Forms for Electronic Books

by

Steve Gano

A.B., Indiana University
1976
B.S., Purdue University
1980

Submitted in partial fulfillment
of the requirements for the
degree of

Master of Science in Visual Studies

at the

Massachusetts Institute of Technology

June 1983

Signature of author . . . .

Department of Architecture
Monday, 24 January 1983

Certified by . . . .

Professor Andrew Lippman
Thesis Supervisor

Accepted by . . . .

Professor Nicholas Negroponte
Chairman, Departmental Committee for Graduate Students

Forms for Electronic Books

by
Steve Gano

Submitted to the Department of Architecture on January 24, 1983
in partial fulfillment of the requirements for the degree of
Master of Science in Visual Studies.

Abstract

The book is proposed as a candidate form for new electronic information systems. The electronic book casts the powerfully expressive new media of computers and video in the accessible forms of the printed book. It is a testbed for developing novel methods of interactive information retrieval and perusal that are as useful and intuitive as those of the printed book.

The evolution of the book in a history of technological innovation is examined, and the electronic book is established as a logical and adaptive next stage in that evolution. The influence of media forms on the information they carry, and on the way we think, is also considered. The movie manual, a videodisc-based prototype for an electronic book, is presented in a photo-demonstration. Videodisc production and the design and implementation of an object-oriented software system for editing and viewing the movie manual are described. The final section proposes some directions for further work on the movie manual, and looks at possibilities for forms of future books.

Thesis Supervisor: Andrew B. Lippman
Title: Assistant Professor of Media Technology

The work reported herein was supported by the Office of Naval Research under contract number N 00014-81-K-0436.

Table of Contents

Introduction

What if books were as difficult to use as most computers? Before you could look inside, you would need a secret password that unlocks it. Once inside, there would be no table of contents outlining all it has to offer, and no obvious way to flip through to find something interesting. Pictures would be rare. Text would be printed with characters in crude patterns of dots. You would never get to see more than 24 lines at a time. No indication of a page's length until it has already scrolled off the top of the screen; no turning back, except by starting over from the beginning. To write in it, you would use a mysterious set of commands. One color. No script.

In contrast, consider the power of a book that incorporates features of new electronic media. Open it to find pages of full color video images, which spring into action when you touch them. At the title page the author tells you, in his own voice, how the book is organized. Should you have trouble with a difficult passage, simply ask the book for specific examples, and to explain unfamiliar terms. Since the book's organization is stored as a computer database, you can add to the database new pages composed of text and pictures from other pages, and new chapters that collect together frequently used material for quick access.

This thesis concerns the book as a candidate form for new electronic information systems. The electronic book casts the powerfully expressive new media of computers and video in the accessible forms of the printed book. It is a testbed for developing novel methods of interactive information retrieval and perusal that are as useful and intuitive as those of the printed book.

Both books and computers are media that can carry virtually any message. The printed book comes by its expressive power through a rich heritage of innovative forms and distribution spanning more than 400 years. A book makes any idea that can be expressed in words and pictures accessible to all in an inexpensive and portable format. But despite the range of disparate messages that books carry, most people have a unique sentiment about the medium itself. Almost everyone enjoys curling up with a "good book"; the "good"-ness is individual, but the "book" is universal.

The computer promises to be a communication medium many times more potent than the book. By design, it provides intelligent access to vast stores of information. It is an active presence that can respond to the user, and modify its presentation to suit the user's needs. A growing flurry of activity in computer graphics and interactive video has added a new dimension of presentation. But the sentiment of most non-expert users toward the medium is more of intimidation than affection. The computer still feels like a keyboard, and perusing is like wandering through the Louvre in the dark with a flashlight. The computer design effort has concentrated on making it more powerful, not more approachable.

The electronic book presents a model for a computational video information system that is both complete and economical. The full capability of the system is expressed by the model; all materials are presented as elements of a book, and the methods for retrieval and perusal are like those for a book. Because of the ubiquitous cultural presence of books, the model is also quickly assimilated; its operation can be explained in a few simple statements, and is consistant with the user's everyday experiences.

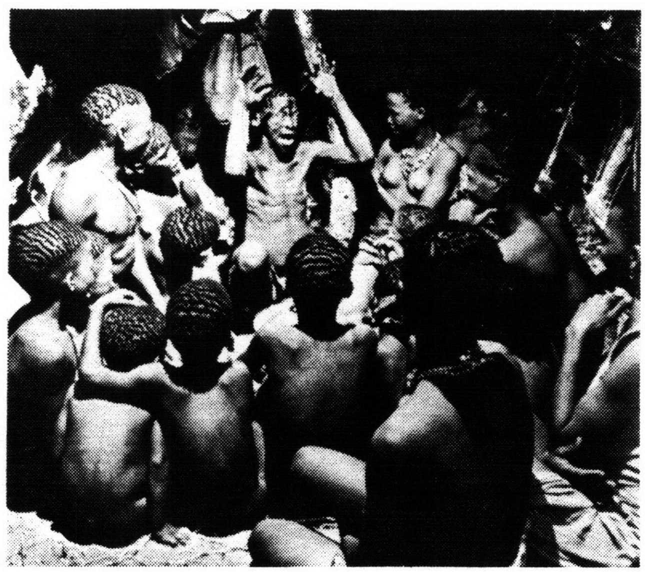
The following section examines the evolution of the book in a history of technological innovation, and establishes the electronic book as a logical and adaptive next stage in that evolution. The influence of media forms on the information they carry, and on the way we think, is also considered. The third section describes the Movie Manual, a videodisc-based prototype for the electronic book. Videodisc production and the design and implementation of an object-oriented software system for editing and viewing the movie manual are described. The final section proposes some important considerations for further work on the movie manual, and looks at possibilities for forms of future books.

Chapter One

A Chronicle of Book Forms

1.1 The Spoken Word

Before typographic technology made the written word commonplace, the knowledge base of a culture was archived in the orator's rhetoric and verse. There was no practical, portable storage technology other than human memory, nor publication beyond performance and conversation. Had the bards of pre-alphabetic Greece stopped singing, their entire culture might have vanished in a few generations.

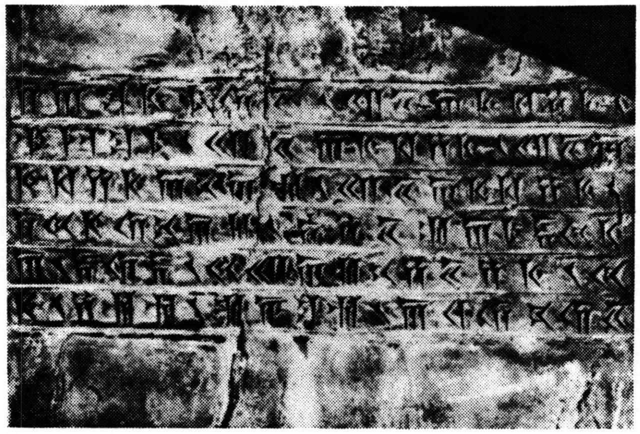


In Bechuanaland (Botswana)
(Nat Farbman, Time Inc.)

It is difficult to imagine living in a time when you could not look up the answers. Education in an oral culture is a process of direct participation, in deed and in the spoken word. No written manuals exist to explain the crafts of the culture, only the learned men and women. Walter Ong, in *The Presence of the Word* (p.28), posits the role of Homer as both entertainer and encyclopedia to the Greeks: "Although craftsmen could build ships and did, we have no evidence that anyone could actually verbalize directions for shipbuilding more informative than the *obiter dicta* on these subjects in the *Iliad* and the *Odyssey*."

Instruction in oral cultures occurs in face-to-face encounters, events where the gesture and spoken word have immediate and personal signification, where thought and action are simultaneous. The communications channels can be characterized as having a very high bandwidth, containing information to touch all senses, but also very local -- nothing is received that does not originate within the personal perceptual space of the individual. The information environment was personal, immediate, and participatory. For an apprentice craftsman, this meant that there was nothing to learn about his craft that lay beyond his personal perceptions and experience. Everything that he needed to know could be seen, felt, and experimented with in his own hands. Skills were remembered in muscular and perceptual traces of exercising the skill, and as the admonitions of the master craftsman.

1.2 The Written Word

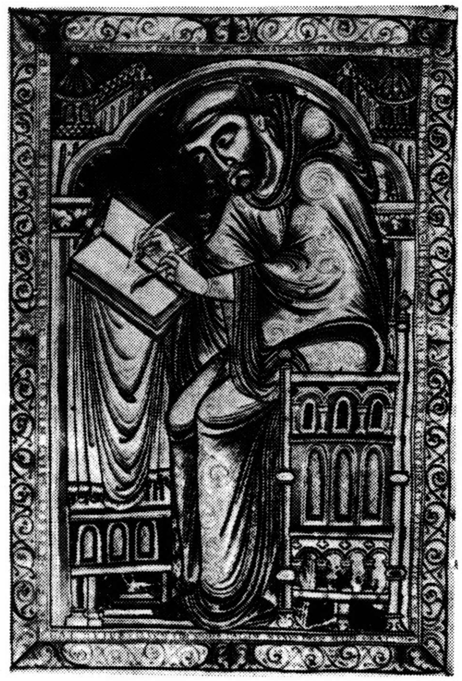


The Behistun inscription, which commemorates Darius the Great (521-486 B.C.). It is written in a Persian cuneiform script of the sixth century B.C., a phonetic syllabary of 41 signs, close to an alphabetic form of writing. It reads: "Darius, the great King, the King of Kings, the King of the countries, son of Hystaspes, the Achaemenid, who built this palace." (from [Diringer 69])

The invention and spread of phonetic alphabets permitted the fixing of verbal thought into a visual code -- static and precisely repeatable, but stripped of the dynamics of acoustic space. Marshall McLuhan distilled the sense of written speech in *The Gutenberg Galaxy*: "The invention of the alphabet, like the invention of the wheel, was the translation of a complex, organic interplay of spaces into a single space. The phonetic alphabet reduced the use of all the senses at once, which is oral speech, to a merely visual code."

For many centuries, though, reading reconstituted written thought with action. Literature was produced primarily for public recitation. Even when in private, literate people regularly read aloud, pronouncing the words one at a time, feeling the texture of the sound. Reading aloud restored the personal significance of the word, making it an active object in acoustic space, a physical event. Jean Leclercq discusses the central importance of reading aloud in memory and meditation in the Middle Ages in *The Love of Learning and the Desire for God* (pp. 18-19):

> ...in the Middle Ages, as in antiquity, they read usually, not as today, principally with the eyes, but with the lips, pronouncing what they saw, and with the ears, listening to the words pronounced, hearing what is called the 'voices of the pages'. It is a real acoustical reading; *legere* means at the same time *audire*; one understands only what one hears...when *legere* and *lecito* are used without further explanation, they mean an activity which, like chant and writing, requires participation of the whole body and the whole mind. Doctors of ancient times used to recommend reading to their patients as a physical exercise on an equal level with walking, running or ball-playing.



A self-portrait of the scribe Eadwine in the Christchurch, Canterbury, scriptorium, from his own manuscript, the *Canterbury Psalter*, c. 1148-49. (from [Evans 66])

The art of memory is essential to an oral culture. In the medieval university, reading aloud and writing were vital to the training of memory. Since there there was yet no way to reproduce and distribute multiple copies of a written text, the simplest way to acquire a text was by copying it, either from another text, or from a teacher's dictation. McLuhan notes (p. 95), "the medieval student had to be paleographer, editor and publisher of the authors he read." The manuscript was indeed a unique, personal product. It recounted the student's intellectual development. A degree candidate of the medieval university might be required to present the books he produced as evidence of his training.

Theologian Amaury de Bene in a classroom of the University of Paris, with four student clerics; from *Cronique de France ou de Saint Denis*, early 14th century. (from [Evans 66])

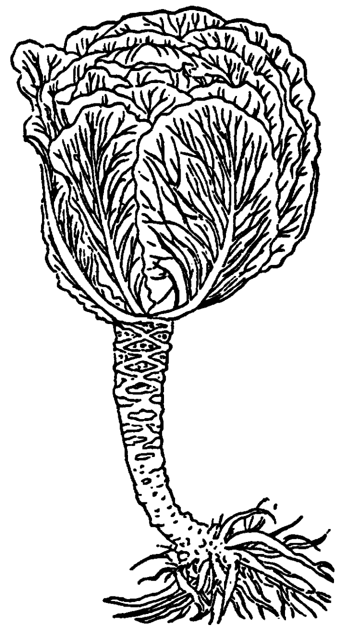
The manuscript was, like oral disputation, a mosaic of opposing points of view, usually from a variety of sources, probing the subject from many angles. It was in fact conversational -- conversations which the scribe created by juxtapositing arguments, and in which the reader participated while reading aloud, resounding the "voices of the pages". Reading and writing were adjuncts to oral discourse, where wit, as ingenuity and resourcefulness in argument, was the measure of intellegence. Ong notes (p. 59) "...the art of structuring thought was taken to be dialectic, an art of discourse, rather than pure logic... . Thought itself was felt to take place typically in an oral exchange."

