DIGITAL MULTIMEDIA:
YESTERDAY, TODAY AND TOMORROW

Glorianna Davenport,
Assistant Professor of Media Technology
Massachusetts Institute of Technology,
Cambridge, MA 02139

INTRODUCTION:

Today, the expression "Digital Multimedia" triggers a range of overlapping dreams and expectations. Cognitive biases, emerging computational techniques, prototype applications and available products stimulate a constant stream of definitions. From "Sight, sound, motion –it's as simple as that" to "Hundreds of megabytes of content –it's a fundamental paradigm shift," promoters herald a revolution in communication. [1] Minimally digital implies a computer–driven system which manages storage, retrieval and display of information across media types. In concert the content and platform should be capable of generating conversations between the user, the machine and chunks of information. Looking out past today's limited applications, with their somewhat clunky sensual and cognitive transitions, into the crystal ball of future information demands and technologies, we anticipate the realization of computational television. Responding to a range of stimuli — including voice, eye motion and gesture –this incarnation of an electronically networked all–digital media system should be able to generate automatic on–the–fly selection and compositing of information segments from multiple data sources, as well as virtual representations of objects and motions structured to mirror known physical and paraphysical behaviors.

THE DESKTOP AND BEYOND

Familiar and comfortable, the paradigm of the electronic desktop suggests access to and manipulation of information of any known datatype – text, numbers, sounds, still and motion picture segments, computer programs, transparent input and appropriate output devices. As we move into the 1990's three compelling engineering challenges will mark our progress toward digital multimedia of the future: the

development of complex document architectures and protocols which are platform–independent; the development of compression and decompression algorithms for digital video; and the generation of computational agents capable of orchestrating story structures and meanings.

Of the mentioned media types, video is currently the most recalcitrant. Capable of providing an "almost like being there" experience of real people, places, events and processes, the bandwidth of motion picture information still technically eludes the capabilities of our digital world. The issue here goes beyond any argument for prettier pictures. Both the temporal nature of motion picture/sound information and the multiple levels of content which are reflected or interpreted generate complex requirements for representation as well as for overall information management.

In terms of story generation, representation and recursive modeling are key. Ron Evans, a native American, tells a story about an African tribal chief who was given a television by a missionary. For some weeks the chief played the television every night, and his community gathered around to watch. Then the missionary left. When he returned months later the television was gone and, as in days gone by, a human storyteller provided entertainment each night. Puzzled, the missionary asked the chief why he had stopped watching television. The TV had many stories, the chief acknowledged, "...but my storyteller knows me." [2]

It is this attribute of virtuoso performance in combination with a compelling level of personalization that we seek to emulate in interactive television. The analogy is to conversation, where transfer of information is shaped by fluid conceptual links which allow for dynamic structuring of meaning. To accomplish this goal, a multimedia publication must forge a symbiotic partnership between the author, multi–layered representations of content, the overall machine architecture, and the user.

EARLY EXPERIMENTS AND PARADIGMS

The laser videodisc player was introduced in the US in 1979, one year after the introduction of the now ubiquitous VHS videotape recorder. For the first time in the history of motion pictures, relatively rapid random (frame–accurate) access to still and full motion video was possible. Except for a few extraordinarily suggestive prototypes which were developed in a handful of research laboratories, the power of random access video did not really catch hold until Apple Computer introduced HyperCard for the Macintosh in the fall of 1987. In the space of two years, multimedia/

hypermedia has become a buzz word across industries and professions whose lifeblood is information, education, training, and/or entertainment.

Much of the early research in the field of interactive video took place at MIT's Architecture Machine Group. Two quite different approaches were evident in the beginning. Both the surrogate travel/flight simulator model on which "Aspen" (1979–1981) was based and the electronic book paradigm on which "The Automatic Transmission Manual" (1982– 1984) [3] was conceptualized drew from the world of games and cinema; both assumed that the viewer would enter into the project framework via familiar imperatives of the form. In contrast, both "Archfile" and the later "Picassofile" were based on an information science approach. The videodisc provided storage for up to 54,000 still frames. The videodisc player became an image server when hooked up to a relational database via a device driver. The tacit assumption of such an information–rich environment is that the user will chase information about which s/he knows enough to request from the system. The database used standard attributes to describe the images (place, date, architect, name etc.), facilitating retrieval of a single image or classes of images. As the project matured, so did the graphical interfaces which were designed to promote user confidence and convenience. [4] The two approaches –video as story and video as information – merged in the mid–80's as Project Athena established a campus– wide network to support undergraduate education at MIT. [5]

