MILESTONE

Computer Orchestrated Asynchronous Sound and Picture Editing

by
David Lyn Kung

B.S., Architecture
Massachusetts Institute of Technology
Cambridge, MA
1993

Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning, in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Media Arts and Sciences
at the
Massachusetts Institute of Technology
June 1995

Signature of Author . . . . . .
Program in Media Arts and Sciences
May 12, 1995

Certified by . . . . . .
Glorianna Davenport
Associate Professor of Media Technology
Program in Media Arts and Sciences
Thesis Supervisor

Accepted by . . . . . .
Stephen A. Benton
Chairperson
Departmental Committee on Graduate Students
Program in Media Arts and Sciences

MILESTONE
Computer Orchestrated Asynchronous Sound and Picture Editing

by
David Lyn Kung

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning on May 12, 1995
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Media Arts and Sciences

Abstract:

Previous research in digital video orchestration has demonstrated the computer's ability to edit sequences based upon a cooperation between video databases and encoded story knowledge. These systems, however, manipulate sound and picture as a single unit. The next generation digital video systems will need to combine the encoded story knowledge with the ability to manipulate audio and video separately. This thesis creates the infrastructure to support computer orchestrated editing in which sound and picture are treated as separate, complementary tracks.

MILESTONE is comprised of two modules. A video annotation component which builds upon previous research in video databases provides the means to create a knowledge representation. A robust set of editing functions provide editing functionality. Potential applications of and arguments for similar applications to MILESTONE are discussed in light of theoretical inquiries into hypertext, cinematic narration, and video orchestration.

Thesis Supervisor: Glorianna Davenport
Title: Associate Professor of Media Technology
This work was supported by the News in the Future Consortium

MILESTONE
Computer Orchestrated Asynchronous Sound and Picture Editing

by
David Lyn Kung

The following people served as readers for this thesis:

Reader . . . . . .
Hans Peter Brøndmo
Former Director of Interactive Technologies
AVID Technology, Inc.

Reader . . . . . .
Henry Jenkins
Professor, Director of Film and Media Studies
MIT Film and Media Studies Program

ACKNOWLEDGMENTS:

This document is a result of a lifetime of collaborations and friendships with colleagues, teachers and friends here and abroad. Thanks and respect to all.

Thanks first and foremost to the Interactive Cinema Group, past and present. Glorianna, for a quarter of my life, you have been an inspiration, mentor and friend. As for the rest of the IC gang, the exact reasons for the following thanks are too long and too convoluted to discuss here, but these thoughts come to mind: Thanks to Lasky for inspiration, Bruckman for mountain screams, Hans Peter for Mimato(which is how this whole thing got started), Liza for conquering fears, Natalio for a late night revelation on engineering, Eddie the "BullyMan" for not seeing me, Thomas for convincing me to drop a class, Stephan for helping to get cable, Mark for putting up with me, Ryan for Petes, Betsy for chips and for being so goshdarn awesome, Kevin for his insights and experience, Gilberte for maturity, Scott for taping my triumph on NBA Jam, Dave for *Speaking Parts*, Lee for Indy, Tag for his far from dogmatic character, Mike for late night walks and conversations, and all the UROPs and affiliates who let me be a MacNomad on their machines.

Thanks as well to my friends and colleagues around the lab, namely John and Wendy, Ravi, Robin, Freeze-Frames, Linda, Santina, and the old tech service gang, BK, Greg, Stuart, Josh, and Ben.

Thanks to my readers, Hans Peter Brøndmo and Henry Jenkins, for your keen insights and criticisms.

Respects and debts to the gang back in Carbondale, IMSA, MIT, NYU, Monaco, and everyone everywhere in between, namely, John, John, Steve, Andy, Andy, Tony, Jen, Karen, Kelly, Dave, Emma, Babs, Terri, Stark and Slaney, Joe, FJ, Colonel Sanders and BB, Jill, Golan, James, Sharon, Gene, "Beb" Bradley, 3W, Jody, Drew, Gary, Gerald the "G-Man" Belkin, Madame Belkin and extended family, and all the teachers and makers.

Thanks and love to Mom, Dad, Sam and Jon.

TABLE OF CONTENTS

PART ONE: INTRODUCTIONS

1. INTRODUCTION

The partnership between computers and cinema bridges many traditions, art forms and professions. Computers create complex visual effects like the "morph," which cannot be created in any other media. Screenwriters compose screenplays on word processors. Soundtracks are recorded and manipulated digitally. Digital video has enabled true nonlinear desktop video editing systems. Do these systems have any real knowledge of cinema? In order for computers and cinema to become more intimate partners in storytelling ventures, computers must gain a greater understanding of the fundamental traits of narrative cinema and how it narrates.

Editing(montage) is a fundamental cinematic technique which has evolved over the last 100 years. The ghosts of Bazin and Eisenstein can debate forever whether or not editing is the essence of cinema, but it is true that editing was born from, and is a unique child of, cinema. Editing compresses or expands time, builds virtual spaces which do not exist: editing creates meaning. Editing is a subset of the greater language of cinema. If we can encode the language of cinema into a computer language, imagine the possibilities! It will be the start of a beautiful friendship.

If we define a shot as a continuous strip of film, then we can define editing as the joining together of two shots. Textbooks on film and filmmaking often describe editing in a similar fashion. This definition ignores the fact that, since the 1930's, editing often happens "on top of" audio. In fact, the dialogue, music, or ambient sounds on the audio track may dictate how the picture track can be edited. Furthermore, the audio track is usually edited as well. In order to encode an editor's knowledge into a computer editing system, the editing system must have all the capabilities of a film editor. Thus far, almost all research in video orchestration has edited sound and picture synchronously. A computer editing system which only joins shots remains a powerful tool, but it falls short of replicating the full functionality of an editor. Asynchronous sound and picture editing places another tool in the hands of the computer orchestrator and provides, if not the knowledge, the capability to edit more flexibly.

Commercial applications such as Adobe Premiere or AVID VideoShop allow for asynchronous editing of picture and sound. These applications, however, lack orchestration abilities or any narrative or cinematic knowledge. The desktop human editor retains all knowledge or reason behind the application's tasks. Narrative knowledge can aid the editor's task of organizing footage or may even provide templates to allow an editor to generate multiple iterations of an edit without having to explicitly create each one. Cinematic knowledge is knowledge of cinematic narration, how cinema tells stories. This knowledge is the most valued knowledge an editor has. This knowledge is required for an editing system to take its knowledge of narrative and express it within the constraints of a given set of footage. A computer editing system, with at least some knowledge of story and cinematic techniques, greatly empowers a user and may help a fully computer-realized synthetic cinema to evolve.

In much the same way electronic text made hypertext a reality and created new forms of interactive fiction, as cinema transitions to the digital domain and departs from its celluloid medium, new ways of experiencing movies become possible besides watching the projected images of celluloid frames. Hypertext and cinema will share the same medium — the computer. Hypertext fictions may no longer be stories told strictly through the language of text, they may become stories told through the language of cinema. Such mergers will require tools, video orchestrators which can provide the infrastructure to deliver and tell such stories. These orchestrators will require knowledge: knowledge of stories and how to tell those stories cinematically.

1.1 Overview

This thesis tells two stories. One story concerns cinema, its past, present, and potential future. The other story is the story of a tool, MILESTONE. This thesis is structured in three sections. The first four chapters comprise an introduction to salient issues which influenced the development of MILESTONE. Chapters five and six pursue a theoretical analysis of hypertext, interactive narration, and their impact on and potential for cinema. The remaining chapters discuss the implementation of MILESTONE and potential applications in regard to the previous two sections.

Chapter One: Introduction
The document begins with an introduction to relevant concepts.

Chapter Two: Approach
The MILESTONE system is briefly described through analogy.

Chapter Three: Background
The history of cinema and the development of editing tools are investigated. An overview of related research provides a glimpse at the current research which has influenced MILESTONE.

Chapter Four: Video Orchestration
Video orchestration is an editing strategy only possible through computing. The pros and cons of orchestration are considered in light of current editing systems.

Chapter Five: Interactive Cinema: Hype or Text?
What does hypertext offer cinema? Hypertext and potential mergers between hypertext and cinema are explored as are their reliance upon the video orchestration techniques discussed in chapter four.

Chapter Six: Interactive Narration
Interactive narration posits a form of interaction which has been assumed or ignored. Issues regarding cinematic narration and what roles interactivity may play in them are discussed.

Chapter Seven: MILESTONE
The MILESTONE system is described and discussed.

Chapter Eight: HyperActivity
Hypertext and interactive narration are revisited in regard to MILESTONE.

Chapter Nine: Conclusion
The document ends with a look at future research and a summary of key points.

Chapter 10: Bibliography

2. APPROACH

MILESTONE is comprised of two modules. Video annotation forms a representation of content. This module builds upon a foundation of previous research in video databases. A set of editing functions provide the basic editing functionality of the MILESTONE system. These modules furnish the requisite components to build an intelligence module. This module may apply expert knowledge about editing structures and techniques to match story goals to specific edit type. To help elucidate the modules and their functions, the following spreadsheet analogy is provided. Although a spreadsheet may be a convenient way to describe MILESTONE, it should not be assumed MILESTONE looks or functions exactly like a spreadsheet. What follows is strictly an analogy.

2.1 Rows and Columns

In a spreadsheet, the user attributes meaning to rows or columns and then fills in numbers accordingly. The organizational structure of a spreadsheet provides a means to describe the numbers. It creates a computer representation for numbers — for content.

The same process applies to video. Before a computer can manipulate video, it must "know" what video is. A computer representation provides a way to describe video to the computer. The representation may take the form of a name or perhaps an object centered approach as in computer programming. Whatever the form, the purpose remains the same; provide a structure to describe the content.

2.2 Adding and Subtracting

Numbers in a spreadsheet provide only a limited knowledge base; after all, we put numbers into spreadsheets so we can manipulate them. We manipulate numbers using elementary mathematical operators like addition, subtraction, division and multiplication. We manipulate cinema using elementary editing operators like straight cuts, cutaways, video overlays and L-cuts.

Digital video affords the ability to manipulate picture and sound in much the same way a word processor enables a user to manipulate text. Picture and sound

can be cut, copied, pasted and altered in any number of imaginable ways. Of course, cutting and pasting alone does not an edit make. The edits must occur across a strict timeline to maintain continuity and synch. Just because we wish to edit sound and picture asynchronously does not mean our sound and picture will always be asynchronous.

2.3 Equations

Numbers or mathematical operators are not what make spreadsheets valuable. The equations which manipulate the numbers and mathematical functions are what make spreadsheets valuable. Equations create new content.

Just as we can build a representation of content, we can also build a representation of knowledge. Previous research in artificial intelligence provides means to do so. In particular, MILESTONE can be used to employ techniques borrowed from expert systems and story understanding/generation to encode cinematic and story knowledge. This knowledge can drive a storytelling system, can create new content.

3. BACKGROUND

MILESTONE gathers inspiration from a diverse collection of related histories. An examination of these histories situates MILESTONE within an evolution of cinematic storytelling techniques and technologies and constructs a foundation upon which to view and prepare for the future.

3.1 Cinema, Editing and Sound

We sit in darkness, eagerly anticipating a shaft of light which will spear the darkness of a theater. Flashing patterns of colored light splash across a screen and tell us a story. Editing is the means, and some would say process, by which we understand much of what we see on the screen. Editing constructs story.

D.W. Griffith is credited by many film historians with establishing cinema as a narrative form. Griffith, however, did not invent editing; historians document editing as a narrative apparatus as early as 1903. Although he may not have invented the fundamental techniques of editing — constructing sequences by intercutting shots of different framings and lengths — Griffith honed the use of these techniques to tell stories. Through his body of work, Griffith evolved the cinematic language like no filmmaker before him. Accelerated montage, parallel editing, and use of close-ups for emotional effect, represent only a few of the many methods Griffith borrowed, expanded, and refined to increase the ways in which cinema could narrate.

Contrary to their name, the experience of watching a silent films was far from silent. Films were often accompanied by live musical accompaniment, narration, or sound-effect machines. Musical scores were composed solely for films as early as 1907. Although sound accompanied films during projection, these sounds were not captured at the same time as the images. Technology had yet to be invented to allow for synchronous recording and playback of sound and picture.

The Vitaphone was the first system featuring synchronized playback to be used commercially by a large Hollywood studio. Although the first Vitaphone film, *Don Juan* (1926), succeeded commercially, studios declined to immediately

transition to sound films. Hollywood's trepidation grew from an uncertainty as to whether Vitaphone films represented a passing fancy or a legitimate call to action. Furthermore, economic factors, and the lack of a firm technological standard to support sound, prevented the Vitaphone from immediately becoming an industry standard. Perhaps more importantly, sound films threatened the entire film industry. All stages of production and distribution, acting, shooting, projecting, etc., required substantial technical overhauls to allow for sound. Hollywood feared it could not afford to convert to sound films. The economic costs were staggering, total costs amounting to approximately $300,000,000. (Cook, 243) When *The Jazz Singer*(1927) and *Lights of New York*(1928) proved enormously successful, Hollywood realized it could not afford not to convert.

Film scholar Thomas Elsaesser has argued that the continuity system of editing solved the problem of maintaining narrative and cognitive momentum across shots of varying spatial and temporal natures. (Elsaesser, 293) The expressive capabilities of sound solved the problem of maintaining this cognitive cohesion by purely visual means. Sound allowed dialogue, music, ambient noise, etc., to contribute to the storytelling process. Dialogue rendered "title cards" for dialogue unnecessary, allowing scenes to play without constant interruption. No longer burdened to act singularly, the visual and editorial components of cinema were free to cooperate with audio in expressive techniques and to create new forms of expression previously undiscovered or unexplored.

The symbiotic relationship between picture and sound, however, came to fruition slowly. Technological limitations regarding sound recording initially hampered cinematic techniques. Immobile cameras and largely immobile microphones meant predominantly immobile scenes. Editing suffered in sound-on-disc systems because it was impossible to edit the disc. (Weis, 386) Even when Hollywood adopted sound-on-film systems, a single audio track governed the film. Studios shot scenes from four cameras to provide different angles and coverage for a scene, but this footage was synchronized and locked to this single audio track. Lighting technicians had to light for four instead of one camera and subsequently forfeited more expressive lighting for more practical lighting. Synchronized picture and sound also prevented cameramen from dynamically

changing camera speed. (Dickinson, 55) These prohibitive factors and the public's fascination with speech threatened to return cinema to its origins — "canned theater." (Ellis, 165)

Technical, artistic and individual efforts pushed the medium forward, integrating sound into its storytelling repertoire. Mobility held one of the keys to growth. Portable sound-proof camera casings allowed the camera to move more flexibly without impacting the soundtrack. Booms followed actors, capturing speech as they moved about the frame. Sound-on-film recording became a standard. Picture and sound were recorded and edited on separate rolls of film. Once audio mixers, which could integrate multiple audio tracks into a single track, gained use, filmmakers realized they had ultimate control over editing of picture and sound asynchronously.

Editing techniques promoting narrative and built around sound now grew into practice. Although editors had previously edited around titles, now they had to edit around dialogue and music. Filmmakers applied dramatic tactics which had formerly been used to control the dramatic flow of chases, to control the dramatic flow of dialogues. The thirties saw the rise in the number of reverse-angle shots, as well as refined methods for editing dialogue. Use of the "dialogue cutting point" — a technique named by Barry Salt to describe how a cut from one speaker to another could be concealed better by cutting on the last syllable of a sentence or word — rose to prominence as editors explored how to make cuts in a dialogue as transparent as possible. (Salt, 285)

As cinematic technology in visual and audio recording developed, film theorists debated and anticipated the impact of sound upon cinema. A recurrent debate regarded synchronous versus asynchronous sound. Most theorists, Eisenstein, Pudovkin, and Clair, among them, praised asynchronous editing of sound and picture. (Weis, 76-77) They recognized that cinema did more than represent reality. Cinema, through editing, constructs realities.