1.3 The Printed Word

A printing shop of the late sixteenth century.
"On the left the typographer is working the press so as to force the paper against the inked types... On the right a scholar is checking the printed sheets while above another is examining the proofs. Between them a compositor is setting up the types of a new text, which he is copying from the book he has by his side." (from [Eco 61])

Two almost concurrent innovations of the mid-fifteenth century mark the beginning of the demise of Western oral culture. The printing press made the book the first uniformly repeatable, mass-produced commodity. Do-it-yourself book production of the manuscript era was replaced by a consumer-oriented culture that invented the author and his audience, roles that are conceivable only when multiple copies of the same book are possible. Moreover, the nature of reading itself changed. The eye can speed over the orderly, uniform space of the printed page at a rate much faster than the eye and ear can digest the irregular script of handwritten books read aloud. The disputational "voices of the pages" fell silent to the single, unified tone of the author that could be absorbed in a glance. There was no reason for students to make their own books, to participate in the process of exploration and discovery experienced by an author, for the books were ready-made.

The printed book dissociated the reader from the interplay of observations in oral exchange. Similarly the new method of perspective drawing reduced the sense of spatial presence to a single point of view. The method proved to be particularly powerful for making informative technical illustrations, and was soon the norm for all forms of illustration, just as equitonal prose eventually became the norm for printed text. The geometric rationalization of pictorial space finds a corollary in schematicized illustrations, the first of which can be found in Fuch's 1545 herbal, *De Stirpium Historia*. The woodcuts were general statements of plant form, not the particular form of an observed plant [Ivins].

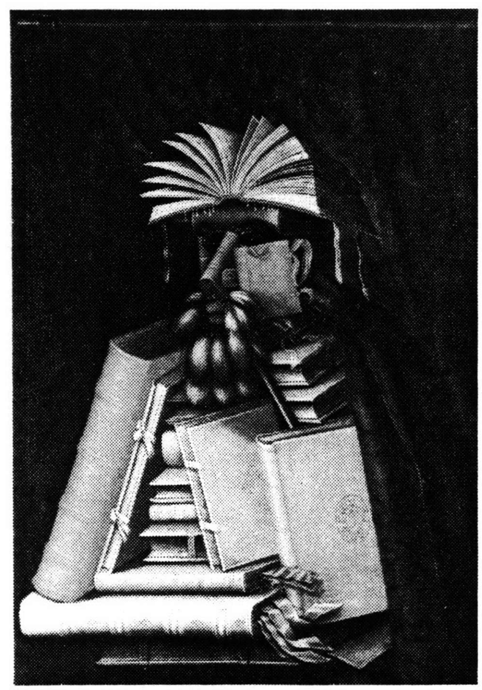


Woodcut of "Kappiskraut" from Fuchs's *De Stirpium Historica*, 1545. (from [Ivins 69])

Visual perspective and schematic illustrations give generalized, homogeneous information about the space or object they depict. An explicit, fixed point of view isolates vision from the dynamic interplay of the senses involved in direct perception of space.

The reduction of the "complex, organic interplay of spaces" into a flat, homogenized visual world inspired an entirely new way of thinking about the world, and man's position in it. Ong writes (p.73): "This neutralized, devocalized physical world has in a profound sense moved out of relationship with man's own personal, social, vocal world. It is really beyond man's ken. Henceforth, man will be a kind of stranger, a spectator and manipulator in the universe rather than a participator."

In the supersonic world of print, the old classroom methods of oral discourse were too slow and contentious for the practical-minded, who, alone with an inexpensive book, could gather facts at a much faster rate. The conception of knowledge itself changed from something characterized by dialectic wit to something more like a commodity, a body of facts, embodied in the book. It was in this new wave of the ubiquitous book that the classroom as we still know it today first took shape in the programs of French philosopher Petrus Ramus (1515-1572), who denounced the Humanist tradition in education, and proposed that learning be centered on the book as teaching machine.



The Librarian, Arcimbaldo

1.4 A Textbook Classroom

The Ramist classroom was not a place for lively debate. Instead of the probing of "numerous simultaneous vistas of any topic whatever" was the lecturer reading from notes to a silent audience. In *The Gutenberg Galaxy* (p. 146), McLuhan considers how these new educational methods paved the way for industrialization:

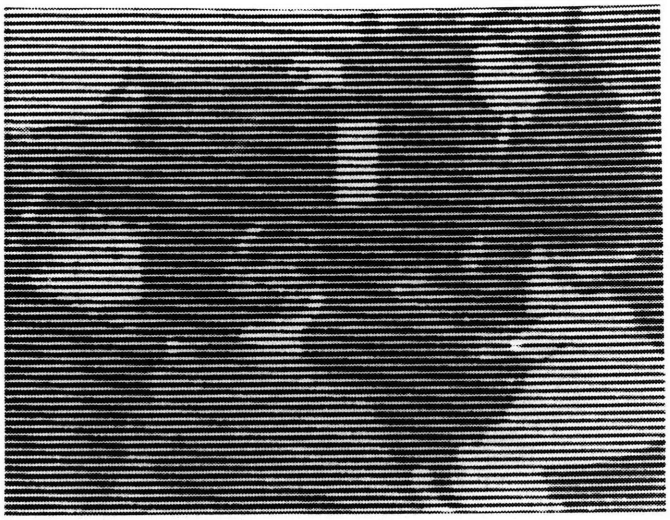
> Ramus was entirely right in his instance on the supremacy of the new printed book in the classroom. For only there could the homogenizing effects of the new medium be given heavy stress in young lives. Students processed by print technology in this way would be able to translate every kind of problem and experience into the new visual kind of lineal order. For a nationalist society keen on exploiting its entire manpower for the common tasks of commerce and finance, of production and marketing, it needed very little vision to see that education of this kind should be compulsory. Without universal literacy it is hard, indeed, to tap the manpower tool.

El modo da insegnare compitare
(The manner of teaching how to count)
Engraving, Florence, 16th century.
(from [Eco 62])

For individuals from primarily oral backgrounds, entering then Ramist classroom, or its industrial corollary, the assembly line, entailed a dramatic sensory reorganization. McLuhan continues: "Napoleon had great trouble in getting peasants and semi-literates to march and drill, and took to tying their feet with 18-inch lengths of rope to give them the necessary sense of precision, uniformity and repeatability." There was nothing in the past experience of these recruits that could have prepared them for the abstract sensorimotor task of rank-and-file marching. It was only through extreme means that the necessary order could be imposed.

Widespread literacy, for which the Ramist classroom was largely responsible, engendered an explosion of practical thought and applied knowledge the likes of which had never before been experienced. A profusion of inexpensive, illustrated, informative manuals laid bare the mystery of technological crafts that before had been locked up in the exclusive craft guilds. From the literate man was born the self-educated man, mechanically ingenious, the inventor of the industrial age.

Arrival of the newspaper
Engraving, Paris 19th century.
(from [Eco 62])



J. L. Baird demonstrating his television apparatus, 1932. (from [Eco 62])

With the industrial age came faster and more efficient machines, and more precise machines and tools that could not only perform man's work for him, but also perform work beyond human physical and perceptual capabilities. Microscopes and telescopes extend human vision into macro- and microcosms that before only existed hypothetically on paper. Jet travel and telecommunications have extended localized human presence to a global one, overlapping many cultures. The new media environment might be characterized by two of its most pervasive elements, the computer and the television. These two repetitious, sequential scanning machines perform at such a super-human rate that they obliterate any sense of sequentiality. Like primitive man, we live in what McLuhan calls a world of "allatonceness".

1.5 A Village Again

Today, it is difficult to imagine a time when you could not change the station, or talk with someone half-way around the globe. In its shaping of our sensorium, the information surround today is much more like that of the pre-literate tribal village than that of the eighteenth century town. While the literate mind must tune out of acoustic space to focus on the visual flow of thought, men of the tribal and global village have sharp ears for the sudden crack in the distance, forwarning of battle. The difference is that tribal man's awareness is local, bounded by direct perception. For the global villager, the crack is of gunfire thousands of miles away, microwaved into his home. But the implication for both is the same -- they are affected by that battle, like it or not.

In Bechuanaland (Botswana)
(Nat Farbman, Time Inc.)

In the tribal village, wits are honed by experience. Children become hunters by hunting, and by preparing for the hunt in play simulations and rites. So, too, does the child of the global village sharpen his wits in play and ritual derived from warfare and commerce acted out on the television screen. The pre-school child of today has seen and heard more of the world than his great-great- grandfather could in a lifetime. He enacts guerilla wars from other lands, or distant galaxies in his backyard. His capacious, orally-trained memory keeps track of television schedules long before encountering the multiplication tables, and he can recite television commercial jingles from memory like encantatory chants. In short, children have, as they have always had, an enormous capacity for learning. They rapidly and without apparent effort acquire the materials and methods of their culture.

1.6 Classrooms of another Culture

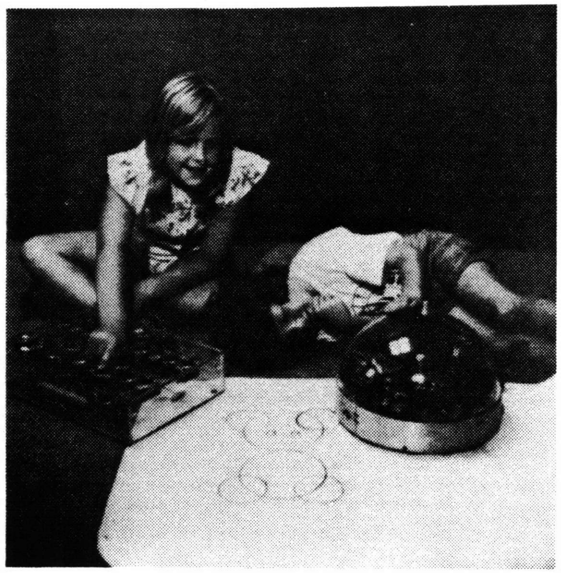
But the classrooms today are still largely products of Ramist principles, centered on textbook procedures and written examinations. Knowledge is something that is acquired through "book learning". When a child enters the classroom, he leaves behind the self-structured aural-tactile world of electronic media for a bewildering world of subjects and class periods in single-file. If television or movies are brought into the classroom, it is usually as "visual aids" to textbook procedures, much as the first printed books were "aural aids" to the oral procedures of the medieval classroom. Early attempts at computer-aided instruction cast the computer in the role of super-textbook, pretending that individual score-keeping and a variety of ways to say "wrong answer" made instruction personal. The idea that an instructional environment consists of orderly rows of desks and books, pigeonholed subjects and alphabetical roll calls comes from another culture. The television-educated child must have some unspoken empathy with the foot-tethered soldier in Napoleon's army.

In his book *Mindstorms*, Seymour Papert discusses the failure of instructional methods that are not "syntonic", that is, not compatable with the cultural roots and self-perceptions of the student (p. 172):

> Educators sometimes hold up an ideal of knowledge as having the kind of coherence defined by formal logic. But these ideals bear little resemblance to the way in which most people experience themselves. The subjective experience of knowledge is more similar to the chaos and controversy of competing agents than to the certitude and orderliness of *p's* implying *q's*. The discrepancy between our experience of ourselves and our idealizations of knowledge has an effect: It intimidates us, it lessens the sense of our own competence, and it leads us into counterproductive strategies for learning and thinking.

1.7 Objects to Think With

Papert proposes computer programming and Turtle geometry (a computational geometry) as a new mathematics for teaching mathematics. Using the Turtle -- a robot-like mechanical or graphical object that draws figures under the control of computer programs -- as an "object to think with", the student acquires an intuition about mathematics that simply can't be conveyed with abstract written notation. Children are encouraged to "play turtle", to act out the movements of the turtle when solving a difficult problem.



Children with LOGO turtle (from [Papert 80])

They are able to project their own acquired navigational skills onto the turtle, and then to the commands that direct the turtle's navigation. In this way, turtle geometry is participatory and body-syntonic, consistant with the child's body-knowledge. The child uses materials and methods acquired from his surroundings as building blocks of his own intellectual structures; and he learns that these structures can be extended to new problems, reconfigured and debugged to meet the demands of his problems.

This paradigm of instruction has parallels in the oral culture of the ancient and Middle Ages, where the art of discourse was central to education, and thought itself was something that was believed to occur in oral exchange rather than in silent study. The building blocks of discourse were commonplaces, or *loci communes*, prefabricated themes and formulas that could be used to support positions taken in debate. Disputation required on-the-fly compositions of these commonplace arguments, and the student who could thread together themes with ease and eloquence was praised for his wisdom. The achievement of *copia*, abundant and fluent discourse, was a prized ideal in the oral culture; it represented mastery of the culture's primary means for storing and organizing knowledge. In *The Presence of the Word*, Ong tells us (pp. 81-82) that writing helped foster copia by permitting the personal accumulation of more commonplaces than could be kept in memory. The written commonplace book, or copybook (from *copia* book) was an "object to think with", which Papert defines as an object "in which there is an intersection of cultural presence, embedded knowledge, and the possiblity for personal identification" (p. 11).

1.8 "As We May Think"

Copia remains a highly prized ideal. But the curriculum of the textbook classroom does not reveal how it may be achieved. There, textbooks and lessons are objects to consume rather than to think with. To fluently express and juxtapose ideas requires agility in retrieving and structuring information. But, as Papert points out (note 4, p. 223), our culture is relatively impoverished of materials for thinking and talking about the process of structuring information. We are voracious consumers of facts, but our techniques of organization and retrieval are ineloquent, and difficult to teach to children.

