Realtime Generation of Customized 3D Animated Explanations for Knowledge-Based Learning Environments*

William H. Bares and **James C. Lester**
Multimedia Laboratory
Department of Computer Science
North Carolina State University
Raleigh, NC 27695-8206
{whbares,lester}@eos.ncsu.edu

Abstract

Rich 3D animated explanations can have a powerful impact on students interacting with immersive knowledge-based learning environments. By generating 3D animated explanations in realtime, a learning environment can create engaging explanations that are tailored to individual students. This paper presents the *immersive explanation planning* framework for generating pedagogically-customized 3D animated explanations in realtime. In this framework, an explanation system selects 3D models and their relevant behaviors, creates camera shots that most clearly depict complex phenomena, constructs a temporal organization that synchronizes narrative utterances with visual elements, plans the movement of the virtual camera that "films" the explanation, and incorporates specialized visual effects to focus students' attention on the most salient concepts. The framework has been implemented in RAPID, an explanation system that plans and renders customized 3D animated explanations of dynamic phenomena in realtime. Results of a focus group evaluation of RAPID are encouraging.

Introduction

Rapid advances in graphics technologies have paved the way for explanation systems that can automatically generate 3D animated explanations. In recent years, a number of projects have investigated the generation of 2D multimedia presentations (André *et al.* 1993; McKeown *et al.* 1992; Mittal *et al.* 1995) and the generation of behaviors for animated agents (André & Rist 1996; Lester & Stone 1997; Stone & Lester 1996). Foundational work has also begun on static 3D illustration generation (Seligmann & Feiner 1993), intelligent camera planning for 3D graphics (Christianson *et al.* 1996), and automatic creation of 3D animated movies (Butz 1997; Karp & Feiner 1993).

One of the most promising opportunities introduced by these developments is the prospect of creating immersive knowledge-based learning environments that generate 3D animated explanations of complex phenomena in response to students' questions. By enabling students to interactively explore 3D environments depicting complex biological, electronic, or electro-mechanical systems, immersive learning environments could provide rich pedagogical experiences with unparalleled visual impact. Operating in conjunction with an immersive learning environment, a realtime 3D animated explanation system could create customized 3D animated explanations that are tailored to individual students.

To illustrate, suppose a student is interacting with an immersive learning environment for the domain of botanical anatomy and physiology and requests an explanation of the physiological mechanisms by which the plant manufactures and distributes glucose. The explanation system could determine the concepts to be presented (glucose, water, chloroplasts, phloem) and their relevant behaviors (the events of photosynthesis and glucose transport). After composing the elements of each scene by selecting relevant 3D models and their behaviors, it could construct a customized visual and temporal structure for the animated explanation that includes camera shots featuring unfamiliar concepts, visual markers on important objects and actions, and well placed pauses to give the student additional time for reflection at critical junctures. If the student were unfamiliar with the interior of leaves, the explanation system could select a cutaway microscopic 3D model of the leaf showing significant anatomical detail. It could show how cells in the leaf perform photosynthesis to produce glucose, demonstrate how glucose is transported throughout the plant by cutting to a macroscopic leaf model, and pan across a cutaway model of the stem to track a highlighted glucose molecule as it makes its way toward the roots.

However, automatically creating 3D animated explanations poses significant challenges. Given a student's

---

*Support for this work was provided by the IntelliMedia Initiative of North Carolina State University and Novell and equipment donations from Apple and IBM.

query, an explanation system must perform a series of complex computations to synthesize animated explanations that are both clear and visually engaging. It must determine the content to be communicated and retrieve 3D models and object behaviors for explaining the concepts. It also must plan camera shots and camera motion paths, introduce cinematic transitions for visual continuity, and create specialized visual effects that focus the student's attention on the most salient objects and behaviors. Critically, it must operate within the realtime constraints imposed by highly interactive learning environments.

