **Integrative Natural Language Generation:** To verbally communicate the selected integrative content, the natural language generator (a) sends the propositions to a sentence planner that creates full sentential specifications, and (b) sends the resulting sentential specifications to a unification-based systemic-functional surface generator.

To construct explanation plans, the system consults the discourse history, extracts relevant propositions from the knowledge base, and organizes them into an explanation tree. The discourse history is represented as a sequence of the n previously generated explanation plans; centrality concepts simply specify a particular concept. The planner first selects the concepts for the target network as described above. Next, it consults the discourse history and the current centrality concept. Together, these are used to identify candidate concepts for the base network. It then inspects the triggering connections to identify the most useful connective relations between concepts in the discourse history and concepts in the target network. The net result of this phase is a media-independent explanation plan.

Finally, the 3D animation specifications and the natural language specifications of the explanation plans are passed to the media realization engines. The 3D animation specifications are passed to the animation renderer (Bares & Lester 1997; Towns, Callaway, & Lester 1998), while the linguistic specifications are passed to the natural language generator's sentence planner (Callaway & Lester 1995) and surface generator (Elhadad 1992). The text constructed by the latter is passed to a speech synthesizer. Visualizations and speech are synchronized in an incremental fashion and presented in atomic presentation units as dictated by the structure of the initial media-independent plan. Both are presented in realtime within the 3D learning environment. The discourse history is then updated to include the newly constructed explanation plan. Both the animation planner and the natural language generation system are products of long-term research in our laboratory. Below we briefly overview their operation.

Integrative Animation Planning

Given a set of `show` communicative goals in an explanation plan, the animation planner must construct an animation that visually presents the propositions in the target, base, and connecting networks. Planning animated explanations is a synthetic process of selecting and organizing 3D models, the 3D behaviors they will exhibit, and the camera planning directives that will be used to plan the movements of the virtual camera that films the action. All of these decisions must take into account the content and organization decisions made during the integrative explanation planning phase:

1. *3D Model and Behavior Selection:* Given a query which specifies a question type, e.g., `(explain-role ?T)`, and a target concept, e.g., `stomata`, the animation planner uses the ontological indices of the knowledge base to retrieve the relevant *concept suite.* Indicating the most relevant visual elements, a concept suite is defined by a sequence of concepts, each of which is either an object, e.g., `Chloroplast` or a process, e.g., `Photosynthesis`, annotated with their associated 3D models and 3D behaviors. 3D models indicate the geometric properties of the objects, and also contain annotations indicating texture maps. Behavior models specify the motion paths that objects can follow in a scene relative to other objects.
2. *Designing Focus Effects:* Enabling a student to focus her attention on the most salient characteristics of the new knowledge presented in an animation is essential for pedagogical reasons. This central feature of integrative explanations is provided by the animation planner's ability to introduce two categories of *visual markers*: (1) it highlights objects depicting new concepts, and (2) it introduces onscreen labels to indicate the nomenclature.
3. *Cinematography Planning:* Through judicious camera shot selection, explanations can direct students'

attention to the most important aspects of a scene. For example, while high and far shots present more information, close-up shots are useful for centering on a single subject (Mascelli 1965). To provide visual context, it initially selects far shots for unfamiliar objects, unfamiliar processes, and tracking moving objects. It selects close-ups for presenting the details of familiar objects.

Integrative Natural Language Generation

To ensure that natural language expressions clearly communicate the integrative content specified for the narration, the natural language generator uses specifications issued in the `say` nodes of explanation plans. The propositions selected for generation include concepts and relations in the target, base, and connecting networks, which are then used to construct sentential specifications by the sentence planner (Callaway & Lester 1995). Because lexical choice has been performed by this point, sentential specifications include all of the open class lexemes and their linguistic roles. After the sentence planner constructs functional descriptions, it passes them to a unification-based surface generator (Elhadad 1992) to yield the surface strings, which are then passed to a speech synthesizer. and delivered in synchronization with the actions of the associated 3D visualization.

