Interface Issues for Interactive Multimedia Documents

Robert B. Allen
Bellcore
MRE 2A367
445 South Street
Morristown, NJ 07960 USA
rba@bellcore.com

Abstract

While extensive standards have been developed for the representation of objects within complex multimedia documents, much less attention has been paid to the usability of these documents. Key features of electronic text-document browsers are identified and the ways these might be applied to multimedia documents are discussed. Some of these features include Tables of Contents (TOCs), linking, and indexing. Several novel interfaces for digitized multimedia lectures demonstrate how TOCs can provide access to high-level multimedia structures.

1 INTRODUCTION

1.1 Multimedia Documents

Documents may be distinguished from collections of loosely-linked information objects in having a high-level organization. A logical high-level document structure offers many cognitive advantages for the reader as compared to presentation of disjoint information objects. Ideas organized by an author's understanding may be easy for a reader to comprehend. In addition, the author's conceptual structure may be acquired by the reader and may provide a framework for that reader to organize additional information.

Recorded lectures are particularly good examples of multimedia documents because they are lengthy, coherent, and may incorporate a wide variety of media. Moreover, lectures may be enhanced with other materials such as student notes, simulations, associated texts, and problem sets. Indeed, the lecture itself might be enhanced, for instance, by providing multilingual versions. Because these enhancements provide many advantages over live or even videotaped presentations, online multimedia lectures and textbooks may greatly improve the quality of education.

Other examples of complex multimedia documents are radio and television news programs and documentaries. For instance, an electronic newspaper could contain a wide variety of material such as images, video, maps, and historical texts. While there has been much discussion of personalized news media, (e.g., [1]) a strong case can be made that it is too difficult to anticipate users' interests and that a consistent document structure (like a traditional newspaper) and powerful document browsing tools will be more effective in giving the user access to relevant material.

1.2 Multimedia Browsers

Current multimedia viewers take little advantage of high-level structure for presenting information objects. Many of these systems simply display the objects retrieved from a database. Even among those systems that present connected material, there is little structuring by semantic content. For instance, one common interface style for presenting the content of a video simply tiles a screen with frames. A second common type of interface uses variants of the video streamer [8] that presents overlapping edges of video frames. Only minimal semantics of the content are used in organizing the material. Timelines (e.g., [4, 5]) have also been widely employed to allow users to browse multimedia presentations but, again, they generally incorporate only minimal semantics.

Streams [4] is a multimedia viewer which presents thumbnail views of several concurrent media channels and lets the user focus on one of them. Thus, a lecture recorded with Streams might include a video of the lecturer, a video of the audience, and a channel showing the lecture's view-graphs. The implementation of Streams provides timelines and a video streamer for indicating position. However, Streams was designed so that no human analysis of the video was necessary; thus, there is no semantic searching capability, no high-level semantic organization such as a table of contents (TOC), and no linking across time between objects.

1.3 Internal Representations

In much of the previous work on multimedia documents, the emphasis has been on the representation of the structure of the documents rather than on the accessibility of the information. Hypertext interchange models such as Dexter [10] deal primarily with simple information objects. HyTime [13], an extension of the Standard Generalized Markup Language (SGML) for linking hypermedia objects, and MHEG [14] allow very general representations of multimedia, but they do not prescribe how these objects may best be arranged for user access. While internal document representation is important in providing efficiency, the emphasis in this work is on the design of usable interfaces.

2 PRESENTATION ISSUES FOR MULTIMEDIA DOCUMENTS

Capabilities such as linking, searching, and TOCs have proven important for presentation of text documents. For instance, the SuperBook^{TM} document browser [7] uses these features to allow readers to move around the high-level structure. However, the SuperBook browser does not manage multimedia as effectively as it manages text. When it launches a multimedia diversion, the user cannot return from the multimedia diversions to other parts of the text nor does it synchronize several concurrent channels of multimedia events as Streams does.