In this paper, we propose the *immersive explanation planning* framework for dynamically generating customized 3D animated explanations in realtime for immersive learning environments. The framework exploits pedagogical 3D animation strategies to guide its composition of animated explanations. It has been implemented in RAPID, an explanation planner that generates 3D animated explanations with dynamic scenes in realtime.[1] RAPID consults a user model as it composes the visual and temporal structure of explanations, plans the virtual camera's pans, cuts, and zooms, creates specialized visual effects for focusing students' attention, and synchronizes narrative utterances complementing the rich visualizations. To study its behavior, RAPID has been used to create a prototype 3D animated explanation generator for the domain of botanical anatomy and physiology. Results of a focus group evaluation of students interacting with RAPID are encouraging.

Requirements for 3D Animated Pedagogical Explanations

To communicate effectively with students interacting with immersive learning environments, 3D animated explanation systems must generate explanations whose conceptual content, visual structure, and temporal organization are pedagogically appropriate. In addition to the realtime performance requirements of learning environments, they should generate customized explanations that satisfy the following criteria:

- *Visual Focusing:* Animated explanations should focus students' attention on the most critical concept at each moment in the explanation (Rieber 1990). Visual focusing can be accomplished by employing close-up shots, incorporating visual markers such as pointer arrows to highlight objects and actions, and anticipating upcoming actions by transporting and reorienting the camera to the location where an important event is about to occur.
- *Pedagogical Chunking:* To combat complexity, animated explanations should decompose complex concepts into sequences of visual "chunks" (Rieber 1990). An important corollary is that explanations should provide students sufficient time to comprehend complex visualizations, particularly when introducing new objects into scenes (Windschitl 1996). As explanation systems plan the content, visual structure, and temporal organization of explanations, they should consider students' familiarity with the structure and function of concepts presented in explanations.
- *Continuity Preservation:* During camera movements such as panning and zooming, the camera position and orientation must be carefully chosen to avoid jarring visual discontinuities (Mascelli 1965). Planning camera shots and camera positions while preserving continuity requires solving precisely the same set of problems that are faced by cinematographers, with the additional constraint that they must be solved dynamically.
- *Realtime Performance:* To provide the responsive interaction required by interactive learning environments, explanations must be generated in realtime. No more than a few seconds should pass between the time when a student poses a question and the generator plans and renders an explanation.
- *Explanatory Immersivity:* To maximize the immersive aspects of students' experiences, animated explanations should "visually cohere" with the student's current position and orientation in the 3D environment. Rather than introducing visually disconnected "movies" on the screen, explanation systems should create explanatory animations that seamlessly unfold within the current 3D environment.

Generating 3D Animated Explanations

Explanation generation begins when a student exploring an immersive learning environment poses a query about a particular concept. Given a query and a 3D world model that represents the current state of the learning environment, the explanation system (Figure 1) exploits pedagogical animation strategies to construct a customized 3D animated explanation.[2] After selecting the relevant 3D wireframe models, behavior models, and narrative utterances from ontologically-indexed resource libraries, the explanation system executes the immersive explanation planning algorithm (Figure 2), which operates in three phases:

1. **Animation Design:** It constructs an explanation plan around a sequence of *focal points*, which represent the single most important object or action at each moment. By considering the criticality of each concept and the student's familiarity with these concepts as represented by an overlay user model (Carr

---

[1] Realtime Animation Planner for Interactive 3D Environments.

[2] It is important to note that this approach employs no canned animations—all animations are composed and rendered at runtime from 3D wireframe models, behavior models, and narrative utterances.