Vannevar Bush addressed this dilemma in his 1945 essay "As We May Think": "The summation of human experience is being expanded at a prodigious rate, and the means we use for threading through the consequent maze to the momentarily important item is the same as was used in the days of square-rigged ships." The maze is a product of an intellect taught to perceive information in the homogeneous space of the printed book, and to organize it in an orderly, hierarchical fashion. The natural life of the mind is quite unlike the narrow paths of the maze. Bush continues:

> Our ineptitude in getting at the record is largely caused by the artificiality of systems of indexing. When data of any sort are placed in storage, they are filed alphabetically or numerically, and information is found (when it is) by tracing it down from subclass to subclass. It can be in only one place, unless duplicates are used; one has to have rules as to which path will locate it, and the rules are cumbersome. Having found one item, moreover, one has to emerge from the system and re-enter on a new path.
>
> The human mind does not work that way. It operates by association. With one item in its grasp, it snaps instantly to the next that is suggested by the association of thoughts, in accordance with some intricate web of trails carried by the cells of the brain. It has other characteristics, of course; trails that are not frequently followed are prone to fade, items are not fully permanent, memory is tranistory. Yet the speed of action, the intricacy of trails, the detail of mental pictures, is awe-inspiring beyond all else in nature.

1.9 A Mechanical Book

What Bush proposed to replace the maze is surprisingly prescient of current trends of research in information science, if not of the technology. He writes at some length to project a technological basis for a "mechanized private file and library", or "memex" device for amplifying knowledge through personalized associative indexing. The workstation is a desk with indefinitely large storage for microfilmed materials, several projection screens, a keyboard and sets of buttons and levers. Buttons transfer the user to common landmarks in the material, such as the head of the index or a personal investigative trail, and levers are used to page through the material at varying speeds.

The design of the memex places great emphasis on recording and retrieving images. Using microfilm and dry processing, the memex user can record notes and correspondances to augment purchased microfilms of books and periodicals. Microphotographs are taken in the laboratory or the field with a camera "a little larger than a walnut" mounted on the forehead, and a viewfinder inscribed in a pair of glasses. Images and text are retrieved by a mnemonic index code, and two items can be associated by joining them with a new code word. As a topic is explored, many associations are made, and investigative trails are built up which include previously published materials as well as new photographs and handwritten memoranda.

New personal trails made through disparate materials in the memex can be viewed and annotated, even reproduced. "It is exactly as though the physical items had been gathered together from widely separated sources and bound together to form a new book." If a trail is pertinent to the research of a colleague, "he sets a reproducer in action, photographs the whole trail out, and passes it to his friend for insertion in his own memex, there to be linked into the more general trail...Wholly new forms of encyclopedias will appear, ready-made with a mesh of associative trails running through them, ready to be dropped into the memex and there amplified."

This type of investigation echoes the personal bookmaking in the medieval university, where, as McLuhan notes, the student had to be "paleographer, editor, and publisher of the authors he read". Reading aloud from and writing in these books were activities that involved the entire mind and body, and helped train the memory, not for facts and words, but for themes and formulas to draw from in scholarly disputation. Similarly, the memex owner uses the materials and methods of the memex as building blocks in the making of investigative trails. Learning takes place while participating in information transfer, and in interactively associating and augmenting the materials.

Dr. Bush can hardly be faulted for casting the memex in a machine-age workstation of levers and buttons.* The consumation of a sudden insight might be achieved in a flurry of lever pulling and button pushing to rival Chaplin on the

---

*He does, however, speculate briefly on interaction with the memex via direct electrical stimulation of the senses: "Must we always transform to mechanical movements in order to proceed from one electrical phenomenon to another?"

assembly-line in *Modern Times.* Even with the relative ease of "pushing data" with a powerful computer, display terminal, and interactive editor, the experience is decidedly less involving than human conversation, and certainly less memorable. The fluency one achieves at the keyboard is of keystrokes and operating system commands, not of ideas.

1.10 New Technologies for Conversation

Recent and continuing advances in media technologies provide the tools for forward-looking prototypes of future instructional and informational environments that are multi-sensory, engaging and responsive -- in a word, conversational. And the development of potential conversational partners is likewise proceding apace. It seems almost trite to point out the ubiquity of small powerful computers in our lives as the cost of executing an instruction becomes negligible. The Japanese now talk of mega-lips (logical inferences per second) instead of mips (millions of instructions per second) in projecting the capabilities of fifth generation computers.

The automatic transmission movie manual videodisc (Architecture Machine Group, 1983)

The optical videodisc, developed as a distribution medium for theatrical films, has proven to be an omnibus medium with unforeseen computational applications. Its unprecedented image capacity of 54,000 frames per side, and still-framing capabilities make the optical disc a very inexpensive frame store for full-color, photographically complex images. Frames can be viewed individually or in sequences at thirty frames per second as sync-sound movies, or at any variable rate in between for slow motion. The inherent random access capability of the rotational medium permits on-the-fly editing of still and moving image sequences directed by the viewer.

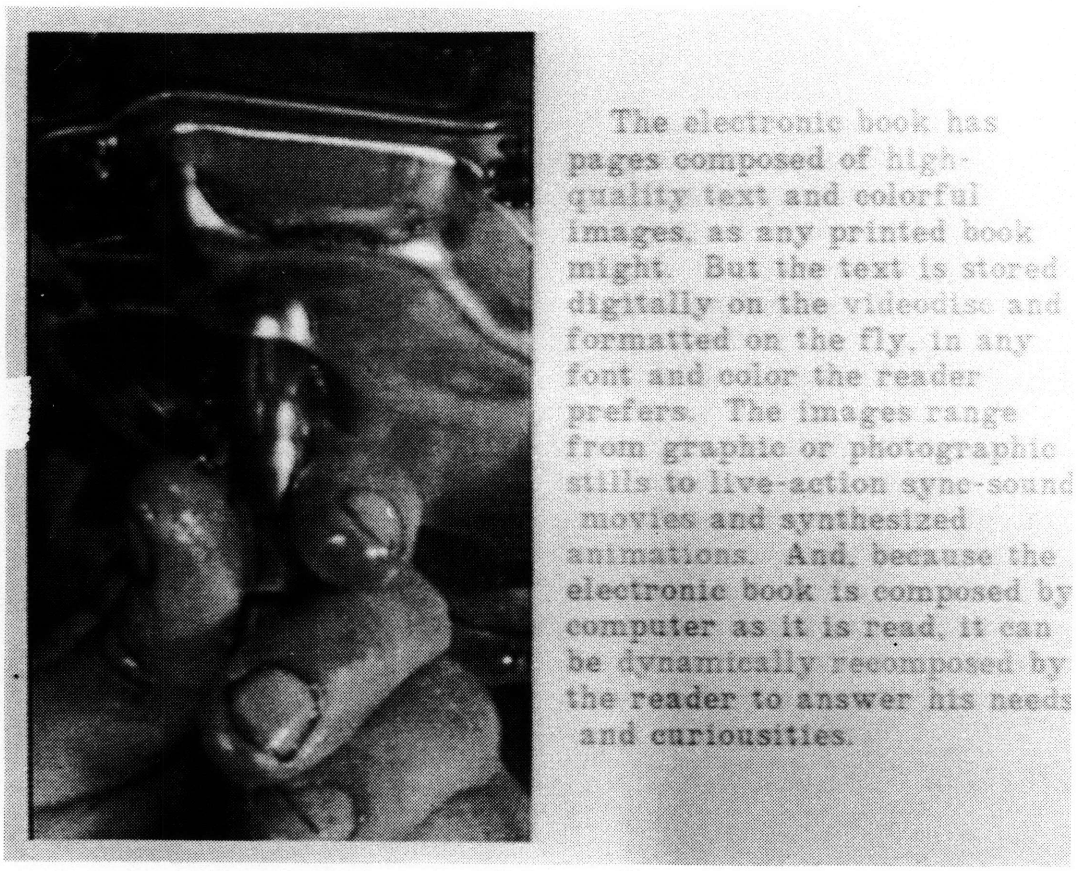
The optical videodisc is also a digital storage medium. And because a disc is stamped from a master, it can be exactly replicated in unlimited quantites. This makes it an attractive alternative to magnetic media, which must be serially reproduced, for distributing software and databases. Augmented by a computer and a simple decoding device, the optical videodisc player makes the television an intelligent, colorful, and dynamic window onto a vast world of information.

Interaction and path-finding through that window should not have to be redirected to a narrow channel of coded keyboard input, but should happen at the window itself. New computer interfaces expand the input-output channel to a human bandwidth, like that of face-to-face communication. Touch-sensitive displays listen to gesture at the window. A stroke across the screen can turn pages, spin movies, or write notes. A simple poke can activate light buttons, or open a new windows onto the space. Speech recognition and synthesis enriches the interactivity even more. A spoken question gives the gesture a context -- "What's that for?", "How do you use that?". The apprentice mechanic with his hands full can direct a very patient instructor by speech alone -- "Stop! Do that again!"

1.11 A New Conversational Book

These new technologies come together in a new medium for information retrieval and perusal, the electronic book. Characteristics of the electronic book may be found in other new information environments, such as the media room [Negroponte 81], the interface as habitat, and teleconferencing in virtual space [Negroponte 80]. The electronic book is distinct in that it is portable and publishable.

A device is portable if, as Alan Kay has noted [Kay 77], "other things may be carried too". Publishability means that the initial material of the electronic book may be produced and distributed in mass quantities. The digitally-augmented optical videodisc is the keystone of the electronic book. It provides a publishable medium for the dynamic imagery and high-fidelity audio essential to an engaging conversation, and the digital database that indexes the material, all in a portable format.



The reader orchestrates the presentation in conversation with the book using speech and gesture. The book display is touch sensitive, which allows the reader to reference objects by pointing, and to make written notes. A variety of global views of the database are available to help the reader select items, and to provide a "sense of place" in the book. Navigational aids, such as a video thumb index, allow the reader to rapidly peruse materials, and get from one place to another.

Like the printed book, the electronic book is fairly "stiff" when first opened. The reader is unfamiliar with the material within, and the book has not yet acquired the personal landmarks that identify its owner. A skeletal structure of overviews, chapters and pages is included, not as immutable form, but as a guidebook to possibilities. The reader quickly learns to negotiate this structure, then venture beyond to create his own. The reader becomes author, drawing together the building blocks of text, image, and sound to create new pages and paths. The book begins to acquire the imprint of its owner through these changes, as well as annotations in word and sound, dogears, and bookmarks; it falls open to page spreads "creased" by extensive use.

Personalization of the electronic book is more than the sum of personal landmarks. It includes inferences that the book makes about its owner from patterns of use. The vocabulary of speech recognition assimilates the reader's verbal style, and incorporates his lingo in its synthesized vocal responses. It acts as a partner in investigative searches. It suggests fertile territory to explore, and filters out irrelevant material based on its evolving model of its owner. As the reader develops a conversational fluency in the book material, the book becomes a mirror of this intellectual development.

Chapter Two

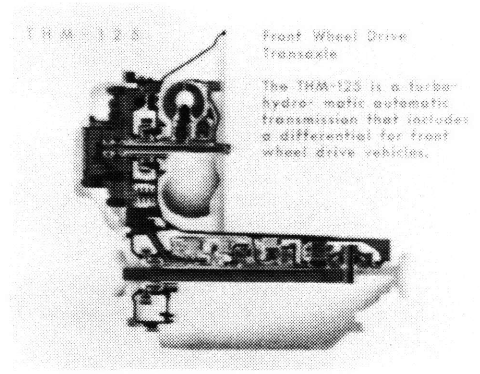
The Movie Manual

The movie manual is a prototype for an electronic book. Its aim is to illustrate features of the evolving form of electronic books for evaluation, and to serve as an extensible foundation for developing new forms. It is one aspect of an investigation into personalized aids for training [Lippman 80, Lippman 82a], in the specific instance of a repair and maintenance manual for automatic transmissions.

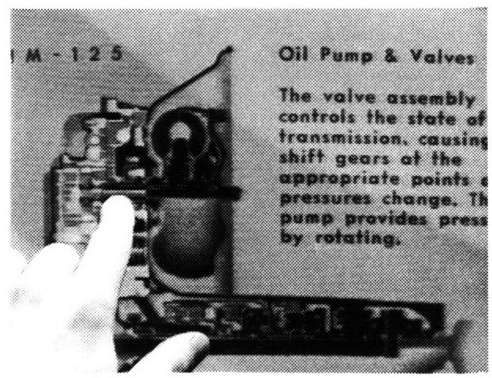
The movie manual addresses the issues of dynamic page composition from page elements of text, image, and sound, all stored on optical videodisc; interaction with these elements through gesture and speech; personalization of the book through annotation; and development of global overviews and aids for navigation through the book. This chapter describes the contribution of this thesis project to videodisc design and production, and to the design and implementation of an extensible software system for editing, viewing, and interacting with the elements of the movie manual.

2.1 A Demonstration

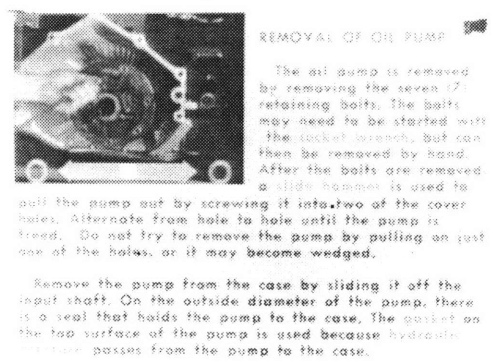
The following pages present a brief tour through the automatic transmission movie manual, to demonstrate some of the features of the current implementation.