2.1 Tables of Contents (TOCs)

A TOC is a composite set of links that provides pointers into a corpus. Unlike the timeline, which provides only limited linking, the TOC has a rich semantics. In addition to helping a user navigate a document, the TOC provides context for understanding the structure of a document. The TOC is probably especially important if there is no semantic searching capability. While fisheye TOCs are fairly common for text-based documents, they have not previously been applied to image and video collections.

As illustrated by the interfaces developed in Section 3, the TOC can be applied to multimedia documents. Indeed, as shown below (in Figure 4), the TOC itself need not be purely text. Even, an audio menu could be considered to be a TOC for an all-audio document. User-placed book-marks can also be organized as a type of TOC. Not only do these allow the user to note and later access sections of a document, but a collection of them can also provide a personalized high-level view of the structure of a document.

2.2 Linking

One interface principle for links is that they should indicate as clearly as possible where they lead. A second principle is that there should be an easy way to undo the link transition. For linking in multimedia documents, there are two additional issues: Linking across spatial and temporal extent. First, for complex graphical material, such as the view-graphs, it is difficult to determine what aspect of the graphic is the link anchor. This may be resolved by hot-spot graphics in which moving the mouse in different parts of the graphic allows relevant links to be highlighted. However, some links may be anchored by several different points on a single graphic. Second, for temporal media (e.g., audio and video) links may be dynamic and may appear and disappear as the material is presented. Moreover, these solutions are not general because a user may wish to get a broader view of the links such as those links that are active in surrounding material when viewing a presentation.

When linking between complex multimedia documents, it may be necessary to provide some of the context for the link anchors. This might be done with a TOC-like link anchor (it is also similar to the graphical link viewer of [9]).

There are similar link problems for accessing user annotations when they are linked to a document. However, because annotations may themselves be multimedia documents there also involve the larger problems for managing links between two multimedia documents.

2.3 Searching

When a user wants to explore a new topic for which a TOC or links are not readily available, a search may be used to identify relevant documents or parts of a document. A few interfaces allow database-style searching across collections of multimedia materials [11, 12]; however, in almost all cases, the queries are stated textually rather than in some other medium. Of course, if available, the full-text transcript of a lecture would greatly improve search quality. Relartively little is known about handling non-textual queries, but it is possible to imagine queries to the audio track or to the visual material. Indeed, complex queries involving several media would be possible, although quantitative comparisons of the similarity of search hits across media would not be trivial. Searching across multiple media introduces difficulties for rank ordering search hits because weights for the contributions of different modalities are unknown.

A useful feature of the SuperBook browser is the display of search hits on the dynamic TOC. In a multimedia document interface, there are a variety of possibilities for how to post search hits on multiple TOCs and how to coordinate browsing of search hits across several media. Yet another challenging problem is how the best "focus of attention" (see Section 2.5) might be selected for presenting the results of a search. For instance, if search hits were clustered primarily in a video, it would make sense to highlight that material.

2.4 Supporting Partially Guided Tours

Guided tours have been widely discussed (e.g., [16, 18]) for controlling the presentation of collections of hypertext objects. Like guided tours, computer based training (CBT) [17] imposes a structure on a user for browsing an information space, although transitions may be a complex function of previous responses.

Rather than strictly guided tours, the approach proposed here may be termed *partially guided tours.* Instead of being read straight through, a textbook is probably more frequently accessed as a reference document in a semi-random order. Replaying a part of a lecture rather than the whole thing, is a simple example of a partially-guided tour. To make multimedia documents effective electronic textbooks, it seems necessary to find ways to support these partially guided tours. Thus, if a user is following a course outline and decides to interrupt the course to browse other material, it should be possible for the user to resume the course at the point he/she interrupted it. In a related vein, the document should support multiple levels of tours. for instance, there could be a quick overview and other, successively more complex views.

2.5 Locking Concurrent Multimedia Streams

Multimedia document interface management gets more complex when a user is able to detach one media stream from other streams. For instance, a user might want to browse view-graphs while listening to the audio track of the lecture. Indeed, there is ambiguity even for the relatively simple interfaces described earlier in this chapter about the interactions of the visual and auditory media. A similar problem occurs if the media streams are related, but do not cover identical material. For instance, a lecture may not exactly follow a textbook. The notion of detaching media suggests that the streams are not "locked" together. An issue is whether TOCs should automatically expand if the audio is in auto-play mode.

