Viewpoints on Demand:

Tailoring the Presentation of Opinions in Video

by

Gilberte Houbart

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in Partial Fulfillment of the Requirement of the degree of

MASTER OF SCIENCE in Media Arts & Sciences

at   the

Massachussetts Institute of Technology

September   1994

Signature of Author
Program in Media Arts & Sciences

August 5,1994

✓ **Certified by**
Glorianna Davenport

Associate Professor of Media Technology
Thesis Supervisor

Accepted by
Stephen A. Benton

Departmental Committee on Graduate Students
Program in Media Arts and Sciences

Viewpoints on Demand:

Tailoring the Presentation of Opinion in Video

by

Gilberte Houbart

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning on August 5, 1994
in partial fulfillment of the requirements for the Degree of Master of Science

Abstract

A model of interaction is presented for a home based system generating evolving documentaries that can be tailored according to the viewer's interests in a given viewpoint or story. This model takes advantage of the strong points in the linear experience that television and movies have traditionally offered by letting the story flow while allowing the viewer's intervention to constrain it.

The journalist works in this personalized context by shaping the material for the video database using graphical annotations for video content and story structure

The framework developed is used to reveal a special angle on the Gulf War: information technologies turned this event into a landmark in the history of not only media but also warfare. The amount of information available to the troops and the public reached levels never seen in previous conflicts.

By adjusting "content knobs" you can, for instance, ask for a presentation combining the perspective of journalists or you might want to hear instead what the former Director of Information for the Pentagon has to say. Maybe you would plan to spend two hours but change your mind in the middle of the documentary, asking for a shortened version. All these adjustments increase control over content. The scenario I just described could be seen as a component of a "smart VCR".

Thesis Supervisor: Glorianna Davenport
Title: Associate Professor in Media Technology, MIT Media Lab

Viewpoints on Demand:

Tailoring the Presentation of Opinions in Video

by

Gilberte Houbart

Thesis readers

Reader _____

Michael Hawley
Assistant Professor
of Computer Science
and Media Arts and Sciences

Reader _____

John S. Driscoll
The Globe Newspaper Co.
Vice President

Acknowledgements

This work is dedicated to the memory of Cristina Bernal, the "grande dame" whose wisdom and creativity I always hoped to inherit.

Contents

1 *What is this all about?*

In May 1968 I was 6 and riots were setting Paris on fire. We were living in a small village near Paris and my parents bought our very first television to watch the reports on the events. Twenty-six years later I was in Boston watching CNN broadcasting live images of the Gulf War. Technology had brought me closer to the action in time for the second event, but I was as shocked to realize that violence is always around the corner. It had been one hour away from me, it was now thousands of miles away. Immediacy was not as crucial as awareness of potential dangers and how they take shape. To me, the real purpose of news is to keep you informed so that you can form an opinion and make informed decisions in your own life at a personal and global level.

Looking back in time is equally important: it is only after it was over that key aspects of the Gulf War were understood, in particular that it could be considered as the first information war [see chapter 5]. Documentaries and history books serve a particular purpose. They give us a long term in-depth perspective that news doesn't have and help us to better analyze news itself. It is to be hoped that the tidal wave of digital media will bring a revival of the documentary. However it would be a mistake to put the technology first. How can we add value to the TV audience of the 500 channels?

1. Technology and motion pictures: a glimpse at the past

A specific story crystallized documentary-film maker Robert Drew's attitude toward the television of the early fifties. Covering for Life Magazine the same stories that television was covering, he marveled at the difference. He was getting real power out of the stories and still pictures while television was getting very dull stuff. He wondered why. A few pictures would capture all the drama. After comparing the results regarding a story about the U.S. government breaking up the DuPont company, he realized that he had been able to get pictures because, he could take a camera anywhere. He finally said:

P.J. O'Connell (1992)
Robert Drew and the
development of Cinema
Verite in America

"It's just a case in which I am almost revolted, and I'm saying we've got to do something about this, we've got to make television system work to capture what's really happening!

I realized that here was a place I could make films. That is, television needed to look into things that tripods and [the film making] systems wouldn't allow at the time."

For Drew, technology was a way to achieve the specific goal of more easily capturing drama by allowing more mobility. Today, digital technology gives even more flexibility to reporters as shown during the Gulf War that was covered in real time. Digital technology also has the tremendous power to allow a sophisticated dialog with the viewer. As information flows into our screen, we need to understand how to satisfy the increasing need for customization of content or what Alvin Toffler calls "demassification".

A.&H. Toffler (1993)
War and Anti War, Survival
at the Dawn of the 21st
Century

2. Viewpoints on demand: the big picture

Today we are informed by subscribing to a newspaper and chasing information here and there on TV and radio. It is a rather *vertical* approach to news that ties one mostly to given newspapers or TV shows that you have discovered and liked. Nobody has the need, time nor money to read all printed newspapers and tune into all broadcast shows. What if there were a customized and easy *horizontal* access to news? What if you could access various interpretations of the same event or events that concern *you*? This means servers of viewpoints accessible to you or your digital agent.

Inevitably in the next 5 to 10 years there will be enough bandwidth on large portions of a computer network like the Internet to carry a video sig-

nal. The Internet already serves as a model for communication and publishing. Information is made accessible to Internet users (published) on servers from which it can be retrieved on demand, or it is broadcast digitally to members of discussion list and readers of electronic publications.

Retrieving viewpoints on a given issue from an online video library will be similar to checking an *encyclopedia* that has a *perspective.* On the one hand one will have the benefit of storytelling and perspective as in a newspaper, magazine or documentary. On the other hand, one will request expert answers with any depth on a given issue as in an encyclopedia.

The ultimate goal is to create virtual characters that embody the perspective of an expert and answer questions (or argue with you) on the fly without pre-recording answers. This means being able to represent someone's viewpoint and knowledge of the world and then have the system infer from that an answer. What if you could request from a virtual representation of Marvin Minsky or John F. Kennedy their view on war-related research and the conflict in Bosnia? Marvin Minsky was probably never interviewed on the topic and JFK is not available anymore for questions. However both have a certain way of approaching an issue, a certain style, various principles and opinions -- a viewpoint in short -- that could be applied to a new situation.

3. Viewpoints on demand: the specific problems addressed

In this thesis, I have addressed one specific aspect: how to dynamically personalize the presentation of a story to put the emphasis on a given viewpoint. At a more global level, it requires an understanding of how to reshape a documentary on the fly as the story unfolds. How can a video story be transformed to adapt to external constraints applied to the content?

An appropriate paradigm might be the smart VCR in the sense that it helps control the linear presentation of video content by making the machine smart about the content itself, not just the physical medium it is recorded on. A standard VCR only lets you fast forward, pause, stop and rewind. A videodisc player will do better by letting you directly access a

chapter or frame number without having to manually search for it, but it still doesn't know anything about the semantics of the content.

"Viewpoints on demand" specializes in the orchestration of a documentary (versus the orchestration of fiction). It has some knowledge about how to rearrange a playback list presenting issues and opinions. It also has knowledge about the basic structure of the arguments recorded by the author during the interviews. This provides the fundamental basis for alternating a linear mode of presentation with a branching mode in which you follow one line of argument or another.

4. Evolving documentaries

Since the point is to change the viewer's experience of a story in a very dynamic way, my approach was to introduce the notion of "evolving documentaries." An *evolving* documentary is a documentary that can adapt to the viewer's input parameters. Parameters are entered at the beginning but can also be changed as s/he watches the story. The story presented will unroll differently depending on their values.

The concept of "evolving" also implies that one starts with a story that will grow into something different, maintaining sequences of the original story but also changing or replacing others. First, this means that the system needs to keep track of what the viewer has seen before a change in parameter values. Second, it means that the system needs to change its plans about what it was going to show next, adapting to this new context.

An evolving documentary can automatically take into account additional content as it is included in the database over days, weeks, months or years. The video annotations I use are such that it is possible to keep adding viewpoints to the database [see chapter on the author]. Thus new clips are added where relevant to one story or another.

In order to develop a system in which the playback adapts very dynamically -- as I just described -- it is necessary to define 1) how the story is generated by the computer, 2) how the multiple constraints defined can be taken into account [see the chapter 4 describing the internal engine and the chapter on creating the database and the role of the author].

Finally -- and related to this -- it is necessary to define an interface for the viewer to control the settings constraining the playback and to control the

progression in the content. There are many ways of applying the principles described.

In the fall of 1993 I gathered information and interviews about the role of information technologies during the Gulf War. This original material was used to create a database of viewpoints and to develop an authoring model for such databases. The model of interaction for the viewer was based on 1) defining levels of interest in viewpoints and issues that can be changed as you watch 2) defining levels of interest in a given headline 3) content-based playback controls 4) browsing arguments. One important part of the work was to create an engaging interaction for the viewer. Another important part was to understand and rethink the role of the author in a personalizable environment.

The rest of this thesis is composed of eight chapters: *chapter two and three* analyze the roles of the viewer and of the author, *chapter four* explains how the engine works selecting the clips and ordering them, *chapter five* is about the content itself (the role of Information during the Gulf War), *chapter six* relates this thesis to previous work, *chapter eight* is dedicated to future directions for this work, the last chapter concludes.

2 *The Viewer*

Presenting Content to the Viewer:

User Interface & Consumer Behavior

Customizing viewpoints & issues

with content knobs

"No interface at all is better than any interface"
                        Jean-Marie Hullot (NeXT)
                        Steve Jobs (NeXT)
                        Nicholas Negroponte (MIT Media Lab)[1]

This chapter takes the point of view of the viewer, explains the design issues about the interface and describes how it works. The first section tries to support five points that determine the model of interaction described in the second section. The third section describes the actual scenario of an interaction. The last section is a series of Questions & Answers.

The five main points supporting the model of interaction are:
1- Smart agents will contribute to eliminate unnecessary interfaces.There still will be a need to control and shape a given program and that task will be handled by interface agents.
2- The common assumption that home and work products should obey different rules of design is wrong.
3- The combination VCR/remote control is a first attempt at constraining the flow of video. It is limited because it lacks knowledge *about* the con-

---

1.   I heard this first from Jean-Marie Hullot around 1987, then from Steve Jobs and finally from Nicholas Negroponte. It is not clear which one influenced the other. Any one of them could have come up with it.

tent itself.

4- Second person voice is good for accessing information. First person participation by the viewer is more relevant to fiction.

5- Content knobs allow linear media to become interactive

1. Design Issues

Nigel Woodhead (1991)
Hypertext & hypermedia,
Chapter 8

The goal of the interface of "Viewpoints on Demand" is to give control over the content without being in the way. *Given a program that is of interest to the viewer,* the interactive application of various constraints gives greater control without requiring extra mental overload as it is often the case in interactive media (Woodhead 1991). It tries to combine power with simplicity.