3.2 Editing Tools

Digital nonlinear editing systems represent the state of the art in motion picture editing technology: the cutting edge, so to speak. Nonlinear editing, however, is

not a new concept; film editing was always non-linear. (Ohanian, 21) It would be a mistake, however, to say the only advancement editing technology has made in almost a century is becoming digital.

Video editing began as a non-linear process. Strips of magnetic tape were spliced and cemented in a similar process to film. Editing video in this fashion continued into the 1970s. The major revolution in video editing came about when editing became an electronic, not a mechanical process. A major player in this revolution was timecode, a signal which recorded every frame on a videotape in hours, minutes, seconds and frames. The Society of Motion Picture and Television Engineers (SMPTE) standardized timecode in 1972.

By providing a time-based numerical reference to video, timecode allowed video decks to rapidly search to precise frames. The mechanics of editing could be automated. Timecode also allowed the process of editing to be recorded and edited. Engineers programmed computers to control video decks, allowing editors to create edit decision lists (EDLs). EDL's consisted of records of the edit points(in/out points on the source/record decks), which audio and/or video tracks to toggle on or off, and the source tapes needed to (re)create an editing session. By storing all the edit points of an editing session, EDLs gave editors the luxury of modifying single or multiple edit points and having these changes applied towards the building of new edits. Except for switching source tapes when required, the mechanical portion of the editing process could be largely automated. Editors could go out to lunch while an edit was being rerecorded. It saved time.

Time saving is a valued commodity in editing systems because a great deal of editing time is spent fast forwarding and rewinding a tape to the correct shot. A nonlinear system saves even more time because less time is spent waiting for a shot to be available. At first, laserdisc systems were experimented with, but they proved too cumbersome and expensive. Like many inventions, the idea existed before the technology did. Thus, the first successful nonlinear editing systems simply multiplied the number of source desks, shortening search time. Editing systems like Montage, Ediflex and Touchvision used different techniques (touchscreens, screenplay based interfaces, etc.) but the underlying principles

were the same. Instead of manipulating shots which exist on different strips of film, these nonlinear editing systems manipulate shots existing on a battery of decks. Once laserdisc became a viable option, these editing systems could integrate the advantages of random access video, saving even more time.

Digital nonlinear editing systems digitize and edit sound and picture. Digital video can be stored on, and randomly accessed from, a hard drive or magnetic/optical media. Digital nonlinear systems eliminated the need for multiple source decks (but not the need for multiple hard drives). Editors now have more control and time to edit a sequence in different ways. Digital systems take advantage of the computer's ability to process, recall, and filter information rapidly. They do not create more options, they allow more options by saving time and granting greater flexibility.

3.3 Editing Intelligences: Review of Previous Research

The story told above is in many ways the story of a dumb tool. To date, digital systems utilize a computer's processing power to render visual effects or to handle list management activities, but not a computer's ability to model concepts. They save an editor a great amount of time by facilitating much of the time consuming manual labor of editing. Can an editing system save an editor time by facilitating the intellectual labor of editing? What kinds of intelligence can an editing system have?

3.3.1 Artificial Intelligence

The following examples of research illustrate a cross-section of research in intelligent editing systems. The underlying thoughts behind these projects originates in artificial intelligence(AI). Artificial intelligence research regarding story involve two key questions: How do we understand stories? How do we create stories? Roger Schank's work in human knowledge structures provides an interesting point from which to begin. His conceptual dependency theory posited a means to represent the meaning of sentences and to embed causal relationships.(Schank, 11) Scripts and plans were used to model knowledge and how we apply knowledge in similar or analogous contexts. These story models function as templates from which to examine different stories. These templates may be employed to summarize or to answer questions about stories.

While Schank's work utilized story models to answer questions about stories, story models may also be utilized to generate stories. Story generation can be considered in different ways. Meehan's TALESPIN involved a simulated world containing characters, goals and interactions. Stories were generated when characters sought to achieve their goals and interacted with different characters and the world.(Schank, 210) Story planners like Lebowitz's UNIVERSE project treat characters in a more arbitrary fashion. Relations between plot points are the dominant narrative construction form. The planner is driven more by story than the characters which inhabit it. (Lebowitz, 3) In 1972, George Lakoff constructed a story grammar based upon a morphology of Russian folktales compiled by Vladimir Propp. (Lakoff, 131) Although Lakoff's grammar was an exploration in linguistics and not artificial intelligence, it is not difficult to imagine implementing a story grammar in code.

Intelligent editing systems may differ depending upon which method they use to model stories. The following experiments, many of which originate from the Interactive Cinema Group of the MIT Media Lab, each tackled one problem, took one step forward towards applying intelligence to editing systems. They all share a common approach. The video is annotated. These annotations provide a way to sequence video, not by sequencing actual video, but by sequencing the annotations used to describe it.

3.3.1.1 *Constraint-Based Cinematic Editing*: Ben Rubin, 1989
Rubin produced an evocative narrative shot in multiple iterations and in multiple ways. For example, some iterations portrayed a scene from different points of view, framings, or character emphases. After annotating each shot with its unique characteristics, Rubin placed plot constraints upon the available video. He could specify qualities in the film such as duration, character emphasis, pacing or narrative significance.

The editing process consisted of three agents. Because many shots represented the same action, the "butcher" chooses the one shot which best matches the given constraints. The "tailor" maintains an overall duration constraint by eliminating shots rated less narratively significant than others. The "doctor" locates and

resolves edits which may be "uncinematic" such as jump cuts. The system then iterates through the tailor and doctor as many times as necessary to resolve any conflicts they incur upon each other and to satisfy all the constraints.

3.3.1.2 *Electronic Scrapbook*: Amy Bruckman, 1991
Bruckman proposed a media form, The Electronic Scrapbook — "an environment designed to encourage people to use home video as a creative medium."(Bruckman, 2) While the interface resembles a family scrapbook, the underlying system resembles a database. The user creates a knowledge representation of each scene, annotating characteristics such as who, what, when, and where. By specifying and combining salient annotations such as the age or actions of characters in the video, the user constructs abstract story models which organize the content into recognizable narratives. The Electronic Scrapbook displays the output of a story model on a scrapbook page for independent or sequential playout.

3.3.1.3 *Homer*: Lee Morgenroth, 1992
Morgenroth designed Homer to generate video stories. Homer generates stories by applying story models to logged video. Story models consist of blocks which corresponded to annotations. Stories are then generated by matching clips from the database against the descriptions requested in the story models.

3.3.1.4 *IDIC*: Marc Davis and Warren Sack, 1993
IDIC generated *Star Trek: The Next Generation* trailers. The trailers were created by a planner based upon Alan Newell and Herbert Simon's General Problem Solver(GPS). GPS based planners hold significance for storytelling because narrative structures can be embedded within the causal relations required for means-end analysis (Norvig, 111).

3.3.1.5 *LogBoy And FilterGirl*: Ryan Evans, 1994
Evans created a bipartite toolkit. LogBoy allows a user to graphically attach annotations to video. FilterGirl orchestrates the sequencing of video by allowing the user to create and revise playout constraints called filters. Filters can change over time through user interaction or dependent structures between filters. As a result, FilterGirl accommodates passive and interactive playout in real-time.

Together, LogBoy and FilterGirl represent a strong foundation for building movies whose playout could vary — multivariant movies.

3.4 Once and Future Thinkings: Review of Related Research

The following research projects are not editing systems per se, but they do offer new and exciting contexts within which to consider the future of cinema.

3.4.1 *Anthropologist's Video Notebook:* Thomas Aguierre Smith, 1992

In creating the Anthropologist's Video Notebook, Smith created a system of video annotation called the Stratification system. This system made use of stream-based annotations. This annotation method treats annotations as objects which have a temporal nature and may be layered on top of each other across time.

3.4.2 *Intelligent Camera Control for Graphical Environments*: Steve Drucker, 1994

Dr. Drucker's work examined computer generated environments. While a great deal of research considered how to move a virtual camera, Drucker's thesis asked, "Why move the virtual camera?" The answers Drucker found were then applied to creating new methods for virtual camera control. Asking why and how to move a virtual camera reveal important lessons for exploring the future of cinematic expression, but, in the context of editing, we should also ask "When do we move the camera?

In cinema, an edit is the joining of two shots. In computer generated synthetic cinema, an edit becomes something different. One equivalent of an edit in synthetic cinema is an instantaneous change in camera position and focal length. Consider the following scenario: You have a computer generated scene of a man throwing a dart into a dart board. You want to "cut" to the dart as it impacts the dart board. The edit happens when we change the camera location. We are not joining shots.

3.4.3 *Media Streams:* Marc Davis, 1995

Media Streams builds upon an exhaustive library of icons to create an iconic visual language for annotating videos. Icons, representing different annotations

such as action or setting served as stream-based annotations. The icons could also be used as a retrieval tool.

3.4.4 *Composition and Search with a Video Algebra:* Weiss, Duda & Gifford, 1995
Video Algebra features a range of video algebra operations used to retrieve and compose video presentations. The expressions of particular interest involve the use of conditional and logic operations(union, concatenation) as a means to organize and manipulate video within a database. While some of the algebraic functions share similar responsibilities as the filters developed by Evans, the algebraic video system also provides a structure to support multiple streams of video and offers a more complete repertoire of navigation and retrieval tools.

3.4.5 *CHEOPS*: MIT Media Lab, work in progress
While Drucker's work examined computer generated environments, the Media Laboratory's Cheops system creates computer orchestrated environments. Cheops' platform supports structured video, "the coding of an image sequence by describing it in terms of components that make up the scene."(Granger, 9) Imagine a paper dollhouse where each element, from Holly Hobby, to a table, to the house itself, can be manipulated and composited separately. In this domain, the concept of editing can change again. An edit may no longer involve where the camera is, when we see and what we are looking at, but what elements are placed in front of the camera at a moment in time. In this case, as with Drucker's work, the process of editing may differ, but the role of editing does not. An understanding of how editing functions will drive these systems and tell them what to do. Furthermore, Cheops deals with sound and picture as separate components which must be combined in playout. Knowledge of how to orchestrate sound and picture asynchronously is required.

3.4.6 *The Media Bank*: MIT Media Lab, work in progress
The Media Bank provides a vision of what future on-line multimedia databases will be. "The bank is a distributed, evolutionary storage environment for audiovisual material accessed both sequentially and interactively."(Lippman, 1) The Media Bank has a four-tiered structure which interacts with, and is hidden from the user's interaction. The application layer concerns itself with how to manipulate elements of the database. The format layer navigates varying data

formats and negotiates between them. The transport layer is then responsible for shipping information back and forth between the user and the database. The content layer is the knowledge representation which supports the Media Bank. How does the Media Bank know what applications to use or even what elements to manipulate within an application? The content layer must know what elements(sound, picture, text, animation, etc.) are available and how to manipulate them synchronously or asynchronously. The content layer must have the brains of an editor.

3.5 Seedy Roms: The Commercial Market

I have noticed a number of CD-ROM titles which advertise themselves as Virtual Cinema or interactive movies. One way or another, these products evoke a cinematic experience. What frustrates me about these systems is that they are often hypertext or branch structured experiences where the user may choose what sequence to see, but not how the sequence may be seen. Cinematic knowledge may have been used to create sequences used in these products, but these systems have little or no cinematic knowledge of their own. Cinematic techniques such as the close-up or montage may be shown, but they are not applied.

3.6 Conclusion: Atomic Video

A disturbing trend I detect in some of the academic research and commercial ventures I have encountered is an atomic model of video. Atom derives from the Greek word *atomos*, meaning indivisible. Almost all of the systems discussed model video as an indivisible union of picture and sound. This model is wrong. History has shown us that the cinematic technique of the early sound era did not fully mature until the tools were available to edit sound and picture asynchronously. No matter how much cinematic knowledge you encode, you can only go so far in cinematic technique if you can only edit sound and picture synchronously. The atomic age must end.

4. VIDEO ORCHESTRATION

The spreadsheet analogy introduced in chapter two basically describes a video orchestration system. A video orchestrator shares qualities with video editing systems, but there exist important differences between the two tools which shed light on their applications. A video orchestrator does edit, in the traditional sense that it joins shots; but it should not be used(at least not yet) as a general purpose video editor. Conventional editing systems, however, lack any orchestration abilities whatsoever. The following discussion will briefly trace the evolution of video orchestration and its applications, and will investigate the pros and cons of video orchestration and video editing systems.

4.1 The DMO

MILESTONE continues the work begun with the Digital Micromovie Orchestrator, also known as the DMO(Davenport, Evans, Halliday, 1993). The DMO consists of two parts: a video database and a layered filter structure. The knowledge representation resides in the video database. "Sketchy" descriptions annotate the video. As the name implies, sketchy descriptions provide a sketch of the video. If a picture is worth a thousand words, a video clip is worth much, much more, and no one would like to sit down and annotate each one. Sketchy descriptions provide the minimum amount of description necessary for a particular use. The filters, in this case, determine the particular use of the descriptions. The layered filter structure of the DMO acts as a query mechanism into the database. It sifts through the database and filters out materials, returning a subset of clips matching a group of descriptions. The two parts of the DMO, the database and the filter structure, later evolved into Ryan Evans' thesis, "LogBoy Meets FilterGirl."

4.2 Multivariant Playout

Evans created his video orchestration system for "multivariant playout." "Multivariant playout is a term which describes movies which can present a different sequence of shots each time they are viewed."(Evans, 12) Because a filter returns a subset of the whole database matching a particular set of descriptions, the probability that a subset contains more than one clip is quite high. In these cases, the system randomly chooses one clip of the subset. This

feature alone creates the potential for multivariance. This feature alone, however, fails to realize the full potential of multivariance. A purely random structure lacks flexibility because there exists no sense of constraints between clips, no sense that the choice(random or not) of one clip may effect the potential clip selection further down the line. Furthermore, we should consider multivariance in the context of interaction.

Although a video orchestrator can create multivariant movies, the key phrase to remember is multivariant playout. The original DMO and the LogBoy/FilterGirl systems can operate "in real-time," editing clips "on the fly." "Look before you leap" nicely sums up the basic way these systems operate. After the current clip plays, the video orchestrator checks the current "state." It may monitor a global variable tied to some interaction and determine future playout based upon the value stored. For example, imagine a "sex and violence" dial on your television which, by turning the dial up and down, would allow you to dictate how much or how little sexual and violent content you or those around you may witness. The video orchestrator can monitor the value of that dial so during the climactic nude gun battle, more or less appropriate clips may play. Human interaction may dictate future playout, but the clips previously shown may dictate playout as well. Just as the orchestrator can monitor some external device, the orchestrator can query previous clips for their descriptions. The choice regarding which clip to show next may be determined by those descriptions.

4.3 Multivariant EDL's

Reel	Source In:	Source Out:
001	01:34:21:04	01:35:01:23
001	01:35:14:15	01:35:45:03
001	01:02:56:12	01:02:58:10
001	01:54:02:22	01:54:12:12
001	01:21:34:08	01:21:42:08
002	02:31:15:14	02:32:05:29
002	02:14:53:12	02:14:55:12
002	02:58:12:11	02:58:18:01
002	02:12:21:15	02:12:27:14

Shot 1	Framing: Establishing Shot Exterior Shot Subject: Diner
Shot 2	Framing: Establishing Shot Interior Shot Subject: Diner
Shot 3	Framing: Medium Shot Interior Shot Subject: Diner Cashier

Figure 4.1 - Sample Traditional EDL(left) and Orchestration EDL(right)

While the previous discussion focused upon multivariant playout, primarily for potential interactive applications, video orchestrators can also create a new kind of edit decision list for non-real-time, non-interactive presentations. Exactly what kind of new kind of edit decision list does a video orchestrator present? On a very basic level, edit decision lists consist of an index of in and out points. A video editing system then takes these numbers and edits a sequence together automatically. Video orchestration works in much the same way, except instead of using numbers to dictate the editing, descriptions dictate the editing. Instead of a system editing a video sequence based upon a set of time code indices, video orchestrators edit a video sequence based upon a set of descriptions. The following scenarios suggest ways in which video orchestration applications can facilitate the use and creation of playout and editing tools.

