Background paper for NCF 90: July 1990

2000 & BEYOND: INTELLIGENT ACCESS DEVICES FOR MULTIMEDIA IN TOMORROW'S TELECOMMUNICATION NETWORKS

Glorianna Davenport
Assistant Professor of Media Technology
Director, Interactive Cinema Group
The Media Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139

ABSTRACT:

A new generation of interactive multimedia applications will emerge as an all-digital video signal becomes accessible via networked telecommunications channels. The issue of how we interact with newly available resources should depend on the nature of the intended experience or structured task. Over the past 10 years, some interactive multimedia projects have focused on the relation between task and input device, while others have extended the language of representation to include graphical cues. These prototype projects have established the desktop and conversation as paradigms for interactivity. This paper uses a thought experiment to explore devices which can extend these paradigms to encompass the complexity of tasks and operations which will define networked interactions for digital media.

INTRODUCTION:

Today the range of overlapping dreams, predictable needs and expectations for "digital interactive multimedia" are expanding rapidly. Specific tasks which must be supported at a network level include directed information retrieval, creative multi-person exchanges, and document construction. Some applications will incorporate links to be accessed on an informal basis (by user selection of a browsing mode or incidental query, for example). In other applications, the interaction itself will be essential to the meaning of the experience. Finally, constructivism, or user manipulation of content, must be viewed as a dominant mode of interaction which overlaps both of the others.

Early experiments in interactive multimedia suggest stylistic dichotomies between the execution of purposeful assignments and casual browsing or cultural gaming. We already have 10 years of experience with LANs; however, we have almost no experience with networked applications for temporal media, such as multimedia archives. Although collaborative authoring programs have been high on the list of "to be solved" problems, we have also had little experience participating in or observing truly collaborative design

or research efforts. Perhaps this reflects the lack, until now, of user- friendly software; or perhaps, as some experts have observed, collaboration is just much more difficult and much rarer than we have heretofore imagined. Some networked information users, particularly in the financial world, have experience in executing individual transactions against a variety of databases; only very recently has the idea of electronic trading become sufficiently real to inspire the design of smart automatic systems which can make multiple transactions against multiple databases without requiring human intervention at each step.

WHAT'S NEW? COMPUTER TELEVISION:

In the year 2000, digital video will be a reality. A digital video signal which is also scaleable means that a fiber-optic communication network will be able to carry lower or higher bandwidth motion-picture signals, at the discretion of the application. Higher bandwidth signals can be displayed at the appropriate resolution and frame rate as determined by the receiver. Digital video goes hand in hand with the idea of computer television - that is, a television which can digitally store information and which can respond to computer programs. The combination of a digital video signal, a method of transmitting that signal from point to point, a means of storing and displaying that signal, and the ability to interact with that signal will make interactive multimedia in the year 2000 very different from any prototypes we have developed to date.

Let us begin with the idea of a teleconference session for automobile being designed for road rally racing. One participant wishes to review international regulations for safety while another participant wishes to review typical use patterns as well as road specifications and construction techniques used in a number of foreign countries. Some of this information is reported regularly via the networks like CNN; other sources of information might include national reports, manufacturer reports, videos produced by tourist agencies and transportation departments, and manufacturer advertising. Archival companies are in business to supply video on demand; you pay based on search time, footage transmitted, and transmission time. The access of a consumer to an archive poses some interesting problems: How do we know what is there? How and at what point in the network is available information shaped into an understandable story relative to the specific request? How are video and computer graphics integrated into the design document and how are the costs of video product billed?

Perhaps more important, given such a scenario, what interfaces shall we use for what tasks? Appropriate input devices are ones which help stimulate interaction while encouraging the user to focus on the experience or task at hand. Despite disclaimers by cantankerous critics, the medium of computational television has already been shaped to some extent by a handful of early interactive video applications. Looking past these early applications, with their somewhat clunky sensual and cognitive transitions, into the crystal ball of future information demands and technologies, we anticipate the evolution of a multidimensional medium which is capable of generating automatic, on-the-fly selection of information segments from multiple data sources in response to a range of stimuli, including voice, eye motion and gesture.

THE DESKTOP AND CONVERSATIONAL INTERACTION:

The two most important paradigms for interaction which have evolved over the past decade are the desktop and conversational interaction. The 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. Conversational interaction implies that an invitation to interact is matched to a specific task or to a general motivation to explore and that the method of interaction is compatible with the task. In addition to the keyboard, reliable input devices include the mouse and the touch screen, while voice, gesture (data glove, eye tracking,) force-feedback joysticks, track balls, etc. continue to be explored in research labs.

MATCHING USER INPUT TO TASK GOALS AND INTERACTION:

The key to conversational interaction resides in the design of appropriate cues for interaction and selecting user input which is appropriate to the nature of the task. In this context, the idea of conversation must be stretched to include a range of gestures as well as voice paired with visual and verbal cues as necessary.

In order to fully appreciate this issue of matching input device with specific user-task, it is worth reviewing a few early projects which expanded our understanding of the interactive experience. "Aspen" (the first interactive video disc project) allowed a user, who was sitting quite close to the display screen, to control the simulated travel experience by touching icons on the screen which indicated forward, left, right, reverse, stop. In contrast, another incarnation of this project invited the user to sit in a chair 10 feet away from a large 10’ diagonal rear projection of the video. In this case, the user controlled their travel experience via a joy stick embedded on the arm of the chair. This example reminds us that the distance of the user from the display is a critical factor in determining appropriate input device and style for interaction.