"Viewpoints on Demand" is a framework for a consumer product and more specifically designed for interactive television. It takes advantage of the processing power of the computer that runs it to provide a rich functionality (the "dumb" TV sets that we have today do not allow this).

1- Do we need an interface at all?

When you plug your computer into a network, the last thing you want to think about is how to configure your modem. Clearly you would rather not deal with that interface. The same happens with content. The more information is available "at your fingertips", the more you are at risk of engaging in various interactions that you would rather avoid. This is why agents are so seductive: they search, select and shape information for you.

If we assume a world in which agents preselect content for us, what interface do we need? What type of dialog do we want? How do we manipulate this personal channel bringing us the content that is most likely to fit what we need and enjoy?

For "Viewpoints on Demand" the hypothesis is that some smart agent has identified an interest in a topic as for instance information technologies, warfare or the Gulf War itself. The selection is listed in the viewer's personal TV Guide. If the viewer happens to actually select the documentary, another storyteller agent acting as an interface agent can then customize the presentation itself. By default it might present a viewpoint that is most likely to please the viewer. However there is still a possibility for him/her to switch viewpoints at any time during the documentary.

"There is always an interface. Some we identify as such, some are so completely part of our vernacular that we do not think about the interface as something other than the thing itself" (Glorianna Davenport 6/93). The current interface for "Viewpoints on Demand" uses a mouse for input and gives feedback visually. Having a conversation instead might feel more natural. Feedback could then be verbal or visual. In all cases the function offered by the interface does not change. The interface still exists.

2- Work and home products don't obey different rules of design

There is a common assumption that in order to understand customer behavior with technology-oriented products it is essential to make a distinction between the home and work environments. A typical argument is that at home we don't have the strong motivation that we have at work to accept complexity. At home we want to relax and use simple and useful or entertaining appliances like food processors or television. At work we have a task to accomplish and we need to be productive.

Paul Saffo (1992)
Consumer Technology
Purchase Behavior in the
1990's - IFTF

Paul Saffo from the Institute for the Future sees a dichotomy between the consumer electronics industry and the computer industry (Saffo 1992). According to him: "We use tools to accomplish tasks, and we abandon them when the effort required to make the tool deliver exceeds our *threshold of indignation* -- the maximal behavioral compromise we are willing to make in order to get a task done". An average consumer has a low threshold while a teenage hacker has a high threshold.
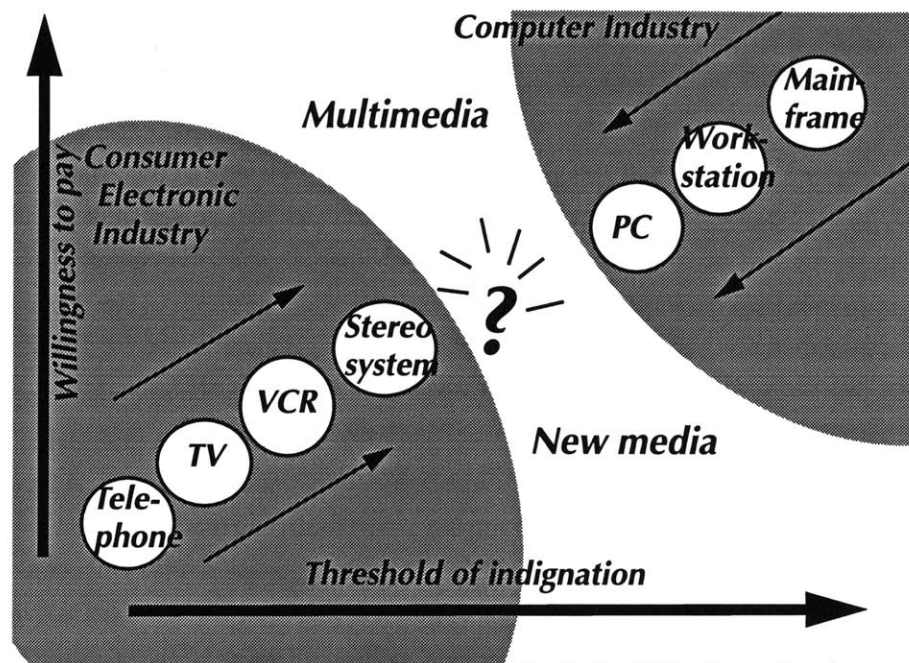
Figure 1 - Paul Saffo's model for purchase behavior in the 1990s (reproduced from his paper)

Saffo sees a trend in which consumer products are becoming more and more complex (VCR, stereo systems) while computer products are becoming simpler (from B-2 bombers to mainframes to PCs).

He concludes that consumer electronics players are more likely to fill the gap between these markets, because it's easier for them to go towards more complexity, and they know the consumer electronic market better. "Consumer thresholds tend to rise over time…Each industry is finely tuned to the specific threshold of indignation of its core customers, and to the price point of its mainline products".

The major problem with these approaches is the distinction between work and home markets or consumer electronics and computer markets. *The customer does not become a different person when s/he comes back home.* More and more people are working at home, using their computer for entertainment as well as for work. The den and the living room are merging. *The real issue is that when we buy and use, we need a reward for our effort, time and money.*

At work, the difference is that we might not have control over what we use.

The reward can be getting work done or being entertained. It does not matter to the customer where the product comes from (computer industry or consumer industry). Increasingly more people use on-line services like America Online or Compuserve, because they have a strong need to communicate. They are ready to go through the technical difficulties inherent in getting connected.

This doesn't apply just to consumer electronics but to all human activities. Hiking or windsurfing also require a tremendous investment from those who decide to participate in them. They do it, because the expected reward is high.

Upside, April 1994
An interview with Eric Nee
of Scott Cook

An essential difference between the consumer industry and the computer industry is *the willingness to take into consideration consumer feedback*. Consumer products are extensively tested and studied in surveys in order to track reactions from customers. Computer products don't go through such thorough testing (see interview of Scott Cook, CEO and founder of Intuit which dominates the personal finance and small business software market). High tech products like the Newton or the NeXT computer did not pay sufficient attention to the customer feedback when they were first brought to market. This is not to say that the technology itself was not good, especially for the NeXT: while the first Newton suffered in its ability to transcode handwritten input which had been marketed as a key feature, the technology of the NeXT was sound. Rather, the market reaction reflects that some key aspects which matter to customers were not taken into account; as a result, it was impossible to turn high quality technology into a successful product.

Work and home products obey the same rules: they must answer a need and provide a reward worth the effort and the cost.

3- VCRs are a step in the right direction but they lack knowledge about the content

VCRs (or video cassette recorders) are basically affordable consumer electronics devices that are hooked to a television and have many interesting features:

• you can record a TV program or movie on tape which implies in turn that:

      - you can watch a program when you want to watch it
      - you can watch it many times

- you can lend it to friends
- you can stop, pause, fast forward and rewind

• you can also manage to remove commercials or other inserts which you do not want in the recorded version.

In fact VCRs are designed to give the users more control over a content presented linearly in time and more freedom by not constraining them to a schedule. These are general rules that can be detached from their actual implementation in a VCR and from the actual hardware.

Videodiscs were supposed to bring to the consumer even more control but the technology never caught on as a consumer product for reasons that I will not debate here. In any case, they provide:

- direct access to a given chapter or frame number
- ability to step through the content frame by frame
- fast forward and rewind by scanning video
- record video but only if you can afford the equipment which is extremely expensive (professional use only)
- some basic programming functions

Another popular device, the remote control, also gives you control over video content (in its TV form) by letting you "channel surf" either to browse whatever programs are on at a given time or to skip commercials.

The advantages of these devices in controlling *linear* presentations should be preserved. VCRs and remote controls are very primitive tools for making TV watching more personal as nobody will watch the same sets of programs at a given time in the same way. One might even say that they are very primitive tools for editing content. How can we reinforce that strength? What all these popular - or not so popular - devices don't have is any *knowledge about the content.* As a consequence they are completely unable to customize content in any way other than with the basic controls provided.

4- Tell me a story: second person is good for accessing information

Brenda Laurel suggests an R&D agenda for interactive movies which calls for first-person participation from the viewer. This makes sense in the world of fiction where the plot may change but does not apply well for the world of news and information.

Brenda Laurel 1989
A Taxonomy of Interactive Movies, New Media News, Vol.3 issue 1

What she seeks to solve is the problem of *"creating compelling interactive plots"*. Her solution is to "endow the system with dramatic expertise and to give it the ability to manipulate, through an internal 'action calculus', the shape of the whole action. Such a system would incorporate the brainpower of a good playwright, writing his script in real time, where one of the characters literally has a life of his own".

With news and information, you do not expect to modify the events you want to be informed about. You want to be able to experience them as an observer, as closely as they actually happen. A second-person type of interaction (or no interaction as all) is more relevant in this context. By no interaction at all, I mean that agents that know you would search information on your behalf.

That is not to say that you don't ever want to simulate your participation in real world events. However, the goal is then different: you want to explore "what if" scenarios, observe the outcomes and eventually compare them to what actually happened. This is extremely appealing for education purposes. If schools had "holodecks"[1] that could simulate the French Revolution, it is expected that more children would, for instance, be more receptive toward history.

5- Content knobs allow linear media to become interactive

We already established that VCRs would benefit from meta-knowledge about the content to adjust linear narrative. What type of interface would allow this?

Before answering the question we should keep in mind that there are alternatives to linear playout of motion picture sequences. We need to understand the type of interaction they provide. Most of interactive media has been based on models in which the experience is constantly interrupted ("suspension of disbelief"). How to characterize the various alternatives?

Brenda Laurel proposes in the same paper mentioned earlier a taxonomy of interactive movies which distinguishes 3 basic forms: *navigational* (like movie maps), *narrative* (like a branching story) and *dramatic* (in

---

1. Virtual reality system used as plot device in a TV show called "Star Trek, the Next Generation".

which the user acts as a real character). Each form is defined by 3 variables:

activity: explore
control
enact

interactivity: frequency (of user intervention)
range   (how much is the user constrained)
significance (effect of a user's choice)

personness: first
second
third

MYST, Broderbund 1994
Rand & Robyn Miller

The Journeyman Project 1992,
Presto Studios

Video games have used physical spaces to take the player through a maze with different dangers to avoid or obstacles to get around. The better you are at this exercise, the better your score. Adventure games like The Journeyman Project or Myst are examples in which the player is given a goal and then explores a space in order to reach that goal. They belong to the navigational form. Hypertext systems also exploit a form of navigation using words as branching points.

Nigel Woodhead 1991
Hypertext & Hypermedia,
Theory and applications

A major downside of branching and navigational structures is that they usually create interruptions by forcing the viewer to constantly make a decision at each node in order to make progress in the narrative. When a given node is reached, there is very limited information on the next possible node. Another major downside is that the user can easily get lost. Many games deal with this problem by providing a map of the space.