4.4 Postproduction

Annotations and filter structures do not require content. An entire knowledge representation and complex filter structure may exist before production moves into full swing. To a lesser extent, this structure exists today. In preproduction, storyboards, shot lists, and shooting scripts dictate the general outline of the editing. Editors often generate a rough edit based upon this information. A video orchestrator can easily assemble a rough edit almost automatically. Video orchestrators "know" what kinds of material a defined sequence requires. Conventional editing systems have no knowledge whatsoever.

4.5 Cookie Cutter Content

A filter structure functions at an abstract level. For example, a user may design the orchestrator to create sequences consisting of beginning, middle, and end clips. While this example is perhaps too abstract, it demonstrates that orchestrators can use abstractions as a template, and can edit multiple iterations of content when the content has a recognizable and repeatable form and structure. These templates are particularly useful when many producers are creating the same type of content. Has anyone noticed how the in-flight video of flying safety precautions looks pretty much the same no matter which airline is showing it? Most film and video productions involve creating industrial and training films and not big budget Hollywood narratives. Because these films do not have to contend with many of the constraints and complications of creating a

particular story and narrative realm, they may be better suited to make use of abstract templates as guides in the editing process.

4.6 Personal Ads

Electronic media is personalizable, or rather, can be personalized. The sex and violence dials described before portray one example of personalized media, one in which the consumer has control over the content. The consumer personalized the content for a particular use. The provider, however, can also personalize content for the consumer.

Whether named personalized ads or interactive ads, advertising which more accurately hits its desired audience will be more effective. Personalized ads already exist to some extent. How often have you received an envelope which boasts in big letters, "(YOUR NAME HERE), you may have already won $10,000,000!!!!!" Television stations broadcast commercials for toys and breakfast cereals during cartoons, while advertisements for shoes and athletic clothes are shown during sporting events. Nielsen ratings, "sweeps" and network programming all factor into a complex equation by which advertisers choose when and on which channel to buy time to show their commercials. The upcoming digital revolution promises many things: five hundred channels, video on demand, home shopping, etc. Video orchestration offers advertising new opportunities and approaches to targeting an audience which may not watch the same shows, or the same kind of shows, or the same kind of the same show, at the same time.

Many ads take a very simple content model, an abstraction of which may look something like the following: shots of people having fun with the product; a product shot; more shots of people having more fun with the product; another product shot; even more shots of people having even more fun with the product; and a final product shot usually sporting a fancy graphic and byline. Usually these shots are accompanied with music of some kind and voice-over narration. The video orchestrator may know the individual watching is a Caucasian male in his early 20's whose shopping reports indicate he usually buys basketball shoes. The orchestrator can then use the same content model but may present clips of Caucasians, and/or Caucasian males, and/or Caucasian male twentysomethings

having fun wearing basketball shoes. An Asian female interested in aerobics may see the same content model, and her instance of the advertisement may feature Asian women wearing aerobic shoes. If the video orchestrator does not know the exact demographics of the individual watching, it will probably know the demographics of people who tend to watch a particular program and can tweak the advertisement accordingly. While this example demonstrates a particularly insidious model for personalized media, this example demonstrates the potential of video orchestration and points towards applications in the realm of sports, news, movies and other fields.

4.7 A Picture is Worth a Thousand Bytes

Multivariance does not require an orchestration system, but orchestration systems facilitate the process of organizing and presenting multivariant films. Furthermore, orchestration saves memory. A video orchestrator does not require all of the different versions of a film, rather, a video orchestrator requires all the elements needed to generate all of the different versions of a film. For example, sometimes a released film may enjoy a release of "the original director's cut." A more common occurrence happens when television networks edit out or dub scenes featuring potentially offensive imagery and language. In these cases, at least two versions of a film may exist and have to be stored somewhere in their entirety. Already, the required storage has doubled and this example presents a very simple scenario. Storage can quickly become monopolized by all the different versions of only one film!

4.8 Critic's Corner

Despite all of this generous hype, video orchestrators do have their disadvantages. One of the main disadvantages originates from one of video orchestration's primary features — sketchy descriptions. Because sketchy descriptions provide the minimum amount of description for a particular domain, sketchy descriptions will tend to fail when applied outside of their domain. A domain may have some success with sketchy descriptions from another, similar domain, but because the descriptions are so closely tied to a unique instance, potential conflicts in annotation and retrieval may present more trouble than it is worth. Sketchy descriptions also fail miserably when applied to the domain of a general purpose video database. This makes sense because the

whole point of sketchy descriptions in the first place was to avoid the problematic issues involved in annotating video for a general purpose video database. Sketchy descriptions avoided the problem, they did not solve it.

Content should not be retrofitted for multivariant orchestration. Any footage cannot instantly be made multivariant. Annotations and filter structures should be created before and/or during the creation of content. Video orchestration is highly authored, as highly authored as a traditional non-variant non-interactive feature. A video orchestrator makes meaningful sequences because at every level, it is designed to do so. The maker ensures that when edited together, the scenes will work. Two jobs must occupy the mind of the maker, organizing the content, and transferring that knowledge to the computer.

4.9 Orchestration vs. Composition

The thought processes concerning an editor flow between two poles, a continuum between conceptual and continuity editing. This continuum exists in writing as well as in film. In writing, we often outline our thoughts. We are concerned with the overall structure and flow of ideas. How does one idea flow into another? Should one point be realized now or later? Once we have a structure, we turn towards issues on a sentence level. Was that spelled right? Is the grammar correct? Interactions at either end has some impact on the other. In cinema, conceptual editing involves the ordering of sequences. The continuity aspect deals with the continuity system of classical Hollywood cinema. Does this scene look and feel right, do my matches work, are the cuts good? Video orchestrators and video editing systems stand on opposite sides of this continuum.

Perhaps the best distinction to make is that video orchestrators orchestrate sequences, while video editing systems edit shots. We use editing systems when composing video, when we do not know for sure what the final sequence may look like or how it may turn out. Orchestration assumes this knowledge is already in place, and applies it. Thus, video orchestration works best when applied to an abstract content model, a level at which orchestration of sequences works just fine. Because they edit sequences, video orchestrators(not that there have been that many) lack fine editing tools. There is no audio scrubber, no

fancy transitions, no sliding of multiple audio, video and effect tracks. On the other hand, video editing systems have no video databases, no annotation systems, and no filter systems. The point here really is not to critique orchestration and editing and debate the benefits of one or the other. Tools and techniques for orchestration and editing should not stand in opposition to one another; they should work in cooperation with one another. We may never have an actual intelligent editor, but we may have an intelligent editing assistant which can help organize and structure content. Digital nonlinear editing systems save time and allow more options to be explored. With intelligence, such options can be suggested.

4.10 Cuts Only

Until video orchestration and video editing tools merge and become one in the same, the best we can do is work on either side of the problem and hope someday to collide in the middle. While video orchestration systems like Ryan Evans' LogBoy and FilterGirl engine do edit, basically, they are "cuts only" video systems. Their editing functionality extends no further than placing one shot after another. Any editor can tell you that you can only go so far on a cuts only editing system. Orchestrators can only go so far if they cannot accomplish a simple editing function such as overlaying video over sound. Providing a greater toolbox of editing tools nudges video orchestrators one step closer to being, if not a video editing system, at least a video composer.

Video orchestrators thus far have demonstrated that orchestrating video more or less at a sequence level works. Why can't video orchestrators also compose those sequences? The inability to edit sound and picture asynchronously may explain why. Without this ability, sequences could not be generated which featured the same video and different audio, or the same audio and different video. Such sequences had to be preprocessed in a video editing application. Furthermore, the constraints required to edit sound and picture asynchronously focus upon the shot level. At which point in a shot should a cutaway take place? Orchestration at a sequence level requires knowledge of narrative — which sequence corresponds to which plot event. Orchestration at a shot level requires knowledge of narration — which shots, when edited together, can narrate a plot

event. These models have yet to be fully researched and developed. Once again, the atomic age must end.

PART TWO: INQUIRIES

5. INTERACTIVE CINEMA: Hype or Text?

Random access video has opened up a new world of possibility for video editors. Random access video also holds great promise for storytellers(not that editors are not storytellers in their own way). Consider the following analogy: a strip of film or a videotape without timecode is to sound and image as the scroll was to text. A scroll presented text as a linear form. Random access was possible, but very cumbersome. Although, in a limited sense, pages allowed random access, the evolution of structures of organization such as chapters, indexes, table of contents and footnotes, made books more easily navigable than a scroll. New forms of recording history and telling stories were enabled and facilitated. Electronic books challenge the physical sequence inherent in print books and have created hypertexts which have changed the way we think of reading, writing, and fiction. We read books. We navigate hypertexts. What does hypertext have to offer cinema and vice versa? What tools will be required to merge hypertext and cinema? Does the fate of interactive cinema lie in the hands of hypertext?

5.1 Hypertext

Hypertext, simply stated in the words of its creator, Ted Nelson, is "non-sequential writing — text that branches and allows choices to the reader, best read at an interactive screen. As popularly conceived, this is a series of text chunks connected by links which offer the reader different pathways." (Nelson, 0/2) Although hypertext does not have to exist in an electronic form, it functions best electronically. Hypertext frees text from the physicality of the printed page, the bound book. Texts can be intricately linked to one another, blurring distinctions between footnotes, marginalia, texts, authors and readers. Hypertext has been applied to diverse fields but is probably most used and known through the World Wide Web. The advantages of using hypertext are particularly clear in navigating "the Web." Can you imagine "surfing the Web" without hypertext?

5.2 Interactive Fictions

A hypertext fiction is a fiction which uses the hypertext form. It is a non-sequential fiction. Hypertext fictions may exist electronically or on printed

pages. Regardless of media, hypertexts are interactive, because users can create their own sequence of text(s) in the act of reading. As hypertext theorist Jay Bolter writes, "The flexibility of electronic text makes for a new form of imaginative writing that has already been named 'interactive fiction.'"(Bolter, 121) For purposes of simplification, electronic/hypertext fictions will be called interactive fictions, though I will expand Bolter's definition to include non-electronic interactive fictions such as *Choose Your Own Adventure* novels.

What aspects of hypertext do interactive fictions exploit? What kind of stories are better told through hypertext? Hypertext, by its nature, is non-sequential. Thus, it would make sense that stories which can be read non-sequentially would be best suited for hypertext. Two story forms lend themselves well to a hypertext environment. Although they will be discussed separately, they may act concurrently.

5.2.1 Multithreaded Stories

> "Agnes Nixon created Erica Kane, who eloped with Dr. Jeff Martin, divorced him, married Phil Brent, miscarried, mental breakdown, seduced Nick, almost slept with half-brother Mark, married Tom, opened disco, escaped murder charge, formed Enchantment, married/divorced Adam, fell for a monk, almost killed by Natalie, married Travis, gave birth to Bianca, adulterer with Jackson, was kidnapped, married Dimitri, found by rape-conceived daughter Kendall (who married Anton, who loves Julie, who's loved by Noah), stabbed Dimitri, returned to modeling.
>
> Meddling millionaire Phoebe Tyler broke up grandson Chuck's marriage to ex-prostitute Donna; Palmer Cortlandt fell for Donna, got dumped, mugged, went broke, worked as busboy, got well, married Donna, learned he's sterile, had fling with Daisy, got divorced, got drunk, shot himself in head, charged with tax evasion, lost it all, got it back, shot, paralyzed, married nurse Natalie, arrested, got hard labor, released, got richer, married Opal, blackmailed Janet from Another Planet.
>
> Joe Martin and Ruth Brent finally married, adopted abandoned Tad, who grew up and smoked pot, ran away, became a cad by sleeping with Liza and her mother, accepted money to date Dottie, fell for Hillary, married pregnant Dottie, divorced, married and divorced Hillary, married Dixie, divorced, had fling with Brooke, saved Dixie from kidnapper, hit head in bridge explosion, lost memory, returned to Pine Valley, found out he had son with Brooke, married her, slept with Dixie, escaped look-alike killer, divorced Brooke, married Dixie, his great love."(*Variety*, ii)

Multithreaded stories often involve multiple characters and storylines which intersect and are woven into a complex narrative fabric. The excerpt above, taken from an advertisement for the soap opera *All My Children*, illustrates an example of a multithreaded story. Hypertext may be a very helpful navigational

tool for multithreaded stories. It could allow readers to follow, swap, or skip storylines as they wish. Although multithreaded stories feature and rely upon intersections between storylines, there are an equal amount of divergence amongst storylines as well. Returning to the advertisement, it presents a very neat, or unthreaded version of the narrative realm of Pine Valley(the setting of *All My Children*). If a reader navigated a hypertext version of *All My Children*, and only followed Erika Kane's storyline, their reading experience would look very much like the first example. It would not be necessary to mention that Janet is Natalie's sister, that Brooke and Erica were both married to Tom and Adam, that Dixie was also married to Adam, and so on and so forth. As illustrated by the advertisement, a reader's experience of following a storyline may be entirely linear, but it would still be non-sequential because the reader omitted other storylines, disregarding the given sequence in lieu of the chosen one.

5.2.2 Nonlinear Stories

The concept of nonlinear stories, when discussed in the context of hypertext, requires some clarification. Many stories are told nonlinearly. Novels such as *Wuthering Heights*, *The Vampire Lestat* or *Waterland* depend heavily upon the literary device flashback. These stories have nonlinear components in that the order in which events are conveyed to the reader do not correspond to the order in which the events actually happened. Although hypertext offers interesting navigational strategies to stories which are *told* nonlinearly, hypertext is best applied to stories which may be *read* nonlinearly.

Milorad Pavic's lexicon novel *Dictionary of the Khazars* traces the fictional history of the Khazars, an extinct nomadic warrior tribe. The story of the Khazars is hidden, distributed throughout the three dictionaries which comprise the novel. The author informs the reader, "No chronology will be observed here, nor is one necessary. Hence, each reader will put together the book for himself, as in a game of dominoes or cards, and, as with a mirror, he will get out of this dictionary as much as he puts into it, for, as is written on one of the pages of this lexicon, you cannot get more out of the truth than what you put into it."(Pavic, 13) Readers are free to read the text in any fashion or order they wish. Terms which occur in more than one dictionary are specially marked by symbols. By cross referencing texts through the three dictionaries, the readers construct their

own history of the Khazars. It is not hard to imagine a hypertext version of this novel which would allow readers to activate hypertext links and jump from one reference to another.

5.3 Spatial Form

Reading a nonlinear story is not an easy task. Readers cannot rely upon an aggregation of cause-effect relationships through time to guide them through the narrative. How then, do readers, come to understand nonlinear stories? In 1945, critic Joseph Frank wrote "Spatial Form in Modern Literature," an exploration of poetry and narrative and how readers come to understand them when the relations within the texts do not correspond to strict causal/temporal connections. The lack of these connections cause readers to see narrative elements "as juxtaposed in space, not unrolling in time."(Smitten, 19)

> "What the concept of spatial form does is to call attention to the departures from pure temporality, from pure causal/temporal sequence. When these departures are great enough, the conventional causal/temporal syntax of the novel is disrupted and the reader must work out a new one by considering the novel as a whole in a moment in time. That is, the reader must map out in his mind the system of internal references and relationships to understand the meaning of any single event, because that event is no longer part of a conventional causal/temporal sequence."(Smitten, 20)

In some ways, encountering spatial forms of narrative resembles assembling a puzzle without knowing what image it is being assembled. We know how individual pieces may form together but we have no knowledge of the overall picture. Once we have a great deal of the puzzle completed, we can better see how pieces relate to one another by seeing how they fit into the larger picture.

