CHI 1990

a report on interface agents in the "Guides" project at the Apple Human Interface Group

ISSUES IN MULTIMEDIA INTERFACE DESIGN: MEDIA INTEGRATION AND INTERFACE AGENTS

Brenda Laurel

Interactivist
13800 Indian Trail Rd.
Los Gatos, CA 95030
408-741-5865
laurel.b@applelink.apple.com

Tim Oren

Apple Computer, Inc.
Advanced Technology Group
20525 Mariani Blvd., MS 76-2C
Cupertino, CA 95014
408-974-3345 oren@apple.com

Abbe Don

Multimedia Designer
3435 Clay St., #4
San Francisco, CA 94118
415-567-8130
abbedon@applelink.apple.com

ABSTRACT

A central challenge in the design of multimedia databases is integrating information from different media sources while reducing the cognitive load imposed on users by the tasks of learning and operating the interface. In light of results from a prototype multimedia project developed at Apple, we believe that an agent-style interface addresses this challenge in several ways. This paper discusses techniques for achieving media integration and details the use of interface agents in facilitating 'navigation', enhancing content through point of view, and supporting users in a variety of instrumental and experiential tasks.

KEYWORDS: multimedia interface, media integration, cross-media links, interface agents, guides, point of view, narrative

INTRODUCTION: A MULTIMEDIA INTERFACE IN PROGRESS

This paper reports a research effort at the phase between evaluation of an initial project and formulation of a new research agenda. The work began as an attempt to create an interface for a multimedia database, satisfying three goals: (1) to integrate searching and browsing activities under a single interface umbrella, (2) to provide simple, effective, and engaging means for users to access the database, and (3) to develop alternatives to spatial metaphors for the organization of information.

Key Issues

The body of this paper is organized around two issues: *media integration* and *the use of interface agents*. Early in our research it became clear that database design and interface design must be approached concurrently [1]. The initial multimedia database was intrinsically 'media-modal'; that is, it tended to segregate information on the basis of its medium of representation, and this structure was reflected by modes of the interface. We see this as a typical problem in current multimedia design. The goal of *media integration* requires a strategy for providing easy and powerful access to information of all media types.

We believe that the value of a topic-oriented database is strongly related to the completeness and variety of its informational content. The greater the variety of form, style, media, and point of view, the greater the potential for the user to explore, penetrate and understand the topic. However, the resulting complexity in the representation of a topic, as well as in the database architecture, threatens to increase the complexity of the interface as well. We have found that the use of *interface agents* reduces the cognitive load for users. Analysis of user testing results [5] suggests that guides can support the user in several additional ways, as discussed in the body of the paper.

Background

The initial project was a prototype multimedia database entitled "The Americana Series: A CD-ROM Sampler of United States History," developed by Apple Computer in conjunction with Grolier, Inc.[1] The database includes encyclopedia articles, historical documents, sounds and music, graphics, and photographic images, treating a variety of topics in the historical period from 1800 to 1850. The interface incorporates several means for accessing information, including timelines, animated maps, an article index, 'tours' (dramatic audio-visual presentations organized around central themes), and 'guides'.

Guides are representations of prototypical characters from the period that function as interface agents [7]. A guide provides navigational assistance by suggesting a next

[1]The initial prototype was developed by Tim Oren, Gitta Salomon, and Kristee Kreitman.

