A MULTI-CHANNEL INFRASTRUCTURE FOR PRESENTING NONLINEAR HYPERMEDIA

TINA SCHNEIDER

South Beach Software
P. O. Box 29395
San Francisco, California, 94129, USA
E-mail: tina@sobeach.com

STEPHEN WILLIAM SMOLIAR

FX Palo Alto Laboratory, Inc.
3400 Hillview Avenue, Building 4
Palo Alto, California, 94304, USA
E-mail: smoliar@pal.xerox.com

While hypermedia is usually presented as a way to offer content in a nonlinear manner, hypermedia structure tends to reinforce the assumption that reading is basically a linear process. Link structures provide a means by which the reader may choose different paths to traverse; but each of these paths is fundamentally linear, revealed through either a block of text or a well-defined chain of links. While there are experiences that get beyond such linear constraints, such as driving a car, it is very hard to capture this kind of non-linearity, characterized by multiple sources of stimuli competing for attention, in a hypermedia document. This paper presents a multi-channel document infrastructure that provides a means by which all such sources of attention are presented on a single "page" (i.e., a display with which the reader interacts) and move between background and foreground in response to the activities of the reader. The infrastructure thus controls the presentation of content with respect to four dimensions: visual, audio, interaction support, and rhythm.

1 Introduction: A New Style of Hypermedia

The **TED9 Trip Report** is an interactive multimedia document that chronicles the events at the TED9 conference attended by James Baker, CEO of FX Palo Alto Laboratory, Inc. The document is nonlinear, allowing readers to traverse the content through a rich variety of paths and to view multiple channels of sound and video at one time. Unlike most documents with a vast amount of information, this document displays all of the information on one "page" and is based upon a continuous, rhythmic temporal infrastructure learned and played out by the reader.

The TED (Technology, Entertainment, Design) conferences, organized every February by Richard Saul Wurman, have been some of the most content-rich and media-intensive offerings of futuristic thinking over the past decade. For approximately half of this decade, Baker has attended the conference, armed with recording devices for video, still images, and audio; and each year he has faced the

challenge of documenting his experience in a trip report. The problem is one of extracting an adequately descriptive account out of an overabundance of source material, a problem that was only exacerbated by a follow-up CD-ROM collection of the talks provided by the conference itself. Baker's goal has always been to communicate his experience, not simply as a chronology of events, but as a means by which a reader could share a similar experience of participation. This goal has proved to be a major authoring challenge, leading to hypermedia documents developed with both an internal authoring system [9.] and PowerPoint; but prior to TED9 (February 1999) each of these attempts was not particularly satisfactory. Consequently, for TED9 Baker requested one of us (Schneider), an artist with background in interactive narrative [8.], to conceive of a new approach to reporting the conference.

Schneider wanted to create an interactive interface that would engage the reader by keeping all of the information tied to a single "page." She wanted to create a sense of a "reading space," where the reader's sense of place would be confined within a single border, steering away from the hyperlink paradigm of navigating through pages to get to the content. Her other focus was to view the reading process as a juggling act between foreground and background information, obliging the reader to deal with more sources of stimuli than his attention can manage. This is a report of the resulting digital document.

2 Background

Hypermedia is generally conceived as a means by which content (usually involving multiple media such as text, images, audio, and video) is embedded in a network structure. Each node of this network represents some unit of content, and the directed edges of the network represent logical associations among these units. These edges are then represented explicitly in the content representation in a form known as hyperlinks.

Since a network is represented in a nonlinear structure, hypermedia tends to be presented as a way to offer content in a nonlinear manner. However, the reading experience of visiting a node and selecting a hyperlink to another node is fundamentally linear: The reader must "traverse" the network by defining one or more paths through it, following each path linearly. Unfortunately, a complex network can impose significant cognitive overhead and may even provoke the disorientation of the reader; and techniques such as scripted documents [13.] have been implemented to make the negotiation of such paths more manageable. However, the reading experience is basically as linear as it is with a printed book, modified only by a different approach to offering the reader choices about what to read next.