Spatial forms of narrative and hypertext complement each other because hypertext allows a reader of spatial forms of narrative to better visualize the juxtaposition of narrative elements. In discussing Joseph Frank and spatial forms of narrative, Jay Bolter writes, "The electronic reader is encouraged to think of the text as a collection of interrelated units floating in a space of at least two dimensions."(Bolter 160) Bolter further argues that because hypertext involves navigation via links, electronic text is far better suited for hypertext than a book because a reader does not have to flip through pages or indices. Electronic texts exist not along a linear path as dictated by the pages of book, but instead, as an intricate web navigated by the reader(and perhaps subject to an author's design).

5.3.1 *Victory Garden*

Stuart Moulthrop's *Victory Garden*, an epic hypertext novel, concerns, among other things, university politics, the Gulf War, and parties. Reading *Victory Garden* initially presents many difficulties. There is a long "ramping up" time during which the reader becomes more aware of the multiple characters and their myriad storylines. At times, the experience proves frustrating because the reasons why the storyline jumps to and fro between storylines and associated commentaries do not appear clear. As a spatial form of narrative, *Victory Garden*'s underlying story structures reveal themselves as the reader reads more. Once the reader attains this "critical mass" of story knowledge, the act of (re)reading the story involves as much discovery of new narrative units as actively searching for more narrative material, and seeing familiar material in new contexts. Furthermore, appreciation of how the author structured the story/interaction functions into the enjoyment of the text.

5.4 Fait Accompli

While Joseph Frank and his colleagues shed some light on how a reader may come to understand nonlinear stories, the author's role should not be forgotten. The order in which a reader may read the narrative units in *Dictionary of the Khazars* is completely arbitrary; this is not the case with *Victory Garden*. In addition to authoring the text, Moulthrop authored the structure of the text, the storylines, their intersections, and how and when a reader could choose to navigate them. The question then becomes, how interactive can an interactive fiction be? The reader exercises some choice regarding the story, but the author provides the choices. In conventional linear texts, however, the author provides no real possibility for interaction. Interactive fictions can at least allow readers of nonlinear and multithreaded stories to follow or pursue distinct story threads.

Some fictions, however, claim to allow the reader to control their destiny as they read. Instead of following story threads, the reader "directs" them. These *Choose Your Own Adventure* or *Twist-a-Plot* books tend to be simple branch structured narratives. While part of the enjoyment of reading these texts derive from a craving to exhaust the possibilities of the narrative realm, these texts also tend to become somewhat boring or frustrating. The reason for this lies in the reader's

illusion of control. The reader and story can only go where the author's branches allow. Besides, if a story could go in any direction, the story would quickly bore the reader because the story would not have to be read; it could be imagined. The author, then, must frustrate the reader's choices to make the story interesting. Basically, the author must, to some extent, make it so that the choices the reader makes hold no relevance to the story. In either case, whether the reader is following a story, or has the illusion of controlling a story, the reader is navigating a narrative structure provided by the author.

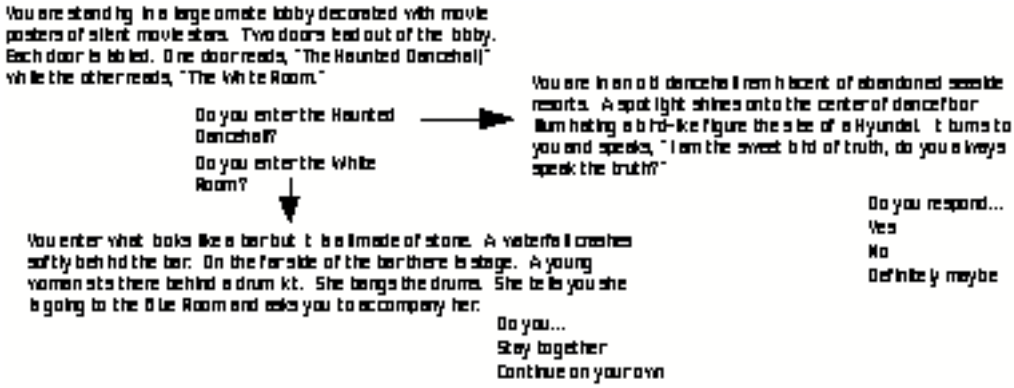

Figure 5.1 - Sample Branch Structured Narrative

5.5 Choose Your Own Misadventures

Instead of a narrative structure, can interactive stories provide a narrative mechanism which can actually respond to the reader's interactions? Branch structured narratives fail to provide such a mechanism. Narrative theorists Marie-Laure Ryan has provided one of the best explanations for why. Marie-Laure Ryan relates *Choose Your Own Adventure* novels to transition networks. Transition networks are graphs or directed graphs of narrative. Nodes on the graph represent narrative units and connections between nodes represent paths the narrative can take.

While transition networks can be used to visualize a branch structured narrative, transition networks may be used to generate stories from an existing or abstracted story grammar or morphology. Ryan argues that transition networks fail as story generators(and by extension, fail as interactive fictions) because they lack "narrative intelligence."

"In this algorithm, the knowledge of the program is a graph of narrative choices. The nodes of the network stand for the events of the story, and the arcs connecting the nodes indicate the logical possibility that an event may be followed by another ... The program is limited to an ability to blindly enumerate all possible traversals of the graph. The fact that all complete traversals result in well-formed stories is determined by the contents of the graph, not by decisions made during the traversal."(Ryan, 234+)

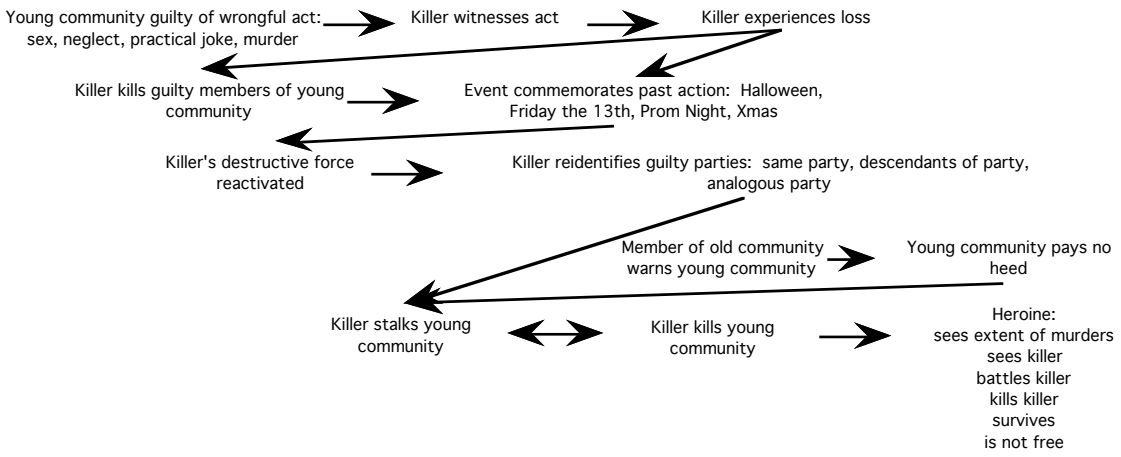

Figure 5.2 - Graph of story morphology of horror films
(adapted from Dika 1990)

5.6 Feedback

Hypertext author and theorist Stuart Moulthrop once described the Achilles' heel of interactive fiction as follows: "The greatest obstacle to interactivity in fiction resides in the text-to-reader leg of the feedback loop."(Moulthrop, 6) Feedback, a term popularized by cybernetics, refers to systems of energy or information in which some of the output is transferred back to input. In branch structured narratives, the author must tightly constrain the narrative design within an immobile structure. The static narrative design prevents the text-to-reader leg of the feedback loop from taking place because the user can only navigate the narrative structure and not influence it.

Stuart Moulthrop wrote *Victory Garden* with and in Storyspace, a hypertext authoring system developed by Jay Bolter, Michael Joyce, and John B. Smith. Storyspace allows authors to define links ("yields" or "words that yield") which may be activated only if the reader has previously read a link's "guard," a prerequisite lexia. In describing Michael Joyce's hypertext fiction *Afternoon*, which was also authored in Storyspace, Moulthrop writes the following: "The system of conditional 'yields' developed in 'Afternoon' employs a very different

kind of structure, one that does contains[sic] a feedback loop: 'Afternoon' is able to modify its current behavior by referring to a record of its previous behavior." (Moulthrop,14)

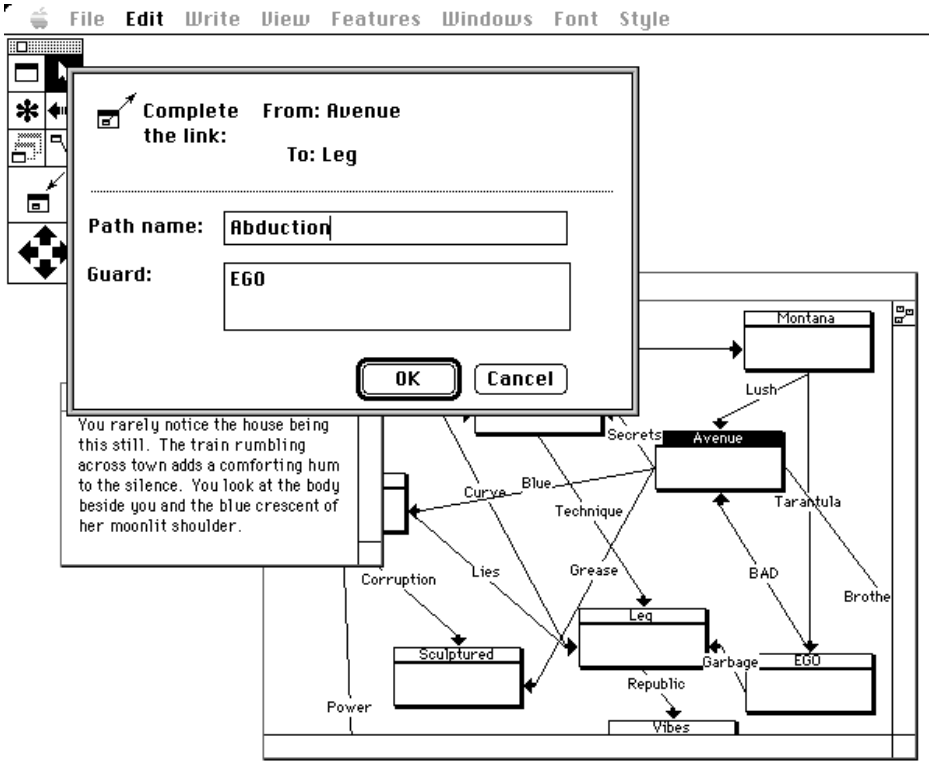

Figure 5.3 - Storyspace (Bolter, Joyce, and Smith)

5.7 Artificial Intelligence and Interactive Fictions

Although Storyspace provides rudimentary tools for narrative mechanisms, it requires the author to explicitly define every possible link. As a story grows larger and connections become more dense, explicit definition becomes a very cumbersome task. Artificial intelligence research in story understanding and generation provides a range of narrative models which may ease the author's task and can more realistically depict narrative structures.

In transition networks, preordained links between plot units determine the narrative. There is a canonical narrative structure. AI research takes a different approach. Stories consist of plot units and the "narrative intelligence" to glue the pieces together. This narrative intelligence takes the form of abstract plot models which graph how individual plot units may relate to one another. Story

understanding research such as Schank's scripts(Schank, 1981) or Lehnert's plot-unit model(Lehnert, 1981) correspond nicely to story generation research such as Meehan's TALE-SPIN(Schank, 210) or Lebowitz' UNIVERSE (Lebowitz, 1985) because in both cases, some kind of plot model is required. In story understanding, the system parses a story based upon models of causation or character and compares it against a predefined plot model. In story generation, a plot model provides the basic template from which stories may result. Although the abstract plot model is explicitly defined, this plot model can recombine a database of plot units in potentially infinite ways without the author having to explicitly create each iteration individually.

User interaction with story generation programs usually extend no further than defining a set of characters, goals, and plans(the means by which goals may be achieved). A story is the result of giving a character a goal, or as Marie-Laure Ryan states, "story = problem + attempted solution."(Ryan, 243) User interaction, however, during the generation of the story is not forbidden. Instead of a user just creating a story by choosing a goal, the user may choose how that goal may be achieved. While a narrative decision made within a branch structured narrative has consequence, the consequence, restating Ryan's words, is determined by the narrative structure, not decisions made interacting with the narrative structure. Intelligent narrative systems, on the other hand, can resolve how to attain a particular consequence(or how that consequence was attained) based upon user interaction and defined models of characters and concepts like motive, action, goals, plans, success, failure, conflict, and resolution. It is this feedback between the story system and the user which makes these story systems more flexible. In narrative structures, user interaction leads to consequence, in narrative mechanisms, user interaction cooperates with the story system to determine consequence.

5.8 Interactive Cinemas

An interactive fiction which uses narrative intelligence is still a hypertext, because there is no fixed sequence to how the plot model can manipulate plot units to generate stories. How can these models be applied to cinema? Many projects in the Interactive Cinema Group of the MIT Media Lab have dealt with issues of interactive narratives in the realms of fiction and non-fiction. These

projects present early mergers between the worlds of hypertext and the world of cinema. They were built upon and influenced the building of the video orchestrators discussed in chapter four. Video orchestration can be seen as the tool which will support mergers between hypertext and cinema.

5.8.1 *Electronic Scrapbook*: Amy Bruckman, 1991

Amy Bruckman's Electronic Scrapbook created topical scrapbook pages of video based upon a limited inference and knowledge system of familial relations and events. A user could request a story and the system would create a scrapbook page which would match that story. For example, a user could request a story of "A Child's Famous Firsts" and the system would generate a scrapbook page featuring video sequences of a child's first words, steps, haircut, and birthday.

5.8.2 *Train of Thought*: Mark Halliday, 1993

Mark Halliday's *Train of Thought* created an environment for multithreaded stories. By combining a database of clips and encoded story structures, *Train of Thought* created a multithreaded interactive narrative which changed over time. Halliday restricted the user interaction to choosing whether or not to view sequences associated to the sequence playing at the time.

5.8.3 *The Collage*: Eddie Elliott, 1993

Eddie Elliott's Collage allows a user to spatially arrange video clips on a computer screen as if manipulating video on a notepad. Arranging clips spatially instead of sequentially creates a loose hypertext structure within which a user may integrate the clips with texts and drawings. If so inclined, a user may construct a narrative with the Collage similar to the outputs generated by Bruckman's system. The Collage, however, provides no structure with which to encounter or create that narrative.

5.8.4 *ConText: towards an evolving documentary*: Work in Progress, 1994

ConText is a multimedia associative browser for "evolving documentaries." Content, whether video, photographs, or texts, are gathered and refined by the author. The author must then annotate the material, creating and refining descriptions for that material. While conventional narrative relies heavily upon causation, Context attempts to create story by association. As its name implies,

context is very important in this system. Content selection, what is shown to the user, is based upon what the user has seen before, what the viewer is currently viewing, and the descriptive context which surrounds the available content selections. For example, when a user views a specific content selection, the descriptions which annotate that selection are presented to the user, providing context. These annotations are weighted. Thus, common descriptions across multiple clips aggregate in value and influence future playout.

The projects discussed above share two common qualities. They all acknowledge that the computer screen, unlike conventional usage of television and film screens, can support multiple streams of video. The intelligent placement of multiple streams of video can express meaning. In Bruckman's work, the scrapbook page provided the topical structure within which to view the sequences. Halliday coined the term "positional editing" to describe how the position of streams of video on a screen can signify meaning. "Positional editing allows the maker to control the layout of the movie elements so as to aid the viewer's story building activity."(Halliday, 71) Elliott's collage becomes a blank canvas upon which a viewer can collect and organize video.