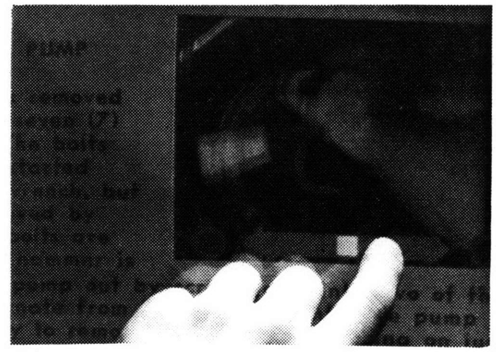
The manual begins with a touch-sensitive table of contents. All of the major components of an automatic transmission are shown in a black and white cut-away illustration.



The reader "browses" over the contents by touching the illustration. The major functional subassemblies are revealed to the touch by highlighting them in full color, and they are briefly described in the text window at right. Each subassembly corresponds to a chapter in the book, and the reader enters a chapter by touching the text window.



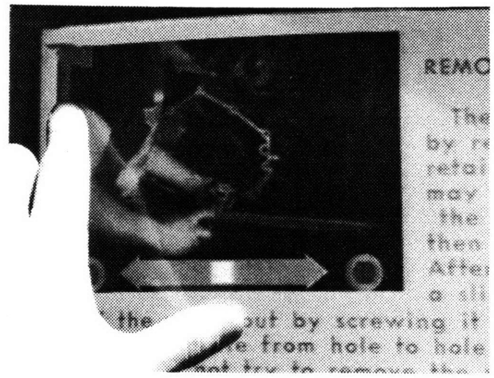
A page in the manual describes a single, complete operation. The text contains words highlighted in red (e.g."slide hammer") which may be unfamiliar to the reader. The directional indexes in the upper corners permit the reader to page backwards and forwards in the chapter.



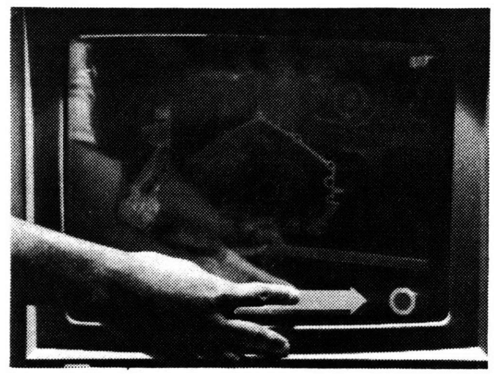
The illustration in this page is in fact a movie. The reader controls the pace and direction of movie play through the graphical controls at the bottom of the window. The round target at left rewinds the movie, and another at right advances to the end. Touching the two-headed arrow in its right half plays forward, at a speed proportional to the distance from center; touching the arrowhead plays full speed. The left half reverse play, and the center square stops the movie.

Some operations were filmed in both left- and right-handed versions. This is the left-handed version of "removing the oil pump".

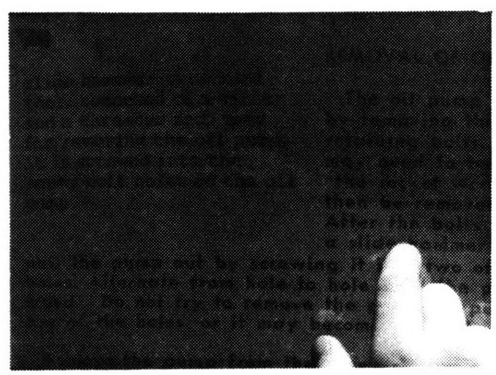
The quarter-frame movie illustration can be "blown-up" to full-frame by touching the "magnify" bar in the corner.



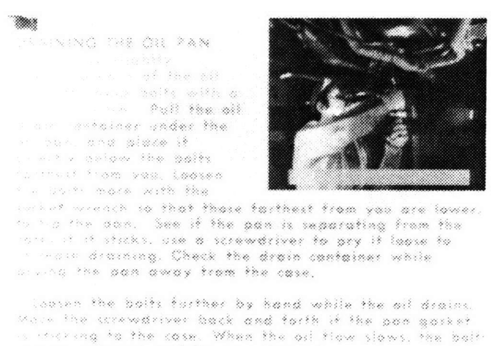
The full-frame version is identical to the quarter frame. The reader returns to the text page by touching the bar in the corner.



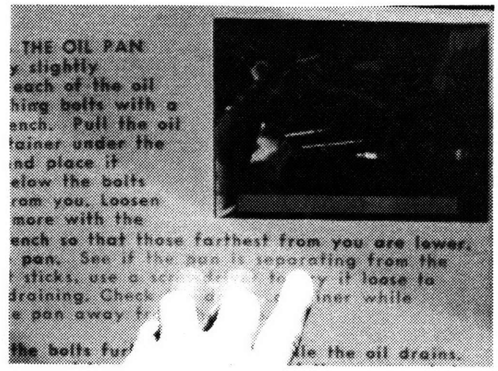
The highlighted words in the text are active page elements. By touching one of these words, the reader requests its glossary definition, which is displayed in place on the page, so that both the definition and the context of the word are presented simultaneously. Highlighted words in the glossary definition may also be interrogated for explanation.



Another page form is the step by step process. Here, text and illustration are segmented into the serial steps of a repair procedure. As the movie or slide show steps through the operation, the text corresponding to the segment is highlighted. In addition to the ambient sync-sound of the movie, this page has narration by the mechanic himself. He gives hints and precautions about each step.

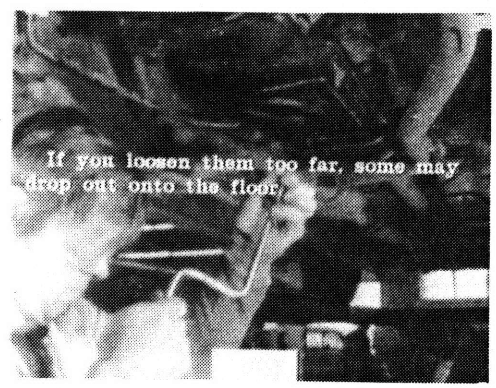


The reader may advance to a particular step, or replay a step, by touching the graphic control strip in the illustration window. Sliding along the strip reveals which text, and how much text, corresponds to each step. The reader may also touch the text itself to select the movie segment or slide illustrating that text.

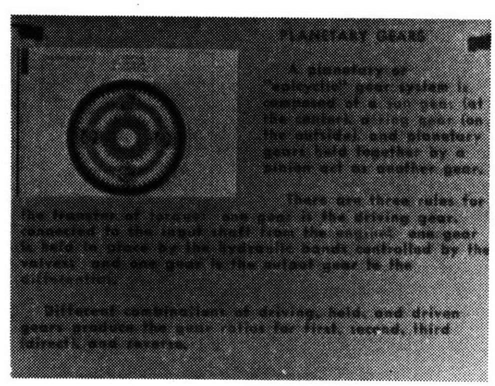


In addition to demonstrating the proper repair methods, the movie manual can also show what happens if the repair is performed carelessly. In this full-frame step by step movie, the text is superimposed, and only the text corresponding to the visible segment is displayed.

All of the live footage was filmed with sync sound, which records all of the important audio cues -- like ratchets clicking and bolts dropping to the floor -- that involve the reader as though he were watching over the mechanic's shoulder. On this page, there is also a voice-over narration which elaborates on the text.



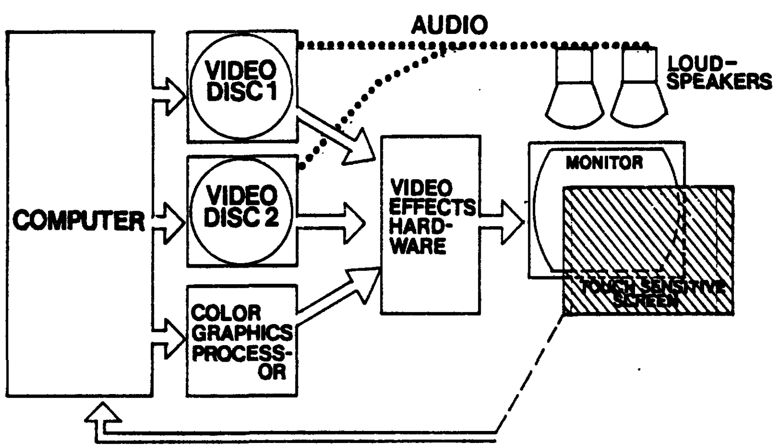
The still-framing capabilities of the videodisc permit a high density of illustrations in addition to the motion footage. The movie manual has slide shows such as this with photographs, schematic diagrams, and detailed line drawings in black and white and full color.



After the last page in a chapter, the manual returns to the table of contents, where the reader may select another chapter.

2.2 The Hardware Environment

The movie manual is implemented in what is now a standard hardware configuration for small interactive systems that include an optical videodisc. One or more externally-controlled videodisc players are used. The output of the players is NTSC video, carrying image and digital data, and two channels of high-fidelity sound. Graphics and grayscale text are generated in a frame buffer, with 640 by 480 resolution. An encoder converts its RGB video output to NTSC, which is keyed over the videodisc output with a computer-controlled switcher/mixer. The manual is displayed on a standard NTSC television monitor.



The movie manual hardware configuration. (diagram by Alyce Kaprow)

The predominant channel of input in the movie manual is touch and gesture, registered by commercially-available transparent touch-sensitive digitizer overlayed on the NTSC monitor. A connected speech recognizer is being integrated into the configuration to augment and disambiguate gestural input. A 4KHz bandwidth digital sound system, built at the Architecture Machine, permits run-time annotation of the manual with spoken or other audio.

All of these devices are peripherals to a 32-bit general purpose minicomputer, running MagicSix, a Multics-like operating system developed at the Architecture Machine. PL/1 is the language used for most systems and applications programs.

2.3 The Software Environment

The movie manual software is modeled on an object-oriented software system, where the elements of the book are independent objects which interact by sending and receiving messages. Objects are instances of abstract data types defined by a data structure and a set of operations, or methods which operate on the data structure. Methods are invoked by messages from other objects, and are the only means for accessing an object. Guidelines for software design were gleaned from the literature on Smalltalk, the software component of the Dynabook, Alan Kay's vision of the personal computer [Ingalls 78, Goldberg 76, Goldberg 79, Kay 82], and on programming languages which incorporate abstract data types, such as CLU [Liskov 77]. Object classes defined include primitive types, such as rectangles and videodisc frame numbers, and more complex, composite types, such as interactive movies and text windows.

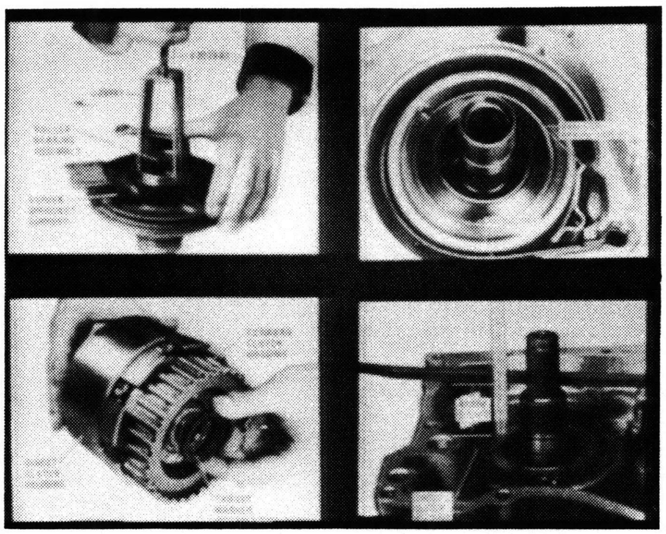
The movie manual author and reader operate in the same environment, using exactly the same software. In fact, the distinction between their roles is intentionally fuzzy. The reader implicitly modifies the page layout by requesting new information which is displayed in overlapping windows on the page. The reader also creates new objects in the book by recording sound annotations or graphical annotations on the page. Index searches can produce collections of objects that are bound together in a new page and inserted into the book. The result is a new book, optimized to the reader's own information gathering and perusal strategies.

2.4 Videodisc Production

The foremost consideration during production of the automatic transmission repair and maintenance videodisc was to include a wide variety of both traditional and innovative presentational styles of material as building blocks for the prototype movie manual. The building blocks fall into the three catagories of primitive data types that can be mastered on optical videodisc -- video, audio, and digital -- though there is naturally much overlap.

2.4.1 Video

General Motors Corporation provided hundreds of photographs and line drawings from their own transmission manuals, including exploded diagrams of transmission subassemblies, detailed photographs of repair procedures requiring special tools, and schematic diagrams explaining both mechanical and hydraulic flow of power through the transmission. These stills were copied onto 35mm film, both as full frames and as optically reduced half-frames, and quarter-frames, four to a videodisc frame.



A frame from the videodisc, with four independent movies quadruplexed on a single frame.

Live action movies were shot of transmission repair procedures in the garage and classrooms of the ITT Technical Institute of Chelsea, Mass. The film was shot in 16mm color, with synchronous sound. The same repair procedures were shot at various distances from medium to extreme close-up, so that the reader might select the most edifying view. Also, the complete disassembly and reassembly of a transmission was shot in a studio on the MIT campus at a rate of one frame per second. This compressed-time recording effects a substantial savings in disc real-estate while preserving the flow of events. Like the stills, all motion footage was mastered on the disc in both full-frame and optically-reduced quarter-frame formats.