The Implemented Generator

The student-sensitive multimodal discourse generation framework has been implemented in CINESPEAK, a generator that constructs integrative explanations with 3D animations and speech.[1] Given queries about physical phenomena, CINESPEAK consults the discourse history, invokes a library of integrative multimodal explanation strategies, and creates animated explanations with accompanying natural language. The integrative multimodal explanation strategies, the media-independent explanation planner, and visuo-linguistic mediator were implemented in the CLIPS production system. The 3D animation planner was implemented in C^{++}. The natural language generator was implemented in Harlequin Lispworks and employs the FUF surface generator and SURGE (Elhadad 1992), a comprehensive unification-based English grammar. The animation renderer was created with the OpenGL rendering library, and the speech synthesis module employs the Microsoft speech synthesizer. CINESPEAK operates in realtime.[2]

[1] CINESPEAK is implemented in a networked computing environment consisting of two PentiumPro 300s communicating via TCP/IP socket protocols over an Ethernet. It's various modules consists of more than 60,000 lines of Lisp, CLIPS, and C^{++}.

[2] Typical explanations last approximately 30 seconds. To construct a typical explanation, CINESPEAK (1) exhibits an initial latency of less than 1 second, (2) conducts animation planning, rendering, natural language generation, and speech synthesis in parallel with explanation delivery, (3) completes all media independent planning and animation planning in under 1 second, and (4) completes all natural language generation activities after roughly 8 seconds.

The PLANTWORLD Testbed

To investigate CINESPEAK's behavior, it has been incorporated into PLANTWORLD, a prototype 3D learning environment for the domain of botanical anatomy and physiology. This domain is challenging to explanation generators because it requires well constructed scenes and camera planning, as well as clear natural language expressions, to communicate the complexities of plants' structure and function. Its knowledge base represents objects (such as stomata, chloroplasts, and xylem), and resources (such as oxygen and water molecules), all of which have corresponding 3D models. It represents processes (such as photosynthesis, water transport, and oxygen efflux), all of which have associated 3D motion paths. Students interact with CINESPEAK by posing questions with a menu-based question construction interface.

To illustrate CINESPEAK's behavior, suppose a student initially sets the centrality concept to be `light reactions` and then asks, "What is the role of photosynthesis?" Because photosynthesis is a superprocess of the light reactions, it begins by noting this relationship. It opens with a shot of the chloroplast and explains, "The photosynthetic light reactions are a step of photosynthesis." It then cuts to an interior shot of the leaf and says, "Photosynthesis produces food sugars for the plant." Next, the student asks about the role of the dark reactions. The system determines that the dark reactions are related to the light reactions by category, i.e., they are siblings, so the explanation begins by noting that both are steps in photosynthesis. With a shot of the chloroplast in full view, the system says, "Like the photosynthetic light reactions, the dark reactions are part of photosynthesis." It then cuts to a shot of a sugar molecule exiting from the chloroplast and says, "During the dark reactions, carbon dioxide and hydrogen are converted into glucose by the chloroplast."

Next, the student indicates that she no longer wishes the light reactions to be a central concept. She then asks what the role of the chloroplast is, and the system constructs a low detail explanation that shows it and briefly overviews its role. She then asks again what the role of the chloroplast is, and the system responds with a detailed explanation as shown in (Figure 1). First, it opens with a shot of the chloroplast and says, "Let's take a look at the chloroplast in more detail. As you saw earlier, the chloroplasts produce food sugars for the plant. Glucose production occurs within the chloroplast." Cutting to a shot of water entering the chloroplast, it says, "Water in the chloroplast is converted into hydrogen and oxygen. Carbon dioxide, hydrogen, minerals and sunlight are converted into food sugars." It transitions then to a shot of a sugar molecule traveling through the phloem and explains, "Food sugars are used throughout the plant for energy."

Finally, the student asks about the role of the xylem. Because the xylem has a role (`xylem-distribution`) that occurs immediately before a role of the chloroplast (`oxygen-production`), the system constructs an explanation that makes the connection between the two processes, and then gives the role of the xylem. With a shot of minerals traveling up the xylem, it says, "Before the chloroplasts can produce food sugars, water and minerals must enter the chloroplast from xylem." It concludes by explaining, "Water and minerals are carried by xylem to the leaf."

Conclusions and Future Work

Dynamically generating rich multimodal explanations holds significant promise for learning environments. Because of the communicative power of engaging 3D animations coupled with well-crafted natural language, multimodal explanations offer the potential of facilitating effective learning experiences. By exploiting a discourse history to determine the content, rhetorical organization, cinematic structure, and spoken natural language narration in a manner that makes explicit connections between new and familiar concepts, multimodal student-sensitive explanation generators can construct explanations that actively promote knowledge integration.