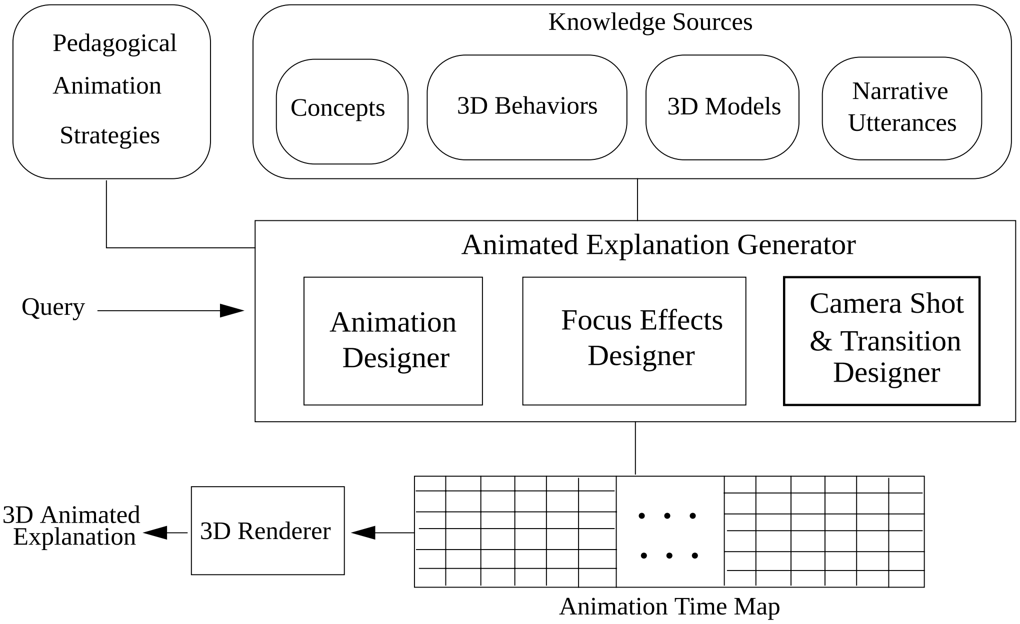


Figure 1: 3D animated explanation system

& Goldstein 1977), it lays out all visual and auditory elements on a multi-tiered *animation time map*.

2. **Visual and Temporal Focus Effects Design:** To direct the student's attention to the most salient concepts of the explanation at each moment, it (a) introduces visual markers, (b) dilates explanatory scenes depicting unfamiliar concepts, and (c) creates specifications for customized camera shots emphasizing important objects and actions.
3. **Camera Motion and Cinematic Transition Planning:** To achieve visual continuity, it (a) plans the tracking, panning, and zooming of the virtual camera that "films" explanations, and (b) creates cinematic transitions (e.g., cuts and fades) to smoothly segue from scene to scene.

When construction of the time map is complete, the explanation system passes it to the renderer. Because the time map houses parallel series of 3D coordinate specifications for all object positions and orientations, visual effects, and camera positions and orientations, the renderer can construct a frame of the explanation for every tick of the clock. These frames are rendered with the accompanying narration in realtime, creating a continuous immersive visualization in which rich 3D explanations mesh seamlessly with the student's exploration of the environment.

Animation Design

Planning animated explanations is a synthetic process of organizing the raw materials of 3D wire frame models, 3D behaviors, and narrative utterances into a coherent explanation plan that is tailored to the individual student. Given a query which specifies a question type, e.g., `(explain-function ?X)`, and a target concept, e.g. `chloroplast`, the explanation system uses the ontological indices of the knowledge base to retrieve the relevant *concept suite*. Indicating the most relevant visual and auditory 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., `Transpiration`, annotated with the following features:

- *3D Models:* Wireframe models indicating the shape and colors of objects in the 3D environment. Models vary in scale, with some providing little detail and others providing great detail. Models also vary in structural completeness, with some representing full structures and others providing cutaway views, e.g., the `Leaf-Cutaway` model.
- *Behavior Models:* The 3D motion paths, initial positions, orientations, and potential mobility of relevant objects. For example, the `Oxygen-Out-Of-Stomata` model specifies the initial location of an oxygen molecule and the coordinates of its motion path as it travels through and exits the stomata.
- *Narrative Utterance:* The relevant narration that can accompany the visual elements of the explanation, e.g., a narrative utterance describing oxygen's release into the atmosphere.
- *Criticality Level:* A value indicating the relative importance of the concept compared to other concepts in the suite.