Quite a different method of interaction was selected in "Put that there." The goal of "Put that there" was to allow generals to move battleships around on a large rear projected electronic map. The project assumed that the task should be accomplished from a distance of about ten feet from the screen with a combination of voice and gesture. The program was trained to recognize both a limited vocabulary and the phrasing characteristics of a command. The battleship object or geographic area was targeted by the user’s finger using a Polemus. "Put that" was synchronized by the program to the object identified with a pointing gesture as the object to be moved. "There" indicated the resultant location which had already been specified in geographic language as "south of Cuba." Today the database for a similar application could well reside miles away from the display.

Another early research area at the Architecture Machine Group focused on the teleconferencing problem of face to face remote. 20 video was considered too banal and unrealistic; instead the "Transmission of Presence" required that mask impressions be created for persons likely to be "telepresencing." Transparent masks were made of a particular person's head and mounted such that they could swivel in three-dimensional space. Positioning information was acquired, again using a Polhemus, from this particular person who was now in quite a different location from the mask itself; this information was used to manipulate the mask. Video of the persons face could be projected into the mask, which added significant detail to the illusion of presence. More recently the data-glove has appeared on our platter of potential input devices. This device extends gestural input to include physical actions and may ultimately affect a wide range of computer-aided training simulations, as well as factory floors, space research, and computer games. Finally, although the research is difficult and results to date are less impressive, we must not neglect the appropriateness of eye-tracking for certain kinds of interaction such as page turning, image scaling, and story selection/modeling.

TEMPORAL MEDIA, DYNAMIC LINKS AND STORY AGENTS:

For dynamic media, such as video and sound, interactive developments have focused less on broadening the range interactive devices than they have on formulating the representations for previewing content, developing icons for user navigation, and building tools for user annotation. For instance, in "The Elastic Charles," the viewer was better able to visualize the result of an interaction through the project's use of motion picture icons or "micons." Micons were used to dynamically link temporal segments, and the user was given the tools to make their own links as they wished.

Created in 1988-89 to explore the concept of a multimedia magazine, "The Elastic Charles" focuses on the Charles River in Cambridge, Massachusetts. 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 links. Such cues could be streamed with the video signal in the case of digital television. "Elastic Charles" is configured using two screens; this would be unnecessary for many applications as micons can be superimposed on the video image itself. 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 they desire a more in-depth or alternative experience. All interactions occur via mouse click.

DESKTOP 2000:

In order to appreciate how we might like to interact with our desktop, we need to look a little more closely at our Desktop in the year 2000. "The Knowledge Navigator"

videotape produced as a promotional by Apple Computer in 1988 is an admirable example of the seductive nature of the conversational paradigm. Primarily through verbal exchange, Phil - the computer - meets every verbal request of the user from finding an article for his class to calling long distance, setting up conference calls and leaving messages for specific people. A picture phone is convincingly added to the desktop and the computer collects knowledge about the voice print of those who call with some regularity.

In actuality, the search scenario is a little simplistic, compared to our automobile design session, but it suggests a device for interaction which I propose could be quite popular in the year 2000: a cellular phone with speech understanding and some form of optical pointer built in. As I promised in the introduction, this is a thought experiment. Let's see what such a device could buy.

In order for archiving to payoff, the user must be able to obtain footage which fits into the needs and expectations of a task. We need to code this footage and stream it while we are shooting the video. A phone type clip on device with speech understanding could be a comfortable addition to the cameraperson's gear. On the retrieving side it would also perform well, saving complex parsing time at the archive end of the connection. The pointing device would be used to create interactions with visual icons whether one is close or distant from the display. Including a pointing or gestural device gives us additional leverage in constructivist areas such as editing, when "cut" watching a real time display would not be precise enough.

References

Apple Computer, Inc. "Knowledge Navigator." Videotape, 1988

Backer, David. "Structures and Interactivity of Media: A Prototype for the Electronic Book," Ph.D. dissertation, Massachusetts Institute of Technology, 1988.

Bender, Walter, and Pascal Chesnais. "Network Plus", for SPSE Electronic Imaging Devices and Systems Symposium, Los Angeles, January, 1988.

Brondmo, Hans Peter, and Glorianna Davenport. "Creating and Viewing the Elastic Charles." Presented at Hypertext2 Conference, York, England, 1989.

Bush, Vanevaar. "As We May Think." The Atlantic Monthly 1945.

Davenport, Glorianna. "New Orleans in Transition, 1983-1986: The Interactive Delivery of a Cinematic Case Study." ICDPT, August 1987.

Davis, Jim. "Back Seat Driver: Voice-Assisted Automobile Navigation." Ph.D. dissertation, MIT, September 1989.

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

"Lippman on Interactivity ”' MacUser, March, 1989.

Schmandt, Chris. "Illusion in the Interface." In The Art of Human - Computer Interface Design, Brenda Laurel, ed. Pub. Reading, MA: Addison Wesley, 1990.