The second quality these systems share also relates to conventional hypertext. These systems manipulate the sequencing or spatial placement of clips of video, yet they do not manipulate the individual clips of video. Conventional hypertexts allow users to navigate complex webs of text, yet the texts themselves do not change. In implementing these systems through video orchestration, because the plot models work at an abstract level, it was easier to equate a video sequence with a plot unit. Although these plot models and video orchestrators cooperate to create interesting stories by manipulation of sequences, plot models and video orchestrators have perhaps unconsciously conspired to create story systems which basically output video instead of text.

5.9 HyperCinema?

HyperCinema may be too much of a buzzword, but it stresses cinema. The term hypermedia tends to lump all media together, stressing an equality between media, and negating the distinct ways in which different media communicate. Hypermedia documents which substitute video for text, with no critical inquiry

into the unique characteristics of video(cinema), fall short of exploiting the medium to its fullest potential. I prefer HyperCinema to HyperVideo, because video tends to connote the medium, video, and not the art, cinema. Perhaps I am over influenced by movies, but I want the cultural baggage of cinema!

HyperCinema adopts the conventions of hypertext and artificial intelligence but infuses the mixture with how cinema narrates differently than text. A plot unit may be equal to a sequence of video, but the expression of that plot unit in cinema differs than its expression in text. A video orchestrator should know to play a particular sequence at a particular time, but it should also know *how* to play that sequence. The orchestrator should know and apply the knowledge, that, in cinema: point-of-view and reaction shots create empathy with characters; that accelerated montage generates excitement; that ominous soundtracks create ominous scenes, etc.

HyperCinema is interactive cinema. Cinema embraces all stories, so interactive fiction is no longer an accurate title. Interactive narrative is no longer entirely accurate either because it disregards how cinema tells stories. HyperCinema needs narrative intelligence for complex stories and interactions, but HyperCinema also needs narrational intelligence to be able to *tell* complex stories. HyperCinema is interaction with cinema at the level narrative and narration.

6. INTERACTIVE NARRATION

"Narration in general is the overall regulation and distribution of knowledge which determines when and how a reader acquires knowledge from a text"(Branigan, 106). Using Branigan's definition as a foundation, this work will examine narration and discuss the question of interactive narration as a paradigm of interaction which has been assumed or ignored.

6.1 The Story Thus Far

The focus of research which follows is based upon the following assumption: story consists of two equal and inseparable components — narrative and narration. Narrative is plot, what happens in a story. Narration is how that plot is presented. Among the realm of buzzwords, the term interactive narrative tends to draw the biggest crowd. The focus falls upon narrative, when narrative represents only part of story. Such works often treat narrative as synonymous with story, narration is assumed and/or ignored. A critical inquiry into narration accesses a relatively unexplored realm for interaction which has potential for makers and audiences alike. Knowledge about narration and narrative makes for better interactive stories. Interactive narration must enter the critical discourse. It cannot be stressed enough, however, that interactive narration is not a new concept. It is an unexplored one.

Narrative and narration cannot exist in isolation, and neither can change without somehow changing the other. For computers to really understand and eventually generate stories, they must have knowledge of both narrative and narration. The history of artificial intelligence research in story understanding and generation has been quite explicit in its focus upon modeling and understanding plot. Because of this, a common critique of research at the time was that computers could not tell good stories. Of course not! No one told the computers how to tell stories. The computers understood a great deal about narrative(plot) but almost nothing about narration. For interactive stories to be truly interactive, the text-to-reader leg of the feedback loop must be dynamic. The computer must be an equal partner in storytelling and not just a novel medium.

6.2 Nth Upon the Times

When and how does a reader acquire knowledge from a text? Let us first examine the when part of the issue. The question of when raises the issue of temporality. Temporality regards the presentation of events in and through time. Temporalilty involves more than just when an event is presented, it also involves the frequency and duration of that presentation.(Cohan, 84) Temporality's role in cinema shows how much the actual duration of an event may differ from its narration. How is it a film can encompass an entire lifetime, yet last only two hours? Why is it a time bomb at five seconds takes half a minute to count down? Obviously, the time it takes to narrate an event does not necessarily correspond to real time. In addition, time does not always move in one continuous direction. Flashbacks and flashforwards transport us through time.

Many interactive experiences have already exploited the temporal aspect of narration. The interactive fictions discussed earlier offer the reader some control over when they read certain passages. Furthermore, an important discussion in the field of personalized media concerns offering consumers control over how long a program may last. For example, you may only have five minutes to watch the news or you may have twenty minutes or twenty seconds(Houbart, 1994). Interactive narration dealing with temporality provides us with an opportunity to control one of our most valued commodities, our time.

6.3 Whose Story is This?

Having briefly discussed the issue of *when* we acquire knowledge from a text, let us now examine *how* we acquire knowledge from a text. "It[Narration] is composed of three related activites associated with three nominal agents: the narrator, actor, and focalizer."(Branigan, 106) "In the strict sense, a narrator offers statements *about*; an actor/agent acts *on* or is acted upon; and a focalizer has an experience *of*. More precisely, narration, action, and focalization are three alternative modes of describing how knowledge may be stated or obtained."(Branigan 105)

In life, we usually encounter only two of the three agents, the narrator and the actor. When our friends convey an anecdote, they narrate. They may recreate

the scene and act within that scene. In narrative cinema, however, such an encounter rarely occurs. A cognate in cinema would consist of nothing more than a talking head relaying a story. Although we often have this situation in documentary, we do not usually find instances in which that is all we see. Usually, we experience what theorists call focalization.

"Introducing the narratological concept of focalization is meant to remind us that a character's role in a narrative may change from being an actual, or potential, *focus* of a causal chain to being the *source* of our knowledge of a causal chain."(Branigan, 101) Focalization takes two forms, internal and external. Internal focalization occurs when the spectator is privy to a view unique to one character. An example of internal focalization in film is a dream sequence or the point-of-view shot. In these cases, we see and hear the story world exactly as the character does. They are one in the same. Horror films such as John Carpenter's *Halloween* use internal focalization to great extent, in an effect sometimes called the "I-camera."(Clover, 113)

External focalization is a view onto a story world that we would witness if we existed in that story world. External focalization prevails in films. We see it when we follow characters, or follow a pan to see the object of their attentions. While we equate the point-of-view shot with internal focalization, we equate the eyeline match with external focalization. Unlike the point-of-view shot, the eyeline match does not correspond exactly to what the character sees, in that we do not see exactly from that character's position in space.(Branigan, 103)

Characters mediate narration. Through their experiences, whether as narrator, actor, or focalizer, we come to know the story world and the world inside the characters. In cinema, focalization dominates our experience. A narrator may frame the narrative, but through focalization do we learn or not learn about the story world. Obviously, we cannot know or experience(at least not yet) everything a character in a film knows or experiences. Furthermore, characters in a film cannot know everything we know(at least not yet). We watch films from a priveleged, though not entirely omniscient view. The view provided us originates from a grander narrator yet.

6.4 Process

Narration is a process. Cinematic narration does not come from one narrator, but rather from a complex system of interactions between scripts, directors, actors, lights, sweat, dreams, pictures and sounds. If a movie is projected in an empty theater, does it narrate? What we see and hear in a theater is what narrates. How did it get there? How does cinema focalize? What makes cinema?

Interactive narration must function at the level of film technique. Bordwell and Thompson's *Film Art* offers a thorough categorization of film technique. They break film technique down into four categories, mise-en-scene, cinematography, editing and sound(Bordwell & Thompson 126). On an abstract level, filmmakers deal with issues like focalization, but on a creative level, close-ups, editing, acting, etc., make focalization possible. The introduction of Bordwell and Thompson's categorization serves, not as an introduction to film technique, but rather to introduce the levels at which interactive narration may take place. Placing somewhat artificial distinctions upon the four categories of film technique, we find two times at which interactive narration may take place. For the most part, mise-en-scene and cinematography belong to the realm of preproduction and production. Editing and sound belong to the realm of postproduction.

6.5 Pre/Post/Pro/duction/jection

When we watch a movie, we watch the final stage of a four stage process: preproduction, production, postproduction, and projection. Making a movie makes a narration. A movie represents one narration(though it can contain multiple layers of narration), a result of a collaboration between filmmakers and their tools. The audience experiences one telling because there exists one film, one long roll of celluloid being churned through the projector. Different versions of parts of a film may exist, but we see one version. Computers make it possible, or just easier, to allow audiences to change aspects or view different versions of a film. Interaction with a film now exists beyond how we choose to interpret it. The age of interactive cinema has arrived.

Interactive narration has two main venues -- what audiences see, and how that material is edited. Changing aspects of a film in preproduction or production involves changing aspects of mise-en-scene and cinematography -- costumes, lighting, composition, acting, etc. Changing aspects of a film in postproduction involves changing the editing of sound and picture. Finally, changing aspects of the film's presentation(projection) involves allowing viewers access to the variations created in the process up to projection.

Although breakthroughs in computer graphics have made it possible to synthesize images not captured on film(special effects), we still have a long way to go before we can synthesize any image and manipulate it instantaneously. Audience impact on interactive narration during the pre/production lacks the technology to be truly feasible. Interactive narration in postproduction however, already exists. This interaction happens in the editing room. Editing sound and picture profoundly affect a film's narration. Assuming enough footage exists, one iteration of interactive narration may involve allowing audiences to impact the editing of the film. In reality, the editor already does this. The editor must edit the film from, and is limited by, the sounds and pictures which been captured.

6.6 I'm Ready for My Close-Up

Different media tell stories differently. An interesting quality of stories worth examining considers how different media convey character. What is going on within the hearts and minds of the characters? Opera gives us the aria, theater, the soliloquy. In cinema, we have the close-up. Perhaps no other cinematic technique pulls our emotional heartstrings more than the close-up.

The close-up, however, does not act alone. The Kuleshov effect demonstrated the power of the close-up even with a neutral reaction shot. Lev Kuleshov writes, "I created a montage experiment which became known abroad as the 'Kuleshov effect.' I alternated the same shot of Mozhukhin with various other shots(a plate of soup, a girl, a child's coffin), and these shots acquired a different meaning. The discovery stunned me — so convinced was I of the enormous power of montage."(Levaco, 200) The close-up exists because of the way it was photographed and because of montage. Montage exists because of the temporal

nature of cinema. In remembering a film, however, we do not recall how the image came about, but we do remember the heartbreaking look on the character's face as he sees his great love for the last time. We remember the product and not the proceses.

Interactive narration operates at multiple levels, not just how the story is told, but who tells the story. Just tweaking whose reactions we see in a film drastically changes our experience of it. By changing whose reactions we see, we change who focalizes the film for us. The main character in a film usually corresponds to who acts as focalizer. Whose reactions do we see? From whose eyes do we see? Editors make these kinds of decisions in the editing room. They will often provide an audience with more close-ups, point-of-view shots, and eyeline matches pertaining to one character, the main character. Interactive narration may allow us to choose or change who functions as the main character; cutting to their close-ups and reactions is one of many cinematic means to achieve such a goal.

6.7 A New Paradigm

Why do we buy soundtracks, novelizations, watch sequels and remakes, wear silly T-shirts, eat Raptor candy and buy action figures? Why do we watch movies more than once? We want to return to that story world. Conventional cinema offers one view into that world, interactive cinema can offer much more. Current interactive movie experiences on CD-ROMs actually mask games or puzzles behind a veneer of cinematic technique. They may create an evocative experience, but for the most part, they fail to create convincing worlds or characters. The titles seem more interested in creating interesting puzzles than interesting stories. The repeatability factor of these projects suffers because of the puzzle approach. Once a user solves a puzzle, they do not want to have to do it again. Interactive narration offers a new paradigm of experience.

Interactive narratives have dominated the discourse of story-based interactive experiences. By definition, these projects allow a viewer to influence the plot. Do we really want to make *Romeo and Juliet* have a happy ending? Would we rather watch *West Side Story*? Interactive narration influences interactive story in two ways. Even if an interactive story focuses upon being an interactive

narrative, the narration must be applied intelligently and artistically to provide a satisfying experience. On the other side of the spectrum, we can explore the possibility of creating an interactive story which focuses upon being an interactive narration.

Interactive narrations may take many forms. We can view a story from multiple perspectives. Each character's view onto a story reveals more about the story and more about the character. Different perspectives reward the viewer for multiple viewings. Referring back to our discussion on the close-up, one reason we enjoy different media forms derives from how these forms explore character. Why? Character frames(focalizes) story. Through character, we gain a greater insight into experience, life. We will never learn the truth of that great abstraction called life, but we can experience different perspectives upon it. This is part of narrative's function. Interactive narration allows stories which respond to our needs as an audience. It harkens back to the oral tradition of storytelling and the infinite "Whys?" a child asks a parent during a bedtime story.

6.8 Interactive Narration is Interactive Cinema

A repeated theme throughout this document attacks the current state of interactive cinema as being nothing more than complex plot models or databases which spit out video with no real knowledge of cinema and *how* cinema operates as a signifying medium. How does cinema operate? How does cinema convey an event? How does cinema narrate? For interactive (cinematic)narration to be truly possible, the computer must have some answer to these questions.

Application of knowledge of cinematic narration extends beyond the realm of story. A great deal of research time, effort and dollars investigates video annotation and retrieval. Video databases and editors share the common trait that both must function within the limitations of what is provided. Editors and databases, however, diverge in purpose. An editor composes, while most databases only retrieve. An editor has cinematic knowledge, most databases do not. A video database should be smart enough to ask, "If a shot does not exist, can I create it?" Recall the Kuleshov effect. A shot of a hungry man did not exist. Montage created it.

PART THREE: IMPLEMENTATION

7. MILESTONE

MILESTONE was created to provide an infrastructure to support computer orchestrated asynchronous sound and picture editing. Implementing MILESTONE involved building from the lessons learned in the development of the video orchestrators discussed in chapter four. To accomplish this goal, it was necessary to build an annotation system for video clips, as well as the means to organize clips into edits and sequences. Furthermore, a battery of editing functions had to be created which could perform the mechanics of editing digital movies, manipulating their sound and picture tracks separately.

Three basic objects form the MILESTONE system — clip, edit and sequence objects. The objects progress in complexity. Edit objects encompass clip objects; and, sequence objects encompass clip, edit, and sequence objects. MILESTONE is built upon a knowledge representation tool, ConArtist, and allows users to create and attach annotations to video clips. Keyword annotations provide the "handles" to allow orchestration, while stream-based annotations provide cues for editing functions. While previous orchestrators limited their orchestration ability to the manipulation of video clips, MILESTONE also allows keyword annotations to be added to edit and sequence objects. These annotations allow the orchestrator to manipulate the video edits and sequences which it generates. MILESTONE also supports the ability to build content and editing templates by example.

7.1 Clip Objects

A clip in the MILESTONE system represents a segment of video, not a shot. A shot is traditionally defined as the result of a continuous run of film through the camera. In Hollywood cinema, the film exposed between the time the director yells, "ROLL CAMERA!" and "CUT!" results in a shot. In home videos, a shot results from the content during one continuous recording session. A clip in MILESTONE can be a shot, but it is not required that it be a shot. A clip may consist of two or three shots, or even a sequence: it depends upon how the user wishes to use the clip. Two forms of annotations may be attached to clip objects, providing descriptions and editing cues.

7.1.1 Clip Based Annotations

The first form of annotation MILESTONE uses is clip based annotations. Clip based annotations hold true for the duration of a clip. The clip is annotated as a whole. Imagine a clip of video about an intersection. What words or phrases would describe this clip? Traffic? Sunny? Carbondale? Annotations can be more than just keywords. Annotations may take the form of a "slot and value" model(see Evans, 1994). In that case, the annotations may be more along the lines of "Location: Carbondale, Subject: Intersection, Traffic, Weather: Sunny, Warm." MILESTONE uses the keyword model.