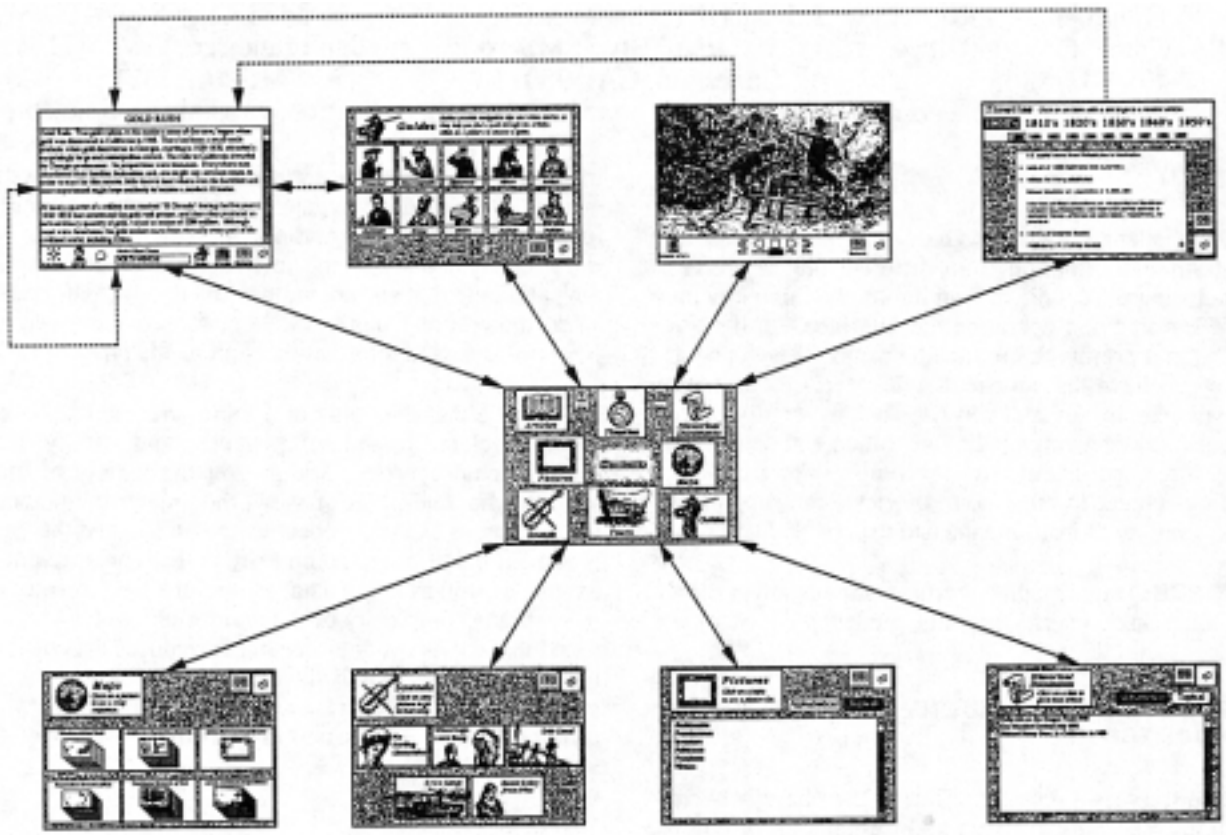

Figure 1. Links in the prototype version of the Grolier project. Arrows indicate the direction of a link. Dashed lines indicate cross-media links.

move in the database in accordance with its 'point of view'. Point of view for each guide was derived from the degree of interest that guide was assigned for various topics in the database [10]. Informal feedback on the first version of the Grolier project suggested several potential areas where the guides could be strengthened or enhanced, prompting the designers to experiment with expanding the guide's role. One guide, the Settler Woman character, was developed to embody expanded functionality in video[2] [9].

Terminology

We use the terms *instrumental* and *experiential* to distinguish two common modes of information access. The instrumental mode is characterized by a deliberate search for information to fill a particular need, for instance, to prepare a report or answer a specific question. The experiential mode is characterized by a more casual approach, often seen in activities such as browsing and exploratory learning, and may imply a more active role for the computer in suggesting interesting information to be examined.

One mode may lead to another. An experiential activity such as browsing, particularly in a database with labeled hypertext links [2] which allow the user to 'see ahead', may become a goal-directed instrumental activity. Likewise the instrumental user may digress into experientially motivated behavior. Recent work suggests that the mode in which a user is operating, and transitions between modes, may be recognized by analyzing actions at the interface [4]. Seamless integration of instrumental and experiential modes was a primary goal in the development of our initial prototype.