Jeff Conklin, 1987
"Hypertext: an introduction
and Survey", IEEE Computer
Vol.20 No 9

This creates another incentive to go back to the linear model that movies and television traditionally have used. Let's answer our initial question: what interface can we build for it? Under that model, content is organized in time rather than in space. The innovation (and answer to the question) is to change *qualitatively* and *quantitatively* the playback by taking into account user input with "content knobs". Nicholas Negroponte in the late 70's had the idea of building a TV with content knobs to control in particular the levels of sex and violence in movies in the same way we can control today the volume level. This is what Ryan Evans calls "fluid interaction" and it has its origins in constraint-based editing.

Ryan Evans, 1993
Logboy and Filtergirl, MS
MIT Media Lab

In this model there are no predefined threads. The final arrangement of the shots (or clips) is the result of applying various constraints to the content and high-level organizing rules. These help select and order the shots (see chapter 4 for more details).

Benjamin Rubin, 1989
Constraint-Based Cinematic
Editing, MSVS MIT

This fits in the narrative form as defined in Brenda Laurel's taxonomy:

• it offers a control *activity* in which the viewer controls the resulting story with content knobs.

• the *frequency of interactivity* depends on how often the viewer wants to modify the parameters of the story; the *range* is constrained; the *significance* is great but not maximal, as switching viewpoints on demand changes the way the story is told but not the events themselves (which is in fact the whole point). Maximal significance would go against the goals of the system.

• The *personness* is second.

2. The resulting model of interaction

As a result of this inquiry, I have built a smart device, a sort of smart VCR, that can adjust linear content based on the various points stated in the previous section. The interface uses content knobs to offer more control over the content without getting in the way when the viewer experiences the story. It provides the means of controlling the selection and shape of the *content* of video pouring on the screen. Video logging makes it possible by providing the system with *information about the content*. Stories are "advertised" to the viewer using **headlines**. The mode of interaction is second person.

In chapter 7 I propose a specific type of branching structure based on text connectors (like BUT, AND, OR) that can be super-imposed to navigate through opinions at the shot level, the shot becoming a node.

The following section describes an actual scenario of interaction. I designed it to be used more specifically in the context of presenting viewpoints on given issues.
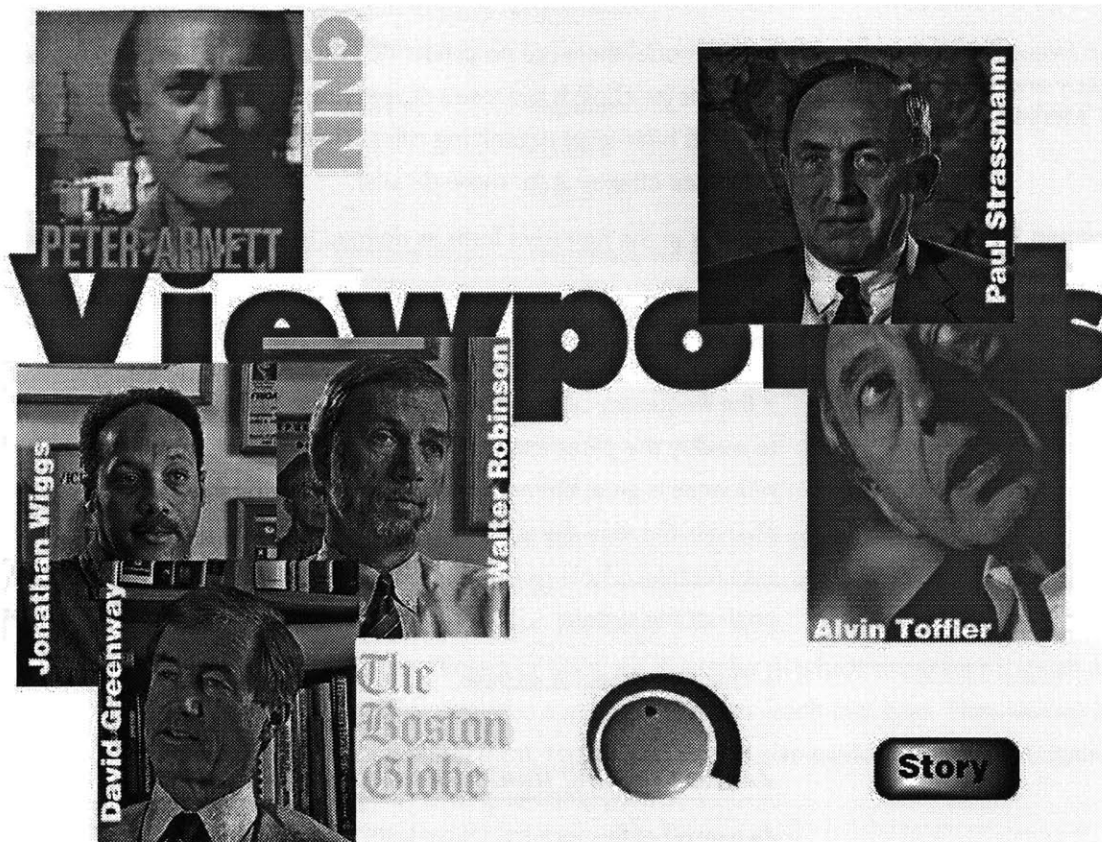
Figure 2 - In the first screen, the viewer attaches levels of interest to one viewpoint or another by using the grey knob. Each photo represent a viewpoint. The size of the photo changes according to the level, giving visual feedback.

3. Scenario of an interaction

The components of the interface

In the current version of the system, the viewer accesses two screens. The first screen [fig.2] displays the different viewpoints available in the database in the form of the photographs of the various people interviewed. The second screen [fig.4] displays the headlines of stories that can be told.

The viewer drives the system by adjusting a "content knob" which defines a level of interest for a given viewpoint and a given story. When the level is set to the maximum, the photo or the text selected become large and vice versa while the others become smaller. This gives immediate visual feedback. Clicking on the "story" button triggers the engine that selects and assembles the clips. When this process is over, a smaller panel pops up [fig.3] which includes the actual video screen and control buttons.

Figure 3 - Once the story is generated, a small panel pops up.

The control buttons look like VCR controls:

play
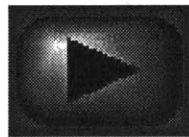
stop
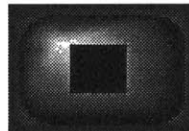
next speaker

previous speaker
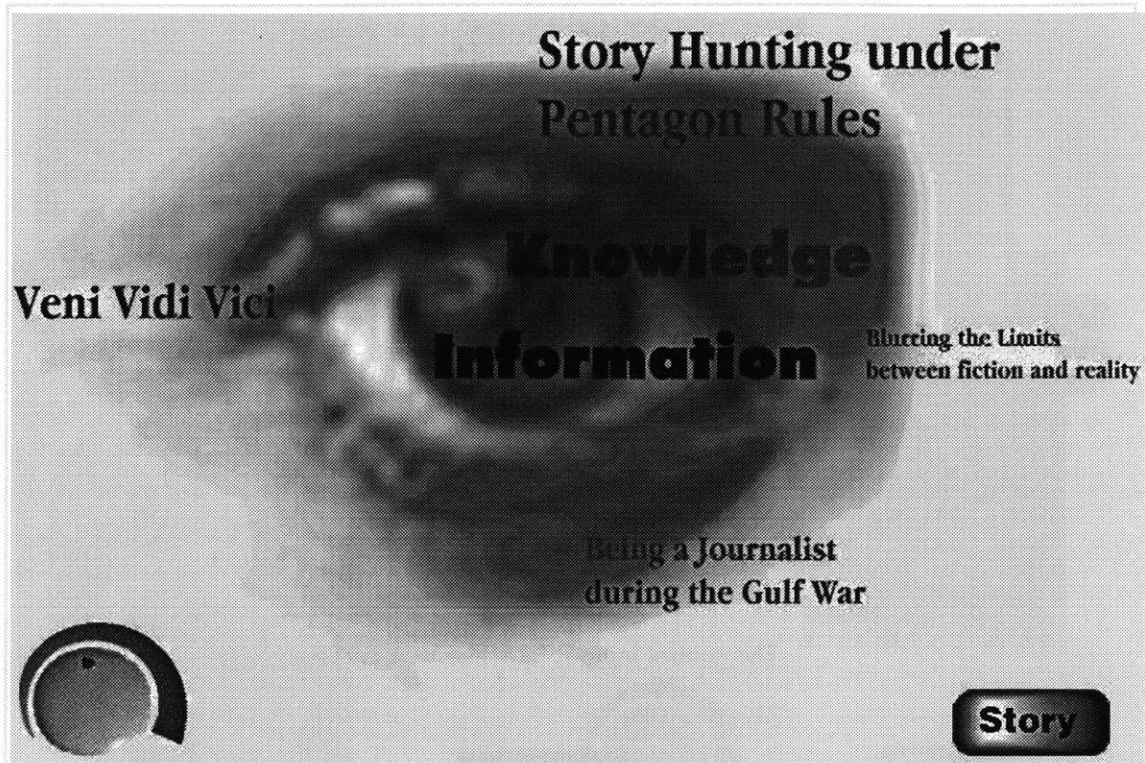
Figure 4 - Second screen available to the viewer. It shows the headlines of the various stories available. The size of the headline changes as the knob on the left is turned up or down, indicating the importance assigned to it by the viewer. The story button on the right side is used to trigger the generation of the story.

See [Chapter 7: Future Work on extensions] for additional-content oriented control buttons.

Stories available to the Viewer

Theoretically with infinite server size, there could be any number of stories. It would only require agents to select a subset. The current system has four.[1]

Introducing the stories and the viewpoints

How should the viewer interact with content in a television story space? Symbols in the Viewpoints screen - like the Boston Globe logo next to the

---

1. There were originally nine stories available to the viewer. A technical problem that is still needs to be resolved brought the number down to four stories. Quick Time - the Macintosh framework for handling time-based data like video - seems to put a limit of about 48,000 frames that can be handled at once in a movie
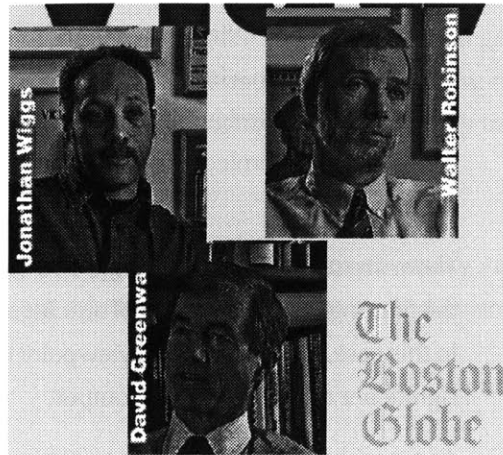
Figure 5 - The Boston Globe logo gives a hint of where the people interviewed come from.

journalists interviewed - give a hint about the person interviewed [fig.4]. Starting with a short version of one of the stories gives the viewer a sense of the contents but does not give background on the various speakers. It is as if you were watching a trailer.