Nevertheless, there are many situations in which content is experienced in a truly nonlinear manner. Driving in a car is an obvious example. The multiple signals that are picked up by the driver are synthesized in a manner that allows the driver to drive, talk, listen to the radio or a passenger, keep the car straight, see the accident ahead, make a phone call, etc. All of these experiences involve multiple sources of stimuli competing for attention. For the driver these sources are in constant motion between background and foreground as the attention shifts from one source to another. What is being proposed here is an infrastructure that abandons the linear traversal required by conventional hypermedia and replaces it with an implementation of this motion between background and foreground that applies to all sources of content.

The **TED9 Trip Report** is thus an attempt to avoid the problems of structural complexity entailed by links. Associations are not implemented explicitly as connecting paths that the reader must select and traverse. Instead they are experienced through the rhythm of the constant changes taking place in the layout. Structure thus emerges from the temporal experience; and, as the structure emerges, so does the reader's sense of a description of the TED conference in both its entirety and the specifics of particularly salient events.

3 Keeping it on the Page: Conceptual Design

3.1 Design Influences and Aesthetics

The primary design priority was to replace the experience of traversing a complex network of hyperlinks with the sense of experiencing the entire document within a single reading space. The document was thus designed with the idea that everything would appear on a single page, but the content of that page would be in constant motion. The challenge was to make the resulting display rich in content without becoming so overwhelming and uninformative that the reader would lose attention and focus.

The second priority was to provide the reader, from the start, with a structural sense of closure. The borders of a digital document are not the same as borders for content printed on paper. The interface constructs of Web pages, such as scroll bars and hyperlinks, frustrate this sense of closure by always threatening the reader with the prospect of more pages to decipher. As an alternative, all the movement on the single reading space is cyclic, allowing the reader to discover a sense of closure within the repetition of the cycles.

However, the movement of the display does more than allow the reader to discover a sense of closure. It is also the primary device for engaging and maintaining the reader's attention. Using movement in this capacity was the result of two influences: One concerns how we process multiple sources of information

vying for our attention (for example listening to the radio while driving a car). The other is the way in which comic book narratives engage the reader.

It is very hard to ask a user to sit still at a computer and read a text document or listen to a long audio file. The reader fidgets, opens mail, and may even move about the room. In a car, on the other hand, a radio provides a background source behind the listener's foreground task of getting from one place to another. Visual movement cues the driver to pay attention at the appropriate times; and, between those times, the driver experiences the "virtual movement" of listening to the radio while attending to any other auditory cues. The result is that the radio signal probably receives more attention within a car than it would if delivered as streaming audio through a computer.

In comic books movement is not only depicted within individual frames but also represented by the layout of the frames themselves [5.]. Large frames dominating a page indicate a longer viewing time by the reader, thus indicating a passage of time. Smaller angled frames can indicate quick takes and flashes in a moment. What is interesting is the way in which movement is prevalent even in a document that is considered flat. While this technique does not translate readily to interactive documents, such as computer displays, it illustrates the need to accommodate movement as a device for maintaining reader attention.

Most current Web documents tend to reflect the layout principles of paper documents. Pages are covered with hyperlinks, images, flashing bulletins, icons and logos, where the "real estate" is the white space on the screen. While comic books can use white space to convey a sense of time and mood, the white space on a Web document is only there to be filled by hypertext objects. Our design challenge was to engage movement in a way that white space on the screen would be not just flat territory but the fundamental instrument of engaging reader attention.

As has already been observed, this problem was solved by designing movement based on *cyclic* patterns. Our hypothesis was that, if these cycles were properly designed, the need for more conventional interface mechanisms, such as buttons or pull-down menus, could be significantly reduced. The movement would both encourage and guide the reader through the navigation of the document's content, and the cycles would engage the reader by shifting the background and foreground information. The reader is thus free to decide, at any time, what is central and what is peripheral, knowing that the content possibilities will always cycle around and make something else eligible for central attention. By being immersed in the unfolding of periodic cycles, the reader also feels situated within the entirety of the document.