The sections below explore the issues of media integration and interface agents in depth. We discuss our work to date, new questions that have arisen, and hypotheses that will be used in our continuing research.

I. MEDIA INTEGRATION

Media Ghettoes

The media-modal nature of the first version of the Grolier's database is largely an artifact of the way it grew. We began with 500 articles culled from the Grolier's *Encyclopedia Americana* (International Edition, 1987) as the core of the database. Grolier's editors provided the topical index and created 'hard' links among articles. Source materials in other media, such as songs, pictures, and maps, were subsequently acquired to flesh out the database.[3]

The resulting database structure and its interface effectively created *media ghettoes* with few direct links between pieces of information of different media types. From the Contents card (the central card in Figure 1), users could enter the database by various 'portholes', but some led to dead ends while others provided the capability to move from one media type to another (see Figure 1). For example, users could move from the Contents card to the Map section to view an animated sequence of the Lewis and Clark Expedition trail; however, users could not move from that map directly to the Lewis and Clark article in the first version; instead, a user would have to go back 'up' to the contents card, 'down' to the article index, and then to the article itself. This implicit hierarchical bias — one that we believe is often present in the design of multimedia databases — may blind designers to the obvious and elegant potential of cross-media links.

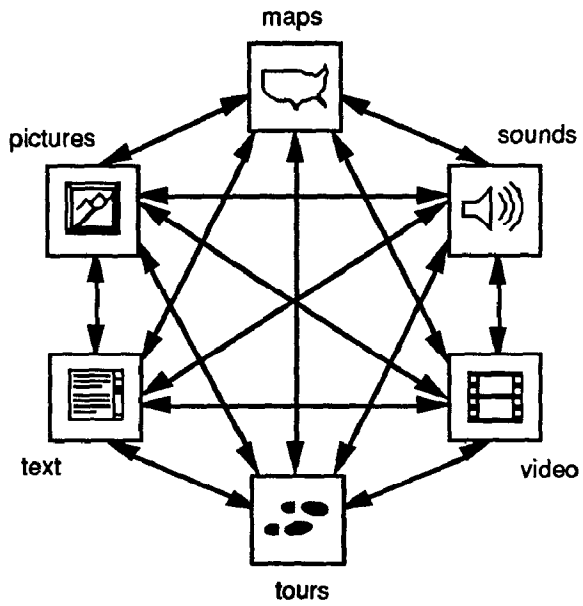

Figure 2. Ideally, two-way links connect information of all media types. Note that a 'map' of such links would probably be too complex to be useful in navigation.

Media Biases

Media segregation is reinforced by the beliefs and attitudes that both designers and users have about different media types: *media biases*. Precisely because our initial prototype was media-modal, users' media biases were thrown into high relief, bringing the issue to our attention. Just as people employ 'naive physics' in

[2]The video guide was developed primarily by Abbe Don, Tim Oren, and Kristee Kreitman.

[3]The R. R. Donnelly Company produced maps of the period and Folkways Music provided songs. Images were provided by the Bettman Archive.

understanding and getting around in the physical world, so they employ 'naive epistemology' in evaluating and using various sources of information. And like naive physics, our naive epistemologies are mixtures of fact and lore. Attitudes about video, for instance, are founded in part on the intrinsic powers of the forms that are commonly represented in the medium (such as drama and documentary). But they are also shaped by our exposure to the uses and misuses of various media in popular culture. In the user testing study conducted on the video version of the Grolier project, responses to the video were both paradoxical and illuminating. A number of users were engaged by the video but questioned its validity. Some users stated explicitly that, because the video segments looked like TV, the information contained in them was suspect — what we might call the 'docu-drama' bias. This reflects a surprising degree of information savvy on the part of our teenage users and has serious implications for multimedia designers. On the other hand, some users felt that, because the first-person video segments were clearly 're-enactments' of historical times, their production values should be higher — real horses and wagon trains in the background, for instance [5].

