FULL SPECTRUM LEARNING

CONFERENCE REPORT

APPLE
MULTIMEDIA
LAB

SAN FRANCISCO

TABLE OF CONTENTS

INTRODUCTION

In "Minds and Media in Dialogue," Michael Mills and Roy Pea provide a theoretical view on how images might be incorporated in learning. As indicated in their title, Mills and Pea emphasize the distinction between internal representations — the minds in which learning is accomplished — and external representations — the media which encourage this learning. They suggest that learning is a result of the interaction (or "dialogue") between these two elements, in a constructivist environment.

This chapter presents a good framework in which to consider a range of related issues, including personal creation as a part of learning, the use of media in education, and the dialogue required for learning (between people or within a person). It also considers the link many "multimedia computing activities" have with this process, with their use of "interactive" imagery — e.g., imagery that responds to individual inquiries and activities.

Both Mills and Pea have continued this line of reasoning since the conference, contributing to the evolution of multimedia computing. Michael Mills is now an Apple employee, developing new tools to handle movies in computer environments. Roy Pea has completed a number of studies on the use of multimedia in classrooms, and has been active in developing tools for multimedia composition.

Mind and Media in Dialogue:

Issues in Multimedia Composition

BY MICHAEL MILLS AND ROY PEA

Part I: Overview of the Key Issues

Mind and media are intimately related, more so than may people imagine. Indeed, as Roland Barthes has observed, the powers of media are in part due to their invisibility once we become familiar with them. No one questions the "naturalness" of the Arabic number system in computations, for example. On the same level, television has become woven into the fabric of our everyday lives, at least in the Western world. This essay is about some of the primary dimensions in these mind-media relationships, and the two-way dialogues that take place between them. We believe that a deep understanding of mind-media relationships will acquire increasing importance as multimedia communications become more prevalent. It is something we already see happening in a society whose messages are increasingly designed and conveyed with computer support. More profoundly, technology is making integrated multimedia — a combination of still or moving images, text and audio — more widely accessible. There are reasons to expect the social and cognitive consequences of such integration to be highly significant.

First we will discuss the key issues and vital questions necessary for creating an educational system that exploits the fertile synergies of mind-media relationships. Since we confront these issues continually in our multimedia research and development projects, we will provide specific examples for illustrating our points.

(Please note that our expositions are offered in the spirit of fueling critical debate about multimedia in education and how it might be used effectively. We do not believe in technology as a panacea for education or any other normative enterprise.)

Throughout our presentation there are four voices which burst forth with varying degrees of prominence, reflecting our own backgrounds and interests. They are:

1. ***The Cognitive Scientist:*** How do symbol systems — such as written language mathematical notations, or programming languages — contribute to the forms, contents and processes of thinking?

2. ***The Technologist:*** How do tools for design and composition shape our ability to create multimedia materials? What new representational tools would better catalyze understanding and creative processes?

3. ***The Instructional Designer:*** What presentations should we create to help minds construct appropriate representations? What activities and supports might foster cognitive growth?

4. ***The Learner:*** What do we need to know to learn from multimedia works? What skills do we need to master to become proficient multimedia composers?

A Historical Introduction

Since one can find enlightening discussions about the impact of writing systems on reasoning, memory and social interactions as early as Plato's *Phaedrus,* we make no pretense of being thorough in this paper. It is but a sampler of questions, containing allusions to some major figures who have written about these issues. It is a framework for discussions among those involved in multimedia compositions in education, and the related issues of "full spectrum" multisensory learning. Our principal focus is on text and images, with sound playing a minor, supporting role.

We will group our discussion themes around three structural headings: *Representations and Mind, Representations and Learning,* and *Representations and Multimedia Composition.*

Representations and Mind

The 20th century philosopher Ludwig Wittgenstein, in his typically aphoristic style, raised a question that cuts to the heart of our concerns: "How do I know what I think until I see what I say?"

Reflect on this statement. Wittgenstein is attributing remarkable mediating powers to the explicit nature of written language forms. Without them, he questions whether he knows what he thinks at all. Somehow, capturing the intermediate steps in the transient, ephemeral, process of human thought in an external medium qualitatively changes the very nature of self-knowledge. Such "epistemic writing" shows that writing is more than communicating ideas held in some reservoir, all cooked-up and ready to express. Like many forms of human action, the design of human knowledge and self-understanding is constructed through a cyclical, revisionary process sometimes described as a "conversation" or "dialectic" with one's materials. Symbolic representational systems, such as written language, flowcharts, diagrams, or mathematical equations, may play a pivotal role in figuring out "what one thinks."

There is a long tradition in asking what the relationships are between symbolic forms and thinking. A large collection of studies addresses *how language shapes thought,* and vice versa. In the mid-20th century Benjamin Whorf and the linguist Edward Sapir captivated many disciplines, from the arts to physics, with their bold conjecture that the categories of language determine thought — for example, that the Eskimo's numerous terms for snow make their perceptions about snow different from those whose language lacks such lexical distinctions. Several centuries earlier, Immanuel Kant's writings opened the door for such speculations by stating that "Percepts without concepts are blind" — that what one "sees" is a function of one's conceptual categories. These conceptual categories can be shaped by the system of symbolic representations a culture has available.

How images shape thought — and vice versa — has been addressed by numerous theorists and scientists. Of particular relevance to multimedia studies are the works of Rudolph Arnheim, in his seminal essays "Visual Thinking," and Kristina Hooper Woolsey in her studies of the functions of diagrams and imagery in mathematical thinking and spatial learning.

We have two observations about the relations between symbolic forms and thought. The first is that *external representations mediate thought.* Cultural inventions such as manufactured artifacts, icons, lists, diagrams, instructions and maps, are not merely ancillary to processes of thinking and reasoning. They play a crucial mediating role in these processes, figuring functionally in their achievement. We find ways, as Don Norman expresses it, of "putting intelligence" into these representations, so that outer objects can elicit appropriate patterns of response.

The second observation is that *the cognitive and cultural effects of new media are unpredictable.* Understanding how new representational media influences minds must be an empirical matter; only after building such media and seeing how they are used can we tell what kind of relations they will find to our minds. Three examples illustrate this point. The first is *written language,* which theorists such as Walter Ong, Jack Goody and David Olson credit with making thought external and thereby an object of reflective analysis which contributed in a major way to the advent of logic and science. The second example is *programming languages,* which no one predicted would lead to artificial intelligence and its influence on manufacturing, design and office automation. The third example is *satellite broadcast television,* which brought the crumbling of the Berlin Wall and the drama of the Gulf War "live" into homes throughout the world. This is a far cry from the early vision of television as a device used for training medical students via better visual access to the surgical ampitheatre.

It is in the context of such examples that one might ask what the impact of computer tools, such as HyperCard, might be on multimedia communications. We are just learning how the use of computers as tools in education — as opposed to computers as electronic page-turners — may influence human minds. And we are only beginning to create a society that aims not only for literacy through print, but through integrated multimedia.

Representations and Learning

How does the form of a representation support the problem-solving and creative activities involved in learning? Our main observation concerning this question is that some representations better "embody" certain activities and are more congenial to their prospective problem-solving processes or creative acts.