This conception of structure emerging from the dynamics of a rhythmic process, rather than the static properties of a data architecture, is actually not particularly new. It is certainly not new to anyone involved with the "lively arts," such as music or choreography. One might even say that a musical experience

involves apprehending a sense of how a given temporal interval is structured as its passing is experienced [10.]. A TED conference involves a commitment of attention over several days, each beginning very early and continuing until very late. It is a temporal experience that may be on the scale of only one of the most massive of musical events, the four operas of Richard Wagner's *Der Ring des Nibelungen*. Making sense out of the passing of time is thus fundamental to the TED experience; so we feel that it is important that a sense of rhythm be fundamental to the communication of that experience. Furthermore, rhythm serves as an "infrastructure dimension" that modulates three other dimensions competing for the reader's attention: visual sources, audio sources, and affordances for interaction. This modulation will be addressed as we now consider how our goals for design could actually be implemented.

3.2 Navigation and Structure

3.2.1 The "Page"

The term "page" is used to indicate that all content is presented through a single bordered environment. In most operating systems this is implemented as a single window that would most likely fill the entire area of the display device. Thus, in contrast to conventional hypermedia, there are not multiple pages that are traversed by following links. Instead, there is only one displayed area, whose content can change both autonomously and in response to interaction with the reader, meaning that the network infrastructure of the hypertext is reduced to a single node and (therefore) no directed edges.

3.2.2 The Layout

The basic layout of the page consists of a grid of windows. The current implementation presents nine such windows in a three-by-three array. Each window displays either a video or a cyclic repetition of a set of images; and it may have a sound track, either directly associated with the video or added as a "narration track." Additional features indicating affordances for interaction may also be displayed. These may include buttons, such as those for the control of a video player, forms for the user to fill in, or even a Web browser. Each of these windows serves as a specific locus of attention.

While each window is designed to be dynamic, the grid itself is static. The reader is thus confronted with a stable spatial array of multiple windows, all in motion and all vying for attention. This video wall approach can take advantage of the virtuality of software, because it is easy for the author to design the grid in a manner most suitable for the final display. This means that the rows and columns do not necessarily have to be even; nor need all windows be restricted to the same

size, dimensions, or even shape. (Thus far, however, all windows have been implemented in a single size with an aspect ratio to support digital video.) However, once the author defines the grid, it remains fixed for the reading of the document being authored.

3.2.3 The Cursor

Reader interaction is controlled primarily through a cursor, which changes shape as it is moved to inform the reader whether a particular window may be manipulated. The shape of the cursor then indicates what sort of manipulation may take place. Manipulation can involve "seizing" a static image and requesting a video "expansion" or automatically controlling the volume on the sound tracks for the different windows (all of which may be "playing" simultaneously). The reader can thus use the cursor to decide which of the multiple sources of content on the page are to occupy the foreground of attention.

3.2.4 Rhythm

Rhythm is provided by the changing of displayed content in each window on the page. These changes may be implemented by either video or a cyclic display of static images or text. The primary control over rhythm involves the rate of change in these latter displays. Different windows may change at different paces. This allows the author to orchestrate the display of information within an infinite spatial-temporal framework. The framework does not support the usual linear protocol of going back or moving forward, such as is found in controls for a VCR. Instead, the framework itself is a continuous loop, creating the effect of an infinite time-line. Spatial-temporal patterns designed by the author and detected by the user are distinguished by the rate of change and by the differences in the visual imagery.

One aspect of this framework is that all windows are in motion. The motion works to keep the reader engaged to anticipate what comes next. It also works to keep information flowing. A single image suggests one intended meaning while several different images relay several pieces of information. Thus, rather than organizing information in terms of paths through a network, this time-based method for displaying content enhances the overall attention of the reader.

Another aspect of this spatial-temporal framework is the control of the rate of change of all information displayed on the screen. This provides a way of directing the user's attention to new information by recognizing a disruption in a current spatial-temporal pattern. An example of such a disruption would be the change of display in a window from a sequence of static images to full video.