When a document interface allows the user to detach and browse different streams, the media may be said to be unlocked. In addition, the diversions in the SuperBook text browser are essentially 'unlocked' from the text. A view-graph launched from the SuperBook browser remains open until the user closes it. This is fairly straightforward for single images, but it may be much more confusing when there are multiple diversions in complex documents. The CD-ROM encyclopedia *Encarta* has a simple locking mechanism; it allows images to be either locked or unlocked from the text.

A user who is following a guided tour and may want to browse other material or to explore linked documents. The simplest solution would be to be suspend the guided tour while the user browses other material and then have the tour reinstated when the user finishes browsing. Indeed, the suspension point could be much like a bookmark and if there were several of them, they could be stored in a list. This bookmark style would be the easiest to implement. However, it does not handle complex, but probably common, cases in which the user wants to do concurrent actions such as browsing view-graphs while listening to an on-going audio presentation.

If media in a document are not locked together, or if several loosely-coupled documents are being browsed, then the user needs to be aware of which streams are being coordinated in actions such as searching. A "focus of attention" could indicate which of the several groups of media is active. It is analogous to the focus window for a window manager. Perhaps, spatial grouping or the coloration of widgets could indicate which streams are locked together.

3 TOC INTERFACE IMPLEMENTATIONS

Four interface implementations, programmed with Motif widgets under the X Window System, are described below.

3.1 Corpora

Two related sets of multimedia materials were obtained. 98 view-graphs for a Bellcore Training and Education Center (TEC) course on Asynchronous Transfer Mode (ATM) and were arranged into modules that formed a simple hierarchy.

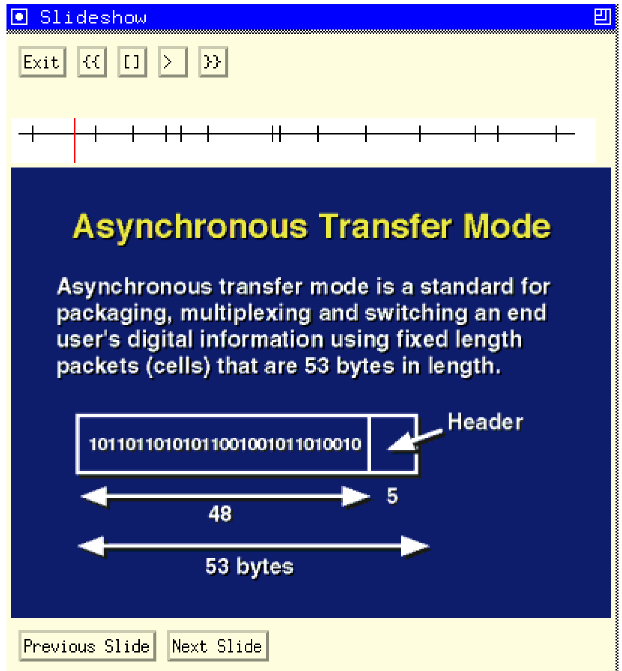

Figure 1 Timeline Slideshow Interface for View-Graphs and Audio.

In addition, a one-hour videotape of a lecture on ATM was MPEG1 encoded and stored. This video was the first part of a three-hour course that roughly paralleled the view-graph course. The images used in the course video that were not available as view-graphs were digitized from the screen images. The audio track of a lecture was μlaw encoded and stored separately. The times at which the view-graphs appeared in the audio (and hence also in the video) presentations were noted and stored.