We may distinguish three forms of congeniality, each of which contributes to the extent to which a symbolic representations system supports learning:

1. ***Congeniality to problem goals.*** It is well-known that certain representational formats are more congenial to a given psychological activity than others. Some prominent examples include *graphs* for projecting trends, *networks* for mapping relations among concepts, *matrices* for statistical reasoning, *flowcharts* for tracking complex processes, and *animations* for comprehending mechanisms.

2. ***Congeniality to learner/user understanding.*** Probably one of the major discoveries of cognitive science with educational relevance is that people interpret new concepts and events by *analogy* to what they believe or know. They attempt to integrate new ideas into existing mental structures. For example, many studies document how robust "naive physics" concepts of how motion and force operate in the world serve as a barrier to understanding formal physics instruction. Graphic representations of problem situations in physics, using arrows to represent force vectors, are not going to be congenial to a learner who has a different set of mental representations for the relations among physical entities.

3. ***Congeniality to learner/user familiarity with the representational medium.*** There is a "learning overhead" to media use and comprehensions. This is most apparent concerning written language, whose conventions are opaque enough to leave a large proportion of learners illiterate, even after more than a decade of formal schooling. But it is easy to ignore the learning overhead for diagrams, pictures and the notational systems employed in mathematics, whether we refer to algebraic equations, graphs and their labelling conventions, or geometric figures. Probably the most educationally-crucial observation we can make about this species of congeniality is that it is more often than not *revisable:* new representational media can be invented that drop the overhead for entry into a community of media users. Cases in point are the shift of multiplication activities, and the shift from the conversational metaphor to an object-oriented, "direct manipulation" metaphor for how humans can give computers instructions.

New forms of representational congeniality

Our experience with multimedia composing tool such as Interactive Videoworks, Super3D, Adobe Illustrator, MacPaint, Guide, LOGO, and now HyperCard, suggests that new interactive multimedia are beginning to provide new styles and dimensions for representing knowledge and processes. One may integrate sounds (natural and musical), with text and imagery; one may use animation to convey processes requiring understanding of dynamics; one may use "hot" pictures or diagrams — which have parts that respond when selected — to provide "ports" into other informational areas, such as explanations, glossary definitions, video clips and animations.

"Hypertext" links provide a new dimension for representing relations among conceptual representations. We are particularly excited at the possibility of putting back together what formal education has torn apart: *taxanomic/categorical knowledge,* exemplified by hierarchical relationships such as class-inclusion; and *experiential knowledge,* exemplified by narrative relationships, "storytelling" about connections between ideas or events. Formal education has played up the taxonomic mode of thought and often cast a negative eye on the narrative mode. For example, word associations such as "runs" to "dog" rather than "cat" have been considered developmentally primitive. Yet as the resurgence of interest in narrative among literary critics, anthropologists, psychologists, linguists and historians attests, there is a certain historical primacy of narrative that cries to be better understood for its distinctive organizational strengths. We can imagine the use of links to readily display these two powerful "paths" through concepts together. For example, a tiger can "participate" in both a knowledge structure of biological taxanomy and as an agent in an archetypal tale of a battle between opposing natural forces.

Representations and Multimedia Composition

We find questions about multimedia composition both exciting and challenging, since integrating the display of images, texts, programs and sounds, across time and space in an interactive environment is highly complex conceptually. There were previously so many technical barriers, for image processing in particular, that these conceptual issues of integrated multimedia design were not salient. We are only beginning to see the possibilities the lowering of these technical barriers makes available to education. We find particular promise in the idea of having students learn by composing multimedia works, not just observing/studying them. But first we wish to raise three general questions under the topic of representations and multimedia composition:

1. How do new representational tools interact with composing?
2. What psychological design processes are engaged in such composition?
3. What skills are required for becoming multimedia composers?

We now treat each of these briefly.

How do new representational tools interact with composing?

We believe representational tools interact in critical ways with composing, according to their structure serving as constraints in either a positive or negative sense. In a negative sense, they can constrain what one wants to do, making what one expects to be straightforward a difficult task to achieve. In the positive sense, representational tools can be constraining in blocking one from making unproductive errors, that is, those from which generative learning would not occur. They can also positively constrain by providing support, or scaffolding, to guide one's activities, that can be turned off when no longer needed.

New tools for integrated multimedia composition have had, in our experience with them, some unexpected consequences which we believe may have general significance. First, they make it *easier to convey multiple perspectives or interpretations of information* — whether it is experimental data in science or psychology, or historical data about a community lifestyle. At Brown University, instructors' experiences with the *Intermedia* system, using hypertext technologies to teach 19th century literature, indicates that teaching which encourages comparative argumentation about different interpretations of a poem is facilitated by the use of hypertext links among documents. It creates a conceptual "web" and displays interpretive "paths" through documents and commentaries.

Secondly, new tools *make easier the encoding of relationships between different representational systems,* which may underlie expert cognition in such fields as mathematics. With integrated multimedia, one can "yoke" representations of an algebraic function, such as a symbolic equation and its graph, and reflect the changes a student makes to values in the equation simultaneously in the shape of the graph. Many mathematics educators believe the cross-representational "intuitions" that can develop from the use of such environments will make higher levels of mathematical reasoning more generally accessible by bringing abstractions to life.

Thirdly, these tools *open up* a *new aesthetics* by creating new kinds of objects of criticism and for creation. For example, we find ourselves and others asking questions such as, "What are *good* web-links in a hypertext environment for a content area?" and "What is *good* interactive fiction?"

Fourthly, use of these tools *forces new attention to understanding the psychological roles of long-used but unanalyzed (and probably underexploted) representations.* When and how are pictures, diagrams, sounds and new dynamic representations such as systems dynamics simulations (for example, *Stella*) functional as "ports" to learning? It is traditionally true that texts and lectures relying on texts are the primary expository vehicle of instruction. There are many reasons to believe that difficulties in transfer of learning are due to the formal character of such presentations, isolated from activities of knowledge application in context. But now a dynamic diagram (for example, of a weight and pulley system) could be both the entry point and home-base for the learner, with text the ancillary medium. Geography could do the same with map representations; learning about trading commodities could be rooted in graphic depictions of action on the trading floor on Wall Street. One can even imagine situation-based representations (for example, in industrial training, engineering, or practical mathemetics) acting as the basis for new learning, approximating apprenticeship-learning in its experiential ties to the desired contexts of knowledge application.

Fifth, multimedia composition tools *challenge "reader-writer" and "author-learner" relationships.* A principal theme in anecdotal reports on the use of hypertext systems is that they begin to change the way one thinks about relations between writers of documents (or composers of multimedia works) and readers of those documents. This is particularly evident in networked hypertext systems. For example, as one begins to make electronic marginalia — a commentary link off of a text passage — that commentary can become part of the electronic community of interpretation for that text passage, to which readers may respond as they would to the original text itself. The embeddedness of such interpretations and their transparent integration in the electronic space hypertext system create could qualitatively change the reader-writer relation. The writer may revise the original text in response to the reader's reaction, and this process may continue its cycle until the boundary erodes between reader and writer, interpretor and creator. Although we cannot pursue the connections here, these technical capabilities underscore related portrayals of the natures of authors, texts and audiences that have emerged in the past decade in literature criticism, hermeneutics, and cultural history (for example, in writings of Foucault).