Video biases may have been exacerbated by the fact that the video segments appeared on a separate TV-like screen. Integrating video and computer-generated materials into a single display may reduce the credibility problem but, as users' differing perspectives on video point out, the differences between such forms as narrative storytelling and dramatic enactment are more than 'screen-deep'; they have epistemological components as well.

Many users believed that the text articles were a more reliable source of "truthful" or "academic" information than video [5]. There are obvious problems of scale when juxtaposing 500 third-person encyclopedia articles side by side with 20 first-person video segments. But a deeper problem results from cultural bias and the "psychodynamic" differences between text and oral media [8]. As Ong points out, in our print-biased culture, a statement such as "the book says" is practically equivalent to "this is true." Electronic media present us with a new set of codes of communication which are often structured liked print but rely on conventions of orality, such as voice inflection and gesture, to convey meaning [3]. Even though orality provides additional channels of information, the *validity* of that information is suspect because it does not match the *media* profile for "truthfulness" — i.e., text.

New criteria for validity of information in non-textual media cannot be directly established through multimedia design, but media integration can combat media biases by avoiding the privileging of text over other media types in the interface. Different media have different representational and informational dimensions. Media integration does not aim to obliterate the differences between media, but rather to optimize the powers of all media types by making them accessible to users with equal ease. The interface should be consistent and flexible, smoothing the transitions between media types.

Cross-Media Links

Since the interface reflects the structure of the database, sequestering information by media type makes it more difficult to accomplish inter-media mobility at the interface. Links work well for text, because users are already familiar with such existing print conventions as glossaries or footnotes. No similar conventions have been established for creating transitions between dynamic media in a computer environment. Links between dynamic media are complicated by issues of control. When moving from one text article to another, we can see the 'linked' text in its immediate context. But linking a text article to a video segment, for instance, presents a host of new problems: should the user be linked to the beginning of the segment or taken to the most relevant point? How can we contextualize a segment of video? Do conventions already exist in video to act as segment markers?

Searchability and Granularity. In order to determine where and how cross-media links can be established, the *granularity* of information must be determined or imposed. By granularity we mean the size of the 'chunks' into which information of various media types can be parsed. With text, we have familiar 'chunks' such as paragraphs and topic headings. Beyond such artificial means as the use of frame numbers, no conventions exist for establishing granularity in video segments; in the prototype version they are treated as single chunks. Because there is only one 'thing' to link to, this lack of granularity radically reduces the potential connectivity of video media in the database. It also limits the ability of users to conduct instrumental searches for chunks of information within video segments, thus reinforcing the belief that video is inherently a less contentful medium than text. Maps, tours, and pictures suffer from the same lack of granularity — and hence connectivity — in the initial prototype.

A solution that we are pursuing is to parse all dynamic-media content topically in order to establish links. Baseline contextualization in video is provided for the user who branches to an intra-video segment by starting the video at the beginning of a topical chunk. This strategy enables users with specific instrumental goals to search with much greater acuity, and it also provides more sensitive and focused links for browsing and exploration.

Interface Consistency. So far we have dealt only with the inter-media dimensions of the interface. Users need to be able to perform the equivalent of 'scrolling around' *within* dynamic media segments as well. Does each media type require a different interface solution? In the prototype database, each media type lives in a separate stack with a unique interface. The advantage is that each interface serves the exact needs of a particular medium; the downside

is a lack of consistency and increased cognitive load in learning and operating a variety of customized interfaces.

We are developing a set of universal controls that is derived from users' needs rather than media characteristics. The controls are designed to enhance users' mobility within and among media types and to facilitate topical searching. While the controls are *represented* consistently across media, their *behaviors* are calibrated to the specific characteristics of media and usage. This strategy requires not only the redesign of the interface but also the re-structuring and re-parsing of the database itself.