2.4.2 Sound

The synchronous sound recorded with the live action footage was predominantly ambient sound, as one would hear while watching the repair procedures in person. This sound provides additional instructive cues, such as the clicks of a socket wrench giving an intuitive sense of the torque required to tighten a bolt.

Also, two forms of voice-over narration were recorded. On one of the videodisc's two audio channels, passages from an automatic transmission manual are read aloud in a neutral voice. On the other channel, the mechanic who was filmed in the ITT garage spoke extemporaneously about the repairs he performed, as he might explain them to a trainee. These verbal explanations are replayed as voice-overs for both still images and movies.

2.4.3 Digital Data

Recent work at the Architecture Machine Group has produced a method for encoding digital data on the standard NTSC video signal [Yelick 82, Brown 83]. The video data stream can be mastered on videodisc, like any other NTSC signal, to add new dimension to the electronic book as an information resource. While the purely digital data videodisc, with net capacities approaching 500 Megabytes per side, suggests the tremendous potential of the optical videodisc as a publishing medium for software and data, the admixture of data with images and sound institutes an entirely new medium, the publishable intelligent videodisc.

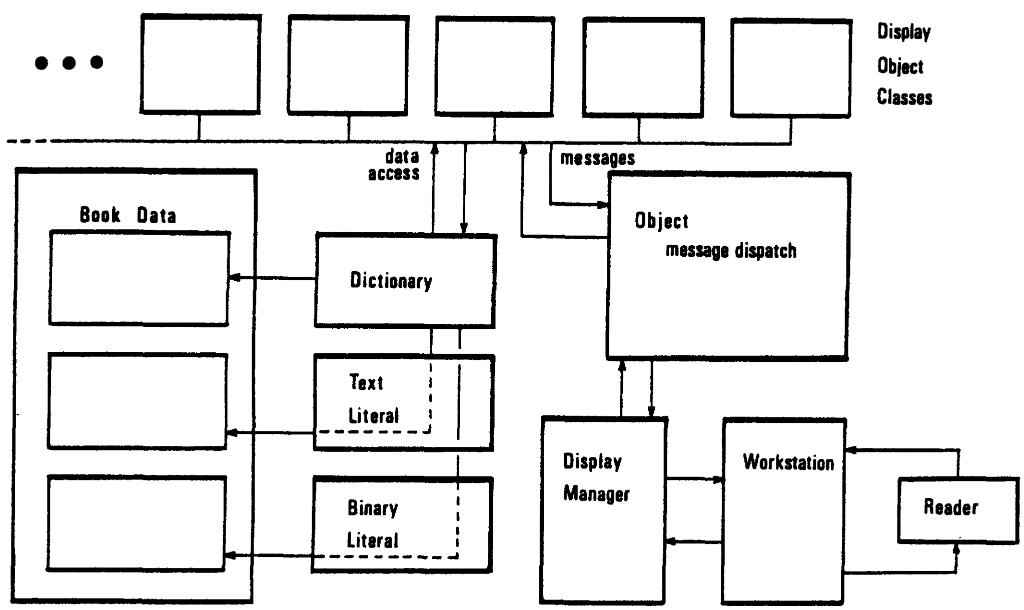
Over a quarter-million characters of ASCII text were encoded and mastered on the automatic transmission videodisc in approximately 25 frames of digital data. The text was the the contents of two automatic transmission manuals, provided by General Motors in machine-readable form. The encoded text can be retrieved from the disc, decoded, and displayed in high-quality grayscale fonts on dynamically-composed pages. The text might also be "read aloud" to the user by an inexpensive phonemic speech synthesizer with builtin text-to-phoneme translation.

Transcriptions of the spoken soundtrack are easily included for alternate presentations of verbal material, for producing hard copy, or for presentation as subtitles for the hearing-impaired. Subsequent versions of the automatic transmission videodisc will also include a descriptive database of the disc's contents, as well as the software for accessing the database.

2.5 The Software System

The movie manual software system adopts some of the principles of object-oriented programming, but is built on top of a data-procedure oriented operating system, MagicSix. MagicSix is a Multics-like operating system, and uses PL/1 as the primary systems and applications programming language. It has a powerful screen text editor, and a large body of support software for frame buffer graphics, including grayscale fonts, optical videodisc control, and device interfaces for speech recognition, touch sensitive displays, and digital sound equipment.

The kernal of the software system consists of four fundamental classes of objects, which oversees a readily-extensible body of display class objects. In addition, there is a display manager which mediates object communication with the reader by allocating and attending input/output windows.



Overview of the software system.

2.5.1 Object

The most fundamental class is the "object", whose instances are the display classes. The object class has methods defined for including new display classes in the software system, deleting display classes, and for enabling or disabling a display class. A display class may be disabled during debugging or modification, rendering all of its class objects "invisible"; these objects are made available again once the class is enabled. No display class may be deleted as long as any instance of it exists.

The most important ongoing function of the object class is its message dispatch method. Any object or process can send any other object a message of the form:

<action> <object> [<modifiers>]

without explicit knowledge of the <object>'s type, or of what procedure will actually execute the <action>. The message dispatcher determines the receiving <object>'s class, and resolves the call to the appropriate class method to handle the call, passing along the <modifiers>, if any. For example, a message to the object class itself might be "create sound note 'bearing noise'". This message tells the object class to create a new instance of the display class "sound note", digitally--recorded sound annotations, with the name "bearing noise". This message might be sent by another object, perhaps a slide of a bearing assembly, as a response to the reader's command "take a note".

The message dispatcher permits display objects, the elements of the electronic book, to communicate with one another by sending and receiving messages. No object requires any knowledge of another's implementation details -- the data structures or operators -- in order to manipulate or request display of another object. In most cases, only the object name is necessary for reference. Probably the most frequent message is "display <object>", often with a window name as a modifier describing what part of the page area to display in. With such a simple interface protocol for object interaction, the software system can easily be extended to include new types of interactive display objects without requiring any modification of existing classes.

2.5.2 Dictionary

The class "dictionary" has but one instance for each book, the data dictionary, which includes an entry for each display object defined for the book. The data dictionary is structured as a binary search tree, ordered on the unique object name. Each dictionary entry contains six elements:

1. the object name: a unique identifier.
2. the object description: an arbitrarily long, free-format text string, which describes the object contents.
3. the object class: a numeric code for the object's class.
4. the object icon: the name of another object, a small digital image that can be used to represent the object in overviews and index listings.
5. the object value: the instance value of the object, a binary string whose format and meaning is determined by the object's class definition.
6. usage statistics: administrative information, such as the name of the user who created the object, and when it was created; when it was last modified, and by who; a reference count, telling how many other

objects refer to this one; a usage count, telling how many times the object has been displayed.

The data dictionary of a book is its index -- an exhaustive listing of its contents, in alphabetical order. Any data necessary to select or preview an object -- its name, description, and pictorial precis -- are immediately available from the dictionary entry, without having to invoke any of its class methods for display or examination. This simplifies index access for overviews and other global navigational aids.

2.5.3 Literals

The object name, description, and value elements of a dictionary entry are in fact instances of the two fundamental classes of literals. All object names and descriptions are instances of the text literal class, and all object values are instances of binary literals. Both types of literals are byte strings, defined by a triplet of positive integers: (file number, byte offset, length). Although both types are essentially identical, separating text from binary data facilitates text string searches.

Every book has its own distinct data directory in the MagicSix file system where literals and the dictionary are stored as multi-segment files. In the literal triplet, file number determines the segment number, byte offset points to the first byte in the string, and length says how many characters are in the string. All literals begin on full-word boundaries. A backpointer to the literal's owner in the dictionary is stored with every literal. A length word is also stored with the literal, so that literals may be examined independently of their owners.

Because all of the descriptive data for all display objects are ultimately stored as text and binary literals, the literal class methods perform all of the storage management duties in the software system, outside of those for the dictionary entries. Literal class operators are defined for inserting, deleting, retrieving, and updating literals, and for typing out both text and binary strings in a readable form.

2.5.4 Display Manager

The display manager class is a class of possible hardware configurations. An instance of the class describes the devices of the reader workstation. The display manager is the one component of the software system which knows the specifics of the hardware environment, such as the dimensions of the frame buffer, polling details of the gesture, point, and speech input devices, and the instruction format for the videodisc players.

With device dependence localized in the display manager, the display object classes can be defined for a virtual configuration, and then transformed to the actual configuration at display time. For example, window boundaries can be defined in a world coordinate space, then scaled and translated to the actual viewing space when displayed. An object may have a small vocabulary of voice commands to which it will respond. If the display configuration lacks a speech recognizer, the vocabulary is simply ignored, or made available for keyboard command input.

Changes in the hardware environment may be incorporated by editing the display manager definition, or by creating a new instance of the display manager. Selecting a new display manager is as simple as selecting a new book; one merely loads in a new profile or dictionary, which is applied to all subsequent display operations.

The present implementation of the movie manual has no general class of display managers, but rather a single display manager written for the hardware configuration described above. The display manager acts much like a window manager of a virtual terminal [Meyrowitz 81]. It maintains a prioritized list of displayed objects, including the region of the display space each occupies and the class methods to invoke when the object is referenced by some input. If objects overlap, the display manager can determine the minimum area of an object that must be redisplayed for it to be completely revealed. When an object is erased, the display manager determines what parts of remaining objects are uncovered, and causes them to be redisplayed.

2.6 Display Objects

The active, visible elements of the movie manual from which pages are composed are all instances of some display object class. A display object describes some parcel of the videodisc -- a movie segment, an audio passage, some digital data representing text or digital image -- as an interactive element of the book. Display objects give the video, audio, and digital materials on the disc names and descriptions, by which they can be retrieved, and forms in which they can be "read", that is, viewed and investigated. The interactive display object is atomic; it incorporates both the videodisc material and the interactive procedures for displaying and manipulating the material.

A display object class defines the structure for data describing an instance of the class, and a set of procedures, or methods, for accessing the data. The internal structure of an object is accessible only to the class methods. An object is manipulated by sending it a message which invokes one of its methods. For example, a page, which is simply a set of object names, selects a subset of objects and sends each a message to "display itself". The reader manipulates objects by implicitly sending messages via the display manager. When the reader touches the screen, the display manager determines which object is visible at the area touched, and sends it a message that the reader is touching it. How the object responds to a touch is determined by the class method invoked by the message.

2.6.1 Class Methods

Messages sent between objects are generally of the form:

<action> <object name> [<modifiers>]

If the <action> is the name of a method defined for the <object name>'s class, that method is invoked for the <object name>, with the optional <modifiers> applied if relevant. If the <action> is undefined, the message is simply ignored.

All display object classes are required to have certain methods defined before they can be included in the software system. This guarantees that fundamental operations, such as creating, editing, and displaying, are valid for all objects. When a new display class is to be added, the Object class is sent a message to "create <new class name>". The Object create method verifies that the minimum set of class methods (create, edit, tty out, display) have been defined, then inserts the new class name in its class list. Display objects of the new class may then be created for any book by the message "create <class name> <object name>".

2.6.2 Creating and Modifying Objects

New objects can be created an any time. The published electronic book contains an initial set of display objects, stored as digital data, which exhaustively describe the contents of the videodisc. The published data dictionary is augmented by local magnetic storage where the user creates new objects for annotations, new or modified groupings of elements into pages and pages into chapters, or entirely new objects and object classes.

Most display objects are created explicitly by author or reader sending "create" messages. In the current stage of implementation, almost all create and edit methods accept instance values from the keyboard. New methods are being developed which allow object creation using touch-sensitive controls and pop-up menus. A graphical control pad for perusing the videodisc was developed as an alternative to terminal keyboard input for specifying videodisc frame numbers.

Some objects are created implicitly, both while reading and during initial database creation. When the reader requests that a sound annotation be made, a new "digital sound" class object is created and entered in the data dictionary. The sound, perhaps a verbal memo, is recorded and stored digitally. The instance value for the object is the storage location and length of the recorded sound, along with other descriptive information which associates the note with a user and the page on which it was recorded.

Objects may be edited implicitly as well. For example, glossary definitions are presented by displaying a text window object named "glossary definition". Whenever a glossary lookup is performed, this object is edited to contain the requested definition. In general, any operation that can be invoked explicitly by author or reader also may be performed in response to messages sent by any object.

2.6.3 Other Common Methods

In addition to the required create and edit methods, "tty out" and "display" methods must be defined for every display object class. The tty out method simply types out an object's instance value at the user's terminal in a readable form. Since every class has a distinct data structure for its instance value, writing a single, general pretty-printer for all object classes, including those yet to be defined, is impractical. On the other hand, the user should not have to memorize the name of a different print routine for each class. The mundane, but vital task of typing out object values illustrates the power of the object-oriented approach. One may examine an object's value, however simple or complicated its structure, simply by sending the message "tty out <object name>".

Displaying an object on the movie manual page is just as simple. A message, "display <object name>", invokes the display method of the <object name>'s class. The message may also contain modifiers which specify how the object is to be displayed. For example, an inherently amorphous object, such as a text window, receives a list of windows into which the text should be formatted. Without the list, the object uses a default window, such as the entire screen, or a default window specified when the text window was defined.