Dynamic changes of content

By hitting *play* the viewer can watch the story generated based on his or her input. If the settings are changed - in particular the viewpoint - the selection will be different, but in the current version overlap cannot be avoided as the system does not keep track of what the viewer has already seen. In particular the fact that the narration remains the same, makes this limitation more apparent.

How different will each playout be as the user changes the settings of the knobs? There are various content-related factors:

- how well a viewpoint is represented in one story (in short how much footage is available).

- how key is one viewpoint versus another in a story (a native from Ecuador can hardly have a position on how life is in Alaska unless s/he lived there). If the viewpoint of a key expert in a story is eliminated, the story might lose its interest.

For instance, in the specific database I created, while all viewpoints are related to the role of information during the Gulf War, different opinions stress different aspects. Two major views are information as seen on the

military side and information as seen on the media side. An expert from the Pentagon has little to say about how the press used satellites but he has in comparison a lot more to say about information warfare. As the system is currently designed, if his/her viewpoint is "turned up", journalists views will still dominate in stories about press and technology for instance.

There are some issues where the two views intersect. A good example is the filtering by the Pentagon of articles and images sent back by the press. In this case, focusing on one viewpoint or another is more likely to trigger significant differences in the output.

As a conclusion, the best way to generate interesting differences in views on a story is to have a well-balanced database in which all perspectives are equally represented. This is not always possible in a practical way - even if you are concerned by fairness - and the viewer should have some visual feedback about the relative importance in size of the different viewpoints.

4. Questions & Answers

Why would someone use such a system?

In the form implemented in the current system, the concept of "Viewpoints on demand" means that you can adjust the viewpoint on a given headline with a given database of interviews.

Pushed to an extreme, it will let you browse the press horizontally and with greater flexibility than it is today. Even if we do get information from multiple sources on major stories, it is still not quite convenient to get a global view of a specific issue across the press. This is a job for agents who can cruise the networks looking for interviews or columns giving a certain perspective on a given issue.

This is currently almost impossible without a serious investment in money and time. In addition, you need good guides to tell you what to check and where (currently humans or magazines serve this function in a limited way).

Does this approach prevent bias?

The approach does not prevent bias. Bias is embedded in the images, the

story structures and the selection of the main topic and of the viewpoints presented. However, this approach can get the audience to become more conscious about the world and the many ways it can be seen. Or one might want to have stories told only in a certain way, with a certain style we like.

Is it different for video or text?

Tailoring viewpoints remains obviously valid for text. It requires different solutions to select, organize and assemble the content.

3 Presenting and Representing Content by and for the Author

Sharing Control with the viewer

"Distributed" Design and Graphical Annotation

1. Redefining the role of the author in personalizable media

... I have brought them as it were into a continued relation, which as I take will be the pleasanter, because you need not trouble your remembrance with looking backe after former matters.

Thomas Gainsford (1623), considered to be one of the first news editors

Bender et al., 1991
Newspace: Mass Media and Personal Computing,
USENIX conference

J.T.Harrigan, 1993
The Editorial Eye SMP

A major research area in electronic publishing is "The Daily Me" (Bender et al.) or how to adapt content to the needs and interests of an individual. This implies collecting information from various sources and doing the layout *automatically*. A great concern in the press is to understand how to redefine the role of the author, and particularly the editor, in helping shape the information delivered to the reader or viewer.

The job of the newspaper (or magazine) editor involves identifying the most important stories according to the interests of the audience (or average reader/viewer). An editor also takes material brought in by report ers and puts it in perspective, fitting it into the overall layout designed to better communicate the message.

The focus of this chapter is to give one example of collecting and editing information for an interactive personalizable medium: what are the tools needed and how can the editor apply a certain pattern while allowing freedom for the viewer? A plausible methodology seemed to be going through the actual process of identifying a potentially interesting con-

tent, investigating it, gathering interviews and shaping the result. Chapter 5 gives more detail about the content itself. The present chapter concentrates on shaping the result and defining the tools. The first section describes in detail the design of the authoring tool used. The second section describes how the tool was used.

Looking for Images

As it turned out, one of the most difficult tasks was to find images to illustrate the interviews. Not only was it of course very expensive to get stock footage but almost impossible. The Gulf War was too recent to be archived by specialized companies. Networks wouldn't have the resources to do a search in their own archives. Not to mention that I would not be able to do it myself because of union rules.

But the major obstacle was that the footage I was looking for just plainly didn't exist. The cameras were not pointing at what I was interested in. Satellites might be the only exception. How could so many troops move so fast in such a short amount of time? The fact that GPS systems provided a mobility essential in the massive move conducted in one night to take the Iraqis' rear by surprise was completely overlooked.

I managed to gather a large amount of images from TV shows, video tapes I rented and a videodisc from CBS/Fox Video. Information technologies didn't get very much attention. A large part of what I saw where images of weapons, planes, ships, tanks. Eventually I found clips of troops using portable radio phones and computers. The Pentagon's public relations office sent me a tape on GPS technology which was very useful. However, I was not - for instance - able to get images of the phone booths set up by AT&T for the troops to call home ("morale calls"). Reusing existing news reports was out of question for the same reason.

This was of course frustrating but it was also stimulating as it reinforced my idea that the subject was original and deserved attention.

2. Representing content: designing a tool for the author

Background

I needed a tool to create a database, represent video content (in particular opinions), story structure and eventually rules (these ended up embedded in the software instead). This form of video annotation (or logging) is seen

here both as a tool for the author to think with and a mechanism for the software to gather information *about* the content. I used a tool implemented by Mike Murtaugh, ConArtist.

M.Murtaugh, 1994
ConArtists, a System for Graphically Representing Story Knowledge, MIT

Mike started the development of this tool in the summer of 93. Until May of 94 we went through many versions of the database as I was providing feedback about the needs and as Mike was coming up with new interface designs. In January of 94 as my interviews were all digitized, I started working on my "real" database and we had to freeze the design of ConArtist. A major obstacle had been that copying one frame from one location to another would take several minutes. Saving and loading the database would take up to 10 minutes. This was optimized in the version I stuck to.

K.B.Haase, 1993
FRAMER Reference Manual Internal Report, MIT Media Lab

ConArtist was written on top of Framer, programming language designed to be used for knowledge representations. In the manner of object oriented languages (OOL), Framer uses a hierarchy of "objects", known as *frames,* but it does not apply the classical class/instance model. In the OOL model, classes describe the general attributes of an object. All the instances of a class share the same attributes (color could be an attribute of the class car). Classes at the bottom of the hierarchy inherit attributes from upper classes (*Mercedes* would inherit from *car*). In Framer, attributes of a frame are other frames (the frame car would be annotated by the frame color which would in turn be annotated by the frame brightness). Inheritance can be provided by specifying certain frames as being prototypes: several frames can then inherit from a common prototype. Framer has also built some inference mechanisms that make it particularly well suited for Artificial Intelligence applications. Finally all Framer objects are *persistent* which means that they can be stored in a database file and reloaded.

A graphical interface to "draw" annotations

In ConArtist all video annotations are graphical. Taking advantage of that I defined a graphical language to categorize the content, describe relations between clips and the structure of the story itself. It is a language using links and nodes. It is applied at two levels: to define a story structure and to define relations between clips.

"It was a Knowledge War"

Contents | Settings | Links

Viewpoints

Basic Topics

Commercials

Prototypes

Editing rules [1]

Demo frames

Editing rules [2]

Story Hunting under Pentagon Rules

Veni Vidi Vici

Being a War Reporter in the Gulf

TV News: Blurring the limits between fiction & Reality
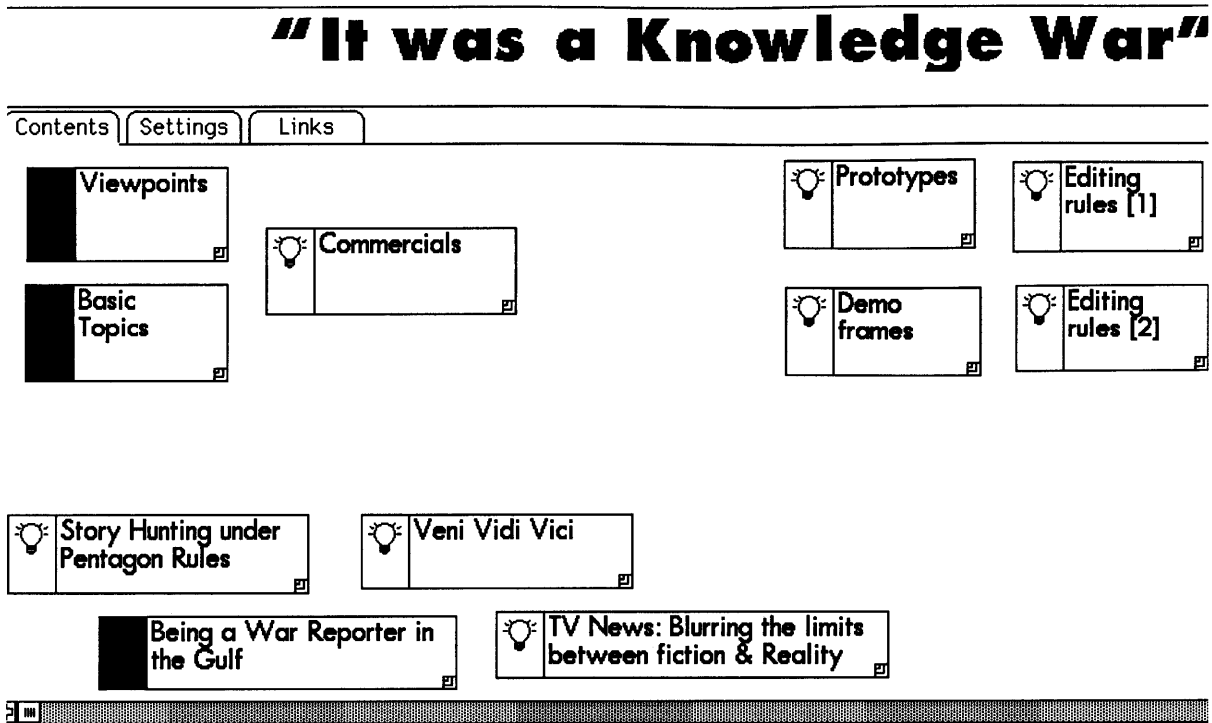
Figure 1 - Database main screen • Shows the frames at the top of the hierarchy

The nodes in a story structure are the "chapters" of the story. The links are potential links between chapters which come from the catalog [fig.2]. At the clip level, the node is the clip itself (each one is a piece of answer from the interview). The links are the various links of the catalog. This language can be seen as using video clips as pieces of sentences, logical links as connectors between these pieces. Sentences are grouped in *chapters.*

Content is also categorized hierarchically. Video clips are the leaves of the various trees (hierarchies). There is one hierarchy of viewpoints and one hierarchy for each story created (plus one hierarchy for commercials).