The reading experience thus depends on the perception of a rhythm of all stimuli, rather than the focused perception and cognition of a single source of stimuli. Such a reading experience may be compared to that of driving a car. It

involves shifting focus from the road ahead to the other signals coming in from the car or the window. Reading also requires the opportunity to shift perspective.

The looping of images and video also contribute to the sense of rhythm. Once the loops are recognized, the reader can also recognize the opportunity to see all of the information without having to manipulate the screen. This provides that sense of closure that many hypermedia documents, particularly those based on complex network structures, lack. Just as the reader of a book can quickly leaf through the entire artifact from beginning to end, this reader can use the temporal loops to get a summary of what the content is about without having to go deeper into the document by losing any sense of place. The reader is thus provided with a "virtual border" and is relieved of the problem, imposed by many hypermedia documents, of having to navigate from page to page to determine where the end is located.

3.2.5 Text

Text is presented to the reader primarily in the form of pop-up windows. Such a window supplements information associated with the location of the cursor. This could involve a simple description of what may be achieved through interaction, or it could be a menu presenting alternatives for interaction. It was the intent of the author to keep the text moving so that it did not become wasted space in the sense that text becomes noise on a page unless it is read. Once read it only occupies empty space. This space could be used more efficiently. Having the text loop consistently like the images would still be considered noise because the text is only valuable when it is read. Thus, it just appears as shifting noise on the screen whereas the visual information given off by the pictures and video provide an immediate sense of the data they represent. Single words flashing on the screen could be used as icons; but they must still be "read" by the user, which may ultimately be distracting or unnecessary. Text thus appears only when it is explicitly requested by the user.

3.2.6 Sound

In this document multiple sound files can play simultaneously. However, the user can exercise control over the volume of one of those files; and this control cues the volume levels of the other files to be automatically lowered for clarity. In this way, the listener can peruse a sound file to determine if she wants to bring it to the foreground of her attention without necessarily sacrificing the other sources competing for that attention. This approach also provides a certain amount of spontaneity and another level of discovery to the reading experience. While looking at the images and moving the cursor about the screen the listener can pick up sounds that may direct her attention to a particular section of a window that she otherwise might have missed.

4 Related Work

Viewed as a means to access a rich body of content, this work bears at least a family resemblance to video indexing and may thus be compared with systems that offer image-based cues to video content, such as [3. , 4. , 6. , 12.], as well as a system that explicitly involves an index structured in comic book form [11.]. However, these indexing approaches basically provide *links* to video content, rather than integrating the video content directly into the user interface. The index structure itself is thus a static collection of links, rather than following our approach to an interface based on movement and rhythm.

Another approach to presenting video content is the navigable movie, such as Tirtza Even's CityQuilt [1.]. The idea here is that the reader should be both empowered and encouraged to navigate within the borders of a video, rather than respecting its *a priori* linear temporal sequencing of events. CityQuilt provides another example of how a reader can interact with video directly in a seamless interface that does not require buttons or pull-down menus as interface devices for navigation.

The idea that reading should be a participatory experience is familiar to the world of hypertext. The reader who is obliged to explore the text by selecting and pursuing links is engaged in active participation in a manner not generally afforded when the text is rendered on a printed page. However, the hypertext experience that probably comes closest to providing a "TED experience" is HyperCafe [7.]:

> *This program places the user in a virtual cafe, composed primarily of digital video clips of actors involved in fictional conversations in the cafe;* HyperCafe *allows the user to follow different conversations, and offers dynamic opportunities of interaction via* temporal, spatial-temporal *and* textual *links to present alternative narratives. Textual elements are also present in the form of explanatory text, contradictory subtitles, and intruding narratives.*

Like the **TED9 Trip Report**, HyperCafe used a spatial-temporal framework to help design the interface. However, a major distinction is in the display of the various stories taking place. In HyperCafe images move across the screen as static pictures which act as icons for audio files and movies that relate to the particular story. Movement is constant throughout the piece in that the viewer selects images that move across the screen. In the **TED9 Trip Report**, on the other hand, the windows are static; and the movement occurs only within the windows.