7.1.2 Stream -Based Annotations

Stream -based annotations, the second form of annotation in the MILESTONE system, have a temporal nature. They are annotations within a stream of video. Using the traffic example above, the annotation scheme might be more like the following: "1:00-1:30 Traffic, 1:00-1:15 Red Light, 1:16-1:30 Green Light, 1:04-1:20 Red Pickup Truck." Stream -based annotations provide a finer amount of descriptive detail. For example, if a traffic clip with a red light were needed, the retrieval mechanism would search for temporal durations which shared the annotations for "Traffic" and "Red Light."

The stream-based annotations in MILESTONE consist of edit points and edit durations. Although we primarily use annotations for retrieval, MILESTONE uses stream -based annotations for editing. How long should a cutaway last? When should the L-Cut begin? The use of these annotations will be part of the discussion regarding edit objects.

7.1.3 Variable Clips

Each clip has a "clone to variable" function. Cloning a clip creates a "variable clip." A variable clip is not a clip, rather the collection of keywords which describe that clip. When variable clips are used in edit and sequence objects, the variable clip acts as a query into the database, returning a clip with a matching set of annotations. Variable clips may be honed, so their queries strike at greater accuracy. Annotations may also be added or subtracted to the clone without impacting the original. Furthermore, one clip may have multiple clones for use in different contexts.

7.1.4 Discussion

In the current implementation of MILESTONE, the user provides the stream-based annotations. An audio parser or computer vision system, however, could conceivably generate a selection of good edit points or durations for the editor. These systems succeed because there are tangible artifacts which the computer can detect, a pause in speech, an increase in volume, perhaps a gesture. Can the computer do more? Can it tell which of the edit points is the best edit point to use? Because the reasons why an editor chooses one editing point or duration is based more on intangible artifacts, it is highly doubtful a computational system could choose "the best" edit point. A system , however, could be built to cut on Salt's "dialogue cutting point," but that instance is only one instance of the many ways an editor may cut a scene.

Will there ever come a time when a computer can edit a sequence as well as and in as many ways as a human editor? The answer to this question may depend upon whether or not the "AI problem" will ever be solved. Furthermore, the issue at hand is a complex representation problem. Models for plot as well as models for cinematic narration must be created and explored. There must be representations for how picture and sound act autonomously, as well as how picture and sound act cooperatively.

MILESTONE does not use stream-based annotations in the traditional sense, because MILESTONE was designed as an orchestration system and not as a general purpose retrieval system. Stream-based annotations often serve a descriptive function: in MILESTONE, the stream-based annotations serve editing functions. Although there are certainly times when annotating edit points and edit durations would be beneficial(and will be discussed later), questioning and exploring the issues involved in descriptive stream-based annotations fell outside the realm of this thesis.

7.2 Edit Objects

MILESTONE orchestrates sound and picture asynchronously. Sound and picture do not have to be edited at the same time; they can be separated and edited individually. This separation happens in the edit object. Editing sound

and picture asynchronously allows the orchestrator to generate video overlays and L-cuts "on the fly" without requiring the maker to preprocess such edits in another application. All but one of the editing functions require two clips(the one anomaly requires three), the two clips which will create the edit. Like twins, these two clips are unique yet share a common bond in creating an edit. In MILESTONE, these clips are represented by a BASE clip and its AFFIX twin.

A common technique in film and video editing is to edit for audio first, and then to edit video "on top of" that audio to maintain or create continuity. This technique is useful because audio, especially dialogue, is more temporally constrained than video. For example, when cutting sound, an editor must ensure the dialogue is sensible and must cut around sentences or phrases. In video, the editor is much freer to edit footage together and must only be careful when cutting around zooms, pans and camera movement. Furthermore, visual cuts can be made less obtrusive by cutting on action. Because audio is less temporally constrained than video, MILESTONE adopts the model of editing video over sound. In particular, the audio in a base clip remains constant. MILESTONE then orchestrates video on top of this audio, using video from the base and affix clips.

7.2.1 Editing Functions

One benefit of the MILESTONE system over commercial editing systems is that the editing functions are predefined. The system "knows" what an L-cut and a video overlay are and how to implement them. These editing functions are common, yet all current commercial applications do not know how to do these edits automatically. In Adobe Premiere, creating a split edit is a three step process; on the D/FX, it is a seven step process, involving toggling video and audio tracks on and off and redefining in and out points. Mechanically, the process is quite cumbersome, yet, conceptually, it is quite easy. "Start the audio for the next clip now, before the video cut." MILESTONE still steps through and toggles the video and audio tracks etc., but this process is accomplished in software. In the digital domain, there is not much reason why the editor should have to worry about the mechanics of how to accomplish every edit. The editing functions defined in MILESTONE are as follows:

7.2.1.1 OverLayHead

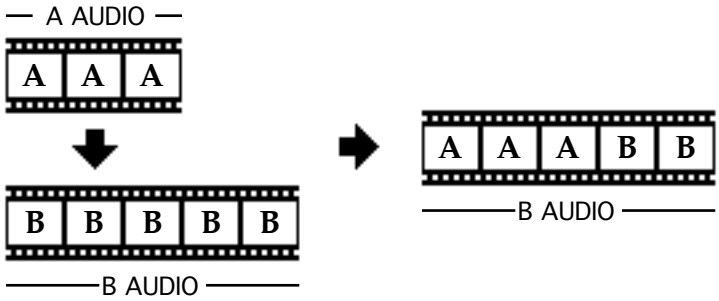

Figure 7.1 - Illustration of editing function OverLayHead

This is simply a video overlay over sound. The video from the affix clip is laid over the base clip, starting at the beginning of the base clip, such that the beginning of the video of the affix clip and the beginning of the audio of the base clip are coincident. The video from the affix clip covers the "head" of the base clip.

7.2.1.2 OverLayTail

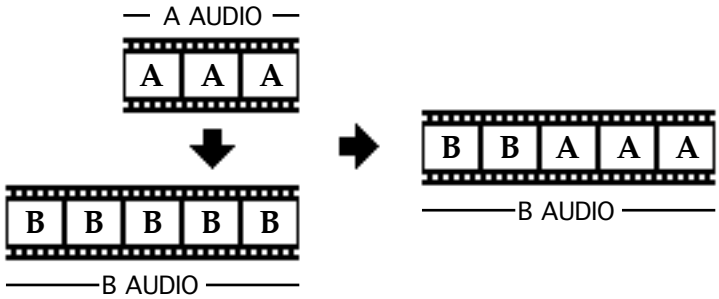

Figure 7.2 - Illustration of editing function OverLayTail

This is simply a video overlay over sound. The video from the affix clip is laid over the base clip such that the video of the affix clip and the audio of the base clip end at the same time. The video from the affix clip covers the "tail" of the base clip.

7.2.1.3 OverLay

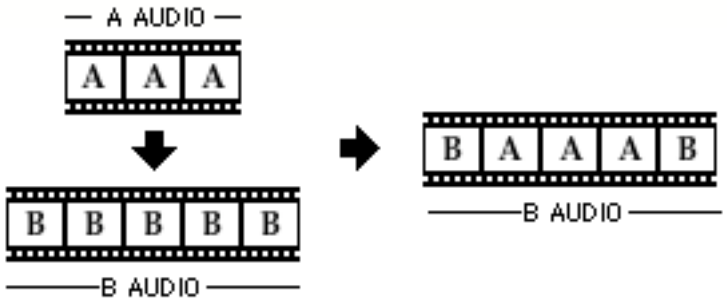

Figure 7.3 - Illustration of editing function OverLay

This is simply a video overlay over sound. If the user does not define a time duration, the video from the affix clip will be centered temporally over the base clip.

7.2.1.4 L-Cut Audio

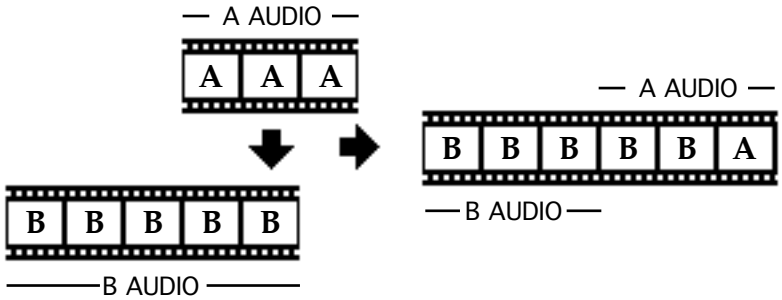

Figure 7.4 - Illustration of editing function L-Cut Audio

An L-cut, also known as a split audio cut, or split edit, is a common editing technique in which the audio for a clip may proceed its visual cut. The base clip is the first clip in the edit, the affix clip is the second. If a time index is not provided, the L-Cut begins at the three-quarters mark in the duration of the base clip.

7.2.1.5 L-Cut Audio Mix

This is the same as above except the audio is mixed.

7.2.1.6 L-Cut Video

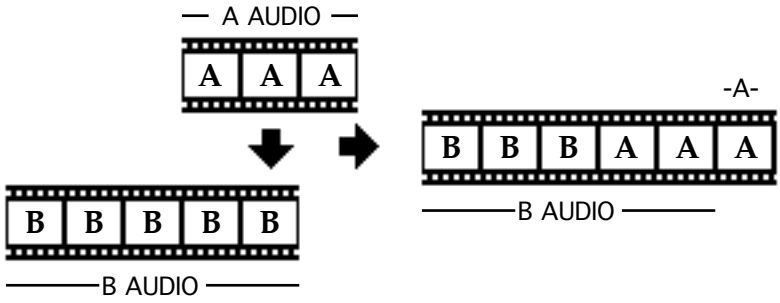

Figure 7.5 - Illustration of editing function L-Cut Video

This is similar to an L-Cut Audio, except the video leads the audio cut.

7.2.1.7 Straddle Cut

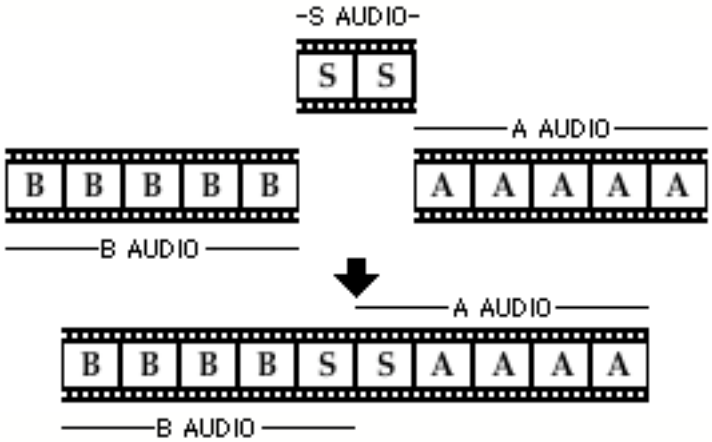

Figure 7.6 - Illustration of editing function Straddle Cut

A straddle requires a base clip, an affix clip, and a straddle clip. No time indexes are needed. This edit "straddles" the straddle clip over the straight cut between the base and affix clip.

7.2.1.8 Straight Cut

A straight cut is a standard edit. This function was created to allow editing of subclips. By dragging over a duration to a base or affix clip, you can edit subportions of the base and affix clips together.

7.2.2 Discussion

Although variable clips may be used as a base clip, this practice can lead to unpredictable results. This side effect occurs because the stream based annotations only mark time indexes, the time indexes themselves are not annotated. Therefore, when a variable clip retrieves a clip with matching descriptions, there is no way for it to know which of the many time indexes to use. The edit object will use a default value which may not be appropriate. Furthermore, time indexes serve a particular editing function unique to the clip to which it belongs. One clip cannot successfully use another clip's time index because there is no guarantee, and it is virtually impossible, that two clips will have exactly the same time indexes.

7.3 Sequence Objects

The sequence object is simply a linearizer. It takes in a sequence of objects, "renders" them and sequences them together. Sequence objects may contain other sequence objects. This interface allows a user to breakdown an edit into small parts which can be worked on individually without having to visualize a whole. Although timeline views are beneficial, it is sometimes cumbersome to scroll through an entire edit to reach a certain point. Although editing systems allow nonlinear editing, for the most part, their presentations are still linear.

7.3.1 Discussion

When attempting to create multivariant sequences with the LogBoy and FilterGirl system, users would often approach the system with a template, a sequence already in mind which they planned to vary. Users often found themselves retrofitting their logs and filters to create that one sequence. There was no way to build filters by example. Users could not create a sequence and use that sequence as a model to build multivariant structures. MILESTONE supports building by example. By substituting clips in a sequence or edit object with their variable clones, users define a two-fold template. The descriptions provide a content template while the editing functions provide an editing template. The system knows what kind of clips to sequence and how to edit them. The templates allow users to build multiple iterations of an edit from one example. In essence, the user is building filters by example.

7.4 Interface

7.4.1 Clip Objects

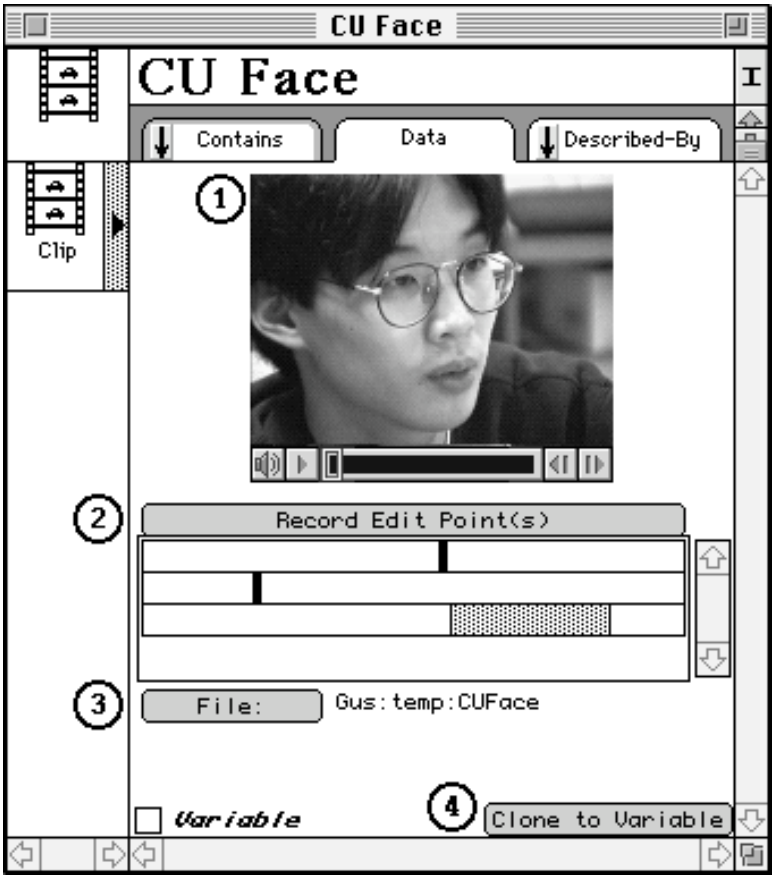

Figure 7.7 - Sample Clip Object

1. This is the clip. The movie controller allows the user to scan through the movie. By holding down the SHIFT key, the user can drag the controller to select a duration.

2. Record Edit Point(s)
This button stores edit points or durations in the scrolling field below it. The scrolling field presents a graphical display of the edit point or duration.

3. File:
This button imports the selected movie into the clip object. The pathname of the movie is stored in the field to its right.

4. Clone to Variable/*Variable*

This button "clones" the current clip and creates a variable clip. The variable clip is an exact replica of its originator except the *Variable* radio button is defaulted in the on position. Variable clips may be edited without impacting the original clip.