To most clearly communicate the selected knowledge, the explanation system must focus the student's attention on the most salient concept at each moment

1. Retrieve (a) a *suite* S of concepts relevant to topic T and question type Q and (b) the sets W, B, and U, containing the 3D wireframe models, 3D behaviors, and narrative utterances relevant to S.
2. For each concept C in S do
    - (a) Decompose the explanatory structure into a series of *focal points* $f_1 \ldots f_n$, one for each critical object or concept in S, and lay them out on an animation time map M.
    - (b) Lay out all non-focal objects in W, non-focal behaviors in B, and elements of U associated with C in parallel with focal point objects and behaviors on M.
    - (c) Create focal effects (visual markers, cognitive pauses, customized camera shots) for the object or behavior associated with each focal point f_i and incorporate them into M.
    - (d) Plan smooth camera trajectories of panning, tracking, and zooming to transition from each f_i to its successor f_j and incorporate them into M.
3. Create an introductory segue from the current state of the 3D environment to the first shot of f_1 and incorporate it into M.
4. Render 3D animated explanation by iterating across the visual and auditory specifications of M.

Figure 2: Immersive explanation planning

of the explanation and carefully shift this focus as the explanation unfolds (Rieber 1990). This is accomplished by structuring explanations as a series of *focal points*, which represent the most important concepts to be communicated at each instant. The explanation system creates focal points for each concept C_i in the concept suite by inspecting the user model and the availability of relevant narrative utterances:

> **Focal Point Creation:** If C_i is marked in the user model as being unfamiliar or has an associated narrative utterance, then if it has a high criticality rating, create a focal point f_i and install it in the scene forming around C_i.

The explanation system employs the focal points to guide its organization of visual and auditory elements. A concept for which the explanation system has created a focal point will be explained in a manner that makes it the sole focus of attention. If more than one focal point exists in a scene, the explanation system distributes the visual presentation of the concepts across time by selecting a *serial* presentation strategy, which specifies laying out the focal points in successive order on the animation time map. For example, suppose a student who is familiar with the leaf interior but not with the stomata or chloroplasts asks a question about the internal structure of the leaf. The explanation system marks these critical concepts as focal points and lays them out for serial presentation (Figure 3). During the camera planning phase, this will cause the camera to first focus on the stomata and then shift to the chloroplast. In contrast, a concept for which the explanation system has not created a focal point can be presented simultaneously with other concepts. In these cases, the explanation system selects a *parallel* presentation strategy, which specifies that the 3D models and motion paths be presented simultaneously. Non-focal concepts explained in parallel with focal concepts serve (quite literally) as the backdrop against which focal concepts are presented. As described below, the system employs focus effects to ensure that focal concepts are brought to the center of attention.

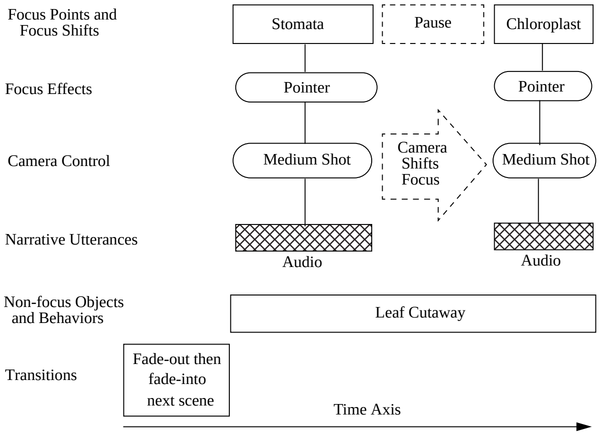


Figure 3: Segment of animation time map

As it creates focal points and lays out the visual elements, the explanation system lays out the narrative utterances on the audio track by associating the utterance with its corresponding concept on the time map. The system synchronizes narrative utterances with the appearance and movement of visual elements by using the length of an utterance as the minimum duration for presenting the concept and then scaling the duration of animations it generates. In a delayed commitment fashion, the system does not fix the actual starting times or durations of explanatory actions until the next two phases to preserve maximum flexibility.