In a similar vein, these tools may change author-learner relationships. What if, rather than read and be tested on a static text on the biologic taxonomy of primate evolution, I had the opportunity to create a multimedia work explaining those relations? What if, to understand the mathematics and physics of motion, I were to craft a program that animated a looming object — instead of memorizing formulas? While the notion of learning by doing is not novel by any means, we believe it takes on a new level of intrinsic motivation in integrated multimedia environments, where one can not only feel a sense of achievement over making a creation one's own, but come to a deeper understanding of the concepts and skills that played a functional role in that construction.

What psychological design processes are engaged in such composition?

Multimedia composition is a process more like film-editing, animation and/or choreography — except that one also wants computational/procedural control over the objects and events of one's design (whether these are text units, visual scenes and objects, sounds, etc.). There are important roles for *new notations* in creating, manipulating and representing events. Scores for coordinating displays of sound, image and text over space and time are needed, and much thought needs to be given to providing variable grain to the units of these scores, and the links possible between them.

Interactive multimedia compositing is even more complex. For who has a theory of interactivity? What does it mean to imagine an audience for interactive multimedia compositions, particularly if more than simple branching contingencies are anticipated?

What skills are required for becoming multimedia composers?

When we list what is involved in multimedia composition, we are struck by the initial complexity of requirements. These include: scripting, graphics design, film-editing procedural thinking, animation, multimedia choreography and — last but not least — what we call *metamedia cognition.* By this term, we refer to a reflective understanding of such things as:

1. The strengths of specific media for specific purposes.

2. How to integrate media for psychic effects such as suspense or explanation.

But while this may be a list that seems to extend the requirements of written literacy beyond accessibility by the general public, or our educational systems, we predict otherwise. More is not always more. We find that these activities are so intrinsically captivating and empowering that they may actually help bootstrap the acquisiton of text literacy — which is often isolated from communicative expression and social interaction. This isolation will be hard to replicate for integrated multimedia literacy, and may be so engaging that it changes the mix of attention, cognition and action that embodies the cultural practices of literacy.

Part II: Insights from the "Active Eye" Project

We have introduced a number of key issues in Part I as a backdrop for thinking about mind-media relationships. In this section we will continue to examine connections between mind and media by focusing on examples from a specific project at New York University called "The Active Eye."

The main goal of The Active Eye project, headed by Michael Mills, was to develop courseware to help students explore the topic of visual perception — how their eyes and minds interpret the visual world. The courseware runs on the Apple Macintosh family of computers, and contains experiments and perceptual demonstrations relevant to a wide range of interests in the visual arts, graphic design, architecture, machine vision, computer graphics and computer interface design. A second goal of the project, in collaboration with Roy Pea, was to use multimedia experiments in perception to stimulate students' interest in science.

We will draw on our experience with The Active Eye to explore three of the major themes from Part I, namely:

1. How do representatons affect learning?

2. What psychological processes underlie the authoring of multimedia events?

3. How to build an authoring environment for designing multimedia events?

Representations and Learning

In Part I we asked, "How does the form of a representation affect one's ability to learn?" We do not pretend to have a full-fledged psychological theory of the role of representations in learning. However, the task of designing Active Eye courseware sensitized us to a number of topics which any theory of multimedia learning will have to confront. They are:

1. Navigation and Orientation
2. Dynamics and Understanding
3. Graphics Overlays and Visual Transitions
4. The Use of Precipitating Events
5. Metacommunication

These categories are not mutually exclusive, nor are they exhaustive. They are put forward as a convenient starting point for discussion.

Navigation and Orientation

One of the major tasks facing authors of multmedia courseware environments is to help users answer questions such as: "What are the contents of the database? What is its conceptual geography? How will I know where I am, where I've been and where I want to go next?" Designing effective representations for helping users navigate through vast informational worlds is a very complex task — especially when one takes into account user-dependent variables such as prior experience, level of expertise, and task goals. Many solutions have been proposed to deal with problems of navigation and orientation in information systems. A partial list includes tree structures, menus (text, iconic, pull-down), maps, and spatial metaphors (such as desktops, rooms, and so on).

In addition to traditional methods, one technique we found useful in the Active Eye project was the use of a "quicktour sampler." Reminiscent of old Hollywood movie "trailers," selecting "Quicktour" triggered a dynamic multimodal preview — a rapid-fire sequence of sixty moving-image clips each lasting less than a second — which sampled the entire set of Active Eye demonstrations and tutorials in the database (see Figure 1). The "Quicktour" was not really intended to be an efficient search or navigational mechanism. It did, however, provide a dynamic, temporally compressed experience of the sights and sounds in the database. Although it has never been formally tested, "Quicktour" seems to give users an overview impossible to achieve through a pull-down menu or tree structure.

The "Quicktour" example reminds us that it's possible to provides information about a database at different levels of abstraction; not only through taxanomic descriptions, network diagrams, maps and spatial metaphors — but at the level of *phenomenal*

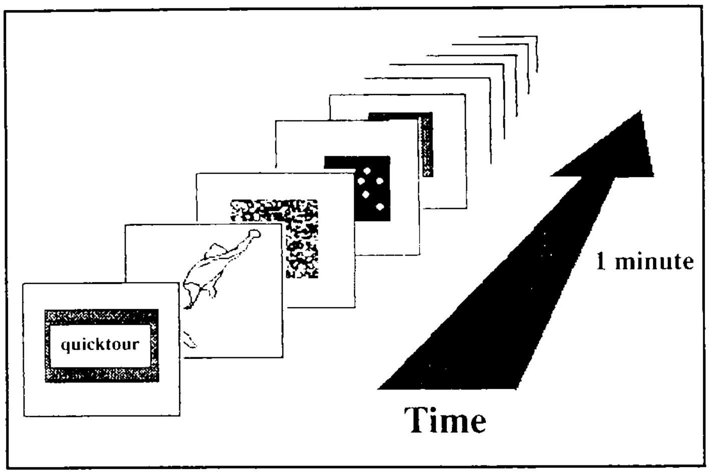

Figure 1. Dynamic multimodal preview. Opening the Quicktour file plays a rapid-fire sequence of 60 moving-image clips sampling the entire set of "Active Eye" demonstrations.

experience (sampled sights and sounds). Designing representations which provide views of a database at different levels of abstraction is likely to become an important issue in HyperCard environments, where information is not only richly interconnected, but exists as text, sound, graphics and digitized imagery. An interesting possibility, in this context, would be to explore *hybrid* representations. Imagine, for example, a tree structure showing the hierarchical organization of a database, but with "leaves" of the tree as icons. A mouse click on an icon, at a particular node, could trigger a sampler of that level of the tree.

Dynamics and Understanding

One of the thorniest questions for students of perception is the following: "How do we build a stable interpretation of the world out of the continuously changing stimuli arriving at our sensory surfaces?" How the perceptual system deals with change is a central issue for perceptual theory. Consequently, the ability to see and hear objects and events changing dynamically was one of the major reasons for building the Active Eye courseware. In general, the ability to provide dynamic displays — to show transformations across time and space — is believed by many educators to be an effective way to increase understanding of many phenomena, especially in the biological and physical sciences.