3.2 Timeline-based Audio-Slideshow Browser

Figure 1 shows an interface for presentation of the lecture audio and view-graphs. The audio of the lecture can be played from any point with the view-graphs presented at appropriate points. Alternatively, the view-graphs can be browsed without accompanying audio. A timeline of the lecture is presented with view-graph transitions marked along it. The user can skip to the points in the lecture by clicking on the timeline. The view-graph associated with that interval is then displayed. This interface is similar to Streams [4] in that the audio and associated view-graphs are always locked together. Cassette-like audio controls are shown at the top of the widget. In this case, { means jumps back, [] means stop, > means play, and }} is jump forward. The view-graphs are controlled by the buttons at the bottom.

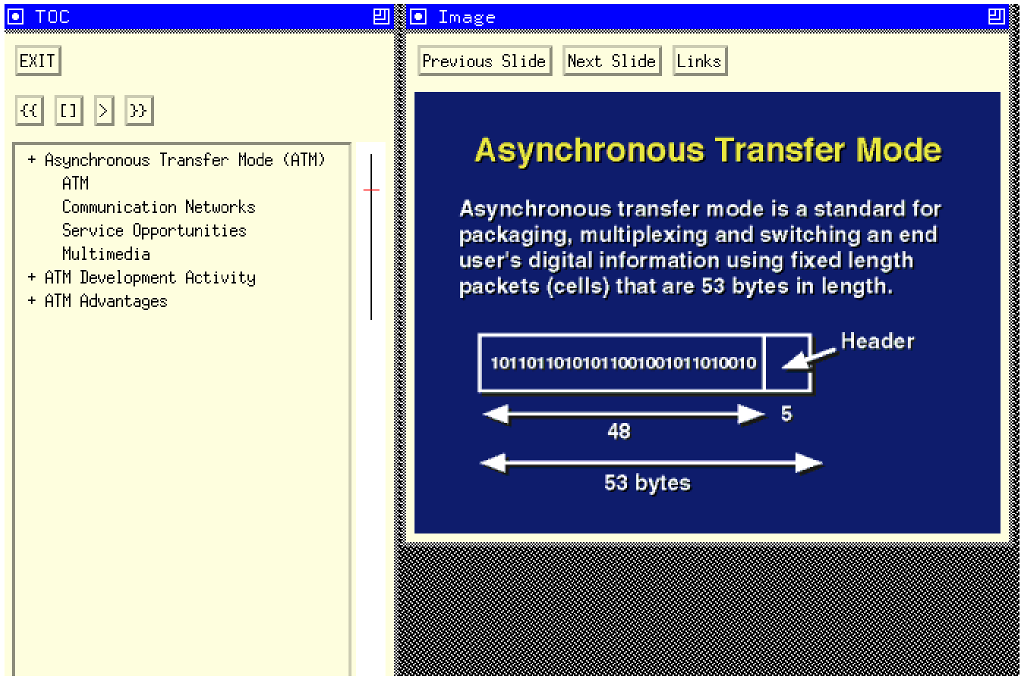

Figure 2 Dynamic TOC for Controlling Audio and View-graphs.

Several modifications to the basic interface would be possible. For instance, some of the structure of the document could be displayed by marking the timeline with different colors. In the implementation shown in the figure, the audio and view-graphs were locked so that changes to audio position updated the view-graphs and vice-versa. However, users might want to clone the view-graph display so that they could browse the view-graphs while not interrupting the ongoing audio presentation.

3.3 TOC-based View-graph and Audio Browser

The timeline in the audio/view-graph interface presented minimal semantic information about the content of the view-graphs. A text TOC may have semantic labels that can help the user navigate the lecture. Figure 2 shows a hierarchical TOC for controlling audio and view-graphs. The text part of the TOC widget resembles the TOC used in the SuperBook document browser. Entries that have subnodes and can be expanded are marked with a "+". A single click on a TOC node opens the subnodes and displays the associated view-graph. A second click closes the subnodes.