The interface takes advantage of the possibilities offered by ConArtist: objects and links between objects are drawn in the same they are in a program like MacDraw. The user simply selects a type of object (MOVIE frame for instance) in the palette and then draws it. Links between objects can be drawn as easily. Objects can be contain sets of objects recursively.

The following are the frames that appear in the front screen and that are at the top of their hierarchy [fig.1]:

VIEWPOINTS frame: contains all the clips categorized by viewpoints and
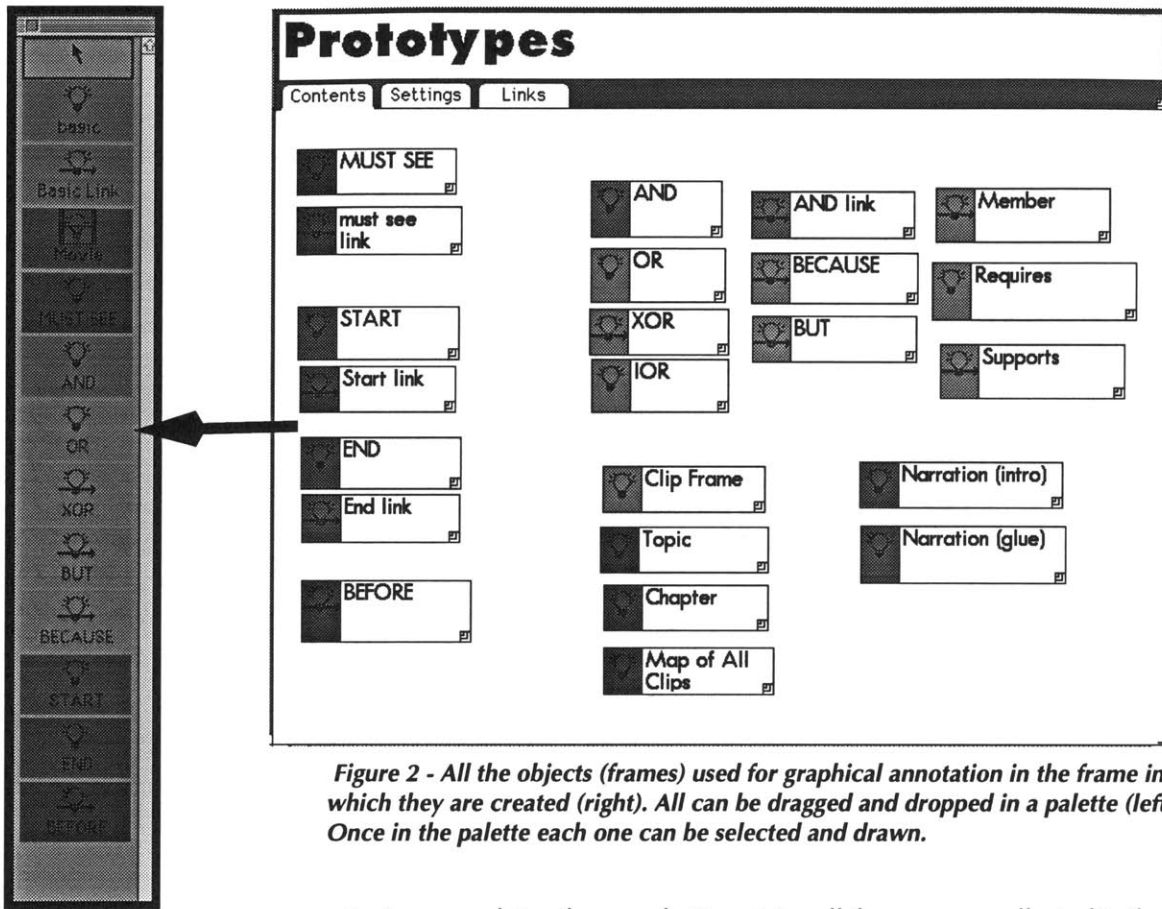
Figure 2 - All the objects (frames) used for graphical annotation in the frame in which they are created (right). All can be dragged and dropped in a palette (left, Once in the palette each one can be selected and drawn.

topic covered. In other words, it contains all the answers collected in the interviews. Answers are grouped in topic frames that contain in turn the clips.

STORY frame: contains the frames needed for the design of the story as well as the graphical description of the story.

COMMERCIALS frame: contains all the commercials available by category.

PALETTE ITEMS frame: contains all the graphical symbols that are used to annotate the video by tagging, grouping clips and drawing relationships between them.

3. Definition of a catalog of graphical annotations

This section describes each notation used: links, containers or labels. They are all created and stored in a special frame [fig.2]. Once created or updated ConArtist lets you select them and drag them onto a palette

to make them more accessible. Each anotation described is a framer frame.

"Editorial" annotation
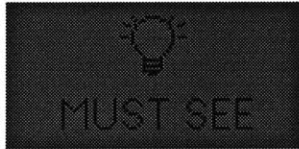
MUST SEE (frame & link): indicates that whatever the other constraints are (time, viewpoint), this clip has to be selected. It is a simple but important tool for the editor to provide perspective: it means that according to his knowledge of the content, this is an information that has priority over the others. It could be because it's an important piece of information or eventually because it's particularly funny or convey very well the personality and position of the speaker.

Structural annotations

Structural annotations help the system identify the various objects (basic objects and sets of objects) it needs to work on as well as how they relate to each other.

MAP of ALL CLIPS frame: is used by the author (and in fact ignored by the program assembling the story) in the design process to import all the relevant clips from the Viewpoints frame (where they are originally stored). Clips can then be grouped and organized spatially on the screen in order to better visualize how they can relate to each other. Making the database system smarter about this layout would make this task easier as the screen becomes easily crowded.

CHAPTER frame: is the principal device used to structure a story. It is used to define the clips available for each step in the story.

NARRATION frame ("intro"): used systematically to introduce a chapter. It is equivalent to the "lead" in a newspaper story.
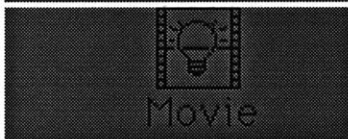
Note: Narrations are repeated even if the viewer has already heard them. [In chapter 7a solution is proposed to avoid this and make the system evolve better as settings are changed].

NARRATION frame ("glue"): used when narration helps make a connection between two sequences or groups of clips.

TOPIC frame: in a chapter, groups clips on the same topic. For instance all clip citing "heavy bombing" as an explanation for the short duration of the war.

MOVIE frame: contains the actual clips that can be played back and all the information about the clip. In particular other places (frames) where it might be used.

CLIP frame: contains a movie frame and provides a text summary of its content.

Scheduling annotations

While a set of rules are embedded in the software in order to handle the ordering of the sequences, it is sometimes useful to impose a certain ordering:

END: To make sure a given clip is the first one shown in a chapter (if necessary).

START: To make sure a given clip is the last one shown in a chapter (if necessary).

BEFORE: In some cases, an order between clips has to be imposed. A typical example is the following: the speaker uses a pronoun ("*It* was important", "*she* knows what is important"). It is necessary to explain first to the viewer what "it" and "she" are referring to. [see chapter 7 on Future Work for a possible way of generating this automatically].

Logical annotations

Logical annotations serve two purposes:
1) to give the scheduling/selecting program an indication about how clips relate to each other logically.
2) to suggest to the viewer possible ways of branching to follow a thread of opinions or get additional detail.
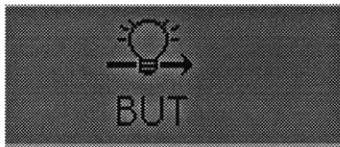
BUT link: links two clips showing opposite arguments, a contrast between opinions or situations:

"The American army uses a more democratic communication system than the Russian army" **BUT** *"the Pentagon set very strict rules to censor the press, preventing the American public from being completely informed"*

"It was very difficult to file articles" **BUT** *"The military offered their planes to send back to the US papers and photos"*
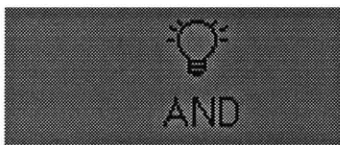
AND frame: is used to group clips for various purposes

OR frame: is used to tell the program that two clips are "equivalent". They can be equivalent because:

- they bring up a similar argument (for examples two clips stating that the Gulf War was finished quickly because of heavy bombing). One can be used to reinforce the other.
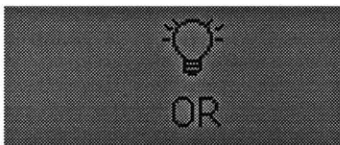
- they show alternative views. Two people might agree for different reasons (for example a TV journalist and a photographer explaining how they both experienced the pools)
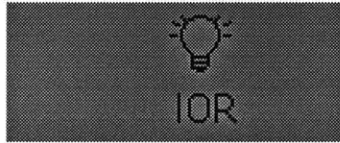
Note: in a chapter, the default is that if there are no visible links between clips, the clips are connected by ORs.
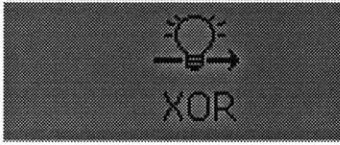
XOR link: is used to tell the program that it can use either one clip or another but not both.
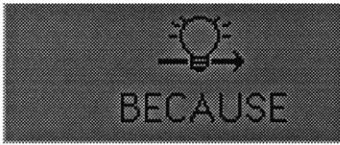
BECAUSE link: links a clip stating an opinion or describing a situation to a clip explaining the situation or justifying it.

"It was very difficult to communicate in Kuwait" **BECAUSE** *"all communication lines were broken and journalists had a difficult time finding phones".*
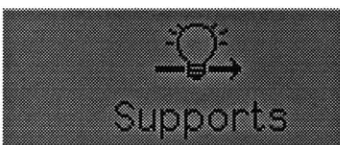
SUPPORTS link: links a clip that supports an opinion or argument stated in another clip

"There is a sort of canard in the military that the press was responsible for the low morale" **SUPPORTS** *"What counted was the amount of casualties as it did in Korea"*

Note: When focusing on a viewpoint, other viewpoints might be selected as well that support it at least *on some specific points*. It might be *composed of a collection of individual views* that converge (right wing view,

left wing view for instance).

4. Inspiration from print media

Much of this authoring tool is inspired by print media.

Stories are announced with **headlines** asin a newspaper. In an earlier design, the viewer would specifiy a story by selecting a combination of words scattered across the screen. These keywords would then be used by the system as ingredients to produce a story. This was abandoned as keywords would lack context and the risk was that the viewer would select randomly words. This would not give the user a sense of being in charge. Headlines give the viewer a better hint at what to expect.

Stories are decomposed in **chapters** like a book. Chapters make it easier to set up interactivity;: the author uses them to define the general structure of a story; a chapter is like a box in which relevant clips are dropped; the final ordering and selection is not decided by the author (but the story structure *constrains* the final output).