But what do we really mean by *dynamics?* For example, we can disinguish between a *dynamic representation* (the portrayal of a smoothly rotating cube on a computer display) and the representation of *dynamics* which may or may not be, in itself, dynamic (for example, a cartoon strip or series of still photographs where, by convention, space is a metaphor for time). In fact, questions about dynamics get even trickier when we talk about *internal representations,* the mental product of an interpretative act. For example, the mental representations of a dynamic situation (a falling object), may or may not in itself, possess dynamic properties.

We raise these questions lest educators have a knee-jerk response that "dynamics is always best." It is important to remember that the best means of comprehending a dynamic phenomenon, in some cases, may not be a moving image, but a static image or sequence of images which allow careful inspection and comparison of changes across "states." A classic example is Muybridge's experiment to discover "if a galloping horse ever lifts all four hooves off the ground at once?" Muybridge made a "reverse animation," turning a dynamic real-time event — the horse galloping — into a series of still-frame snapshots. The resulting sequence revealed that, indeed, a galloping horse does lift all four hooves at the same time.

So there are situations where the portrayal of dynamics, in the sense of perceiving real-time continuous transformations, may not be the optimal representation for learning. However, the reverse is also undeniably true. One good example from the Active Eye project is an experiment called Camouflage. A viewer sees a field of random dots (Figure 2). Suddenly, a random-dot square "pops" out and is seen as a figure moving against the random-dot background. If the display stops moving, the square fades into the homogenously textured background, and is no longer visible.

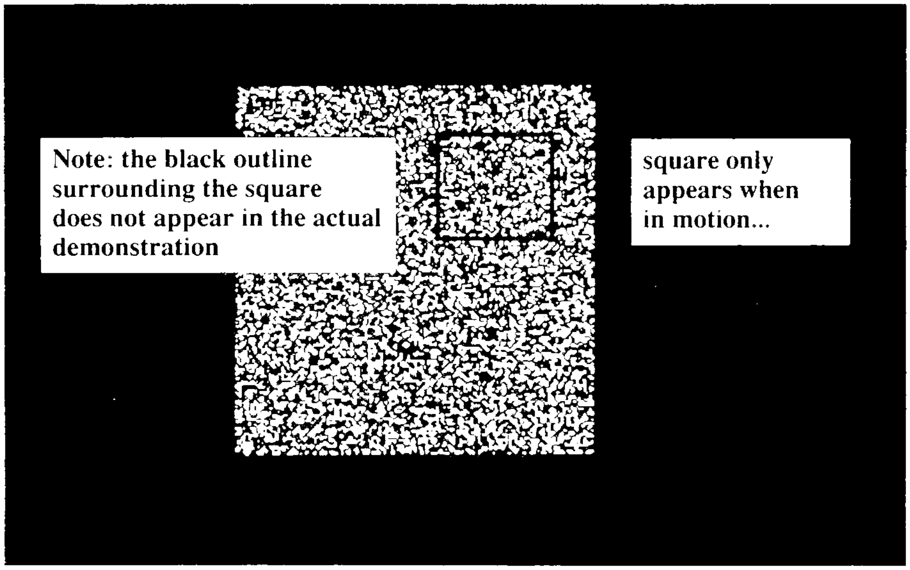

Figure 2. *Static media cannot convey certain perceptual phenomena. When the "square" stops moving against the field of random dots, it "fades" into the background. Source: A frame from* The Active Eye Project *(Mills, 1986) based on early work by James J. Gibson.*

This experiment is modeled after a series of films J. J. Gibson made in the 1950s to highlight the importance of temporally changing patterns in perception. According to Gibson, the ability to see the "square" in this demo depends not on registering discontinuities in the light reflected from the scene (edges, contours), but on perceiving a higher-order dynamic variable — an abstract pattern of "occlusion-disocclusion" of textured elements.

For our purposes, it is sufficient to note that although you can get an intellectual grasp of Camouflage through words and diagrams, you cannot *experience* the phenomenon without a dynamic medium. Moreover, we have noticed that getting a student to actually experience the phenomenon serves not only to stimulate curiosity but can also motivate further inquiry — the search for alternative explanations, for example, such as the Gestalt psychologists' notion of "common fate" or correlated change leading to perceptual grouping.

A second example from the Active Eye project which highlights the importance of dynamics are the experiments in biological motion perception. In these experiments, following on the work of Gunnary Johannson, you looked at a static display consisting of seemingly random points of light against a dark field (see Figure 3). Suddenly the lights began to move and you perceived a human shape performing familiar activities — dancing, walking, running, swinging a baseball bat, etc. As with Camouflage, this Light-Point display reveals some of the rules — based on dynamics — that the perceptual system uses to organize complex input into a meaningful pattern. Other experiments by Johannson show how these complex biological motions can be analyzed using mathematical verctor-theory. In summary, the relationship of dynamics to understanding will be very important to full-fledged multimedia learning systems. What is needed, however, is not lip service to the importance of showing things dynamically, but to specify what concepts can best be conveyed through transformations — including situations where dynamics are bext understood through the medium of still images.

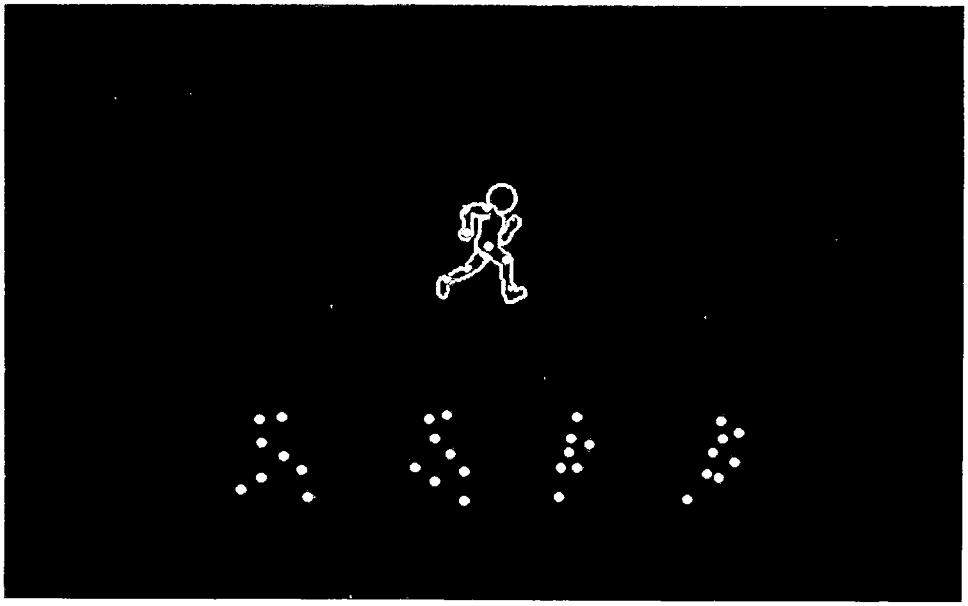

Figure 3. *Four frames from a demonstration of biological motion perception adapted from the work of Gunnar Johansson. When the display is static, you see a meaningless configuration of lights. Add motion to the display and you immediately see a "running man." Source:* The Active Eye Project *(Mills, 1986)*

Visual Transitions and Graphic Overlays

Showing events and processes dynamically is one area where multimedia educational systems may shine. But just portraying dynamics — showing objects and events changing through time — is only part of the story. The production toolkit of multimedia will also include techniques from film and the graphic arts, such as the ability to show visual transitions, and the use of graphic overlays. Although filmmakers, animators and graphics artists have long possessed a strong intuitive grasp of the cognitive and aesthetic effects of transitions and overlays, we do not yet have a detailed psychological theory on how these techniques might work to enhance learning in computer environments. (Notable beginnings can be found in the work of Rudolf Arnheim, Kristina Hooper Woolsey, and Julian Hochberg.)