In the prototype implementation, display method communication with the display manager is optional. Ideally, all display methods would negotiate use of the display devices with the display manager, since it alone would know the details of the hardware configuration. The display manager also fields system interrupts signaled by the input devices, or by the videodisc players when an audio or video segment has finished, and sends messages to the appropriate object. If an object plays some segment of the videodisc, it sends the manager a message telling it what method to invoke when the segment is done. Any object that responds to the user's touch sends a message to the display manager telling it what region it occupies, and what method to invoke when the user touches the screen in that region. If regions overlap, the manager can determine which object is "on top", or most recently displayed.

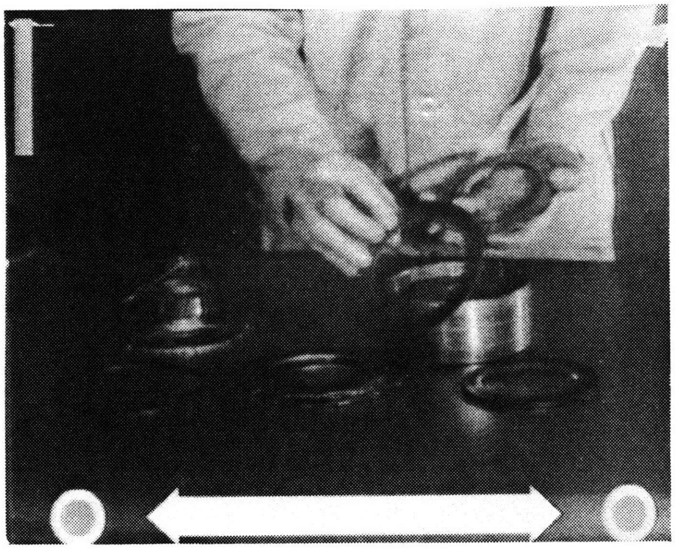
Object display regions may overlap when a temporary window, such as a glossary definition or an index entry, is displayed on the page for immediate reference. Or, the reader may wish to display more objects than will fit on the display surface without overlap. When an overlapping object is moved or erased, the uncovered objects should be restored. Most object classes, then, have a "redisplay" method which regenerates the object in a given region. However, the entire object is not redisplayed, since it may still be partially obscured by other objects; also, for some objects such as text windows, regenerating the entire display region may take an annoyingly long time. Therefore, the display manager determines the minimum portion of the uncovered objects that must be redisplayed, and sends a "redisplay" message to those objects, along with a list of the regions that must be regenerated.

When the reader moves from page to page, the entire display is erased and the next page is constructed on a blank surface. The page object's display method sends the display manager a message to re-initialize the display map. However, when a single object is removed, it must release whatever system resources which it has claimed, and be erased from the display. A class's "remove" method does just that. Along with whatever class-specific termination functions that must be performed, the remove method tells the display manager to erase the object, which causes the redisplay of anything that object obscured.

2.7 Some Display Classes

2.7.1 The Basic Movie

The basic movie is a simple object in appearance and behavior. It is a sequence of contiguous frames from the videodisc, displayed on the page in a window of a prescribed size and location. Along the bottom of the window, a graphic control is presented, through which the reader changes the direction and speed of movie play. The round target at left rewinds the movie to the first frame, and the target at right advances to the last frame. If the first and last frames are the same (i.e., if the basic movie is in fact a still), no movie controls are displayed.



The basic movie, with graphic controls generated in the frame buffer, and keyed over the videodisc image.

The basic movie begins on the first frame, and remains frozen until the reader touches a control. The two headed arrow controls the speed and direction of movie play. Touching the right arrowhead plays forward at full speed; the pace is slowed by touching closer to the center, and stopped by touching the center square. Likewise, reverse play speed increases as the touch moves closer to the left arrowhead.

Basic movies that are smaller than the entire page are usually displayed as movie illustrations on text pages. Often those movies appear twice on the videodisc, in both reduced and full-frame sizes. The basic movies which have full frame versions will also display a "magnify bar" in the upper left corner. When the reader touches this bar, a full-frame replica of the movie is presented, temporarily obscuring the text page. This full frame expansion has the same graphic controls, scaled to the larger size. The vertical bar in the magnified movie returns the reader to the text page with the reduced version of the movie.

The data structure for the basic movie is simple. It includes the frame numbers of the first and last frames of the movie sequence; which synchronous audio track(s), if any, should be played; and the name of the window in which it is to be displayed. Like all display classes, the basic movie has methods defined for creating, editing, and displaying objects, and for typing out their instance values. A "touch" method is also defined for handling reader interaction with the graphic controls.

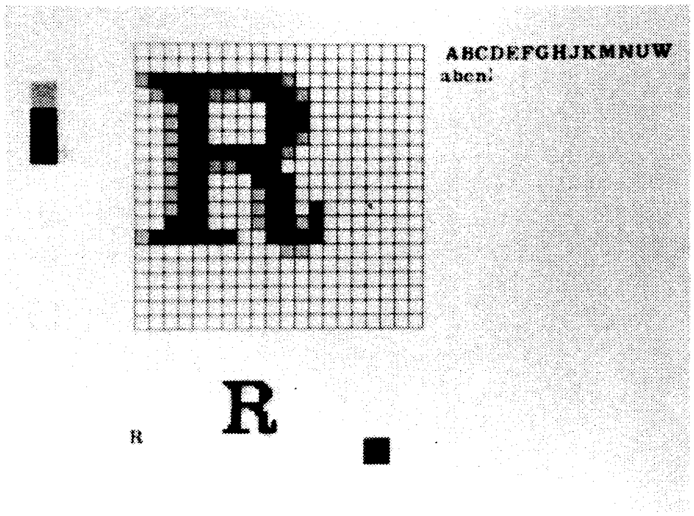
A constraint on basic movie display, a constraint shared by all videodisc image classes, is that the videodisc image is not relocatable on the display. It's position on the frame is fixed when the videodisc is mastered. This constraint is alleviated somewhat by repeating the images in different windows of the videodisc frame. This was done for the automatic transmission videodisc with all of the single frame images; each slide appears on the disc five times -- as a full-frame, and four quarter-frames, one in each quadrant. All of the motion footage appears on the disc at least twice, once as full-frame, and again in quarter-frame. The redundancy affords some flexibility in page makeup, but nowhere near the complete freedom for conventional book designers. Perhaps the engaging feature of book pages with movies as illustrations compensates for limitations on page format. And, with continuing innovations in digital television, real-time video transformations -- scaling and translating the image to appear anywhere on the page -- may be a plausible option in the forseeable future.

2.7.2 The Text Window

ASCII text is the most plastic of representational forms in the electronic book. With the small set of standard digital codes, a vast range of video and audio presentational forms is possible. Even the least expensive microcomputers and calculators can display clearly legible, if tiresome, character sets on CRT and LCD displays. The very same digital text can be routed for hard copy output in fragmented dot matrix or crisp daisy wheel characters. And this text has still another life as the spoken word, through inexpensive phonemic speech synthesizers such as the Type 'n' Talk*, which has an inboard ASCII-to-phoneme translator.

The text window class of the movie manual presents digital text in dynamically-typeset blocks, composed from attractive grayscale fonts, and which respond to the reader's touch with new, or newly-presented information.

*TM Votrax, A Division of Federal Screw Works, Troy, Michigan

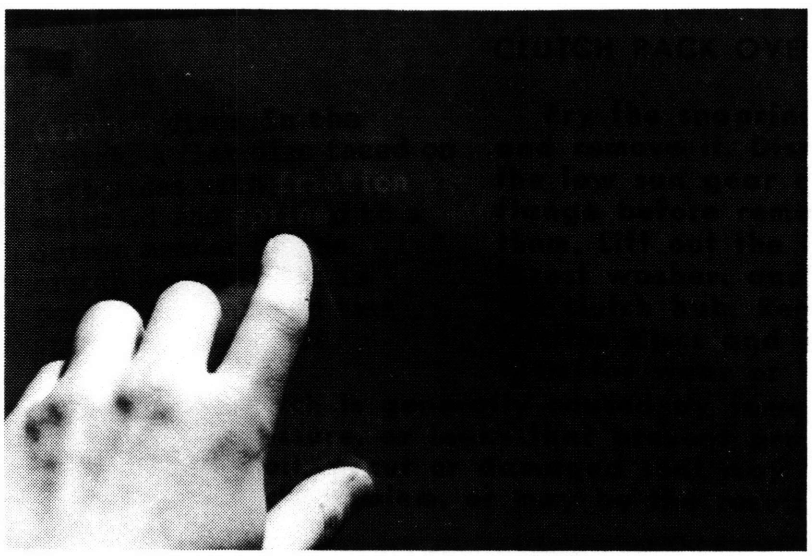


A "soft font" character, magnified to show the gray levels which smooth the edges of the character image.

The grayscale, or "soft" fonts were developed by the Architecture Machine Group for dense display of near-print-quality text on color television [Schmandt 80]. Characters are displayed as two-bit images, blurred at the edges, producing a smooth composite video profile that even inexpensive television receivers can follow and reproduce accurately. Intermediate gray tones demarcate regions on fractional pixel boundaries, producing what is perceived as a high- resolution image on a low-resolution display. And, perhaps even more significant is the wide gamut of colors available at no additional cost or computing overhead.

Color is used in text windows, both to please the eye and as a "subcarrier of information". Text is displayed in cool, low-saturation colors, on white backgrounds. When a text window is displayed over a videodisc object, bright, neutral colors on a transparent background are most legible.

Color is also used to signify active elements of the text. Words or phrases that have glossary entries are highlighted in a different color than the surrounding text. When the reader touches a highlighted word, a temporary text window is composed with the glossary definition and displayed on the same page, in a distinct font and color. The glossary definition may also have highlighted words to be interrogated for definition. The glossary window implements a very useful, and novel feature of juxtaposing reference and context in a book. The reader can lookup the meaning of an unfamiliar word without giving up the sentence that triggered the query.



A glossary definition, displayed on a page where the word is used. The definition also has highlighted words which may be looked up in the glossary.

The text window class was one of the first to be implemented, and its data structure reflects the author's incomplete conceptualization in the early stages. The text window instance value consists of a MagicSix file system path name to the text file, the name of the font with which to display the text, and the background, character, and highlight colors to use. In restrospect, it seems more useful to assign font and colors at display time, based on global considerations, such as consistancy with other recently-displayed pages.

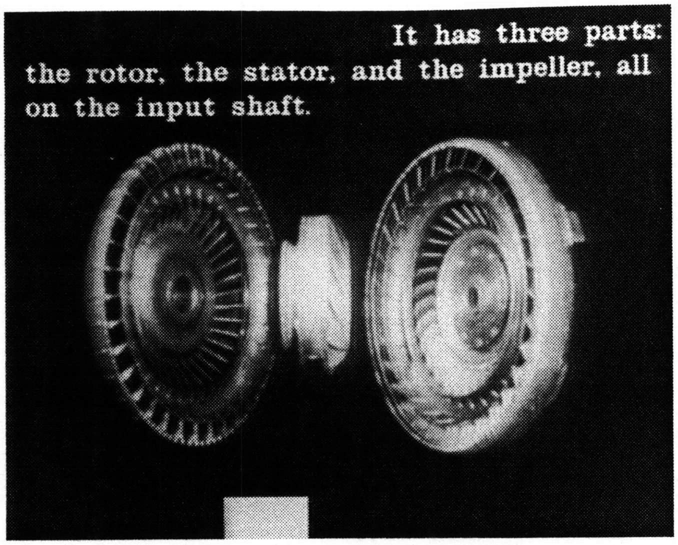
On the other hand, the text filename itself is not a sufficient handle on directing the appearance and behavior of the text window. The text window display method uses a formatting and display software package developed for another project. This package initially allowed a limited degree of format variability through control characters embedded in the text. As the capability of the text window was enhanced, the display software was modified to permit more embedded control of text display and response, to the point that the software became unweildy and occasionally unreliable.

Consequently, a new text display package is being developed to provide more flexibility, and extensibility, to the text window class. The text window object will be the text file itself, containing an abundance of embedded control information -- what Lippman has termed "illuminated text" [Lippman 82b]. Meta-syntactic characters in illuminated text have many analogs in print. For example, asterisks and numbers denote footnotes, brackets surround bibliographic citations, and parenthesized pointers (e.g., "(see figure 2)") associate text and image. The electronic book incorporates these asides and amplifications in the interactivity of the text window.

2.7.3 The Step by Step Process

A most common form of exposition in repair manuals is the numbered list of instructions which lead the reader through a repair process. Each step may be accompanied by a photograph illustrating an intermediate stage of the process, or a diagram with labelled parts and precautionary notes (e.g., "Do not remove this screw!").

In the automatic transmission repair movie manual, this form is embodied in the step by step object class. A step by step object combines a text window object with an ordered list of videodisc segments containing movies and/or stills. The videodisc image is displayed with a narrow rectangular control strip, scaled to the size of the image window, and divided into subrectangles, on for each text/video segment. Quarter-frame image windows appear as insets in the page, with text displayed on a white background. Full frame images have the accompanying text window superimposed through a transparent background.



A step by step photo-animation, an exploded diagram of the torque converter.