II. GUIDES

Problems with the Hypertext Model

A hypertext-style database, with each information item linked to others which are most relevant, seems to offer a convenient model for browsing by representing connections, but it falls prey to two difficulties. First, as the number of items in the hypertext grows, the potential connections also expand. The prospective paths from a given item quickly exceeds the optimal "seven plus or minus two" choices to present to a user at a given time. This leads to a search subproblem in which the user must determine the optimal next choice from the connection list or "embedded menu" [12].

The second hypertext problem is disorientation due to a rapid succession of jumps between items which may appear widely disparate to the naive user. To overcome this, a number of systems reify a unifying information structure for the database as a 'map', often in an explicit spatial metaphor [13]. Each information item has its context represented in the map, and it may provide a framework for reorientation. However, not only do such displays become extremely complex and difficult to understand in large databases, but viewing of the 'map' often requires manipulation of a distinct set of interface controls, presenting yet another subproblem.

Both the extended-menu-selection and map-browsing subproblems increases the user's cognitive load. Occupying cognitive 'bandwidth' with an interface task is surely limiting to users' goals, whether experiential or instrumental. Thus hypertext — and its common spatial metaphor — has serious deficiencies as an interface model.

The Guides' Role: Integration through Point of View

When information is assembled and presented to a user on the basis of her or his needs and interests, as approximated by the guides interface, that information is likely to be experienced as a series of events unfolding in *time* rather than a fixed set of articles that exists in an information *space*. A temporal metaphor such as that embodied by narrative may be a better match for the user's experience [3].

In their role as companions who suggest next moves in the database, guides link information together for the user in narrative and temporal ways. The narrative form of the Settler guide's video stories further emphasize the temporal aspects of information flow. In the initial prototype, guides provided suggestions for next moves only within the text portion of the database. It seems clear that guides can serve to achieve greater media integration at the interface by enabling them to suggest next moves of any media type. A lecturer, for example, often employs various media to make a point — maps, slides, and so on. The pivot point for the structure and representation of the information is the *speaker* rather than the media type. Guides perform a similar function as interface agents.

Likewise, a lecturer (or any other information-presenter) organizes information in terms of point of view. Even when information is delivered in the most 'objective' possible format (such as the way most newscasters try to avoid indicating how they *feel* about the stories they report), point of view is implicit in the selection and arrangement of information that is delivered. Selection and arrangement are guided by the information provider's notions of order, relatedness, causality, salience, relevance, and importance. Point of view is an important way in which a speaker, as in the example above, provides unity and continuity in the presentation of information. This notion is also applicable to the design of the guides.

Beyond its use in providing continuity to the flow of information, we have another, somewhat more idealistic, reason for emphasizing point of view. As we noted in the section on media biases, third-person text (such as encyclopedia and newspaper articles) is quite often assumed to be 'objective' and therefore *devoid* of point of view. Indeed, consumers of mass-produced information are in danger of losing touch with the notion of point of view altogether. Developing an awareness of point of view as an aspect of information, regardless of its medium of representation, is a crucial component of information literacy. Using guides to embody point of view across the whole spectrum of media, information types, tasks, and goals can help users to form new concepts and intuitions about information.

Characterization

An obvious way to create the guides' characters was to base them on actual historical personages. The problem was deciding, in the absence of complete data, what a given individual might have known about a given topic or exactly which topics she or he would have been interested in. Fabricating point of view (or its expression) for actual people seemed to smack of dishonesty. But how does one go about inventing a character from scratch?

One of the reasons why interface designers have tended to shy away from the notion of anthropomorphic agents is the assumption that the task somehow involves the modeling of a complex, human-like personality — a seemingly impossible task. The notion of *dramatic character* rather than human personality as a model for interface agents is both more appropriate and more tractable [7]. A dramatic character can be defined as a bundle of traits that are of two types: *actionable traits* — predispositions to behave in certain ways, and *semiotic traits* — the 'outward and visible signs' that allow us to infer those predispositions [6].