There has also been a study of the effects of viewer cognitive responses to various interface issues such as whether text or images dominate a page and how multiple sound files can be disturbing to readers [2.]. This study addressed issues of concern in making the TED9 project. Although we did not have access to the study before designing the piece, it was interesting to note some similarities between their discoveries and our design model for the trip report.

5 The Actual Implementation

Preparing the trip report for TED9 involved working with an abundance of media material, both recorded during the conference and subsequently provided on a CD-ROM by the conference organizers. The authoring problem was one of deciding how this abundance could be presented to the reader of the trip report without becoming overwhelming. Since the entire TED experience is one of competition for attention, the infrastructure was applied to provide the reader with a sense of that competition. Explanatory material was provided in the form of narration tracks, provided by James Baker and John Doherty, both of whom were at the conference; but much of the content also "speaks for itself" in the form of "live recorded" video sound tracks.

The layout of the trip report consists of nine active windows in a three-by-three grid, all displayed on one page. The upper left and lower right windows display one-minute videos. Each video consists of four segments, each of which is a

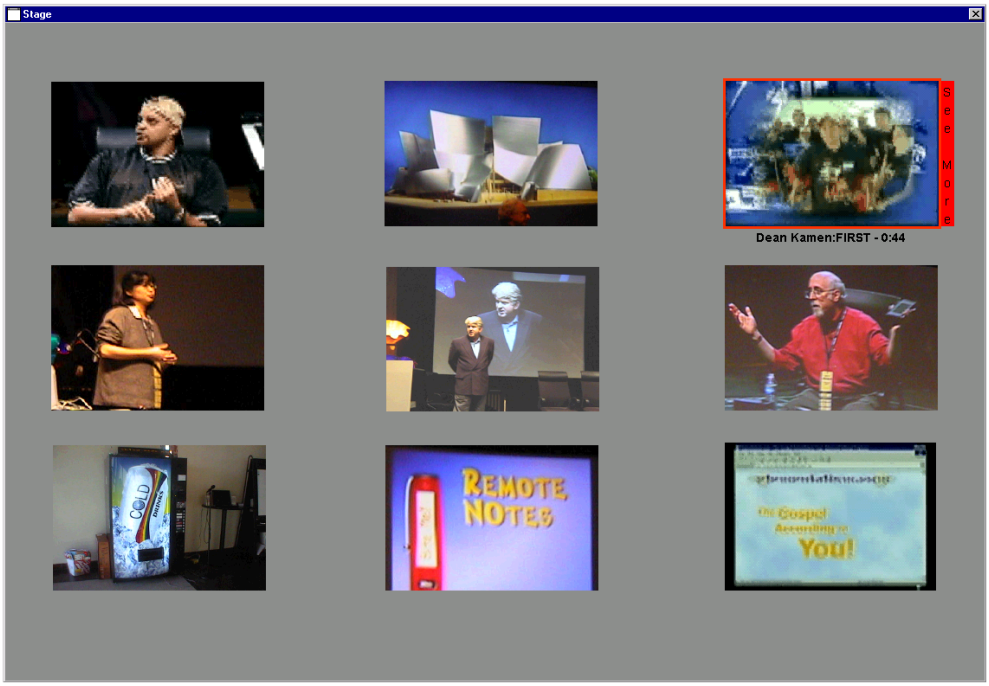

Figure 1: Example of selected story.

recording, approximately twenty seconds in length, of a different TED event. Both videos are looped and act as background information. This information gives a sense of the tone of the conference without needing to become the center of attention. The looping gives the reader the opportunity to catch a particular scene while browsing the other windows.

Each of the other seven windows contains two stories each. Each story is represented by a static image, and the images alternate roughly every five seconds. Each story also has a narration track that is approximately 30 seconds in duration. This narration can be played when the cursor is placed over the image of the story and the mouse is clicked.