7.4.2 Edit Objects

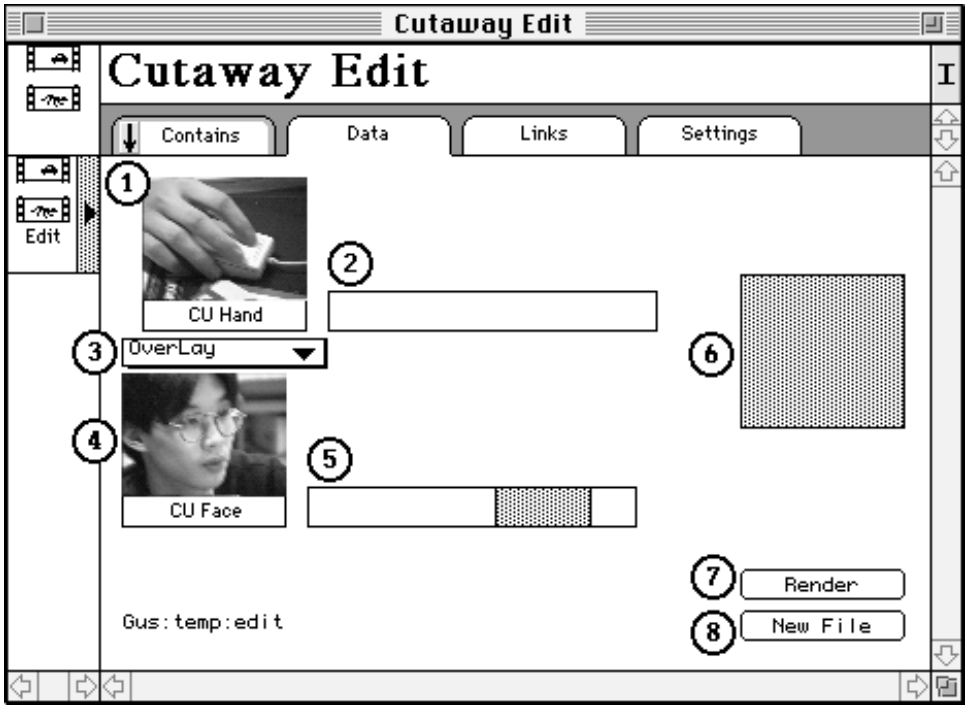

Figure 7.8 - Sample Edit Object

1. Affix Clip

The affix clip is affixed to the base clip. The video which overlays the base clip originates from this clip. In L-cuts, straight cuts, and straddle cuts, this is the second clip. Affix clips may be substituted with variable clips.

2. Affix Clip Time Index

This is where time indexes for affix clips are dragged and stored.

3. Edit Button

This button is a pull down menu from which the user chooses an edit function.

4. Base Clip
The audio from this clip serves as a base, though its video is mutable. In video overlays, it the video from the affix clip which is laid over this clip. In L-cuts, straight cuts, and straddle cuts, this is the first clip. Base clips may be substituted with variable clips though the results may be unpredictable.

5. Base Clip Time Index
This is where time indexes for base clips are dragged and stored.

6. Straddle Clip
Straddle clips are only needed for the Straddle Cut edit function. This clip "straddles" the cut between the base and affix clip.

7. Render
This button "renders" the edit.

8. New File
This button allows the user to specify a name and pathname for the clip created by the edit object.

7.4.3 Sequence Objects

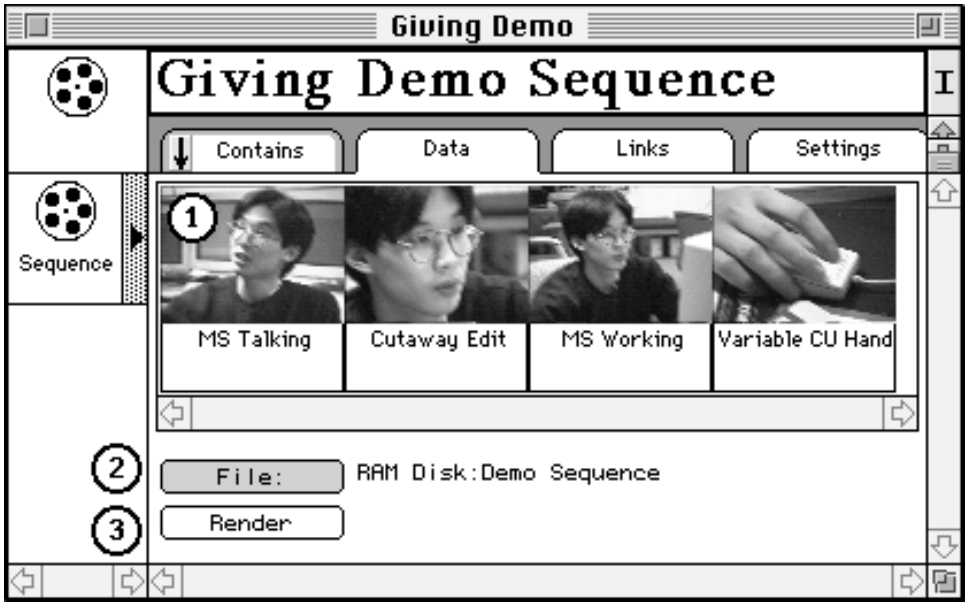

Figure 7.8 - Sample Sequence Object

1. Sequence Window
Clip, edit, and sequence objects are dragged to this window.

2. File:
This button allows the user to specify a name and pathname for the clip created by the sequence object. The pathname is stored in the field to the right.

3. Render
This button "renders" the sequence.

7.5 Implementation

Programming MILESTONE required the cooperation of many tools. Some aspects of MILESTONE were programmed from scratch while other aspects depended upon pre-existing tools. These tools, in turn, required other pre-existing tools. Because of this recursive nature, perhaps it is best that I start from the beginning.

7.5.1 Software

FRAMER, a persistent knowledge representation system, provided the database component of MILESTONE(Haase, 1993). FRAMER also served as the database for the theses of Halliday(1993), Evans(1994) and Houbart(1994). FRAMER, however, lacks a visual interface. ConArtist, a tool developed by Michael Murtaugh as part of Gilberte Houbart's thesis, "Viewpoints on Demand" provided this interface. ConArtist, in addition to being an interface to FRAMER, is an extensible interface tool which allows users to create and customize their own interfaces. MILESTONE was built on top of ConArtist. As with ConArtist, MILESTONE was implemented in Macintosh Common LISP. The editing functions in MILESTONE were written from scratch and called QuickTime system traps to manipulate and edit the video and sound tracks.

7.5.2 Hardware

MILESTONE was developed on a Macintosh Quadra 850 with 40 megabytes of RAM and operating under system version 7.5.

8. HYPERACTIVITY

This chapter attempts to merge the previous discussions of interactive fiction and narration. The common thread with which I bind the discussions grows from the following proposition: an interactive text is a hypertext. Interactivity implies choices: branches from a singular fixed sequence. The following pages do not describe another new media or story form; rather, they describe a way present tools and research point towards a particular way to read, watch, and interact with particular texts and movies.

Although hypertext may eliminate or suppress linearity and cause-effect relationships, non-sequentiality does not require that these sacred cows of story become sacrificial lambs at the alter of hypertext. An interactive story allowing a reader to choose which story lines to follow functions at the level of narration. Narration(how the story is told), not narrative(what the story is) is affected. In this case, the chosen story line may still follow all the old, familiar conventions of linear texts we have grown accustomed to reading. Because no one path exists, the story becomes non-sequential, hyper.

8.1 Steadfast Lexia

How can we improve the text-to-reader leg of the interactive story feedback loop? Spliting the atom of story into its components, narrative and narration, we can examine one scenario in which feedback amongst the reader, narrative and narration may function. Most hypertexts base themselves about Roland Barthes' term *lexia*, a block of text.(Landow, 4) In most cases, each lexia remains immutable. Stuart Moulthrop's *Victory Garden* follows this example. In essence, the reader navigates through the narrative. It is not an interactive narrative because the reader can not change the narrative, though the reader may experience multiple narratives which may contradict each other. It is, however, an interactive narration, though a very limited one, because the reader only has a small measure of control over the narration: basically, only the temporal component.

Interactive fictions often present the same lexia within different story lines. These intersections provide one means of allowing the reader to jump from one

story line to another. Part of the fun of interactive fictions arises from reading the same event within different contexts and from different points of view. Because the lexia remain static, however, regardless of context, the level of feedback amongst the narrative, narration and reader is very low. Static lexias make interactive fictions appear more like reactive fictions. An interactive story system equipped with some intelligence about narrative and narration may raise the level of feedback in a storytelling system. When a lexia(a plot unit) is approached, the system should know the current context or dominant point of view and change the lexia, change the narration of that plot unit to reflect that context and point of view.

8.2 If the Story Fits, Tell It

As stated above, this investigation addresses particular kinds of stories. What particular kinds are there? Almost all stories could benefit, or at least, would suffer no detriment, from the smallest of interactive/hypertext components. Some stories, however, suit hypertext and interactivity better than others, just as some stories suit novels better than short stories, or some stories make better plays than they do movies, etc.

The works of "postmodern" or "metafiction" writers like Coover and Borges and older colleagues like Joyce and Sterne anticipate hypertext and interactive fictions in their structure(or lack thereof). Through techniques like digression and associative instead of causal connections, these authors have explored fiction, not as a unitary sequence, but as a multiplicity of events existing simultaneously, not in a two dimensional sequence of lexias, but in the multidimensional space between the author, the text, and the reader. These texts, however, existed as books before their appropriation as hypertexts. They are not hypertext fictions. Hypertext, however, can facilitate both the act of reading these works and recognizing the alternative structures of narrative and fiction within them.

Readers of less difficult and more conventional texts(though I extend the notion of text here to specifically address television shows) could also benefit from interaction. Here I address multithreaded stories and/or serials which involve multiple characters involved with multiple story lines with multiple intersections

between them. Soap operas such as *All My Children* probably represent the pinnacle of this form, though most television series, comedic and dramatic, also qualify. In these cases, we all have our favorite characters and our least favorite characters. We have our favorite story lines and our least favorite story lines. Currently, the only interaction we have is to switch channel or not to watch when the characters and plots subject to our indifference dominate the story. Furthermore, we often encounter scenes in which a character we like may be in a scene but is not the center of attention. We may pay more attention to that character and perceive that character as having a greater role, but this perception can not be actualized.

8.3 The Magical Mystery Tour

While the previous examples focused upon how interactive narration affects how we may read a text, interaction can also affect what we read. Take a mystery, for example. At an abstract level, a mystery presents the reader with an effect: a murder; but no cause: the motive. We read and enjoy watching the detective's investigation, piecing the mystery together and imagining the possible worlds of different suspects. We follow the detective's story line. The detective focalizes our experience. What if we followed the murderer's story line? What if the murderer or the suspect focalized the story? We may no longer have a mystery, we may have a chase movie. Many mysteries have elements of a chase film, though these elements lurk in the background until the climactic conclusion of the film, when we know the identity of the murderer. The film *The Fugitive* combined aspects of both the mystery and the chase film because the prime suspect, Dr. Kimball, was being chased at the same time he was trying to solve the mystery of the murder of which he had been accused.

Disparities and hierarchies of knowledge is how Branigan describes the way narration functions. "Narration comes into being when knowledge is unevenly distributed —when there is a disturbance or disruption in this field of knowledge."(Branigan, 66) He explains how disparites and hierarchies of knowledge influence a reader's response to a text; and, though he falls short of extending his example to encompass how narration may impact or cooperate with narrative to suggest genre, an interactive story can allow us, if not to experience such a reading, to experiment with one. Again, we are drawn to the

idea of feedback, that narrative and narration impact each other in a complex system known as story. These examples narrated thus far exploit hypertext as a navigational tool, however, and not as an input tool, a mechanism for feedback. The plot elements narrated before should influence what and how future plot elements are narrated.

8.4 Implementation: A Pulp Fiction

Implementing an storytelling system requires many elements, though I will specifically address issues of hypertext, dynamic lexia, and computational models of narrative, narration, and the reader's experience/expectation. We begin with plot. The research in artificial intelligence described before explored different methods for generating and understanding computational models of plot. For all of their impressive and complex algorithms to create narrative intelligence, these plot models often assume an omniscient narration. There is no "narrational intelligence," no model for how each character conceives of the world and how each character will narrate the same event in different ways. Narrative intelligence provides the knowledge as to whose story is being told, what has been told and what the reader knows versus what the characters know. Narrational intelligence should utilize this knowledge to change the presentation of each lexia, each plot unit.

The film *Pulp Fiction* featured two instances of repeated scenes, though each iteration differed slightly from the other to reflect the story line which surrounded it. It is my hope that an interactive narration can come close to something like the experience of watching those segments of the film, a sense of fascination and reward from seeing the same parts of the narrative narrated differently.

In discussing the works of Alfred Hitchcock, Branigan writes the following:

> "Using the example of a bomb placed in a briefcase under a table, he[Hitchcock] explained how he could create feelings of suspense, mystery, or surprise in the audience. If the spectator knows about the bomb, but not the characters seated around the table, then the spectator will be in suspense and must anxiously await the bomb's discovery or explosion. If the specator and a character both know that there is something mysterious about the briefcase, but do not know its secret, then the spectator's curosity[sic] is aroused. Finally, if the spectator does not know about the bomb or the briefcase, then he or she is in for a shock. Hitchcock recognized that these effects can be intensified according to what we know about a character and our emotional involvement with him

> or her. He realized that there is a close relationship between a spectator's wish to know, and his or her wishful involvement with situations and persons in a film."(Branigan, 75).

AI story generators tend to output a plot unit, a lexia. Often the determination as to which plot unit to output is dependent upon requisite preconditions. These preconditions are necessary for "means-end analysis," a problem solving technique based upon an assumed solution and an examination of the means necessary to acheive it. A solution to a problem, then, can be seen as a chain of cause-effect relationships — a narrative. The next step is to take the resulting "narrative" and use the same techniques used to generate that narrative, to generate a narration.

Assume a plot generator has output a variation on the event described by Hitchock. There are two people seated in a restaurant and below their table sits a briefcase containing a bomb. How should a storytelling system narrate and create this scene? The following falls somewhere between pseudo-code and description, but it does provide a basic model for discussion:

Pseudo-Code: Plot Element

Plot Element	Bomb in Restaurant
Characters	victimX, victimY
Setting	restaurantX
Pre-Condition	(is-present victimX victimY)
	(bomb-present restaurantX)
Goals	(kill victimX victimY)

Pseudo Code: Narration Elements

Narration Element	suspense
Pre-Condition	(knowledge-of bomb audience)
Result:	Start scene on a close-up of the briefcase, then pull back
	Accompany scene with suspenseful music

Narration Element	mystery
Pre-Condition	(knowledge-of-briefcase audience)
Result:	Same as above
	Accompany scene with mysterious music

Narration Element	surprise
Pre-Condition	
Result:	Ensure briefcase is seen in establishing shot, otherwise ignore

Narration Element action

Pre-Condition (knowledge-of-bomb audience)
(active-story victimX)
Result: Crosscut between victimX to briefcase
Accompany scene with exciting music

The narration elements described above contain three parcels of knowledge, a model of audience response; a model of how to affect narration based upon narrative knowledge; and a model of how to acheive the narration cinematically. First, examining the cases of suspense and mystery, the narration element uses the pre-condition that the audience knows about the bomb in building the audience model. Whether or not the audience knows about the bomb can be queried from a plot model of what has been narrated to the audience. Finally, the result proposes one way which the scene may be shot or edited to achieve the desired effect. Furthermore, the examples take advantage of an "audio Kuleshov effect." Often, just playing more suspenseful, happy, or exciting music can affect how we experience what we see on the screen. In the mystery example, because the audience does not know the briefcase contains a bomb, we have a more mysterious(yet still suspenseful) situation. The action narration provides an example exploring the issue of interactive narrations which allow the spectator to follow different story lines. Assuming that the spectator is following victim X's story and probably has some sympathy for the character, the crosscutting can accelerate the suspense to a level of greater action.