Consider first the use of visual transitions in film and video: cuts, zooms, pans, dollys, etc. We experience them all the time in movies and television. But how do they function perceptually and cognitively? What mental operations do they evoke? How do we specify their impact on our interpretation of an audio-visual sequence? These questions are meaningful not only for the filmaker, but for the peception psychologist, cognitive theorist, and the designer of human-computer interfaces.

From a purely perceptual point of view, these are questions about our ability to handle "cuts" — abrupt transitions from one spatial layout to another — without getting disoriented. (This questions also applies to abrupt acoustic transitions.) How to ensure smooth transitions in the restricted spatial environments of the computer displays is a critical, underexamined theme of interface design.

Even less understood is how optical or acoustical transitions can do more than show changes in spatial layout, and be part of a complex *communicative* act. For example, a change from a global long-shot to a close-up is not only the author's way of getting the viewer to register local detail in a scene. It is the author's way of saying, "This is important! Notice the detail."

In fact, there are many techniques — which Gerald Millerson calls *production rhetoric* — used by filmakers to engage the viewer in a subtle communication grame. For example, "objective" changes in image quality (brightness, contrast, resolution, duration, cutting rhythm, etc.) are used to convey complex moods and shading. In addition to visual/acoustical transitions, graphic overlays can be powerful methods of guiding scenes and events. Again, the detailed psychological issues are not worked out, but we can give a couple of sketchy examples. At the most basic level, graphic overlays perform a "foregrounding" function; they help the viewer "parse" the elements of a display into "figure" and "ground." Consider, for instance, the use of the oval overlay in Figure 4, a frame from a tutorial on expert perception in sports.

The function of the oval is not really to call attention to itself. Its particular shape or color is not critical in this instance. Rather, the oval serves as a marker of sorts, signaling the viewer to allocate processing resources to the regions it surrounds (the raised arm of the diver), while telling the viewer to temporarily filter out the rest of the image.

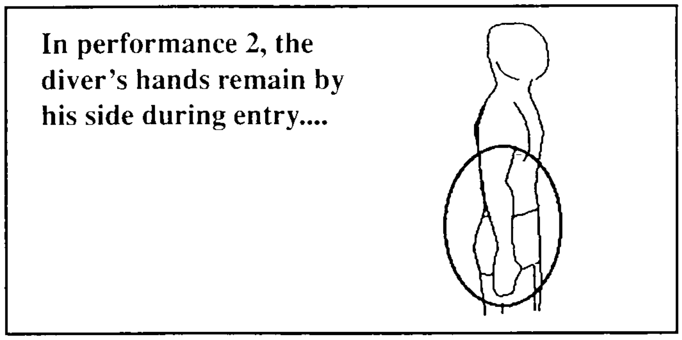

Figure 4. *Use of a graphic overlay to induce* Foregrounding. *Frame from an interactive animated tutorial on expert perception in sports. With permission of the Ontario Science Centre.*

Foregrounding is a relatively low-level cognitive operation induced by graphic overlays, and may be accomplished in a variety of ways — color changes, scaling, brightness, etc. Most likely, there is a hierarchy of cognitive functions which overlays can evoke. Consider, for example, Figure 5, which is part of a tutorial on the mechnics of a swing in gymnastics.

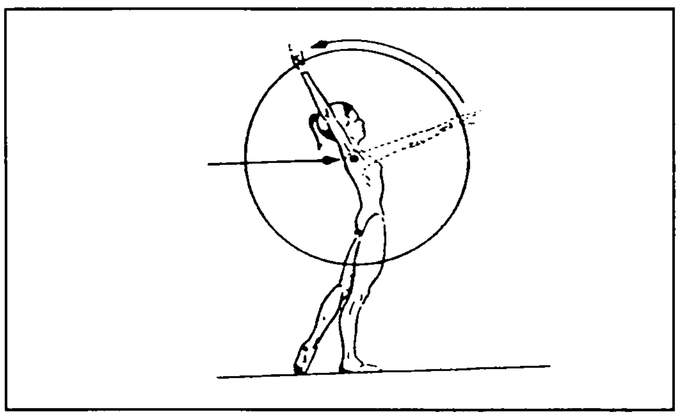

Figure 5. *Diagram from a tutorial on the mechanics of swing in gymnastics. How do the diagrammatic elements induce a mental restructuring of the "event"? Source: G.S. George,* Biomechanics of Women's Gymnastics, *Prentice-Hall, 1980.*

How do the diagrammatic elements — the circle, the arrows, the dots, the hashed lines of the arm — function together with the accompanying text to induce a mental restructuring of the static image? What cognitive operations are evoked, and in which order? A fine-grained information-processing model of how people understand diagrams is beyond the goals of this paper. But here is a partial "script" showing some of the requisite mental transformation activated by the text and diagrams:

1. The straight arrow must guide the viewer to see the dot as the location of the shoulder joint.

2. The dot, hence the shoulder joint, must be seen as being the center of a circle.

3. The gymnast's arm, beginning at the shoulder, must be perceived as forming the radius of the circle.

4. The curved arrow is an "operator," and directs the viewer to execute a 2D rotation whereby the arm is mentally rotated backwards, following a circular path which passes through the intermediate position of the extended hash-lined arm.

Like foreshortening, *mental rotation* in the 2D place is a rather low-level operation induced by an overlay. Higher-level operations might involve using diagrams to induce mental rotation along the Z-axis (in mental 3D space), fitting formal *X-Y* graphics to realistic situations, finding analogies across representations at different levels of abstraction, etc.

Helping Students Fit Diagrams to "Event Structures"

We have begun to look more systematically at the kinds of cognitive processes involved in understanding and using diagrams and figures, especially the kinds found in traditional scientific texts. There are many difficult issues involved here, and a detailed discussion is not our intent. One approach, however, is to assume that most scientific diagrams are meant to portray, not static scenes or situations, but well-formed "events."