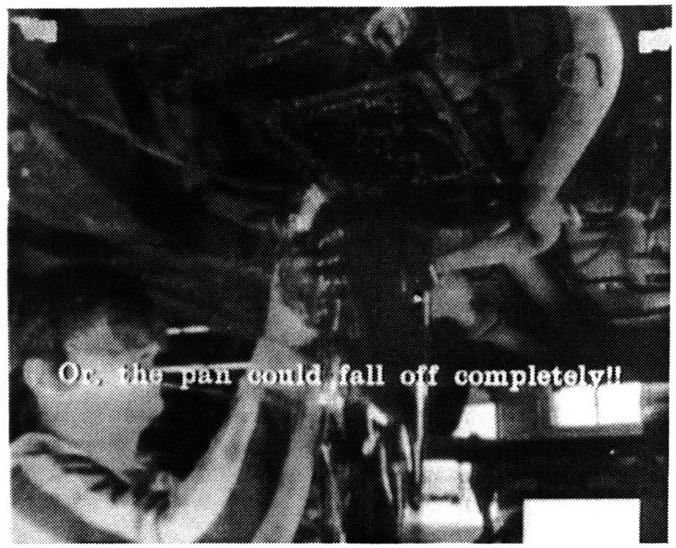
Display begins with the first image or movie segment of the list, with its corresponding text segment highlighted in a different color. A movie segment may be accompanied by synch sound from the same videodisc. Asynchronous sound, such as narration, from a second videodisc player may be played with either movie or still segments. When both audio and video portions of a segment have finished, the step by step object advances to the next one; the highlights in the text and control strip are also advanced to the next segment.

At any time, the reader may intervene to replay a segment, or to skip ahead. By sliding along the touch-sensitive control strip in the image window, the reader may select any one of the steps in the list. The highlights in the text and in the slider follow the reader's touch, revealing the division of material, and how much text accompanies each step. The text is also touch-sensitive. The reader may touch any passage of text, and the entire segment containing the passage is highlighted, along with the corresponding section of the slider. When the reader stops touching, the selected videodisc segment is played.

The ASCII text file for the step by step object's text window contains control characters which delineate the segments of text. The segments need not be ordered, or even contiguous in the file; a unique control character identifies the subsequent text as belonging to a particular segment number. The step by step object's instance value contains the name of this text window object, and an ordered list of videodisc segments, and the name of a window in which to display the videodisc images. Each video segment is described by a start and stop frame number, and the play speed, ranging from zero (for stills) to 256 (for full speed, 30 frames per second). Which sync audio track to use (left, right, both, neither) is also specified. If simultaneous, non-synchronous sound is to be played, it is also described by a start and stop frame number, and which audio track(s) to use.

The step by step display class is a very flexible form for conjoint text and image display. It has been used for slide shows explaining principles of planetary gear operation; for dynamic "exploded" photodiagrams of transmission assemblies; and for labelled animations explaining special tool usage. It is quite effective as the electronic book analog of the printed manual's step by step procedure. And, with live sound movies as illustrations, it can demonstrate aspects of repair procedures which the printed manual can only describe. For example, a step by step object was created demonstrating what can happen if the transmission oil pan is removed carelessly. The point of the lesson is brought home when the errant mechanic is drenched with transmission fluid.

A step by step movie, demonstrating the wrong way to drain the transmission fluid.



2.8 Higher-order Objects

Any object of the movie manual may be displayed and examined in isolation. But it is the use of the objects as building blocks for larger semantic constructs that crystalizes the notion of the electronic book as an "object to think with".

In the printed book, an author develops his viewpoint in the semantic forms of print -- the sentence sustains a thought, a paragraph collects thoughts into a coherent point. An illustration amplifies an idea in counterpoint to the text. Sections and chapters gather ideas around a central theme or topic.

> But there are also syntactic boundaries tied to the medium of presentation more than to the information content. The *page* is an excellent example. The ideas in a letter or book clearly do not fall on page boudaries. But one remembers a three page love letter or 900-page novel partly as such... [Negroponte 79a].

With inventing the electronic book comes the opportunity to reorganize semantic and syntactic constructs in useful alignments, and even to invent new ones. Moving illustrations and touch-sensitive text are novel extensions of familiar book elements. The electronic book must also have classes of objects for collecting and juxtaposing these elements in semantic constructs that encapsulate viewpoints -- both the author's and the reader's -- on the object universe. These aggregate objects are both capsules, for what they contain, and filters, for what they leave out. They must be permeable as well, to accomodate changing viewpoints as the reader makes the book his own. And, they must be wholly integrated into the uniform environment of the electronic book. They interact with other objects, including the objects they contain, by passing messages.

The homogeneity of the object universe has implications on the overall form and use of the electronic book. There is no inherently heirarchical taxonomy of object classes. Grouping objects may be nested. Elemental objects may reference higher-order objects as asides. This environment may be too flexible, and frustrating for the reader. Thus, the movie manual is initially presented as a conventionally-structured book, with the promise of free reign when the reader becomes more comfortable in the environment. Issues of global views to give the reader a sense of place, aids to navigation through the book, and the nature of grouping and regrouping objects are all larger concerns of the movie manual project. Presently, the manual includes two familiar grouping constructs that provide a familiar structure to the manual, and serve as models for implementing more unconventional groups later.

2.8.1 The Page

The first grouping object class implemented for the movie manual was the page. Like a page in a printed book, the movie manual page is a syntactic unit. By definition, it covers the entire display; only one page may be displayed at a time. But, unlike its print counterpart, the movie manual page is a semantic unit as well. It presents completely a single topic or idea, in one place.

The data structure for the page is simply an unordered set of display objects of various types. It may contain more objects than can be displayed at once. The excess may be redundant along some dimension; for example, the same illustration may be represented both as a photograph and a schematic line drawing. The reader may introduce new materials onto the page, such as a glossary definition, or a full-frame blow-up of an image. The characteristic feature of the page is not the initial selection and arrangement of objects on the display, but rather the page as a frame for a topic -- to borrow from Papert, a "mind-sized" bite. It is a workspace for viewing objects which share a common attribute, for studying that attribute, and ultimately as a base for finding other objects with related attributes.

Page display begins with a blank background, and the name of the page written in the upper left corner. Objects are selected from the page's set until the display resources are exhausted. Currently the selection algorithm is minimal. A single videodisc image object is selected at random from those available in the page. If display space remains, a text object is selected, also at random. As items are selected, an outline of the impending page format is drawn. Finally, each object selected is sent a message to display itself.

The page class was implemented before the display manager, and performs many of the same functions, such as fielding reader interactions and videodisc management. It is an obvious place to handle this overhead, but is less general than the display manager. As new forms of grouping object classes are defined, these administrative tasks won't have to be replicated for each.

2.8.2 The Chapter

The chapter object class collects pages into linear lists, and instigates their display one at a time. Chapters provide an entry into the electronic book software system from some "front end", or table of contents program.

Currently, the table of contents is a black and white cut-away illustration of an automatic transmission. The program allows the reader to "browse" through the table of chapters by highlighting the major subassemblies of the transmission in color as they are touched. A chapter is associated with each subassembly, and its contents are briefly described by a small text window displayed with each highlighted region. The reader enters the chapter by touching the text window. The table of contents program sends a display message to the chapter corresponding to the selected assembly.

When a chapter is displayed, it merely sends a display message to the first page in its list. Once the page is formatted, the chapter display method places graphic directional icons in the top corners of the page, with which the reader may select the next or previous page. The first and last pages display a different icon which indicates that going backwards or forwards, respectively, will return to the table of contents.

The chapter as linear list of pages emulates the familiar structure of a printed book. But in operation, it is more like a strict tutorial than a book, since it confines the user to that linear path. Another form of chapter under development presents the reader with a local table of contents, with a concise graphical description of each page, and lets the reader select his own path. Just as any display object may appear in many pages, a page may be included in any chapter. Thus, the reader's traversal can easily cross semantic boundaries as necessary to develop an investigative path.

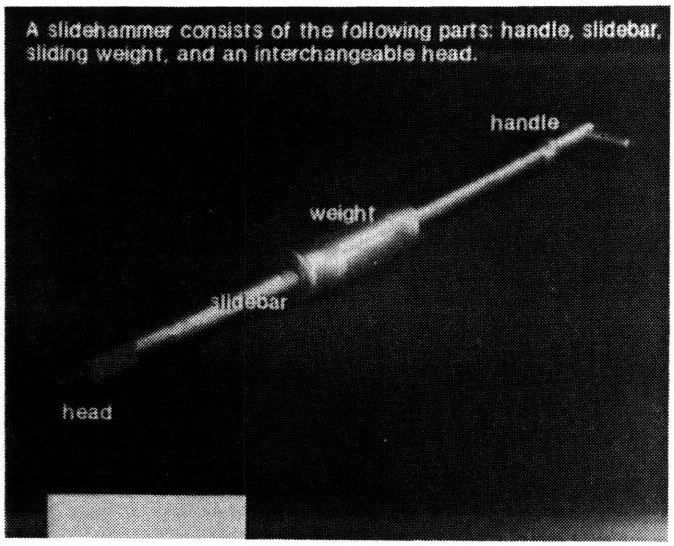
Chapter Three

Present and Future Electronic Books

3.1 Continuing Work

The present state of the movie manual demonstrates a number of salient characteristics of an electronic book. Perhaps the most striking feature, judging from reactions of visitors to the lab, is the display of interactive movies on the printed page. At the reader's touch, the familiar literary tableau of text and pictures springs to life. The reader looks over a mechanic's shoulder, as an apprentice might, to watch him perform a repair, and listens to him expain the procedure in his friendly, casual parlance. With a simple gesture, the reader can ask the expert to show and tell again and again. And he can refer to a concise written description, presented simultaneously with live demonstration, and interrogate it for explanations. The page engages the reader's eyes, ears, and hands; he is once again a participant in a complex interplay of sensory spaces.

The movie manual has also shown the utility in defining syntactic building blocks smaller than a page. It is a simple task to create or modify a page, even with the basic toolkit available now. The uniformity of the object-oriented environment has been reassuring to inexperienced users of the system. They learn a small set of powerful operations that apply to all classes of objects, and in a short time they can create new display objects and pages for the manual. Some of the most inventive pages have been the first made by a novice user; this has been most gratifying for both the user and the system's programmer.



A step by step slide show which describes the slide hammer, a special tool used to remove the oil pump. This was the first page made by a new user of the movie manual software.

But the current state of the movie manual is merely a foothold, albeit a firm one. As novel forms of information display and interaction are developed, they can be integrated into the system and evaluated without impact to the existing display classes. Digital sound and digital paint object classes are high-priority items. These will permit the user to make informal audio and written annotations in the book, as well as prepare more formal sync-sound graphic objects such as video chalkboard lectures which augment the text and film materials.

A new movie type, the flipbook. The reader flips through the frames of the movie by stroking the control strip at the bottom. This control is most useful for live footage recorded at less than 10 frames per second, or for flipping through a slide show.

The potential of more intelligent, illuminated text is in the early stages of investigation. In addition to being a more elegant approach to text windows, it offers new possibilities for establishing the kind of associations between text and image, and text and text, that Bush proposed for the memex.

New algorithms are necessary for selecting the elements from the page object and displaying them in an attractive format. These methods might draw upon an evolving model of the reader's preferences and needs, and explicit requests, as they are monitored.

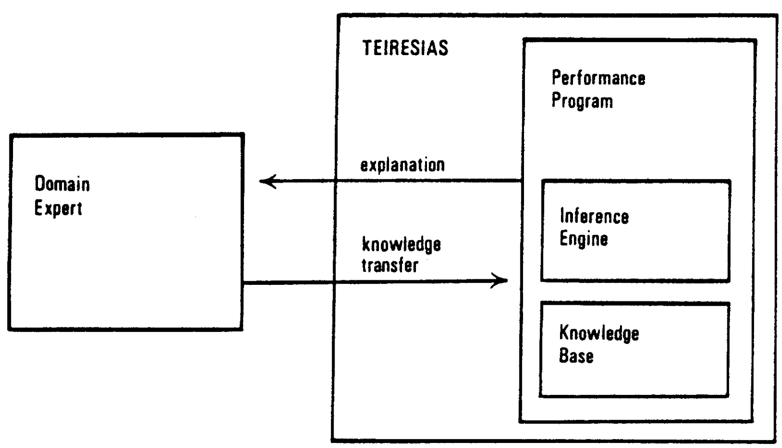
Providing the reader with a sense of place in the electronic book, and the means to get from here to there, are probably the most important and challenging issues yet to be addressed, and will receive their due attention in further research [Backer 82a]. Negroponte describes these media-inherent bases of orientation as "media fiducials" whose functions must be reinstated in new media forms for books. "What becomes of the bookmark as the indicator of progress through material? One need only recall one's reading of a novel such as Tolstoy's *War and Peace* to recall with appreciation the familiar, the friendly presence of this guide through the vast succession of pages." [Negroponte 78]

Finally, the user interface to authoring tools must be moved from the keyboard to the display, the reader's workstation. Some initial work has been done on graphical tools for perusing the videodisc. But ultimately, all of the methods for creating objects should be as simple and obvious as the methods for displaying them, and should operate in the same speech and gesture domain. An object should be edited by directly changing its appearance on the display, so that the result may be immediately viewed and evaluated. For example, the page display might be equivalent to an edit buffer for text windows, where attributes and associations can be added and modified by user interaction. New text could be added by keyboard, or by a user-trained handwritten character recognizer.