An important result of experimenting with the interactive shaping of the content was that customizable content calls for "distributed design". By distributed design I mean that a clip might end up used in different contexts in different stories. It might have various meanings depending on the viewpoint. As the content was to be tailored to the viewer on-demand, I needed to be able to *visualize* various possible results. I needed also to instantly jump from one set of clips to another as the first set would inspire ideas for the second. ConArtist and Framer allow multiple copies to be updated in parallel.

These are characteristics of design general. When editing a traditional movie, an editor might want to visualize multiple possibilities prior to fixing a sequence. In non linear media supporting this process becomes a necessity as the final choice is specified by the actions of the user.

In this system, the editor does not focus anymore on what is the best final edit but s/he defines a general shape or structure for the stories and how they need to be constrained.
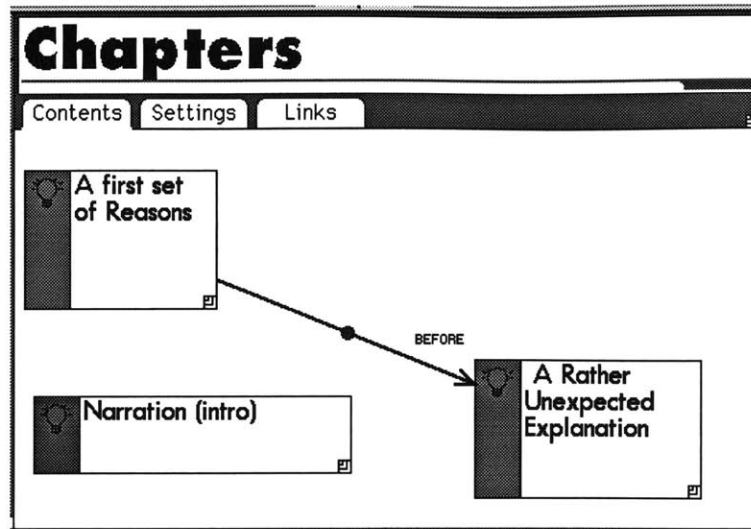
Figure 3 - There is no order set on the chapter

5. How the stories were annotated

The following sections describe the various stories available to the viewer and their annotations.

"Veni Vidi Vici"

This story explores the reasons why the Gulf War was so short. Answers vary from the use of heavy bombing to the destruction of vital information centers early on in the war. The story has a very simple structure with two chapters, each one presenting two very different sets of answers.

This is a good example of how the author is able to impose a certain perspective which shows the contrast between two very different types of answers.

An alternative would have been to merge both chapters and let the viewer alternate viewpoints. The current solution shows at least one answer from the first set, even if the emphasis is on the second.

A first set of Reasons

Contents | Settings | Links

Narration (intro)

Low Moral

Bombing

Answer 8.a
Robinson 1
04:14:17.25
04:14:43.18
Clearly the bombing
had an impact, may
be more by its
effect on moral.
Most of the
casualties on the
Iraqi side were due
to desertions before
the ground war.

Answer 7 Robinson 1
04:11:41:19
04:13:47.27 complete
Answer. See below for
elements of the
Answer.

Answer 7.d
Robinson 1
04:13:15:04
04:13:47.22
conclusion: so all
this combined...Very
few casualties.

Answer 7.a
Robinson 1
04:11:41:19
04:12:01.21
underestimated
the equipment,
...

OR

XOR

Answer 19.3
Greenway
03:26:29:25
03:26:57.03
The impact of
air power was
completely
underestimated.
Desert with no
trees.

BEFORE

Answer 7.c
Robinson 1
04:12:38:09
04:13:15.00
Always better
to overestimate
the enemy's
strength. It
seems like the
Middle East is a
place where we
are not able to
gather a lot of
information.

Answer 19.2
Greenway
03:26:19:21
03:26:29.11
We never
knew how
low the
moral was.

Answer 7.b
Robinson 1
04:12:01:23
04:12:36.27
more of the
same (low
moral, bad
training)
Concludes: so
their defense
crumbled.

BEFORE

BEFORE

OR

Answer 19
Greenway
03:25:54:19
03:27:09.13

XOR

Answer 19.1.2
Greenway

Figure 4 - Sample of the clips in the first chapter. Two OR frames indicate that two clips can be decomposed into several clips, giving more flexibility to the story generator. BEFORE links force an order on some clips.

Looking at the links existing between clips in the first chapter [fig.4], we note that some clips can be decomposed (OR links). In some cases, an additional ordering constraint is imposed (BEFORE links). There are two subtopics under which answers can be grouped: Low Moral and Bombing.

"Story Hunting under Pentagon Rules"

Any journalist remembers from the Gulf War the rules imposed on the press by the Pentagon in order to filter out any crucial information. This story tries to answer various questions: What were these rules? How were they applied? How did journalists react to them? What was the motivation of the Pentagon?
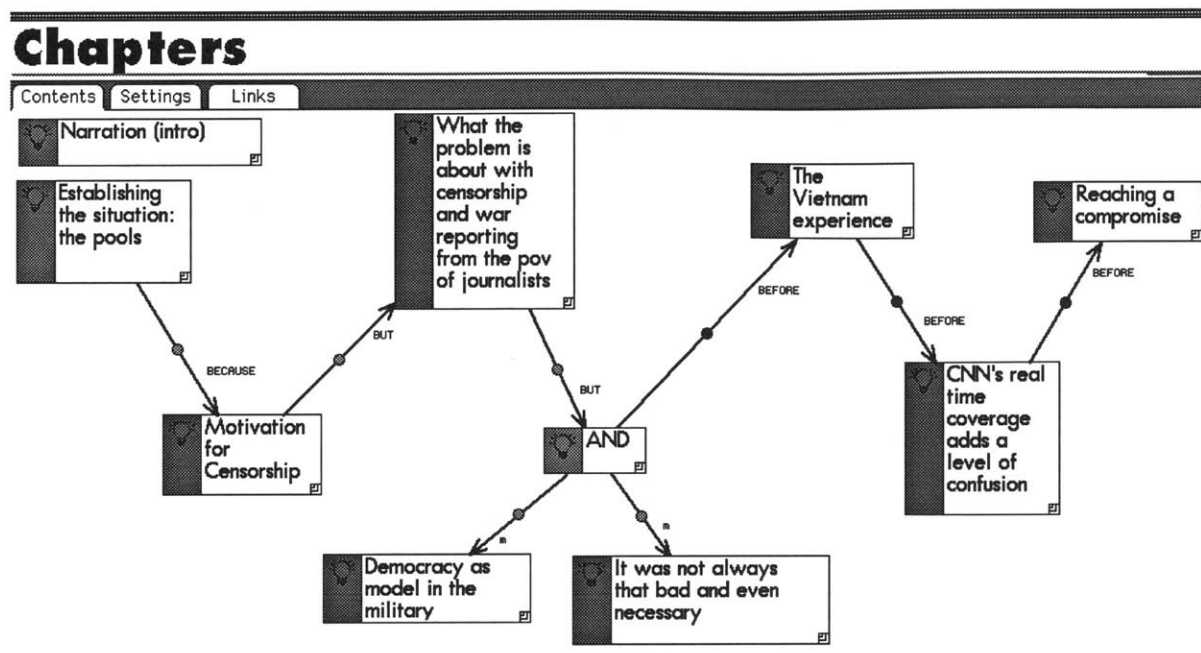
Figure 5 - The 8 chapters and the links connecting them

The story structure includes 8 chapters [fig.5]:

1) Establishing the Situation: the Pools

The first chapter explains what the pools were and how they worked as they were the major instrument of control.

2) Motivations for Censorship

The reasons as seen by Walter Robinson and Paul Strassmann who have very different viewpoints. Unfortunately this chapter contains only two clips. Having a voice representing the Pentagon's view would have been a valuable addition.

What the problem was from the pov of journalists

Contents | Settings | Links

One thing we learned about the GW is that we are at the mercy of the Pentagon. The american people don't get an honest look at what is going on (Pentagon continues following success in

tendency in the military to blame the messenger so they have an interest but the military know they are in a democracy so they have to put up with it.

Answer 15.1 Greenway 03:22:31:27 03:22:51.00 (Were rules stricter than in Vietnam?) [Oh, I'm sure that's true - cut out]

There is a sort of canard in the military that the press was responsible for the low moral. I think it is vastly

Answer 9.3 Greenway 03:13:21:14 03:13:35.05

Control of where TV crews could go was sometimes obsessive.

Same as precedent but beginning is cut out [But the point you see is that]

An Greenway 03:22:52:13 03:23:15.28 What counted in fact as it did in Korea was the amount of casualties.

Answer 4.a Robinson 1 04:05:49:00 04:06:10.17 what people saw was what the armed forces allowed to be filmed.

Supports

Supports

Supports

Supports

Supports

Supports

BECAUSE

BUT

Answer 16.2 Greenway 03:24:12:18 03:24:17.17 Public opinion doesn't play much of a role in Iraqi politics

Answer 22 Robinson & Wiggs 05:09:33:15 05:10:20.06 [difference in censorship between Iraq and the US] The difference between the Iraqi public and the American public is may be that the Iraqis knew they were taken whereas the Americans didn't know.

Answer 14 The answer is of course they are more vulnerable. Hussein had complete control over the information that was sent to the Iraqis but Bush had no control over the information sent to americans. Pstrassmann.1 01:21:40:26 01:22:15.23

Democracy: Iraq vs US

Figure 6 - The clips of the chapter "What the problem is about" and existing links between some of them.

3) What the problem is about

Presents reactions explaining why the press rejects partially or in total the rules defined by the Pentagon. Figure 6 shows clips supporting the point made by other clips ("supports" links). This is to be used in partic-ular to support a given opinion with a combination of answers from dif-

ferent persons.

4) Democracy as model in the military

The previous chapter criticizes the Pentagon for censoring the Press and puts in question its concern for democracy. It gives a negative image. This one shows in contrast opinions expressing that the American army uses a democratic model of communications. It gives a more positive image.

5) A less radical approach in the press

This chapter also contributes to counterbalance the harsh criticism of the Pentagon by showing less extreme positions. Journalists explain that it was still possible to go hunting for a story on your own. Some show a relative understanding for the rules while still criticizing extreme cases.

6) The Vietnam experience

To add a level of perspective, the chapter makes a comparison with Vietnam.

7) CNN's real time coverage adds a level of confusion

By broadcasting in real time, CNN provided the world but also Iraqis with information that was difficult to control.

8) Reaching a compromise

As a form of conclusion, this chapter shows how a difficult compromise was reached between the press and the military.