Consider, for example, Figure 6a, a diagram from a physics text showing "the converging of light by a curved mirror." We assume that part of the task before the student is to understand the "story" of this diagram — to map this diagram into an "event structure" where the diagrammatic elements are assigned various roles in an event schema (a cognitive structure with "actions," "agents," "instruments," "objects," and so on). Seen in these terms, the role of multimedia systems becomes clearer. It is through using such devices as dynamics, visual transitions, and overlays, that the student makes the appropriate mappings between the diagram and its underlying "event." It helps parse the diagram into forces, causes, and/or instruments where the semantics of the diagrammatic elements (such as arrows, points, etc.) are unambiguous. One advantage multimedia systems have over static text is the ability to show a student how a diagram such as Figure 6a got that way — by using, for example, a dynamic sequence which makes explicit the transition from a realistic situation (Figure 6b) to its abstract diagrammatic counterpart (Figure 6a).

To sum up, we suggest that a deeper understanding of the cognitive effects of graphic overlays and visual transitions will yield payoffs, not only in terms of better understanding the role of representations in learning, but also in terms of designing more effective multimedia learning environments.

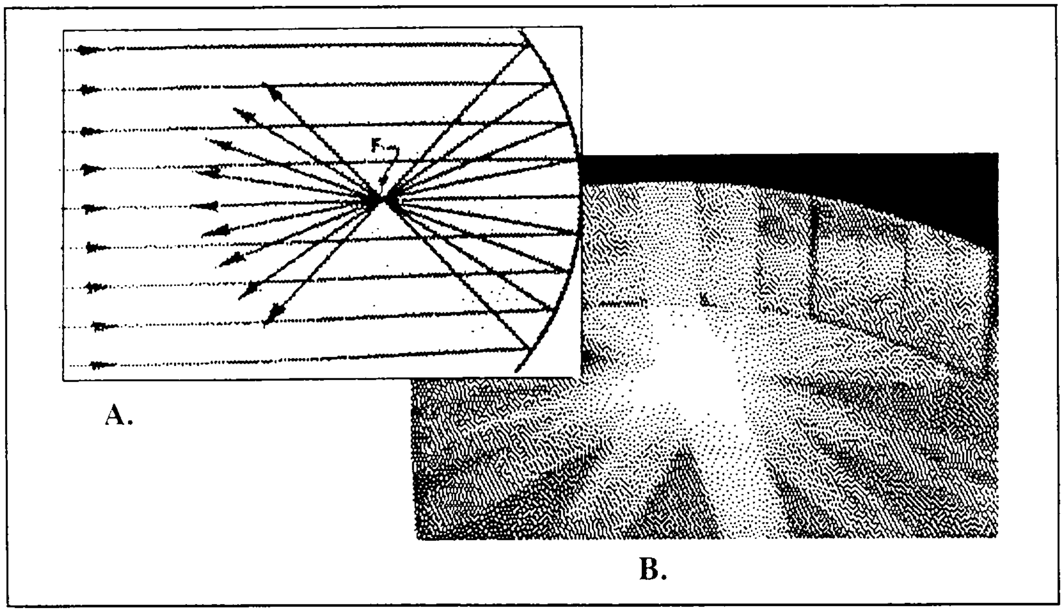

Figure 6. *(A) A diagram from a physics text showing the "converging of light by a curved mirror. (B) Part of a dynamic sequence showing the effects of increasing the number of angled plane mirrors on the light rays. How could one make a dynamic sequence showing the gradual transition between the realistic situation (B) and its diagrammatic abstraction (A)? Adapted from: PSSC* Physics *(1960).*

Precipitating Events

Another important theme raised in Part I was that a powerful hypermedia system — where students have access to a vast web of interconnected facts and phenomena— can be a double-edged sword. On one hand, a powerful browsing/authoring medium such as HyperCard gives students undreamed of access to a vast reservoir of potentially relevant learning experiences. On the other hand, the same medium — where everything can be linked to everything else — may cause severe information overload and educational paralysis if *constraints* are not built in. This is also a difficult theoretical issue facing cognitive psychologists. How does the *human mind* — the ultimate interconnected database — manage to constrain its own creating/gathering information activities? How does it avoid getting lost in a vast web of interconnected ideas? (In fact, we all know people who have problems "getting to the point" for precisely this reason!)

One attempt to build some associational constraints into the Active Eye database is the use of what we call *precipitating events.* The idea is straightforward. It involves designing demonstrations in the form of "experiments" where apparently similar situations — the same in all respects but one — produce strikingly different perceptual effects. The student is then left with a clear question: "Why did the effect occur in one case but not the other?" They are then presented with a set of related demonstrations or further experiments to conduct — a possible path through the database — which could help provide answers to their question.

An example of a precipitating event from the Active Eye project is the experiment called "Through the Eye of the Needle," modelled after a well-known experiment by T. E. Parks. In this demonstration, people try to identify a familiar object based on changing contours appearing within a small aperture or slit (see Figure 7). In one case, the slit remains stationary while the object's contours move back and forth beneath it. In this "stationary slit" condition, the object is very hard to identify. The second condition is the same as before, except that this time the slit moves back and forth and the object beneath it remains stationary; in this case, successive views of the object are not only presented through time, but in adjacent spatial locations. Most people quickly perceive the outline of a familiar object (a camel) in the "moving slit" condition. So here is a precipitating event — a puzzle. Why is it easier to identify the camel in the case of the moving slit?

A. B.

Figure 7. *When the observer's gaze is fixed, it is very hard to identify the camel at (B) based on successive views of the contours appearing within the stationary "Slit" at (A). If the slit appears to move back and forth rapidly across the stationary outline, then the camel at (B) is much easier to identify. Adapted from Parks (1956) and used with permission of the Ontario Science Centre.*

Text cards provide a preliminary discussion of the effect, and highlghted glossary items serve as gateways to further tutorials and experiments (see Figure 8). For example, one aspect of the "needle" phenomenon has to do with *retinal persistence.* Clicking on this glossary item provides a pathway to additional experiements involving retinal persistence. Other pathways leads to theoretical perspectives of the "needle" phenomenon: for example, how we integrate the results of momentary gazes of the world — high in detail, but fragmentary — into a coherent scene.

The use of precipitating events involves adapting one traditional method of scientific inquiry — the notion of a controlled experiment — to build in some constraints to the large number of potential associational links within the Active Eye database. Admittedly, the content area of perception lends itself well to this approach; however, the use of demonstrations or experiences which lead to clear questions — and constrain the kinds of answers — may be a useful strategy for other domains as well.