This work is designed for rapid deployment into the classroom. Hence, a critical next step in its evolution is to empirically investigate the efficacy of the integrative explanation generation methods in full-scale learning environments with actual students. The prototype systems on which this work is based have all been subjected to such studies and revealed the strong need for the customization capabilities provided by the strategies proposed here. While we have experimented extensively with the integrative strategies in the implemented prototype learning environment, studying them under controlled conditions in the classroom is sure to yield significant insights. We will be exploring these issues in our future research.

Acknowledgements

The authors gratefully acknowledge the following individuals' assistance: Stuart Towns, for his work on early versions of CINESPEAK; Bradford Mott, for his assistance in preparing the manuscript and for comments on earlier drafts of this paper; and Dr. Loren Winters of the North Carolina School of Science and Mathematics for collaboration on PHYSVIZ, the second learning environment in which components of CINESPEAK are being studied. Support for this work was provided by a grant from the NSF (Career Award IRI-9701503), the IntelliMedia Initiative of North Carolina State University, the William R. Kenan Institute for Engineering, Technology and Science, and a gift from Novell, Inc.

References

Adams, M., and Bruce, B. 1982. Background knowledge and reading comprehension. In Langer, J. A., and Smith-Burke, M. T., eds., *Reader Meets Author / Bridging the Gap: A Psycholinguistic and Sociolinguistic Perspective.* Newark, DE: International Reading Association. 2–25.

Bares, W., and Lester, J. 1997. Realtime generation of customized 3D animated explanations for knowledge-based learning environments. In *Proceedings of the Fourteenth National Conference on Artificial Intelligence*, 347–354.

Butz, A. 1997. Anymation with CATHI. In *Proceedings of the Ninth Innovative Applications of Artificial Intelligence Conference*, 957–62.

Callaway, C., and Lester, J. 1995. Robust natural language generation from large-scale knowledge bases. In *Proceedings of the Fourth Bar-Ilan Symposium on the Foundations of Artificial Intelligence*, 96–105.

Christianson, D.; Anderson, S.; He, L.-W.; Salesin, D.; Weld, D.; and Cohen, M. 1996. Declarative camera control for automatic cinematography. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, 148–155.

Elhadad, M. 1992. *Using Argumentation to Control Lexical Choice: A Functional Unification Implementation.* Ph.D. Dissertation, Columbia University.

Green, N.; Kerpedjiev, S.; Roth, S.; Carenini, G.; and Moore, J. 1998. Generating visual arguments: a media-independent approach. In *Proceedings of the AAAI-98 Workshop on Representations for Multi-modal Human-Computer Interaction.*

Karp, P., and Feiner, S. 1993. Automated presentation planning of animation using task decomposition with heuristic reasoning. In *Proceedings of Graphics Interface '93*, 118–127.

Mascelli, J. 1965. *The Five C's of Cinematography.* Cine/Grafic Publications, Hollywood.

McKeown, K. R.; Feiner, S. K.; Robin, J.; Seligmann, D.; and Tanenblatt, M. 1992. Generating cross-references for multimedia explanation. In *Proceedings of the Tenth National Conference on Artificial Intelligence*, 9–15.

Mittal, V.; Roth, S.; Moore, J. D.; Mattis, J.; and Carenini, G. 1995. Generating explanatory captions for information graphics. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 1276–1283.

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

Seligmann, D. D., and Feiner, S. 1993. Supporting interactivity in automated 3D illustrations. In *Proceedings of Intelligent User Interfaces '93*, 37–44.

Suthers, D. D. 1993. *An Analysis of Explanation and Its Implications for the Design of Explanation Planners.* Ph.D. Dissertation, University of MA.

Towns, S.; Callaway, C.; and Lester, J. 1998. Generating coordinated natural language and 3D animations for complex spatial explanations. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence*, 112–119.

Wahlster, W.; André, E.; Finkler, W.; Profitlich, H.-J.; and Rist, T. 1993. Plan-based integration of natural language and graphics generation. *Artificial Intelligence* 63:387–427.