"Blurring the Limits between Fiction and Reality"

Many considered that the theatrical coverage of the Gulf War on television contributed to turn it into a big show, almost like an action movie. This story has a simple structure [fig. 7]. It contains only two chapters. Two clips have been labaeled as "must see" clips so they get selecte with higherpriority.
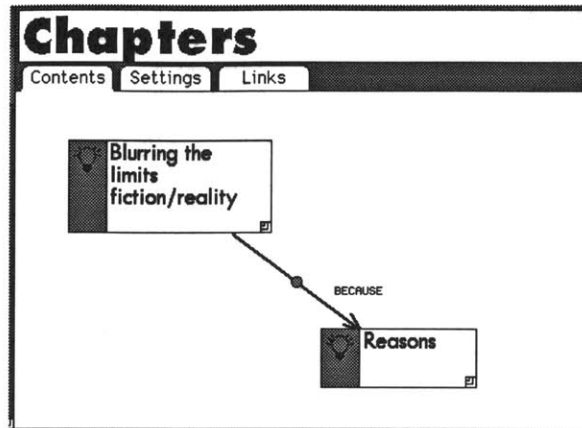
Figure 7 - Story structure

Figure 8 - Clips of the first chapter

"Being a Journalist during the Gulf War"

This story has also a simple structure with two chapters:the first one shows how information technologies were used by journalists, the other one reports anecdotes about safety issues. It does not matter in which way they are ordered.

Figure 9 - Samples of clips from the chapter about technology

Clips from the chapter about the use of technology by the press [fig.9] show how a contrast between two facts is represented using BUT and AND annotations: Jonathan Wiggs says that on the one hand there were problems, but on the other hand he had solutions like sending back photos using AP or REUTER.

6. Why would an author subscribe to this?

Documenting a story may involve following its progress. If the issue is complex, more than one story are likely to be created. Making them all available is a motivation. There are other motivations:

Answering the need for customizing video

The need for the viewer to customize video has already been mentioned. Content providers will want to answer that need.

Answering the need for access to multiple viewpoints on an issue

A publisher of a viewpoint server might want to offer a platform for various viewpoints to be published and accessed. There are two publishing options:

1) Access viewpoints on a topic, independently of one another. This means that they are not presented in relation to each other in a story. Multiple sources will feed the server. A digital agent provided by the publisher might even go look for more viewpoints on other similar servers. You would get a collection of interviews and articles by various columnists on the selected topic.

2) My model of presenting viewpoints in relation to each other, in the context of a story defined by an author.

Concern for more participation from the viewer

Giving the audience more control over content might motivate many authors, specially if there is a market for it.

Linking a database to a discussion list

Such a framework would be of particular interest to a person (journalist?) creating a discussion list associated with a database of content. A tight relationship could tie together the database, the playback engine attached to it and the discussion list.

The Internet has been very successful at spawning virtual communities and discussion lists (the so called "news groups"). It has demonstrated that a visceral human need was not completely fulfilled, the need for communicating and exchanging opinions freely. A major problem faced by these discussion lists is the lack of tools for browsing a line of argument. It is also difficult to go back in time and browse the archived discussions. In addition, these groups are today limited to text because of the bandwidth.

The database could contain either past discussions or the equivalent of a documentary gathered by the owner of the list. In the second case it would serve as a base for discussion and actual links could be created between the database and the personal views expressed (or even gathered) by the members of the list.

7. Are we fooling the viewer?

One general concern might be that the viewer is fooled into thinking that because there is interaction and control of viewpoints, there are less chances for being misled and more control over information.

This is entirely in the hands of the author. The model for creating databases of viewpoints that I propose leaves an option that current media don't offer: interactive access to several viewpoints on a particular issue and the possibility to focus on one or another. However, the structure of the story, the selection of the content, the relations represented in the database between the viewpoints, how they are articulated one in reference to the other, are still in the hands of the author. The narration also

is written by the author.

8. Conclusion: Editors as server publishers and administrators?

The deluge of information as well the pressure to access it is such that there will be a constant effort to build better and better systems for assembling stories and presenting information automatically. Carefully crafted documentaries and articles will be available as well to satisfy a need for greater quality, following the same behavior and motivation that entice us to buy a paper back book, a hard cover book or an art book.

A probable scenario for the next decades is that there will be an increasing number of journalists *authoring databases* and using high level tools for organizing, shaping and structuring their content. By high level, I mean that they are not intended for manipulating video frames or words but viewpoints and issues, stories, very much like an editor does in a newspaper. The difference is that content will be shaped and annotated in such a way that it can be usable by customizing software and agents. The first to take this direction will be reporters and editors who are willing to take a risk and explore new venues. If CNN was to make available its archives of news footage, some type of editor would have the task of setting up the server by organizing its content and annotating it with advanced annotating tools. This is already happening among the community of journalists: many are now setting up on the net Gophers and Mosaic nodes, exchanging pointers to other servers on the Internet through discussion lists like nit@chron.com[1] or CARR-L (Computer Assisted Reporting and Research)[2]. Collecting agents may soon be available in crusing the net and searching databases. Powerful tools are likely to make editing a less tedious task and make video editing accessible to the public.

1. nit@chron is a discussion list exclusively for journalists
2. CARR-L is a mailing list created by Elliott Parker, Journalism Dept., Central Michigan University

4 Behind the Scene: the Mechanics of the Playback Engine

Knowledge representation of opinions & viewpoints

First steps towards a storyteller agent

This chapter explains how the story generator works and describes the knowledge representation it relies on. The final section analyzes its weaknesses. Alternative solutions to improve the results will be described in the chapter on Future Directions.

1. Knowledge Representation

"Good representations make important objects and relations explicit, expose natural constraints, and bring objects and relations together."

"Once a problem has been described using an appropriate representation, the problem is almost solved".
P.H.Winston (Director of the MIT Artificial Intelligence Laboratory)

Using a semantic net representation

I used a semantic net representation, because it met the needs for expressing relations between various opinions. Each clip in my system is a comment, answer or bit of answer given about an issue. Several answers by one person (or group) contribute to build in the mind of an observer a model of that person's (group's) viewpoint. Answers or comments are denoted by nodes with links between nodes representing how the nodes can be *coordinated* to articulate opinions. These links need to be labeled and eventually handled as objects themselves. A semantic net also lends

P.H.Winston, 1992
Artificial Intelligence, 3rd
Ed.

itself to a graphical representation of knowledge.
Any representation consists of four fundamental parts:

1• A *lexical* part that determines which symbols are allowed in the representation's *vocabulary*

2• A *structural* part that describes *constraints* on how the symbols are arranged

3• A *procedural* part that specifies *access procedures* that enable you to create descriptions, to modify them, and to answer questions using them

4• A *semantic* part that establishes a way of associating *meaning* with the descriptions

Lexical part

In a semantic net the nodes denote objects and the links relations between objects. Link labels denote particular relations. The chapter on the author described the graphical symbols (objects and links) used for annotating the content.

Structural part

Semantic nets connect nodes to each other with links appearing as arrows. Here the semantic net diagram displays nodes as rectangles (virtual boxes) that can contain recursively other semantic nets [fig.]. The result is a hierarchical structure with lateral links. The hierarchy tells how viewpoints and stories are decomposed into the clips at the lowest level. Lateral links show logical, structural and temporal connections within the content. BUT, AND, OR, BECAUSE, BEFORE links are lateral links. (See also the chapter on the author).

Procedural part

Sections 2 and 3 explain the software that accesses the database, actually puts together the movie, and adapts to the viewer's requests. Some of the constructor, reader, writer or erasure procedures used were part of the database software itself. The others were part of the story generator.

Semantic part

The meaning of the nodes and links as well as their labels have been defined in order to let the system associate story and content information with clips. Story information was created as knowledge about how the content is orchestrated. Content information was focused on relations between clips that *coordinate elements of opinions*.

2. The Basic Principles

The current software presents viewpoints in a way that makes sense, informs the viewer and adapts to the parameters set by the viewer (given the eventual constraints imposed by the material available in the database). Clips are organized in a way that is inspired by TV news editing: a narration introduces an issue, asks a question and is followed by various reactions or answers. The main difference is that the narration is much shorter in order to leave more room for the actual expression of viewpoints. The story progresses through various "chapters". Each chapter may contain sub-topics. A selection of clips is based on the following rules or constraints:

- "must see" clips are all selected
- the order of the chapters is followed
- for each chapter the narration comes first and is followed by a reaction from each available speaker
- if there are several clips for a given speaker, they are grouped
- in the case where a chapter includes subtopics, the same rules applied at the chapter level are applied at the topic level
- the length of story requested by the viewer is taken into account
- two key organizing rules allow the focus on viewpoint:

A first, simple and quite powerful rule is to spend more time elaborating the viewpoint. This can be translated into a rule such as the following:

■ IF *A is the major viewpoint* THEN *maximize time allocated to A*

Another rule is to intercut characters' comments in a way that supports a given viewpoint. This can be translated into a rule like the following:

■ IF *A is the major viewpoint,* AND *B supports A* AND *C opposes A* THEN *juxtapose (A,B)* OR *juxtapose (B,C)*

The following example shows a context for applying this rule:

A: *"It was not possible to let the journalists free to roam around."*

B: *"The American public saw what the Pentagon allowed them to see. Photos taken from airplanes showed successful bombing, Patriot missiles were shown as being effective, no images of civil casualties were broadcast."*

C: *"That night we moved many miles west to take their rear by surprise. Anyone close to where this movement was happening could have detected that a major move was taking place."*

A, B, and C express three viewpoints:

A: A Pentagon official

B: A journalist

C: A soldier who took part in the action

C can reinforce A and make B less convincing if juxtaposed to A or B. It can have the inverse effect if it is missing.

Here is another example:

◆ **Let's assume first that the message to convey is** that the Pentagon was mostly engaged in propaganda and marketing the war, with little concern for troops and civilians. This gives a negative image of the Pentagon.

How to translate that? Images of successful bombing can be intercut with interviews of civilians whose houses were destroyed and families killed. The effect is that the Pentagon was lying about the reality of the war. A journalist lists the various ways in which the Pentagon was controlling the press. In a public speech Bush announces that the Iraq invasion of Kuwait will not be tolerated. This cuts to various interviews of peace activists that picture President Bush as having more concern for protecting the oil supply than for Kuwait. Shots show various anti-war demonstrations across the country. Interviews and images illustrate that Patriot missiles were far from being as successful as it was said during the war.

◆ **Let's assume now that the message is** that the Pentagon wanted *by all means* to protect the troops on the battlefield and maintain high moral. Without popular support it would have been impossible. This gives a positive image of the Pentagon.

How to translate that? The same interviews of civilians are used with a

comment saying that *only* a few bombs missed their targets killing civilians. Losses could have been much more important without constant concern for avoiding highly populated areas. Some shots are intercut to show phone booths allowing the troops to call home free. Troops see images on television of supporters saying how much they are proud of them. They cheer. Images of successful bombing are shown with a voice-over commenting that the coalition was very successful at destroying the Iraqi's defense and indeed the war was finished in the wink of an eye.