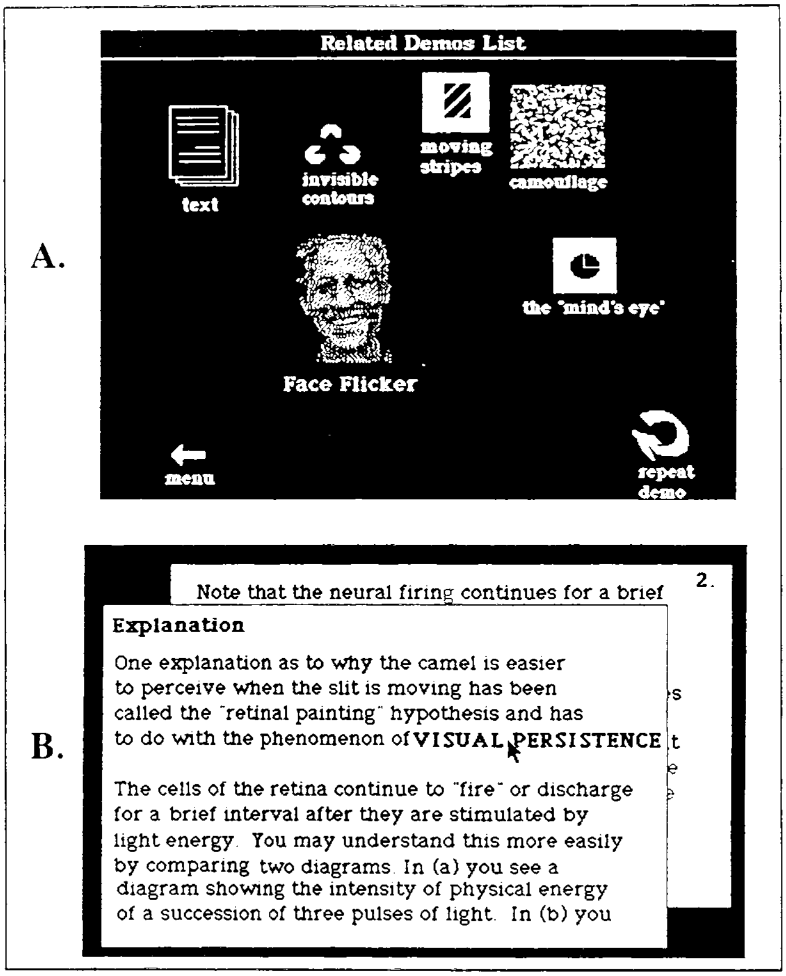

Figure 8. *(A) Pathways to further demos and experiments spawned by the "Eye of the Needle" phenomenon; (B) Use of highlighted glossary items which activate animated tutorials.*

Metacommunication

In his book *Steps Toward An Ecology Of Mind,* Gregory Bateson discussed the dual nature of human communication. He tells us that any communicative act transmits not only an explicit *content* (what the message is about), but also an implicit *relationship* between the sender and receiver (its metacommunicational aspect). It is usually the gestural and vocal characteristics of the message which carry the meta-communicational information. When your boss says, "You're a terrific writer!," you may or may not feel terrific, depending on how he or she says it. In other words, it's not just what you say but *how* you say it that counts. John Searle calls this phenomenon "speech acts."

Builders of information systems are becoming increasingly aware that metacommunicational issues permeate virtually every aspect of computer system/software designs. These range from the choice of font, page layout, and the wording of "help" messages, to the actual configuration of hardware elements that make up the system. Perhaps the Apple Macintosh computer, more than any other machine in history, has successfully anticipated how all aspects of its design establishes a relationship between the user, the system designer, and the company.

Metacommunicational issues will be no less important to the design of effective multimedia systems in education. We tried to be conscious of these issues in designing the Active Eye database. For example, we explicitly injected elements of humor, whimsy and surprise from time to time into the database (a digitized image of the author appears as stimulus in various demos) as a way of establishing a kind of relationship between ourselves as teachers/designers and the student. Although it is hard to state exactly, the relationship we sought was quite subtle: namely, that while the content of demonstrations is important and serious, it can also be fun — and is nothing to be afraid of or feel stupid about.

A word of warning about the use of humor and whimsy. If overdone, it can backfire. Everything becomes a joke and nothing gets learned. There is probably no guidebook that can dictate the right mix of humor and seriousness in the design of educational software. Nonetheless, the designer should be aware of this fine line.

Psychological Issues in Authoring Multimedia "Events"

Let us move now from issues of learning to those of authoring or composing multimedia documents. First, what does it mean to "compose" a multimedia document?

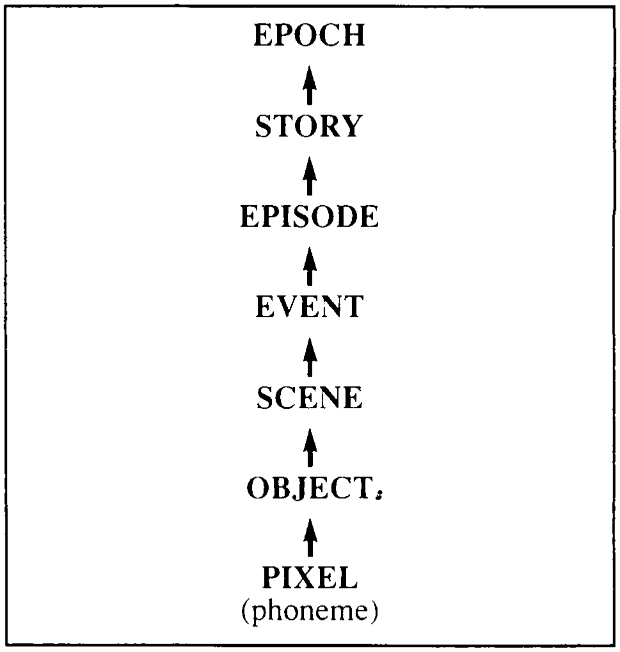

Figure 9. Multi-media authoring can be thought of as occurring at a number of hierarchically nested levels.

It may be useful to think of multimedia authoring tasks as those involving operations — both cognitive and physical — on a number of hierarchically nested units (see Figure 9). The smallest unit for computer generated graphics and text is that of the *pixel* (or *phoneme,* in the auditory domain). *Objects* are formed by groups of pixels. Hence, if an author's attention is focused, not on individual pixels, but on objects — for example, when moving the image of a cup around the screen — we might say the author is working on the next higher level unit. Objects are units which can become embedded in static *scenes* in which the author considers the relationships between several objects. When the objects begin to change in relation to each other — to move through time or space — we get an *event,* a critically important psychological unit in multimedia authoring. Several events may be thought of as constituting an *episode,* yet another, higher level description. Taken together, a number of episodes may form a *story.* Finally, many stories may be woven together over great stretches of time to form an *epoch,* a very global unit indeed.

In addition to hierarchical units, multimedia authoring can also be thought of as requiring the managing of data across different sense modalities: visual and acoustical being the major ones at present, although kinesthetic, olfactory and gustatory data may be added at some point in the future. Moreover, editing operations in each sense modality can occur at different hierarchical levels.

Building Systems for Composing Multimedia "Events"

We believe that a better understanding of the psychological processes underlying multimedia composition might lead to new creative environments: hardware and software tools that facilitate the physical and conceptual operations involved in multimedia authoring. Can we already foresee what some of these tools might be? It is probably fair to say that no single system, at present, is the ideal multimedia authoring environment. Apple Computer, Inc.'s HyperCard possesses many of the earmarks of the ideal multimedia authoring environment. It has tools and representations not only for creating sounds, images and text, but also for helping the author perceive potential links and associations among these elements. Moreover, HyperCard designers have the ability for "computational control" over elements at different conceptual levels through the use of the HyperTalk programming language. Even HyperCard, however, has not fully exhausted the list of tools which one might want to author or choreograph complex, dynamic multimodal *events.*

Here it is informative to look at another authoring environment, Macromind's "Videoworks Interactive." This system, initially designed as a program for doing 2D animation, has recently been extended to provide tools for creating interactive animations. One of the most fascinating features of Videoworks Interactive is what the call "the score." The score is an ingenious notational system designed to facilitate what we might call event processing. Figure 10a shows some features of the Videoworks Interactive interface. Here one sees the score window, as well as two other windows: the "control panel" which has buttons for playing, stopping, stepping through the animation; and the "cast" window which shows seven of the 256 possible "castmembers" or graphical objects which can be brought onto the "stage"— the actual area where the animation will unfold.