Twelve of the fourteen stories also offer supplementary video, recorded at and provided by the TED conference. This video can be played when the reader clicks a "See More" text window that appears beside the story window. Figure 1 illustrates an example of the layout where the reader is selecting a window with a "See More" option.

When the user clicks the "See More" bar, the associated video will then replace whatever is currently being displayed in the window. Unlike previous displays, this video has longer duration (up to fifteen minutes) and appears with a control bar allowing the viewer to scroll through the content. This change in display (by allowing the control bar to appear) also sets up a visual distinction among the other stimuli on the page.

The only other text that is displayed appears as a caption below any window in the grid. (See Figure 1.) The caption provides a title of the image being displayed and informs the reader of the duration of its associated narration track. However, the user is free to browse the rest of the grid while the video is playing. In Figure 2 we see an example in which a video is playing in one window while the reader is discovering another video in a neighboring window.

All of this material is prefaced with an "overture" consisting of an audio file of people talking. An image of the conference site appears in the center of the screen with the title "**TED9 Trip Report**." When the cursor is over the image it turns to a hand indicating that it is to be selected. Once selected a video starts playing. The text at the bottom of the video indicates that it is the introduction by Jim Baker and is approximately 40 seconds long. As the introduction is played, background music from the conference starts playing and increases in volume as the introduction finishes. Meanwhile the nine windows begin to populate the screen one at a time, displaying pictures of all of the speakers that Baker will be discussing. By the time the introduction is finished, all nine windows will be actively looping; and the cursor will indicate that the images can be selected, displaying text description in response to being moved by the reader.

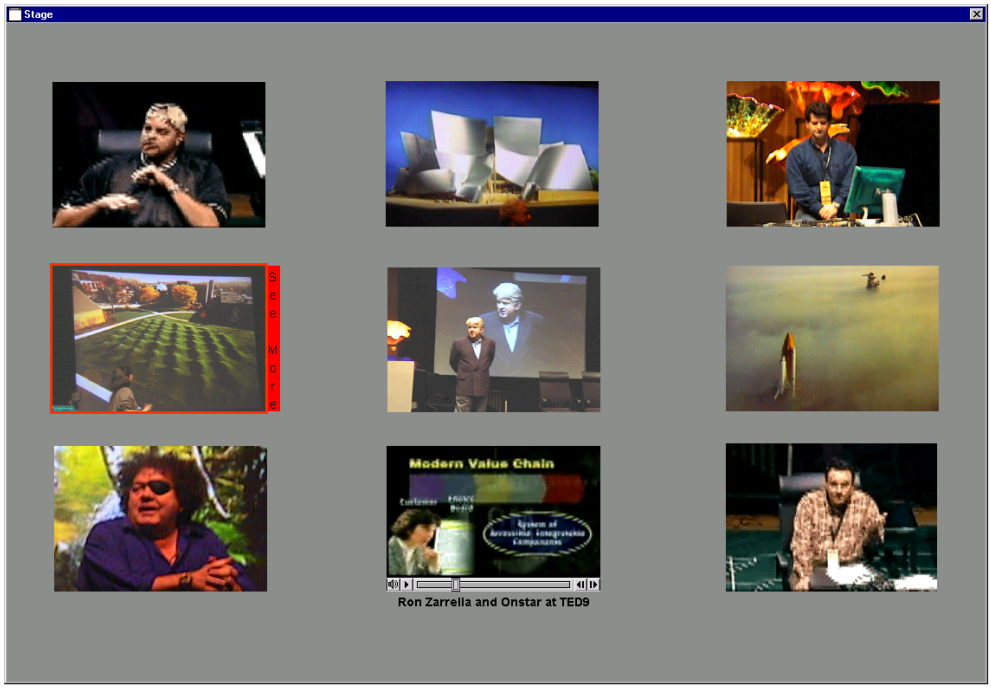

Figure 2: One story and "See More" video playing simultaneously

6 Future Research: The Narrative Question

It should be noted that this infrastructure was not intended to support a catalog for a large amount of data. There were many more speakers at the TED conference than were represented in this system; but the document was meant to be Baker's trip report, rather than an exhaustive description of the entire conference. The result thus reflects those events that Baker experienced that left the greatest impression. Whether or not the same infrastructure could support that more exhaustive description remains to be seen, but we feel there are possibilities. Our focus was on developing a reading environment that provides closure for both the author and the reader; and, as the subject matter gets more extensive, additional interface devices may be necessary to provide that sense of closure.