Designing Focus Effects

By structuring explanation plans around focal points, the explanation system creates a framework for incorporating visual focus effects that will direct the student's attention to the most important concepts at each juncture of the explanation. For every focal point created in the animation planning phase, the explanation system considers incorporating each of the following three focus effects for accentuating focal objects and processes and creating smooth focal shifts:

- **Visual Markers:** If a focal point object is the sole object in motion in a particular scene, no visual marker is needed since its motion already draws attention to itself. For example, if a water molecule is rising through the xylem but there is no other visible movement in the environment, no marker is required. In contrast, if a focal object is stationary, then its *explanation history* is examined. The explanation history records the frequency with which a particular object has been presented in previous explanations. If a focal object has never been presented or has been presented only once, the explanation system introduces a visual marker that points to it. For example, when a stationary oxygen molecule is a focal point, the explanation system introduces a pointer to it. The explanation system also coordinates the narration with the visual marker by synchronizing the marker's appearance with narrative utterances.

- **Cognitive Pauses:** Because introducing new concepts requires significant processing on the part of the student, animated explanations should include cognitive pauses to provide the student sufficient time to acquaint himself or herself with the unfamiliar objects and actions (Windschitl 1996). Cognitive pauses are particularly critical when explanations include concepts with unfamiliar visual representations and significant accompanying narration, which requires additional processing. The explanation system introduces cognitive pauses by dilating segments of a scene that depict unfamiliar and complex concepts. Let f_i represent a focal points depicting unfamiliar objects or actions or accompanied by narrative utterances. The explanation system dilates segments that occur immediately prior to f_i and and immediately after f_i. For example, suppose an animation time map constructed during the animation design phase presents the leaf cutaway model as a whole, then introduces the stomata, and finally shows a water molecule transpiring from the stomata. If the student is already familiar with stomata but not with transpiration, the explanation system dilates the scene by introducing pauses preceding and following the presentation of transpiration.
- **Customized Camera Shots:** Through judicious camera shot selection, explanations can direct students' attention to the most important aspects of a scene, even in complex scenes presenting a number of objects in motion, and provide visual context. While high and far shots present more information (Mascelli 1965), close-up shots are useful for centering on a single subject (Millerson 1994). The explanation system therefore creates camera views that not only factor in the geometric considerations affecting object and action visibility, but it creates shots that are pedagogically customized. 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. For example, rather than introducing the new concept of xylem with a close-up, it first presents xylem in the larger visual context of the stem and then zooms in to present the details. By the end of visual effects planning, the animation time map has been modified to account for the dilations associated with cognitive pauses and updated to include the visual markers and the customized camera shots.

Planning Camera Trajectories

Visual continuity in 3D animated explanations is analogous to discourse coherence in natural language explanations: both the structure and the content are essential for effective communication. However, achieving visual continuity in animated 3D explanations is difficult, particularly for explanations with moving objects which require a variety of camera shots. Explanations cannot be presented with arbitrarily chosen camera motions (panning, tracking, zooming) without producing visual discontinuities (Mascelli 1965), nor can they abruptly cut from shot to shot without introducing confusing context switches. The explanation system must therefore plan camera motions to smoothly transport the viewer from one shot to the next. To plan the virtual camera's motion from focal point f_i to focal point f_j, the explanation system employs the following three strategies:

- **Anticipation:** By employing the classical animation technique of *anticipation* (Lasseter 1987), it prepares the student for an upcoming action by transporting the camera to the location where the action f_j will be presented a few moments before the action occurs.
- **Jump Cut Avoidance:** To avoid jump cuts, which are jarring cuts from one shot to another that is only slightly different (Mascelli 1965), it estimates the amount of travel distance needed to pan and track the camera from a shot of focal point f_i to one of f_j by interpolating camera positions and orientations represented in spherical coordinates. If the distance is small, then it pans from the shots depicting f_i to that of f_j and introduces a temporal dilation to allow for the camera motion; otherwise, it cuts to the next shot.[3]
- **Pan/Track Preference:** To avoid disrupting the student as it introduces unfamiliar concepts, it pans and tracks from f_i and f_j, unless doing so would introduce a jump cut.