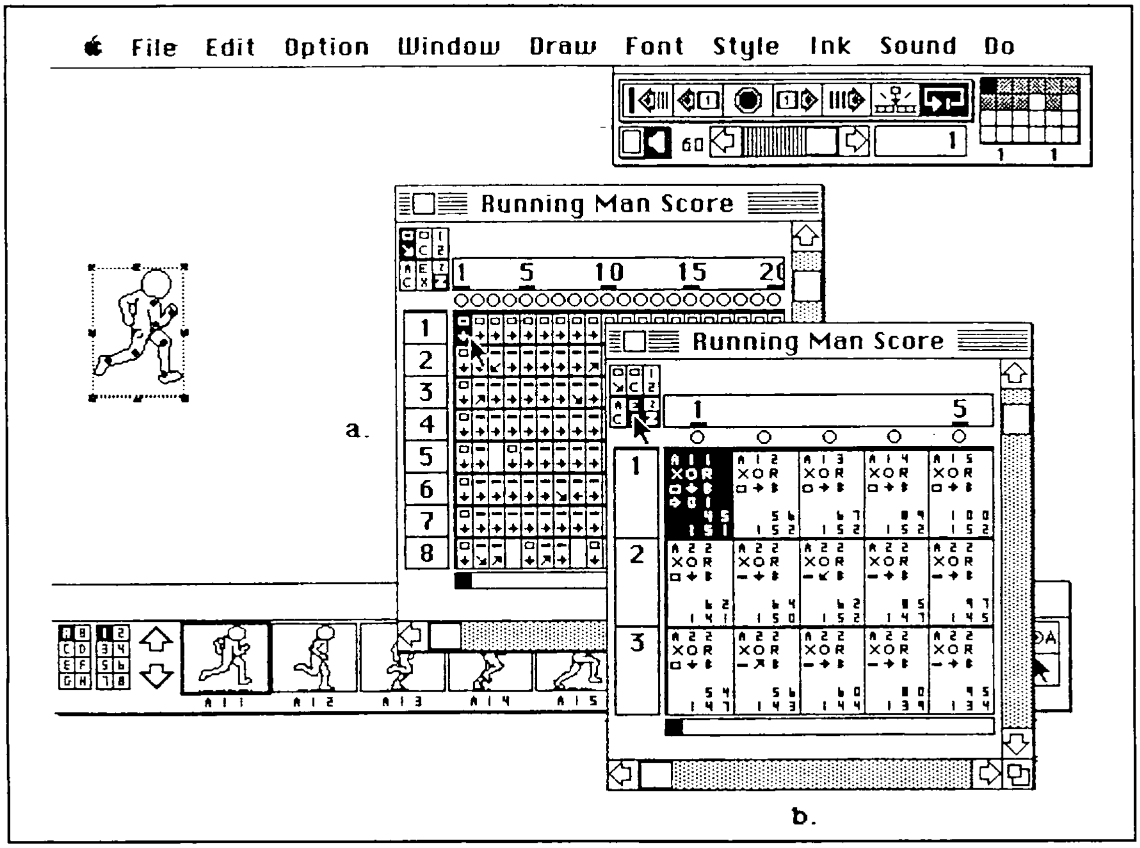

Figure 10. *(A) Part of the Videoworks interface showing the Score window, the Cast window and the Control Panel. (B) Shows an expanded Score from Videoworks Interactive with the symbols used to indicate "event" information (screen coordinates, object-type, relative motion, etc.)*

Look more closely at the score window in Figures 10a and 10b (a zoomed-in detail). By glancing at the symbols in Score, the author can quickly grasp certain kinds of relevant information: which "sprite" is appearing in a given frame, the object type (Quickdraw, primitive or bitmap), its screen coordinates, inking mode, action codes (user-definable actions such as frame jumps), the type of sound, foreground priority, and the like. Moreover, editing operations such as cutting, pasting, positioning, etc. take place by manipulating frames or groups of frames from in the score itself — rather than on the graphical or acoustical objects. The important point is that a notational system — the score — was designed to be cognitively congenial to authoring at the level of an event.

Conclusion

Videoworks Interactive should receive high marks for helping one think about such things as shape, time, motion, and local causality: for example, the "presentational" aspects of events (an event's surface structure). It is less suited, however, to helping one think about hierarchical relations, links, structural views, alternative descriptions, and computational control (deep structures). For the latter, HyperCard seems better at present. This suggests that a productive avenue for future work would be to examine how the best properties of these two systems can be merged and extended in a single system. To accomplish this, however, will require an interdisciplinary approach involving system designers, multimedia composers and psychologists: that is, where a deeper understanding of the psychological processes underlying event authoring can lead to suggestions for new representational tools and, conversely, where new methods for designing and composing events can lead to better theories of human congition, communication and creativity. Fostering cooperation across disciplines may be one of the most exciting by-products of the mind-media dialogueue.

ROY D. PEA

Dr. Roy D. Pea is a John Evans Professor of Education and the Learning Sciences at Northwestern University, and Consulting Professor at Stanford University. He is a cognitive scientist with special interests in integrating research and the design of effective learning environments for science, programming, and multimedia computing. He has published one book and sixty articles and reports on learning with computers, cognitive science, cognitive development, and language learning.

MICHAEL MILLS

Michael Mills is a Principal Scientist in the Human Interface Group within Apple Computer's Advanced Technology division. Since coming to Apple in 1988, Dr. Mills has led a team of programmers, engineers and designers in developing interface techniques for incorporating and manipulating dynamic data (digital movies, sounds and animations) into Macintosh applications. Mills' early prototypes of "dynamic documents" were a primary catalyst for Quicktime™, Apple's new system software enabling applications to include digital movies. For the past year, Mills has been working with the Quicktime team ondeveloping human interface components for capturing, compressing, navigatingand editing digital movies.

Before coming to Apple, Mills was an Associate Professor in the Interactive Telecommunications Program at New York University where he taught courses on computers, cognition and computer animation. An NSF grant allowed him and Dr. Roy Pea to study the role of cognitive processes in understanding scientific diagrams. He also developed interactive computer exhibits in perception for the New York Hall of Science, General Physics Corporation, Apple Computer and the Canadian Government. Mills is the author of *The Active Eye Guidebook and Software* (Lawrence Erlbaum Associates).

From 1975 to I982, Mills was a professor in the graduate program of communications at the Universite de Montreal where he taught courses on cognition, research methods and communications laboratory. At this time, Mills collaborated with the Government of Canada on cognitive and perceptual issues related to Telidon, and published articles in the area of pictorial representation and motion perception. Before coming to Montreal, Mills lived in France and took courses at the Universite Aix-Marseille.