In addition to the audio being coordinated with the view-graphs (and hence the TOC), the audio can also be controlled with the modified VCR-buttons at the top of the TOC. To the right of the TOC, a hash mark indicates the position

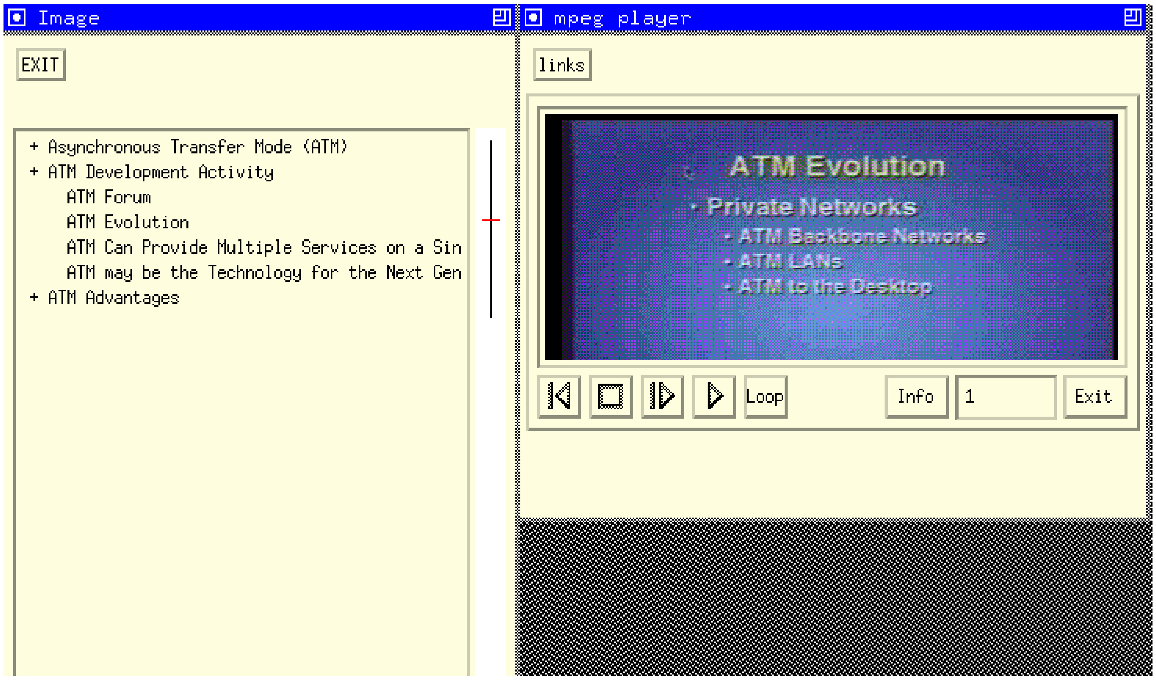

Figure 3 Dynamic TOC for Controlling Video.

of the audio being played. It may be noted that is not an interval representation of the positions of the view-graphs and that its length changes as the TOC changes.

3.4 TOC-based Video Browser

As shown in Figure 3, a hierarchical TOC widget, similar to the one used in the previous interface, was developed to control a video player. The video player was adapted from the freeware UCB MPEG1 player. This interface could subsume the functionality of the view-graph TOC browser by presenting still video frames of the view-graphs. However, some of the video view-graphs were not static and could not easily be displayed in a system that presented still frames.

Video allows advantages that audio does not. It is probably easier to search for events in fast-forward video than in sped-up audio. On the other hand, the lecturer is shown during long periods of the video. Thus, side-by-side displays of both the video and the view-graphs might be useful.

Browsing and searching could be greatly improved by effective automatic semantic processing of multimedia material. For audio, audio processing might improve search, word spotting, and speed play back. The structure of videos can also be analyzed automatically (e.g.,[6]). Sometimes, as in a entertainment video, there is not a clear hierarchical structure in a multimedia stream. Thus,