To achieve visual continuity within an explanation, the explanation system considers the visual and semantic content of the last shot of the previous focal point f_i and the first shot of the upcoming focal point f_j. If the camera distance between f_i and f_j is small, then the explanation system employs smooth tracking and

---

[3] To compute camera trajectories for panning and tracking, the explanation system interpolates the camera's position and orientation, represented in spherical coordinates, between the shots for f_i and f_j.

panning. If the context switch represents a *scale shift* between 3D models, e.g., if f_i presents a macro-exterior shot of an object such the `exterior-leaf-model` and f_j presents a micro-interior shot of the same object such as the `interior-leaf-cutaway` model, then the explanation system introduces a fade-in/fade-out transition. Finally, if there is a great camera distance separating the shots of f_i and f_j or if there is no scale shift, then it introduces a cut. To create smooth context switches from the final shot of the previous explanation to the first shot of the upcoming explanation, it plans camera tracks and pans if the camera distance is small. If the camera distance is large, it introduces cinematic transitions, such as fading to and from black. Finally, it updates the user model by marking focused concepts and then passes the animation time map to the 3D renderer.

An Implemented Explanation System

The immersive explanation planning framework has been implemented in Rapid, an explanation planner that generates 3D animated explanations with dynamic scenes.[4] Rapid's decisions are guided by pedagogical animation strategies described above. It operates in realtime, with rendering beginning less than 1 second after a query is posed. To investigate Rapid's behavior, it has been used to create a prototype 3D animated explanation system for the domain of botanical anatomy and physiology. It employs approximately 15 3D models and 45 narrative utterances to generate explanations about a variety of plant structures (e.g., leaves, chloroplasts), processes (e.g., photosynthesis, transpiration), and resources (e.g., oxygen, carbon dioxide).

To illustrate, suppose a student has just finished viewing several animated explanations concerning the xylem, phloem, and role of these vessels in resource transport. At this point, the student might see the shot of Figure 4 (a), which depcits a cut-away model of the plant with exposed xylem and phloem. The student next poses the question, "What does the leaf do?" Rapid first retrieves a concept suite for leaf function and the relevant 3D models, including the leaf-cutaway, stomata, and chloroplast models, behaviors, and narrative utterance. It then lays out the relative temporal order and focal points. In this case, the stomata and chloroplasts are made focal points because of their high criticality and the fact that a narrative utterance is available. The focal point for stomata is set to precede that of chloroplast to reflect the suggested sequence in the concept suite. Objects not currently in focus, e.g. the leaf-cutaway, are relegated to a subsidiary time map track which keeps them in view for the duration of all focal points in the scene being planned. Rapid now adds a highlight arrow pointer to each focal point object because the student is unfamiliar with the concepts. It sets the duration of each focal point and its highlight to the duration of the corresponding audio. It then inserts a cognitive pause between the two focal points since both are unfamiliar. Since the stomata and chloroplast are both in the same scene and the viewer is unfamiliar with them, it directs the camera to pan smoothly as it shifts from one focal point to the next. It sets the duration of the pause and camera pan to an interval that is proportional to the distance that the camera needs to move between the shots of the stomata and the chloroplast.

Next, Rapid selects a shot angle and distance for viewing each focal object. Because each object is part of a larger object (the leaf-cutaway) and the student is unfamiliar with both, a medium range shot is chosen. The shot viewing angle is fixed to one of a number of frontal angles. Since the previous scene included the tree-cutaway model and the upcoming scene includes the leaf-cutaway, Rapid selects a fade-out/fade-in transition. In cases in which the previous scene included objects in the same locale, a camera pan between shots would have been chosen if the camera travel time were not too substantial.