What would be interesting to see with this structure is how a short story would unfold in this format. Since this document served to describe the events at the TED9 conference, it did not address some of the issues of plot development that

would occur if the content were a comic book drama, short story, or movie. In our next project we would like to apply the infrastructure we have developed to the rendering of a narrative, where the reader discovers the unfolding of the story through the rhythm dimension of the display. In addition to continuing to explore the communicative power of rhythm, narrative content would provide us with an opportunity to experiment with alternative layouts, moving away from the regularity of a three-by-three grid towards the more sophisticated layout techniques found in comic book renderings of narrative [5.]. We feel this would be an interesting alternative to those interactive narratives based on a game framework in which the reader becomes one of the story's characters. We are more interested in how a reader may come to *discover* the story element of a narrative than in providing the reader with the opportunity to *participate* in that story.

7 Acknowledgements

We wish to thank James Baker for providing the motivation, source material, and encouragement without which this project could not have been completed. We also wish to thank John Doherty for his assistance in preparing the source material for the **TED9 Trip Report**. Finally, we must thank Richard Saul Wurman for conceiving an event that has proven to be so resistant to conventional documentation.

References

1. Even, T., CityQuilt. In *Proc. ACM Multimedia '95*, ACM, 1995, pp. 279–280.
2. Faraday, P., and A. Sutcliffe, Multimedia: Design for the "Moment". *Proc. ACM Multimedia '97*, ACM, 1997, pp. 183–192.
3. Jones, R. C., D. Dementhon, and D. S. Doermann, Building Mosaics from Video Using MPEG Motion Vectors. In *Proc. ACM Multimedia '99*, ACM, 1989, pp. 29–32.
4. Kelly, P. H., *et al.*, An Architecture for Multiple Perspective Interactive Video. In *Proc. ACM Multimedia '95*, ACM, 1995, pp. 201–212.
5. McCloud, S. *Understanding Comics*, First Harper-Perennial, New York, NY (1994).
6. Merialdo, B., *et al.*, Automatic Construction of Personalized TV News Programs. In *Proc. ACM Multimedia '99*, ACM, 1999, pp. 323–331.
7. Sawhney, N., D. Balcom, and I. Smith, HyperCafe: Narrative and Aesthetic Properties of Hypervideo. In *Proc. Hypertext '96*, ACM, 1996.
8. Schneider T., "Interactive Digital Narratives," MFA Thesis, San Francisco State University, San Francisco, CA (1998).

9. Shimizu T., S. W. Smoliar, and J. Boreczky, AESOP: An Outline-Oriented Authoring System. In *Proc. Thirty-First Annual Hawaii International Conference on System Sciences*, IEEE Computer Society, 1998, Vol. II, pp. 207–215.
10. Stockhausen, K., "Structure and Experiential Time," *Die Reihe*, 2 64–74 (1958).
11. Wilcox, L., *et al.*, MBase: Indexing, Browsing & Playback of Media at FXPAL. In *Proc. ACM Multimedia '99*, ACM, 1999, p. 204.
12. Zhang H. J., *et al.*, Video Parsing, Retrieval and Browsing: An Integrated and Content-Based Solution. In *Proc. ACM Multimedia '95*, ACM, 1995, pp. 15-24.
13. Zellweger P., Scripted Documents: A Hypermedia Path Mechanism. In *Proc. Hypertext '89*, ACM, 1989, pp. 1–14.