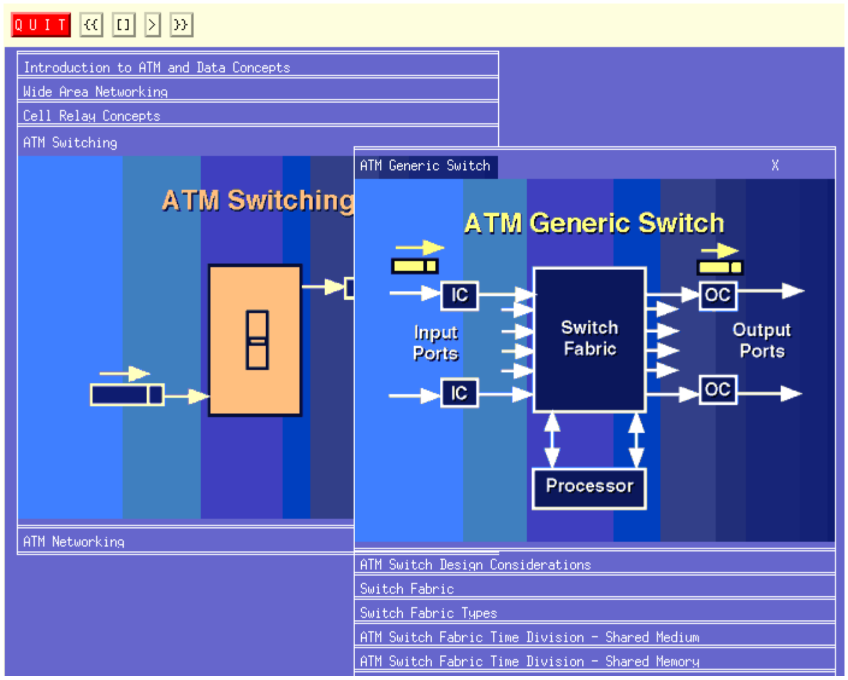

Figure 4 Visual TOC Browser.

videos may be processed automatically to extract hierarchical or other more complex structures such as grammars (e.g., [15]).

3.5 Visual TOC Browser

For materials, such as the view-graphs, which have rich semantic and visual structures, a graphical browser may supplement the purely word-based TOC used in the other interfaces. A prototype interface that merges graphics into the TOC is presented in Figure 4. View-graphs displaying top-level nodes are shown down the left side of the screen. The stack of view-graphs is controlled by a type of "notebook interface" such that clicking on one of the titles raises that view to the top of the stack and displays it. In the figure, the "ATM Switching" view-graph has been opened and the user has asked to see subnodes under it. The "ATM Generic Switch" view-graph is the top of the stack of subnode view-graphs.

4 DISCUSSION

4.1 Other Media, Other Structures, and Other Widgets

The lectures for which these interfaces were developed were relatively simple multimedia documents. Moreover, the view-graph collection used here was able to be hierarchically decomposed easily. Not all multimedia documents will be partitioned as cleanly.

This work has emphasized developing interfaces to enhance the presentation of existing lectures. Of course, these are common and their format is familiar and has a long tradition. Authoring lectures explicitly for hypermedia distribution is probably tricky; but if they were to be widely distributed, it might be worth the effort. For lectures which consist of logical modules, whether explicitly authored that way or not, paths through the course at several different levels could be developed. Alternative views of modular lectures could also be developed. Alternate hierarchies or, possibly, alternate structures such as timelines [3] might be used to organize the material.

An archive of several multimedia courses itself may be considered a large document and individual courses and lectures within those course may be considered to be subdocuments. Indeed, courses themselves may be aggregated and the whole collection may be organized hierarchically in much the way that [2] extended the hierarchical SuperBook interface to the management of libraries.

4.2 Possible Social Impact of Multimedia Lectures

With some planning, videotapes of lecturers can be delivered at almost any time and place. The lectures supplement and sometimes replace live lectures. Moreover, they allow VCR capabilities such as pausing, reviewing, and fast-forwarding. However, digitized multimedia lectures presented across a network with viewers such as those described here can provide substantial enhancements over videotapes. As discussed in this paper, even more can be done if a full range of multimedia document tools are included. Digitized lectures presented via a sophisticated video-on-demand service. For instance, they allow high-level information-access tools such as TOCs, content searches, and hypertext links. A digitized lecture can, of course, also be readily merged with other computer-based training aids such as simulations.