3.2 Future Books

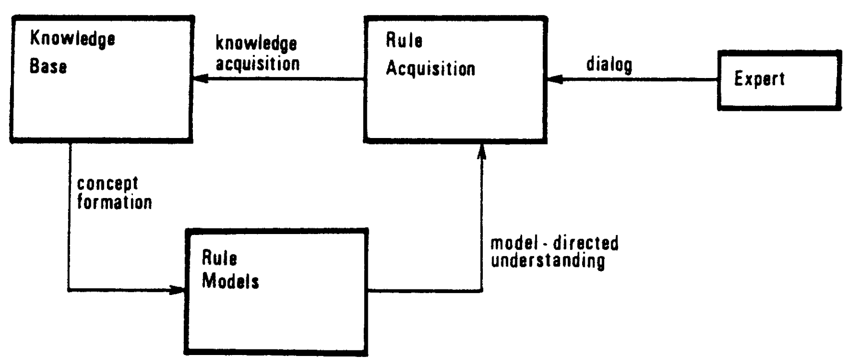
The electronic book is not a design specification. It is a an attitude for designers, a way to conceive the form of an information system. It is a powerful, but concise, metaphor that makes a complex computational environment accessible to inexpert users. The movie manual illustrates aspects of the form in a repair and maintenance workbook. But other contexts are easily imagined.

Consider the expert system MYCIN [Shortliffe 76], a computer system which guides physicians in diagnosis and treatment of infectious diseases. Presently it is consulted through question-and-answer sessions at the keyboard. Diagnositc data collected through new medical imaging techniques, such as CAT scans and ultrasound imaging, must be discussed in quasi-natural language dialogs where much of the visual nuance is lost. Like the first books, MYCIN is a hand-built, expensive, and limited resource. But expert systems are just beginning to emerge from the laboratory. When published on videodisc, as electronic books, they will be as approachable as Merck manuals, and physicians will be able to apply the expertise of the finest medical minds in the world to everyday diagnoses.



The interactive transfer of expertise to an expert system, mediated by an introspective program, TEIRESIAS. (from [Davis 79])

The current research model of expert systems has two components: the knowledge base (what to do) and the inference engine (when to do it). The knowledge base is built in the image of a single expert of the problem domain. The first systems employed a programmer to mediate between the expert and the knowledge acquisition program, but interactive programs have appeared which have the flavor, if not the form, of an electronic book. Randall Davis, the author of one such system called TEIRESIAS, describes the knowledge acquisition as an interactive information transfer: "TEIRESIAS does not attempt to derive knowledge on its own, but instead tries to 'listen' as attentively as possible and comment appropriately, to help the expert augment the knowledge base" [Davis 79]. The human expert challenges the expert system with a problem, evaluates its solution, and then makes any necessary changes to the knowledge base to improve its performance. However, TEIRESIAS also has meta-level representations of the knowledge base, a way of "knowing what it knows", and can challenge the human expert if his modifications are not consistant with existing rule models.



The synthesis of model-based understanding and learning by experience results in a novel feedback loop. (from [Davis 79])

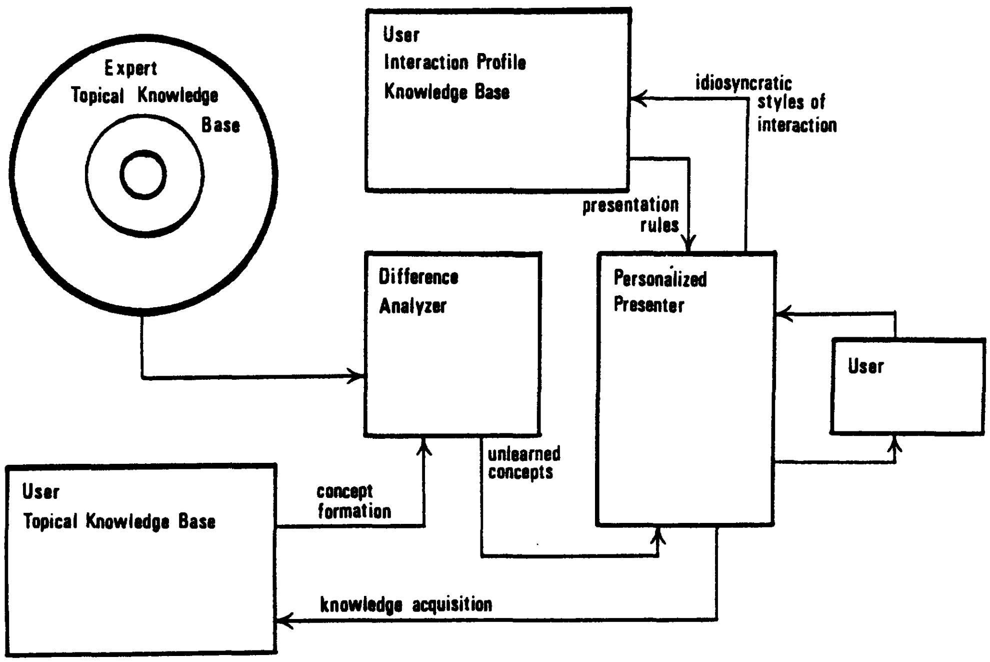
Davis describes TEIRESIAS as a novel synthesis of model-based understanding and learning by experience, and notes that the resulting feedback loop has some interesting implications for the expert system's self-organization. But the synthesis also has far-reaching implications for the future electronic book. It is clear that the goal of TEIRESIAS is to formulate a consistant model of its user's expertise. Let's assume two plausible variations on that goal: first, that a similar system, perhaps distributed among several processors, can maintain several distinct, self-reorganizing knowledge bases; second, that they can accumulate knowledge about different domains of the user's experience, or inexperience.

One of these knowledge bases is the expert system, the electronic book of object-level knowledge on a topic, published on videodisc*. This disc is "plugged in" to the owner's desktop reading station, where there already resides another knowledge base, expert in the idiosyncratic interaction style of the reader. A new knowledge base is created, representing the user's expertise in the problem domain. Initially sparse, it is rapidly enlarged as the reader's common sense and general knowledge is applied to the new domain. A difference engine compares the rule models of the expert system and the reader's knowledge to guide the reader to relevant materials.

---

*One advantage of cleanly separating the inference engine from the knowledge base is, in Davis' words: "...if all of the task-specific knowledge has been kept in the knowledge base, then it should be possible to remove the current knowledge base, "plug in" another, and obtain a performance program for a new task. The explicit division thus offers a degree of domain independence."

A proposed multi-model electronic book.



Presentation rules from the user's interaction profile guide the formatting of materials on the page, and interpret the reader's responses. Because the profile has a model-based understanding of the user, it can disambiguate tentative responses to be consistant with the model. But it can also learn from experience, when the reader challenges its "second guesses" of his intent.

In this system, we have the elements of an electronic book as Socratic dialog, whose aim is the gradual convergence of the user's topical knowledge base to the expert's. It is a rich conversational medium, engaging the reader's attention through all of his faculties.. It is a mentor who comes to know the reader, grows with him, and introduces him to the experts in every field.

Afterword

This document has been my object to think with for almost a year. It is a copybook, in thesis format, which encorporates ideas I've read, discussed with colleagues, and sometimes restated or combined into new ideas.

I thank all of the members of the Architecture Machine Group for their comradeship, their high standards of creativity, and the implicit challenge to produce work of comparable stature.

Andy Lippman has been my thesis advisor, and is director of the movie manuals project, and the group. He has patiently provided me the resources to conduct this work, and the motivation to take it from scribbled notes to demonstrable form.

David Backer has been my partner in research. His imagination, wit, and friendship have been very important to this work.

Jane Nisselson's encouragement and enlightened observations are part of the spirit of this text. Her determination has been an inspiration.

To the designers and engineers of MagicSix, my appreciation. I have never had such great fun hacking a

system. Joseph Goldstone has been a trusted guide through its nether regions, and was instrumental in printing this document. Walter Bender's fine work has provided useful tools and building blocks for the movie manual, and just about every other ArchMach project.

Earlier work and brainstorming with Peggy Weil and Susan Brennan enriched my sense of the possibilities of videodiscs. It was also alot of fun. My methods for realizing those possibilities have profitted from lively discussions with John Correa and Steve Yelick.

My thanks to Dr. Marvin Denicoff and the Office of Naval Research for research support of a style that fosters both productive creativity and personal development.

References

[Backer 82a]
David Backer.
Structures for Interactivity of Media.
PhD Thesis Proposal

[Backer 82b]
David Backer and Steve Gano.
Dynamically alterable videodisc displays.
In *Proceedings, Graphics Interface '82.* National Computer Graphics Association of Canada, 1982.

[Brown 83]
Eric Brown.
Optical publishing.
Master's thesis, MIT, February, 1983.

[Bush 45]
Vannevar Bush.
As we may think.
Atlantic , July, 1945.

[Carpenter 60]
Edmund Carpenter and Marshall McLuhan, editors.
Explorations in communication.
Beacon Press, Boston, 1960.

[Davis 79]
Randall Davis.
Interactive transfer of expertise: acquisition of new inference rules.
Artifical Intelligence 12:121-157, 1979.

[Davis 82]
Randall Davis.
Expert systems: where are we? and where do we go from here?.
AI Lab memo 665, MIT, June, 1982.

[Diringer 68]
David Diringer.
The alphabet.
Funk & Wagnalls, New York, 1968.
Third edition

[Eco 62]
Umberto Eco and G.B. Zorzoli.
A pictorial history of invention.
Weidenfeld and Nicolson, 1962.

[Evans 66]
Jane Evans, editor.
The flowering of the middle ages.
McGraw-Hill Book Company, New York, 1966.

[Goldberg 76]
Adele Goldberg and Alan Kay, editors.
Smalltalk-72 instruction manual.
Technical Report, Xerox Palo Alto Research Center, March, 1976.

[Goldberg 79]
Adele Goldberg and David Robson.
A metaphor for user interface design.
In *Proceedings.* 12th Hawaii International Conference on Systems Sciences, 1979.

[Hadas 54]
Moses Hadas.
Ancilla to classical learning.
Columbia University Press, Morningside Heights, NY, 1954.

[Hamilton 70]
Edward A. Hamilton.
Graphic design for the computer age.
Van Nostrand Reinhold Company, New York, 1970.

[Ingalls 78]
Daniel H.H. Ingalls.
The Smalltalk-76 programming system: design and implementation.
In *Proceedings.* 5th annual Principles of Programming Languages Symposium, 1978.

[Ivins 69]
William M. Ivins, Jr.
Prints and visual communication.
The MIT Press, Cambridge, MA, 1969.

[Kay 77]
Alan Kay.
Microelectronics and the personal computer.
Scientific American 237(3), September, 1977.

[Kay 82]
Alan Kay.
New directions for novice programming in the 1980's.
In P.J.L. Wallis, editor, *10,* Volume 2: *Programming Technology.* Pergamon Infotech, 1982.

[Leclercq 61]
Jean Leclercq, O.S.B.
The love of learning and the desire for God.
Fordham University Press, New York, 1961.

[Lippman 80]
Andrew Lippman.
Movie manuals: personalized cinema as an instructional partner.
Proposal to the Office of Naval Research, MIT, 1980.

[Lippman 82a]
Andrew Lippman and David Backer.
Personalized aids for training: an assault on publishing.
In *Proceedings, 4th annual Conference on Video Learning Systems.* Society for Applied Learning Technology, August, 1982.

[Lippman 82b]
Andrew Lippman.
conversations.

[Liskov 77]
Barbara Liskov, Alan Snyder, Russell Atkinson, and Craig Schaffert.
Abstraction mechanisms in CLU.
Communications of the ACM 20(8), August, 1977.

[McLuhan 61]
Marshall McLuhan.
Inside the five sense sensorium.
Canadian Architect 6(6), June, 1961.

[McLuhan 62]
Marshall McLuhan.
The Gutenberg galaxy: the making of typographic man.
University of Toronto Press, Toronto, 1962.

[Meyrowitz 81]
Norman Meyrowitz and Margaret Moser.
BRUWIN: an adaptable design strategy for window manager/virtual terminal systems.
In *Proceedings.* 8th annual Symposium on Operating Systems Principles, 1981.

[Negroponte 78]
Nicholas Negroponte, Richard Bolt, and Muriel Cooper.
Books without pages.
Proposal to the National Science Foundation

[Negroponte 79a]
Nicholas Negroponte.
Books without pages.
In *Proceedings.* IEEE, 1979.

[Negroponte 79b]
Nicholas Negroponte.
The impact of optical videodiscs on filmmaking.
Technical Report, MIT Architecture Machine Group, June, 1979.

[Negroponte 80]
Nicholas Negroponte, Andrew Lippman, and Richard Bolt.
Transmission of presence.
Proposal to the Cybernetics Technology Office, DARPA

[Negroponte 81]
Nicholas Negroponte.
Media room.
Proceedings of the Society for Information Display 22(2), 1981.

[Ong 67]
Walter J. Ong, S.J.
The presence of the word: some prologemena for cultural and religious history.
Yale University Press, New Haven, 1967.

[Papert 73]
Seymour Papert.
Uses of technology to enhance education.
Memo 298, MIT Aritficial Intelligence Lab, June, 1973.
(LOGO memo no. 8)

[Papert 80]
Seymour Papert.
Mindstorms: children, computers, and powerful ideas.
Basic Books, Inc., New York, 1980.

[Schmandt 80]
Christopher Schmandt.
Soft typography.
In *Information Processing 80.* IFIP, 1980.

[Shortliffe 76]
E. H. Shortliffe.
MYCIN: computer-based consultations in medical therapeutics.
American Elsevier, New York, 1976.

[Yelick 82]
Steve Yelick.
The authoring of optical videodiscs with digital data.
Master's thesis, MIT, June, 1982.