8.5 "Don't Look at the Finger, Look at Where It's Pointing"

Obviously, the examples described above only begin to flesh out a full-fledged storytelling system. As more research examines the issue of narration, particularly interactive narration, better models and a better understanding of narration will contribute towards a better storytelling apparatus. MILESTONE can function as one of many workhorses in the process. A simple example utilizing MILESTONE's features concerns the "pre-commercial cut." Soap operas often feature mini-cliffhangers leading commercial breaks in which a character delivers a particularly intriguing line of dialogue; and, as the music rises, we cut away to another character or stay on that character for a reaction. In the context of MILESTONE, the OverLayTail function creates the edit, while the descriptions attached to clips provide the means to build the narrative and narration models and ensure the edit object manipulates the correct clips.

Sample Scenario:
Mr. and Mrs. Frank and Jan Harvey are seated for breakfast. Behind them, at the sink, is Jan's sister Jill, with whom Frank has been having an affair. We see them all in a wide shot, Jan to our left, Frank to our right, and Jill just left of center behind Frank and Jan.

Frank hiccups through his breakfast. "Scare me," he says to Jill.
Jill responds, "I'm pregnant."

Now, do we stay with the long shot, or cut to Jan, or Frank, or Jill, or Jan and Jill, or Frank and Jill or just Frank and Jan?

Pseudo-Code: Plot Element
Plot Element "I'm pregnant."

Pseudo Code: Narration Element
Narration Element Jill's story
Pre-Condition (active-story Jill)
Result: (OverLayTail wideShot reactionShot-Jill)

Pseudo Code: Narration Element
Narration Element Frank's story
Pre-Condition (active-story Frank)
Result: (OverLayTail wideShot reactionShot-Frank)

Pseudo Code: Narration Element
Narration Element Jan's story
Pre-Condition (active-story Jan)
Result: (OverLayTail wideShot reactionShot-Jan)

8.6 Interface

Because interface depends, or should depend upon the content, describing what a potential interface may look like presents difficulties. We can, however, discuss when the interface appears. Basically, we have two options, real and non real-time interactions. The real-time interaction may look something like *Doom*, *Voyeur*, or a VR scenario where you literally follow characters around to follow their stories, much in the same way the play *Tamara* involves the audience following characters from room to room. The interface could just impact at a surface level: adjustments for sex, violence and language for example; or even more simply, a skip to next chapter feature present on laserdisc systems.

One problem with real-time interactions stems from the fact that a language for interaction has yet to be developed for cinema. How do you signify a (hypertext) link in cinema? An early experiment in the Interactive Cinema Group involved the creation of "micons," moving icons which supported the development of "The Elastic Charles: a hypermedia journal." Micons are short digital movie loops which serve as links from one video sequence to another. Micons had a temporal nature. "When linking from video to video a micon appears on the video screen as the link it represents becomes relevant and disappears once the link is no longer relevant."(Brøndmo & Davenport, 3) While micons presented one potential solution, the problem has yet to be and may never be solved purely within the language of cinema. In hypertext, clicking from lexia to lexia substitutes the process of turning from page to page. In cinema, the cognate of turning page to page is to keep watching. Books require interaction. Hypertext requires interaction. Cinema requires inaction.

Non real-time interactions attempt to deal with this issue by allowing the spectator to specify preferences before viewing. In addition, non real-time interactions seem more natural. Rarely do we switch or interrupt storytellers(except to ask questions) while they speak. We do, however, ask storytellers to repeat stories or parts of stories to us. Real storytellers do adjust their performance based upon interactions with an audience, but these interactions present enough difficulties for the human storyteller, not to mention any computer storyteller.

Finally, a lack of, or a different form of non real-time interface may be required to afford random access story. At times, reading interactive fictions like *Victory Garden* becomes frustrating because the reader must repeatedly click through various story lines to gain access to others. The story starts to stand in the way of itself. This rapid fire clicking does not wholly originate from a hunger to exhaust a database, it comes from a hunger for narrative. As theorists examining spatial form in narrative suggest, the reader eventually works out a syntax for a text, an understanding of how each element relates to its siblings. Once this gestalt view, this critical mass of story knowledge and structure coalesces, the reader no longer requires exposition. This freedom of interactive fictions — to be able to read any lexia at any time but to know its exact function within a narrative

universe — liberates the reader from the printed page and should liberate the electronic reader from repeated mouseclicks through now redundant and unnecessary exposition and story lines. There must be a way to view and access the entire story database when the reader is prepared for it, and prepared to understand it.

This need for random access story currently exists and may be more necessary in hypertext. Often, we may only read or watch the "good parts" of a particular novel or film. One of the benefits of linearity is that we can know exactly where the good part is located. This is not necessarily the case in hypertext. Interactive fictions do suffer somewhat from their electronic milieu in that it is difficult to physically visualize the story because there is no fixed structure. Author Robert Coover describes reading hypertext in the following way: "As one moves through a hypertext, making one's choices, one has the sensation that just below the surface of the text there is an almost inexhaustible reservoir of half-hidden story material waiting to be explored. That is not unlike the feeling one has in dreams that there are vast peripheral seas of imagery into which the dream sometimes slips, sometimes returning to the center, sometimes moving through parallel stories at the same time."(Coover, 10) This dreamworld, however, becomes a nightmare when one wishes to seek out a specfic half-hidden text, because that specific lexia has no specific location, only pathways to it.

9. CONCLUSION

9.1 Future Work

A milestone is not just a marker of how far we have come, it is a marker of how far we have yet to go. The following scenarios suggest improvements and additions to MILESTONE.

9.1.1 Descriptive Stream-Based Annotations

Adding descriptive annotations to time indexes in clip objects would be the most beneficial addition to the MILESTONE system. It would allow the system to orchestrate a larger potential of editing functions and would allow base clips to be variable clips. Queries into the database could request clips and a specific time index based upon its descriptions. A hybrid of MILESTONE's functions and Marc Davis' *Media Streams* would be a potential solution.

9.1.2 Fine Editing Tools

Audio scrubbing and a better means to define and manipulate video selections are two examples of fine editing tools which would improve MILESTONE. Because the focus of MILESTONE was orchestration and not editing, the interface does not support fine editing. It is possible but it is not easy to accomplish. Additional features such as automatic audio crossfades, dissolves and other video effects would smooth transitions between clips.

9.1.3 Automatic Annotation

Currently, all the annotations, keywords and time indexes, are created by hand. Recent developments in speech, audio, and gesture analysis can automate or assist the task of annotating.

9.1.4 More Intelligence

Although the combination of the sequence window and variable clips allow for a quick and rudimentary way to program by example, all they really do is form a template. A more intelligent system and representation could better abstract out a story model and try to match it to similar or analogous story models previously recorded.

9.1.5 Multiple Streams

MILESTONE currently assumes a single playout window, though one advantage of computers and digital video is the ability to support and orchestrate multiple streams of video as demonstrated by Mark Halliday's "Train of Thought."

9.1.6 Variable Clip Feedback

Although variable clips operate as queries into the database, there does not yet exist a mechanism to see all the potential results of that query. Visualizing potential results would assist the editing proceess by showing the editor potential options. Furthermore, the ability to edit the results would allow a user to better hone the results of a variable clip query. For example, one clip may match all the required descriptions, but for some reason, may not work at all in the current editing template.

9.1.7 Interactive Narration

Although artificial intelligence has focused a great deal of attention upon different means to understand and generate narratives, very little research has approached the issue of understanding and generating narrations. It would be worth exploring how models of narration can be built on the computer, or if it is even possible. Can narration models be built from the same tools and techniques used to generate plot models? Will totally new models have to be formed? We will not be able to truly answer the question, "Can the computer tell a story?" until we try to tell computers how to do so. Narration must no longer be assumed or ignored. It must enter the critical disourse.

9.1.8 Multimedia

Although this thesis, perhaps naively, focused purely upon cinema, many of the concepts found here are relevant to multimedia applications which integrate cinematic expression with pictures, sounds, and texts.

9.1.9 Film Technique

In examining interactive narration, this thesis looked exclusively at how editing a scene could affect an audience. Editing, however, is only one of many film techniques. Interactive narrration in postproduction assumes all the requisite

images and sounds already exist. An ideal system would be able to generate and compose the requisite images and sounds for the audience.

9.2 Here's Where the Story Ends

Interactive cinema must have interactions at the level at which cinema operates as a signifying medium. We have only begun to address this issue. The reasons for this vary, but, speaking very generally, the following cases perpetuate a particular notion of interactive cinema which has no knowledge of cinema's unique qualities as a medium:

- AI: AI story models output plot points, events. This focus tends to result in the output of plot events as sequences.

- Hypertext: The conventional notion of the lexia as an immutable object supports the notion of sequence as lexia.

- Video Orchestrators: Because video orchestrators are primarily cuts only systems, it is much easier to have them output sequences than to compose them.

Film historian Jack Ellis claims D.W. Griffith's greatest contribution to cinema was discovering "that the shot rather than the scene [sequence] should be the basic unit of film language."(Ellis, 47) The examples above have yet to truly make, or attempt to make, the same discovery. Until interactive cinema, in all its potential forms, deals with cinema at the shot level, attempts to merge film language with computers and to transfer an editor's knowledge to computers will fail or be crippled.

It is no small coincidence that we encounter more interactive narratives than interactive narrations, that more scenes are orchestrated instead of composed. Narrative may be expressed through a scene, but narration must be composed from(and by) shots. In order to compose narration, we must face the difficult task of translating the film language into computer language. The film language is a language of narration — not just how to tell stories, but how cinema tells stories. Cinematic storytelling systems need to speak the language of film. To

accomplish this task, computers need the requisite tools to speak. Cuts only orchestrators are not enough. Asynchronous sound and picture editing is not enough either, but it gets us close.

9.3 Afterward: Rosebud

Why is this thesis titled MILESTONE? The original title of the thesis was ASPEN, an homage to the ASPEN project during the early days before the Media Lab was even built. Also, ASPEN was a convenient acronym — Asynchronous Sound and Picture Editing Nucleus. I later decided, however, that I wanted a title related to film history, preferably during the early sound era. Gore Vidal, a director during the early sound era was an early favorite, but it was difficult to imagine naming a thesis GORE or VIDAL. Another director noted at the time for his innovative filmmaking was Lewis Milestone. I passed on Milestone until I rediscovered the fact that he had directed *Hallelujah, I'm a Bum*, an amazing musical starring Al Jolson, star of the landmark film, *The Jazz Singer*. I recalled how amazed I was when I saw the film, and so, I named my thesis after its director, Lewis Milestone.

10. BIBLIOGRAPHY:

Abelson, Robert and Schank, R. *Scripts, Plans, Goals and Understanding*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers, 1977.

Bolter, Jay D. *Writing Space: the computer, hypertext and the history of writing*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1991.

Bordwell, David and Thompson, K. *Film Art*. New York: McGraw-Hill Publishing Company, 1990.

Brøndmo, Hans Peter, and Glorianna Davenport. "Creating and Viewing the Elastic Charles - a Hypermedia Journal." *HyperText II Conference Proceedings*, York, England, July 1989.

Branigan, Edward. *Narrative Comprehension and Film*. London and New York: Routledge, 1992.

Bruckman, Amy. *The Electronic Scrapbook: toward an intelligent home-video editing system*. MS Thesis, MIT Media Lab, September 1991.

Clover, Carol. *Fantasy and the Cinema*. London: BFI Publ, 1989.

Cohan, Steven and Linda Shires. *Telling Stories*. New York: Routledge, 1988.

Cook, David. *A History of Narrative Film*. New York: W. W. Norton & Company, 1981.

Coover, Robert. "Hyperfiction: Novels for the Computer." *New York Times*(August 29, 1993): 10.

Davenport, Glorianna, R. Evans and M. Halliday. "Orchestrating Digital Micromovies." *Leonardo,* Vol. 26, No. 4., 1993.

Davis, Marc and Warren Sack. *IDIC: assembling video sequences from story plans and content annotations*. Internal Memo, MIT Media Lab, 1993.

Davis, Marc. *Media Streams*. Ph.D. Thesis, MIT Media Lab, February 1995.

Dickinson, Thorold. *A Discovery of Cinema*. London: Oxford University Press, 1971.

Dika, Vera. *Games of Terror*. Rutherford, N.J. : Fairleigh Dickinson University Press, 1990.

Drucker, Steven. *Intelligent Camera Control for Graphical Environments*. Ph.D. Thesis, MIT Media Lab, 1994.

Elliott, Edward Lee. *Watch • Grab • Arrange • See: thinking with motion images via streams and collages*. MS Thesis, MIT Media Lab, February 1993.

Ellis, Jack. *A History of Film*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

Elsaesser, Thomas and Adam Barker ed. *Early Cinema: space frame narrative*. London: BFI Publishing, 1990.

Evans, Ryan. *LogBoy Meets FilterGirl: a toolkit for multivariant movies*. MS Thesis, MIT Media Lab, February 1994.

Granger, Brett. *Real-Time Structured Video Decoding and Display*. MS Thesis, MIT Media Lab, February 1995.

Haase, Ken. *FRAMER REFERENCE MANUAL*. MIT Media Lab, 1993.

Halliday, Mark. *Digital Cinema: an environment for multi-threaded stories*. MS Thesis, MIT Media Lab, September 1993.

Lakoff, George. "Structural Complexity in Fairy Tales." *The Study of Man*. Vol. 1, 1972.

Landow, George. *Hypertext: the convergence of contemporary critical theory and technology*. Baltimore: The Johns Hopkins University Press, 1992.

Lebowitz, Michael. "Story Telling as Planning and Learning." Columbia University, 1985.

Lehnert, Wendy G. "Plot Units and Narrative Summarization." *Cognitive Science*. Vol. 4., 1981.

Levaco, Ronald. *Kuleshov on Film*. Berkeley: University of California Press, 1974.

Lippman, Andrew. *The Distributed Media Bank*. IEEE First International Workshop on Community Networking, July 13-14, 1994.

Morgenroth, Lee. *Homer: a story model generator*. BS Thesis, MIT, June 1992.

Moulthrop, Stuart. "Hypertext and 'the Hyperreal.'" *Hypertext '89 Papers* (1989).

Nelson, Ted. *Literary Machines*. South Bend, IN: The Distributors, 1987.

Norvig, Peter. *Paradigms of Artificial Intelligence Programming: Case Studies in Common Lisp*. San Mateo, California: Morgan Kaufmann Publishers, 1992.

Ohanian, Thomas. *Digital Nonlinear Editing*. Boston: Focal Press, 1993.

Pavic, Milorad. *Dictionary of the Khazars*. New York: Vintage International, 1989.

Ryan, Marie-Laure. *Possible Worlds, Artificial Intelligence, and Narrative Theory*. Bloomington, IN: Indiana University Press, 1991.

Rubin, Ben. *Constraint-Based Cinematic Editing*. MS Thesis, MIT Media Lab, June 1989.

Salt, Barry. *Film Style and Technology: History and Analysis*. London: StarWord, 1983.

Schank, Roger and Riesbeck, C. *Inside Computer Understanding: Five Programs Plus Miniatures*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers, 1981.

Smith, Thomas A. *If You Could See What I Mean...* MS Thesis, MIT Media Lab, September, 1992.

Smitten, Jeffrey and Ann Daghistany. *Spatial Form in Narrative*. Ithaca: Cornell University Press, 1981.

Variety. January 2-8, 1995.

Weis, Elisabeth and John Belton ed. *Film Sound: theory and practice*. New York: Columbia University Press, 1985.

Weis, R., A. Duda and D. Gifford. "Composition and Search with a Video Algebra." *IEEE MultiMedia*, Vol. 2, No. 1, Spring 1995.