 Within the education context, video — both stills and motion —could introduce students to new places and cultures. Used to enhance language learning, design, politics, and economics, video brought an exciting component to many fields of study. The educational nature of the Athena applications pushed developers to integrate annotation tools into the electronic work place. Under the umbrella of *Visual Information Systems* [6], a group of faculty developers (including Patrick Purcell and the author of this paper) worked with students to build applications such as "Galatea", a networked video server enabling both local and remote broadband delivery; "Light Table", a graphical interface for slide selection and comparison; [7] and an interactive case study environment developed for "City in Transition: New Orleans, 1983–86" [8]. The programs ran video and software, and were used by students as real information delivery environments.

These early experiments gave rise to the beginnings of a theoretical understanding and identification of system attributes which allow an interactive experience to emulate discourse. Andy Lippman, director of the Movies of the Future and Television of Tomorrow projects at the MIT Media Laboratory, includes the following five primitives on his list:

granularity (the chunking of information,) interruptability, limited look–ahead, the appearance of infinitude, and graceful degradation (the ability of the system to find escape routes when the system does not have information which the user wants). [9] In addition, the New Orleans project clearly illustrated the idea that the roles of viewer and editor were merging.

WHAT SHOULD I DO AND WHERE CAN I GO? AN ART OF SIGNS AND SYMBOLS

In the early days of computing, punch–card input affected our intellectual understanding of numbers and our social understanding of mechanization. Those who thought about it thought that computing would come to the desktop; most of those people believed that everyone would want to write programs all the time. [10] Perhaps because it was before anyone really had access to digital machines, Vannevar Bush did not promote this vision in his classic article entitled "As We May Think." Rather, in this article, published in the Atlantic Monthly in 1945, Bush invented *Memex*, a powerful paradigm which explored the idea that computers would be able to augment human memory by generating and maintaining associative and personalized links between chunks of information. Bush described Memex, later acknowledged as the ancestor of Hypertext, as follows:

> Consider a future device for individual use, which is sort of a mechanized private file and library...A Memex is a device in which an individual stores all his books, records and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory.
>
> It consists of a desk, and while it can presumably be operated from a distance, it is primarily the piece of furniture at which he works. On the top are slanting translucent screens on which material can be projected for convenient reading. There is a keyboard and sets of buttons and levers... [11]

While this article was stimulated by the then current microfilm technology, the article suggests later innovations which are part of today's information universe, including the database, digital pictures and sound, automatic interpretation of both text and sound, gestural input devices and storage mediums such as the CD–ROM. Perhaps because television was in its infancy, Bush does not include any references to motion pictures or to the user actively developing computer programs.

However, he does make reference to user cues and mapping, which are considered classic problems for hyper– documents today.

Where should I go —I don't know what's out there. This is interesting —I want to know more. Uh–oh, I'm lost —I want to go back to where I was before. ..

HyperCard was the first commonly available computer application to mimic the potential of Memex. With its release the multimedia world took off. HyperCard's object–like structure, high–level scripting language and graphically familiar emulation of cards and stacks have inspired a wide range of users to generate multimedia documents. The spreadsheets, which provided one of the first important paradigms for computational manipulation of data, are finally being used as a front database display in combination with video. This interface provides a powerful tool for building business case studies. As HyperCard and spreadsheets are used to help authors orchestrate new messages using multiple media — video, text, numbers and graphic data — we begin to reap some of the fruits of Bush's early vision. While delivery platforms still limit widespread distribution of common content, and representation of content lags behind manipulation of the signs, the impact of these new literary forms is already being felt in entertainment, education, and business.

"The Elastic Charles" was created in 1988–89 to explore the concept of a multimedia magazine. Focusing on the Charles River in Cambridge, Massachusetts, the magazine combines half an hour of video with a substantial amount of text. Built on a Macintosh III HyperCard platform, this project was propelled by three design tenets: 1) to explore the concept of dynamic links for temporal media; 2) to generate a design approach for combining text and motion picture–based content such that gentle transitions can be provided as the viewer shifts between watching video segments and reading text segments; and 3) to develop the minimum set of tools so that users can author and annotate this publication or any video (on laserdisc) of their own choice. The magazine was developed in conjunction with a class of 15 students. One hundred copies of this informal media publication are currently in circulation.