Characters in the theatre demonstrate how extremely minimal trait sets can successfully suggest human-like qualities. In theatre as in real life, people are quite skilled at fleshing out a character (or a person, such as a stranger met at a party) on the basis of very little information [11]. Character traits function as a kind of cognitive shorthand for allowing people to make predictions about behavior. Each successive action by a character serves to de-bug and refine our model of it and hence our ability to make accurate predictions.

The prototype features ten 'generic' characters of the period. Each guide possesses a few semiotic traits (e.g., gender, occupation, and costume) and the actionable trait of suggesting next moves in the database based on an algorithmic representation of point of view. Through user testing we were able to confirm that users imbued these simple characters with traits far beyond those actually represented; e.g., emotional qualities such as support and rejection and knowledge about the user's activities and goals [5, 9]. These results suggest several more active and pervasive roles for guides.

What is the best medium of portrayal for a guide? In the prototype, guides are represented in several ways: as graphic images with text, as iconic images, and as a character enacted on video. User testing results show that users perceived continuity of character across all representations; that is, the Settler was still the Settler whether she appeared as a character on video, a graphic image, or an icon. The persistence of character over different representations is another way in which guides can contribute to media integration.

How Guides Can Support the User

Minimizing Cognitive Load. A theme of the Grolier project has been to reduce the cognitive load on users through active behavior on the part of the computer. One way that this is accomplished in the initial prototype is by automatic generation and presentation of a restricted, ranked sublist of the connections available from any text article. The head of the list is presented as a default 'next move' option, creating the possibility of moving through the database by repetitively clicking at a single location. If the user is willing to accept this default, the entire 'navigation' problem is eliminated.

The choices presented by the guide are a ranked subset of the items linked to the current article. The relevance of each article to the guide's point of view is determined by the intersection of the article topic list and the topics which characterize the guide. Guides successfully provided a simple and effective means of representing the complex interaction of an algorithm with a large stored database — a behavior virtually inexplicable to the technically naive.

Navigation and Content. In the first version of the database, guides enabled users to answer the question, "based on where I am now, where should I go next?". User testing revealed that users wanted to know why the guide was bringing them to an article, and once they got to the article, they wanted to know if they were seeing the article from the guide's point of view [9]. The video version attempted to provide answers to these questions through the vehicle of first-person stories drawn from women's diaries of their experiences on the Overland Trail. 'Navigation tip' stories were encountered 'on the links' and 'point-of-view' stories about given articles could be accessed by clicking on a video icon.

Test users of the video version failed to note any functional distinction between the stories and the rest of the database content; the stories were perceived simply as part of the flow of information. Interestingly, however, the navigational questions voiced by users of the first version disappeared. We interpret this collapse of navigational and content frames as evidence that the temporal metaphor has the capacity to reduce cognitive load, at least in experiential and browsing activities.

Instrumental Use and Personalization. While guides were quite attractive to users on the level of engagement and other experiential qualities, they were judged deficient in the context of instrumental usage; that is, users who knew what they were looking for in specific terms often saw the guide as an obstruction rather than an aid. Part of the solution is to develop more powerful behaviors on the part of the guide. For instance, a guide might 'notice' when the browsing actions of the user coincide with its own interest profile and then volunteer itself as a provider of further related information. The system should also be able to track the user's actions by noting the topics browsed, detecting coherent behaviors, generalizing, and then volunteering further correlated items. Such a system capability would support the emergence of instrumental user goals from initial experiential tasks. The cycle is closed if a such an emergent profile may be captured as a new guide and stored for the benefit of later users.

Framing and Help. In the video version, users were introduced to the Settler guide in a segment where the actress appeared out of costume. Because she was charged with delivering information about the operation of the

system and the interface at this juncture, the guide spoke 'out of character'. We discovered through user testing that users did not even identify the guide in this segment as the same *actress*, and that they imbued her with an entirely different set of traits and functions. They expected this guide to provide help throughout their use of the program, and they perceived the guide as being an ever-present 'safety net' and companion [5]. *All of these user responses were based on a single non-interactive video segment of less than two minutes in length.* What began for the designers as a framing device aimed at distinguishing levels of activity quickly became the beginnings of a distinct type of agent: a *frame guide.*

A frame guide portrays those functions which are 'of the system', e.g., help, and those which relate to the function of content guides but transcend any one guide. For instance, the function of noticing a novel emergent browsing pattern on the part of the user and generalizing it to the creation of a new guide is properly at the 'meta-level' of the frame. Likewise, mediating the actions of the content guides as a group — "Don't volunteer until I say otherwise" — is a frame function. A frame guide might also mediate selection of a content expert guide — "I want someone who knows about gold mining," eliminating the need for users to 'interview' a number of prospective guides.