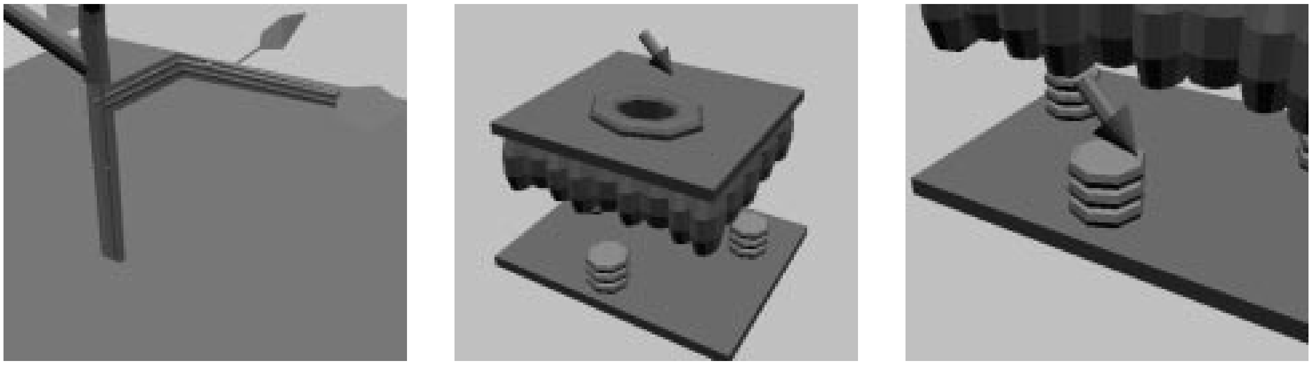
The explanation planner submits each scene of the completed animation time map to the 3D renderer. The resulting explanation opens with a fade-out/fade-in transition to the leaf-cutaway. Each focal point is presented in sequence. The first focal point is the stomata depicted in Figure 4 (b). Once the narration for the stomata has completed, the highlight arrow disappears and the camera pans down to the chloroplast. This pause and pan creates an anticipatory effect in which the student's attention is drawn towards the chloroplast. The camera comes to rest centered on the chloroplast (Figure 4 (c)) immediately before it becomes the current focal point. The explanation concludes by presenting a highlight arrow near the chloroplast with narration describing its physiological role.

Evaluation

To gauge the effectiveness of the immersive explanation planning framework, a focus group observation study was conducted with seven students from elementary school, middle school, and highschool, ranging in age from eight to sixteen. The study was conducted in three phases. First, students observed an explanation generated by a version of Rapid in which all focus effects, customized camera planning, cinematic transitions, and anticipatory strategies were disabled. Second, they observed an explanation generated in re-

---

[4] Rapid is implemented in C++. It consists of approximately 15,000 lines of code, and employs the OpenGL graphics library for realtime 3D rendering. To operate within the hardware constraints of educational settings, Rapid was developed to run on a cost-effective delivery platform, a Pentium 133 Mhz PC. It currently runs at 7 frames/second without the use of a graphics accelerator. Inexpensive graphic accelerators scheduled for release later this year will instantly increase its rendering speed to a full 30 frames/second.



(a) Resource Transport (b) Point out the stomata (c) Identify the chloroplast

Figure 4: Shots from an animated explanation of physiological functions of the leaf

sponse to the same question but by a fully functional version of Rapid. Third, Rapid generated explanations in response to a variety of questions posed by the students. To accelerate the interaction, students stated the questions and the tester constructed the queries. After the final explanation phase, students were asked to summarize their opinions of the software.

The results of the evaluation are encouraging. Rapid was well received, with six of the seven students stating their preference for the fully functional version.[5] The focal point approach proved effective, with several students observing that they liked the pointer coinciding with the narration to guide their attention. Several praised the smooth camera pans. The study also suggested two additions to the strategies. Students recommended including more opening scenes—cinematographers refer to these as "establishing shots"—and suggested including additional highlighting such as textual labels, particularly when the narration describes several objects simultaneously in motion. We will be adding these strategies in our future work.

Related Work

Multimedia presentation systems have received increasing attention in recent years. A number of projects investigate techniques for generating multimedia presentations, including Wip (André *et al.* 1993), which focuses on generating customized illustrated texts, Comet (McKeown *et al.* 1992), which addresses issues in coordinating text and graphics, and Sage (Mittal *et al.* 1995), which generates text, graphics, and explanatory captions. Animated simulations have been the subject of much study in the knowledge-based learning environment community, e.g., (Eliot & Woolf 1996; Hollan, Hutchins, & Weitzman 1987; Woolf *et al.* 1987), and several recent projects have explored animated interface agents (André & Rist 1996; Kurlander & Ling 1995; Lester & Stone 1997; Rickel & Johnson 1997; Stone & Lester 1996). However, unlike Rapid, which dynamically generates and renders 3D animations, these efforts focus on 2D graphics generation or utilize a library of static 3D graphics which are not animated.