The invention of "micons" (low–bandwidth, digital motion–picture icons) and their incorporation in the "Elastic Charles" interface represent a major breakthrough for multimedia navigational cues. In most cases motion picture segments are better represented in pictorial form with text than by text alone. While picture icons are good memory joggers to the individual who selected them, they are not necessarily evocative to a novice or first–time user. On the other hand, as the four–second loop of motion displayed in each "micon" attests, the rendition of motion relative

to location and/or gesture can be compressed to a low band–width and still communicate generic aspects of the story to most users.

"Elastic Charles" is configured using two screens, and the micons are frequently superimposed on the video image. Displayed with or without a title bar, and appearing and disappearing in relation to the video content, these mini–movies cue the viewer to interact while offering some insight into the content of the represented link. A text description of the link segment can be accessed prior to selecting the video. This preview feature is designed to ease viewer frustration and build viewer confidence. Text annotations appear on a segment card and include key words, so the viewer can browse for similar segments if s/he desires a more in–depth or alternative experience. [12]

While micons and dynamic links are striking, they can occasionally prove perplexing to a first–time viewer, who is unprepared for chunky transitions and has no prior knowledge of how this complex information platform is structured. On the other hand, the informational throughput gained from the combined use of text, video and graphics is enthusiastically embraced by most subscribers who have a particular interest in the Charles River, the history of rivers, and/or general urban issue, regardless of their technological expertise.

WHERE HAVE I BEEN? MAPS AND OTHER TALISMANS

A central task of multimedia makers involves inventing appropriate vocabularies which invite users to play while minimizing their anxiety level. As multimedia projects gain conceptual sophistication, new interfaces will integrate static story elements with live data streams.

Geographical information has attracted a range of multimedia experiments (Aspen, the New Orleans Project, the MIT language learning projects, and the Elastic Charles). The attraction exists in part because spatial orientation and/or travel can serve both as a story spine and as viewer reference. Today, extremely sophisticated GIS data sets are available and provide exact referencing for many navigational tools. This year we will be able to buy a CD–ROM which contains all maps from Boston to Washington; next year, some cars will offer a CD–ROM reader option.

At the Media Lab, a program entitled "Back Seat Driver" [13] was implemented to study language cues for understanding directions. The program is twofold. Anyone can call a given number and ask the program for directions from where they are in Boston to anywhere else in Boston; the machine responds with a set of directions. "Back Seat Driver" will

allow the user to hook into the system from a car; the car is tracked by the machine and the synthetically generated directions are give to the driver as the car approaches significant decision points.

Given the reality of the CD–ROM publication, and the availability of a very portable satellite navigation system which could be carried by a pedestrian, we are planning to add mini–movies –perhaps in the form of micons with sound bites –to the digital map data. The challenge here is to keep the segments short but very descriptive. Segments representing points of interest can be mixed with a motion video yellow pages filled with vendor advertisements by location. The project would build up a knowledge base about the urban history and community as well as about travelers' needs. In a sample situation, you are a tourist and want to spend this morning viewing historic sites. The machine would construct a video preview of its proposed tour from Charlestown to downtown Boston. The system calculates that, given your style of investigative observation, the tour in question would take 3 hours. You want to spend more time in the downtown area. The system modifies the route. Or, as another example, you have an hour between your early morning meeting and your luncheon appointment –time enough to get your hair cut and look for that special gift. Informed of this fact, the system then offers you a selection of video ads to browse through before finalizing your itinerary.

The importance of speech as an input and output device in this situation can be clearly appreciated. Indeed, with a generation of low–cost voice recognition systems, voice as input should find its way into many more multimedia applications.

LOOK–AHEAD AGENTS OF THE FUTURE – MY STORYTELLER KNOWS ME

While the movie map project deals with some issues of limited look–ahead and personalization, it still represents a fairly closed system, with the main live data feed being a person's location at a given point in time. On a different tack, both news and business applications direct our attention toward issues of representation and story modeling in order to utilize the power of live data feeds.