LOOKING FORWARD

We are pursuing the following goals in the next phase of our research:

- Develop strategies and techniques for better media integration, including database integration, cross-media links, and the establishment of a consistent interface for dynamic media.
- Develop the representation of point of view by experimenting with characterization and portrayal, enabling the use of multiple guides, and allowing for interaction among guides.
- Expand the functionality of guides as interface agents to include framing, active and passive help, the capacity for personalization through user modeling, and the ability of users to construct 'custom' guides.

A continuing hypothesis of the design team is that multimedia databases and interfaces such as those created for the Grolier project can replace some of the conventional notions of information with a new kind of information literacy. Continuing developments in multimedia database and interface design can reconfigure, and in turn be fueled by, the ways that people use and think about information systems.

ACKNOWLEDGEMENT

This research was supported by the Advanced Technology Group of Apple Computer, Inc.

REFERENCES

1. Akscyn, Robert, Elise Yoder and Donald McCracken. The Data Model Is the Heart of Interface Design. *Proc. CHI '88*, Washington, D.C. [May 1988], 115-120.
2. Conklin, Jeff and Michael L. Begeman. gIBIS: A Hypertext Tool for Team Design Deliberation. *Hypertext '87 Papers*, University of North Carolina, Chapel Hill [November 13-15, 1987], 247-251.
3. Don, Abbe. Narrative and the Interface. In Laurel, B., ed., *The Art of Human-Computer Interface Design*. Reading, MA: Addison-Wesley, forthcoming (spring 1990).
4. Husic, Freda. *A Preliminary Investigation of Navigating Patterns in the CHI '89 InfoBooth.* Apple Computer Technical Report, 1989.
5. Karimi, Shay, Brenda Laurel, Tim Oren, and Abbe Don. *Evaluating Guides: A User Testing Study.* Apple Computer Technical Report, 1989.
6. Laurel, Brenda. *Toward the Design of a Computer-Based Interactive Fantasy System.* Ph.D. dissertation, The Ohio State University, 1986.
7. Laurel, Brenda. Interface Agents: Metaphors with Character. In Laurel, B., ed., *The Art of Human-Computer Interface Design.* Reading, MA: Addison-Wesley, forthcoming (spring 1990).
8. Ong, Walter J. *Orality and Literacy: The Technologizing of the Word.* London: Methuen & Co. Ltd., 1982.
9. Oren, Tim, Gitta Salomon, Kristee Kreitman, and Abbe Don. Guides: Characterizing the Interface. In Laurel, B., ed., *The Art of Human-Computer Interface Design*. Reading, MA: Addison-Wesley, forthcoming (spring 1990).
10. Salomon, Gitta, Tim Oren and Kristee Kreitman. Using Guides to Explore Multimedia Databases. *Proc. 22nd Hawaii International Conference on System Sciences*, Kailua-Kona, Hawaii [January 3-6, 1989], 3-12.
11. Schwamberger, Jeffrey. *The Nature of Dramatic Character*. Ph.D. dissertation, The Ohio State University, 1980.
12. Shneiderman, Ben. User Interface Design and Evaluation for an Electronic Encyclopedia. *Proceedings of the 2nd International Conference on Human-Computer Interaction*, North-Holland, 1987.
13. Yankelovich, Nicole, et al. Intermedia: The Concept and Construction of a Seamless Information Environment. *IEEE Computer* [January 1988], 81-96.