This technology can, potentially, have a positive impact of making exceptional lectures widely available with powerful tools. An open question is whether this will have an effect on live university lecturers. Even if they do reduce the need for some live university lectures, digitized multimedia lectures are unlikely to eliminate the need for the rich interactions among students and professors such as in small group classes or tutorial settings. It also seems likely that universities will continue to be centers for high-end computing and communication equipment which will be necessary for the most effective teaching.

4.3 Envoi

It is hoped that the interfaces described here will enhance the student's understanding of the structure of a document and enhance the ability to access information by providing a coherent framework for that information. Although there are many other aspects of multimedia documents interfaces that remain to be explored, multimedia documents seem likely to be an important framework for information delivery.

5 ACKNOWLEDGEMENTS

I thank Janet Allen, Mike Bianchi, Gil Cruz, Paul England, Selma Kaufman, Dan Ketchum, and Deborah Schmitt for helping to collect and prepare the lecture materials.

References

[1] R. B. Allen. User models: Method, theory, and practice. *International Journal of Man-Machine Studies*, 32:511–543, 1990.

[2] R. B. Allen. Two digital library interfaces that exploit hierarchical structure. In *Proceedings of Electronic Publishing and the Information Superhighway*, pages 134–141, May 1995. Boston, http://superbook.bellcore.-com/PAPERS/RBA/LIBR/libr.html.

[3] R. B. Allen. Interactive timelines as information system interfaces. In *Proceedings International Symposium on Digital Libraries*, to appear. Tsukuba, Japan, http://superbook.bellcore.com/PAPERS/RBA/TL/tl.html.

[4] G. Cruz and R. D. Hill. Capturing and playing multimedia events with STREAMS. In *Proceedings ACM Multimedia*, pages 193–200, Nov. 1994.

[5] M. Davis. Media Streams: An iconic visual language for video annotation. In *Proceedings Visual Languages*, pages 196–202, Aug. 1993.

[6] N. Dimitrova and F. Golshani. Rx for semantic video database retrieval. In *Proceedings ACM Multimedia*, pages 219–226, Nov. 1994.

[7] D. Egan, et al. Formative design and evaluation of SuperBook. *ACM Transactions on Information Systems*, 7:30–57, 1989.

[8] E.L. Elliot. Watch-grab-arrange-see: Thinking with motion images via streams and collages. Master's thesis, Visual Studies, MIT, Feb. 1993.

[9] S. Feiner. Seeing the forest for the trees: Hierarchical display of hypertext structure. In *Proceedings ACM SIGOIS: Conference on Office Information Systems (COIS)*, pages 205–212, Mar. 1988.

[10] F. Halasz and M. Schwartz. The Dexter reference model. *Communications of the ACM*, 37:30–39, 1994.

[11] T. D. C. Little, et al. A digital on-demand video service supporting content-based queries. In *Proceedings ACM Multimedia*, pages 427–436, Aug. 1993.

[12] G. Marchionini and G. Crane. Evaluating hypermedia and learning: Methods and results from the Perseus project. *ACM Transactions on Information Systems*, 12:5–34, 1994.

[13] S. R. Newcomb. The HyTime hypermedia/time-based document structuring language. *Communications of the ACM*, 34:67–83, 1991.

[14] R. Price. MHEG: An introduction to the future international standard for hypermedia object interchange. In *Proceedings ACM Multimedia*, pages 121–128, Aug. 1993.

[15] D. E. Rumelhart. Notes on a schema for stories. In D. Bobrow and A. Collins, editors, *Representation and Understanding*, pages 211–236. Academic, New York, 1975.

[16] P.D. Stotts and R. Furuta. Petri-net-based hypertext: Document structure with browsing semantics. *ACM Transactions on Information Systems*, 7:3–29, 1989.

[17] P. Suppes. Some theoretical models for mathematics learning. In *Journal of Research and Development in Education*, volume 1, pages 5–22, 1975.

[18] P. T. Zellweger. Scripted documents: A hypermedia path mechanism. In *Proceedings ACM Hypertext*, pages 1–14, Nov. 1989.