As we have seen in previous examples, the impact of digital manipulation using multiple data types has brought the roles of viewer and editor closer together. While the idea that each of us becomes our own storyteller can be scary or delightful, it has some profound implications both in relation to time and in relation to meaning/understanding.

Exploration and choice require time; the leisure of not making a rapid judgment is critical if we are to use browsing and associative linking to construct intelligent frameworks for everyday actions.

While we may find the idea of our machine as our storyteller intriguing, it is also unimaginable to most of us. How often do we search for the right section of a book, or debate which is the best route to take, or contemplate how to say something which will have meaning to someone else? How can some set of electrical impulses structure a meaningful story? The future of our information–rich age may offer no alternative. Given the reality of information overload and the power of computation, our task becomes not just how to construct representations and story agents, but also under what circumstances we will trust our story agent to interpret intent, make associative links, and generate representations which promote hands–on and heads–in learning. [14]

News offers an obvious platform for the development of story agents. Already ABC News Interactive is preparing and distributing videodiscs in which, while they edit the source material, they also assume students will make their own clip lists and "integrate excerpts in research and presentations. The availability of BISON will allow television news archives to open their doors to automated retrieval systems. As computer television comes on line, the networks may be able to generate significant income by selling video streams which are computationally assembled based on some combination of content representation, story model, and actively shifting viewer profile. Several projects at the Media Lab are exploring the area of machine–selected news. [15]

As we begin to have more than incidental success in structuring stories on the fly, many new applications in business, industry and medicine are likely to emerge. In closing, I do not want to sell multimedia short. I believe that many complex interactive narrative experiences await the maturing of both technology and the culture itself. Clearly, this medium has incredible potential for entertainment as well as learning —but that is another chapter.

[1] From video interviews with Phil Schiller, Apple Computer and Robb Glazer, Microsoft Inc., in Jonathan Harber, "The Emerging Multimedia PC industry: A HyperCASE," Masters thesis, videodisc portion, Massachusetts Institute of Technology, 1990.

[2] Ron Evans at The 10th Annual Convention for the Preservation and Perpetuation of Storytelling, reproduced in Steve Kostant, "By Word of Mouth: Storytelling in America," videotape, 1985.

[3] The Automatic Transmission Manual, principal investigator Andrew Lippman; David Backer, "Structures and Interactivity of Media: A Prototype for the Electronic Book," Ph.D. dissertation, Massachusetts Institute of Technology, 1988.

[4] The principal investigator for Archfile and Picassofile was Patrick Purcell, Visiting Associate Professor of Computer Graphics, MIT 1981–1990 and Director of Communications and Sponsor Relations at the Media Lab, 1988–1990.

[5] Project Athena was formed in 1983 under a grant from Digital Equipment Corporation and IBM to establish a campus–wide computer network for undergraduate education at MIT. Project Athena supported development of software such as X windows in order to generate coherence for the multivendor platform.

[6] Project Athena awarded grants to faculty for courseware development. In 1986, Patrick Purcell, Glorianna Davenport, Merrill Smith and Frank Miller formed the Visual Information Systems group to continue development of projects begun under the umbrella of the Department of Architecture's Computer Resource Lab.

[7] Patrick Purcell and Dan Applebaum, "Light Table: An Interface to Visual Information Systems," in Electronic Design Studio, MIT Press, 1990.

[8] Glorianna Davenport, "New Orleans in Transition, 1983–1986: The Interactive Delivery of a Cinematic Case Study I" ICDPT, August, 1987.

[9] "Lippman on Interactivity," MacUser, March, 1989.

[10] Robert Wolford, Ameritech Development Corp., in conversation.

[11] Vanavaar Bush, "As We May Think," The Atlantic Monthly, 1945.

[12] Brondmo, Davenport," Creating and Viewing the Elastic Charles," presented at Hypertext2 Conference, York, England, 1989

[13] Jim Davis, "Back Seat Driver: Voice–Assisted Automobile Navigation," Ph.D. dissertation. M IT, September 1989.

[14] Virginia Doland, "Hypermedia as an Interpretive Act," Hypermedia , Graham, England, Vol. 1, no.1, 1989.

[15] Walter Bender and Pascal Chesnais, "Network Plus", for SPSE Electronic Imaging Devices and Systems Symposium, Los Angeles, January

1988; Lee Morgenroth, "ACE: Automated Content Editor", Term Project for Davenport/4.985, May 1990.