Several recent projects investigate issues in intelligent 3D graphics. These have produced effective techniques for constructing static 3D illustrations (Seligmann & Feiner 1993) and simulating humans in a 3D environment (Philips, Badler, & Granieri 1992). More closely related to our work is that of the Esplanade project (Karp & Feiner 1993), the Camera Planning System (Christianson *et al.* 1996), and the Cathi (Butz 1997). Esplanade creates 3D animated movies and the Camera Planning System plans complex camera motions for dramatic 3D scenes. However, in contrast to Rapid, neither focuses on generating pedagogical animated explanations nor creates animations that are tailored to individual users. Esplanade bases its decisions on cinematic conventions rather than the communicative requirements of learning environments. It operates in "batch mode" rather than in realtime and does not consider narrative synchronization issues. The Camera Planning System provides an effective language for camera control of dramatic scenes that is based on film idioms rather than the requirements of pedagogical explanation generation. The Cathi system is perhaps most similar to Rapid. Cathi, which is part of the Ppp project (André & Rist 1996), employs a hierarchical planner to produce 3D animations. It operates in an incremental fashion by interleaving planning and rendering to produce timely performance, and it uses spotlights and color changes to direct the viewer's attention to objects of interest. In contrast to Cathi, however, Rapid employs pedagogically motivated focus techniques and synchronizes narrative utterances with the animations it generates.

---

[5] The other student expressed no preference.

Conclusion

Rich 3D animated explanations can have a powerful impact on students interacting with immersive knowledge-based learning environments. We have proposed the immersive explanation planning framework for automatically generating 3D animated explanations in realtime. By structuring explanations around a series of focal points that direct students' attention to the most salient concepts, an explanation generator can create animated explanations that are educationally effective. This framework has been implemented in RAPID, an explanation system that plans and renders customized 3D animated explanations of dynamic phenomena in realtime. A focus group evaluation of RAPID demonstrates that immersive explanation planning is an effective means for generating animated 3D explanations. Perhaps the greatest challenge ahead lies in coupling animated explanation generation with immersive learning environments that provide direct manipulation problem-solving functionalities. We will be exploring these issues in our future research.

Acknowledgements

The authors wish to thank Patrick FitzGerald and Woody Robbins for their insights on dynamic camera planning. Thanks also to Roberta Osbourne for assistance with the 3D modeling, the students who participated in the focus group study, and Charles Callaway for comments on an earlier draft of this paper.

References

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

André, E.; Finkler, W.; Graf, W.; Rist, T.; Schauder, A.; and Wahlster, W. 1993. WIP: The automatic synthesis of multi-modal presentations. In Maybury, M. T., ed., *Intelligent Multimedia Interfaces.* AAAI Press. chapter 3.

Butz, A. 1997. Anymation with CATHI. To appear in *Proceedings of the Innovative Applications of Artificial Intelligence Conference.*

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

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

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*, 409–415.

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

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.

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

Lasseter, J. 1987. Principles of traditional animation applied to 3D computer animation. In *Proceedings of SIGRAPH '87*, 35–44.

Lester, J. C., and Stone, B. A. 1997. Increasing believability in animated pedagogical agents. In *Proceedings of the First International Conference on Autonomous Agents*, 16–21.

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.

Millerson, G. 1994. *Video Camera Techniques.* Focal Press, Oxford, England.

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.

Philips, C.; Badler, N.; and Granieri, J. 1992. Automatic viewing control for 3D direct manipulation. In *Proceedings of the 1992 Symposium on Interactive 3D Graphics*, 71–74.

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

Rieber, L. 1990. Animation in computer-based instruction. *Educational Technology Research and Development* 38(1):77–86.

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

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

Windschitl, M. 1996. Instructional animations: The inhouse production of biology software. *Journal of Computing in Higher Education* 7(2):78–94.

Woolf, B.; Blegan, D.; Jansen, J. H.; and Verloop, A. 1987. Teaching a complex industrial process. In Lawler, R. W., and Yazdani, M., eds., *Artificial Intelligence and Education*, volume 1. Norwood, New Jersey: Ablex. 413–427.