3. Overview of the Story Generation Algorithm

The problem is to select the right clips and schedule them for playback according to parameters set by the viewer. The environment of the engine is composed of the database (clips and annotations) and the viewer's actions.

The action starts with efficiently parsing a tree: currently, the hierarchy has stories on top, composed of chapters, which are themselves composed clips. Some chapters have sub-topics. Tree search is a mechanism for which Artificial Intelligence has made major contributions, because quite often solving a problem can be translated into a tree search. The selection and scheduling algorithm is implemented as a "greedy" algorithm that traverses the story hierarchy. Greedy algorithms assume that local optimizations provide a fairly good global solution[1]. Since doing a global search of the database tree would take an excessively long time, the alternative is to optimize locally.

The pre-decomposition in stories and chapters helps minimize the breadth of the sub-tree to parse. Another advantage is that the sub-tree is not deep. This means that there is no need to evaluate and apply techniques like depth-first search or breath-first search. What makes the search challenging is the combination of constraints to be applied in the

---

1.   The notion of "greed" as an algorithm comes from a simplistic model of the debates between socialists and libertarians. Socialists consider greed the source of all evil. Libertarians see greed as not necessarily negative. These debates cross over into economics, philosophy, government, and computer science, using jargon from all four fields.

selection and ordering: total length, weight for each story and viewpoint, clips that must be selected, available clips and their duration, eventual clip dependencies.

At the top level, a story is selected, a time allocated, and the story-selection function `plan-story` is called. The duration of the clips returned is calculated and is subtracted from the total time. The time left is then used for the rest of the stories.

At the story level, `chapter-dependents` finds and orders the dependent chapters (those that have ordering links on them) and orders them correctly with `chapter-order`. Currently, the scoring function for chapter ordering is a random number generator. Since this is only used for non-dependent chapters, it didn't seem that it was too weak.

At the chapter level, `must-see-clips` looks first for the clips annotated as "must see" clips. The remaining clips are given to `select-clips` to try to find an optimal selection. This works by trying every combination of clips that are not "must see" clips (essentially, counting upwards in binary). The function `select-score` is used to score a selection, and `check-lst` to make sure the selection is legal. The legal selection with highest score is chosen.

After the clips are selected, they are ordered with `order-them` in a way similar to what is applied to chapters. `Dependents` collects the dependent clips, and orders them in a possible way (note: currently, this ordering is not scored!). Afterwards, the non-dependent are added in a position that will maximize the score of the ordering with `order-score`.

The `select-score` function tries to optimize the two parameters provided by the user: **length** and **viewpoints**. In order to accomplish this, what the program does is multiply the average weight of the frames by a function of the total time:

[Σ [(weight of clip speakers*clip frames)/total frames]] * f(t)

f(t) is a function of the total time.

Currently, f(t) is a step function `stepf`. If the time is within a certain percentage of the desired time, a 1 is returned. Otherwise, the gaussian value is returned (a much smaller number). *This has the effect that if a total time is within the threshold, it will be counted much more than times outside, but the same as all times inside.* Therefore, if there are several selections

within the threshold, the best speaker-weighted one will be chosen. Otherwise, the time closest to the desired will be chosen.

5 "It was a Knowledge War"[1]: Notes about the Content

"When we compare the new features of warfare with those of the new economy, the parallels are unmistakable. The day may well come when more soldiers carry computers than carry guns. Knowledge in short, is now the central resource of destructivity, just as it is the central resource of productivity".

Alvin & Heidi Toffler

1. Information: a special angle on the Gulf War[1]

I became extremely interested in the topic of information and the Gulf War when I read about the impact of CNN's real-time coverage of the war, the heavy debate it triggered, the use of satellites and other information technologies by the media. But what really stimulated me was a conversation I had with Paul Strassmann whom I had previously interviewed in reference to his book, "The Business Value of Computers" .

---

1. The U.S. Army's Chief of Staff called Desert Shield/Storm the "knowledge war"

A.D. Campen (ed.), 1992
The First Information War
AFCEA

A.&H. Toffler, 1993
War and Anti War

Paul Strassmann had spent two years at the Pentagon as Director of Information in charge of a major reorganization of its information system. After that, he taught information warfare at West Point. He made this revealing statement: *It was the control and destruction of the Iraqis information system that was decisive in the outcome of the war.* He suggested I read a book, "The First Information War" which is indeed a rich source of facts. Later on I read Alvin and Heidi Toffler's book, "War and Anti War" which takes a similar perspective on the war.

Barbie Zelizer, 1992,
CNN, the Gulf War, and
Journalistic Practise, Journal
of Communication, Vol. 42
No1

CNN's live coverage had been intensively debated in specialized magazines and books. What was original about what I had just learned was that this media phenomenon, to an extent was mirroring the way the war was being conducted by integrating information technologies. The "Information Highway" as a mass media phenomenon was just about to happen in 1991. Also it seemed to me important for all of us to become aware of the changes in warfare. If warfare is to become centered on information and if the future of terrorism is to destroy information systems, then a new mindset is required on the part of all of us: "We are exposing the world's first information-based society, with 50 percent of the world's computing and 70 percent of its advanced software capability, to enormous dangers. This is the place where the first information Chernobyl will happen" [New York Times, July 21, 1994].

2. Why was information more important than it ever was before?

He who knows the enemy and himself
Will never in a hundred battles be at risk.

...

Complex systems such as battle conditions are rich in information -- information that must be acquired immediately (...) Once the specific time has passed, information loses its strategic function and importance, and at best retains only historical value.
Sun-Tzu, the Art of Warfare, c. 403-221 B.C.

Sun-Tzu, the Art of
Warfare, c. 403-221 B.C.
translation and comments
by Roger Ames, Ballantine
Books, 1993

The importance of information is not new in warfare. It was an important part of the principles taught by Sun-Tzu ("Master Sun") to the various rival families during the Warring States period in ancient China. For him knowledge is all about tracing connections, seeing changing patterns in order to predict what might happen. *"To know is 'to realize', to 'make real'. The path is not given but is made in the treading of it. Thus, one's actions are*

always a significant factor in the shaping of one's world." This approach to knowledge encouraging action and creativity contrasts with its Greek counterpart in which knowledge is about establishing a relatioship between an idea and an objective world. *"One 'knows' the world not only passively in the sense of recognizing it, but also in the active shaping and 'realizing' of it".*

This philosophy of knowledge goes beyond warfare and appears very up to date in our information-loaded world. The major difference between today and the Warring State period is that information technologies now enable us to gather, communicate and process information extremely efficiently. The emerging problem today is to deal with the overflow of information. The difference between Vietnam and the Gulf War is stunning. This is why the Gulf War appears as a turning point.

On the military side

The Gulf War saw what has been called the largest single communications mobilization in military history. Starting with minimal capabilities in the region, a complex set of of interconnected networks were built at high speed. Numbers may be the best way to convey the massive presence of information technologies:

- 700,000 phone calls per day
- 30 million phone calls during the air war alone
- 30,000 radio frequencies
- By the end of Desert Storms, there were more than 3,000 computers in the war zone actually linked to computers in the U.S.

To put it into perspective, 700,000 phone calls a day is the traffic of a small town. There were 500,000 soldiers involved but obviously most of them were more busy preparing to fight than communicating. It was the first major conflict in which individuals brought their laptops to the battlefield. Troops were allowed to call home free (AT&T set up the phone booths).

According to Paul Strassmann, GPS[1] was *the* key technology of the war. It provided the mobility that was essential to the success of the vast operation that consisted in hitting by surprise the flank of the elite Iraqi Re-

---

1.   Global Positioning System: portable device which accurately indicates the locationof a moving object anywhere on the earth surface

publican Guards. The operation required that troops move *144 kilometers in 18 hours* and in proper order. This was possible only because they all knew precisely at all times their position in relation to each other. The information was available to the Command and Control center as well.

Peter Arnett, 1994
Live from the Battlefield,
From Vietnam to Baghdad,
35 Years in the World's War
Zones - Simon & Shuster

Describing his third day of live coverage in Baghdad, Peter Arnett reported *"Late that morning the telecommunications center, one of the larger buildings in the city was reduced to rubble. It just disappeared from the southern horizon"* . For Paul Strassmann the reason why the war was so short was because the Allies strategy was based on attacking and destroying the Iraqis communication centers.

Alvin & Heidi Toffler similarly comment that while the coalition forces were busy collecting, analysing and distributing information, they were also busy destroying the enemy's information and communication capability. *"The task was to disrupt the brain and nervous system of the Iraqi military."* The Pentagon's final report to Congress points out that the early attacks targeted *"microwave relay towers, telephone exchanges, switching rooms, fiber optic nodes, and bridges that carried coaxial communications cable ... On TV the public saw planes, guns and tanks, but not the invisible, intangible flow of information, data and knowledge now required for even the most ordinary military functions"*.

In "The First Information War" Colonel Alan Campen gathers several reports illustrating the key role of information in the success of the war. *"The key roles"* - he writes - *"played by exotic weapons and skilled, superbly trained and motivated people were revealed through television, widely reported in the media and thoroughly documented in after-action reports. Implicit but unseen in those reports are the information systems - networks of computers and communications that synchronized the air campaign and that turned dumb bombs into smart weapons. Without those systems, combat forces could never have been as effective.*

The weapons - and the tactics by which they were employed - were the product of years of research, testing and evaluation, simulations and realistic training. This was not the case for information systems. One observer even noted that 'much of what they did from August through February had not even been dreamed of in July'."

On the media side

"Over the six weeks of the war, more people watched more hours of television than in any time in history." (Former General Perry Smith reporting for CNN).

"The media itself became the star of the spectacle." (Alvin & Heidi Toffler, *War and Anti War*)

"In my wire service days, minutes made the difference between a scoop and being second. In live television, seconds made the difference [...] Superficiality is one of the perils of live television but hesitation is a mortal sin. Still, it is not easy to speak before you think [...] It was my turn again at the open mike. I said that all the bombs seemed to be hitting directly on target. I tried to qualify. We could see a lot from our ninth-floor perch but certainly not everything. We didn't know what was happening out there. I knew that in the critical first moments of the war, public opinion could form on our report. We had an open line to the world; there was no censorship at either end. The Iraqi minders had taken off for the bunkers." (Peter Arnett, *Live from the Battlefield*).

The radically new information technologies transforming the military were also transforming the media. All media increasingly rely on computers, fax machines, satellites, and telecom networks. This trend was illustrated in full scale during the Gulf War.

The media and the military

Control of information means allowing information to flow and controlling who gets what. A major conflict grew between the press and the military during the Gulf War. Many journalists criticized the Pentagon for censoring the information the press wanted to make available to the public.

For both the military and the press, satellites were playing a key role. Many satellites were repurposed to be used in the Persian Gulf. The military even used commercial satellites for non classified information.

3. Viewpoints: who was interviewed?

I interviewed three journalists from the Boston Globe who were foreign correspondents during the Gulf War and the former Director of Information at the Pentagon.

Two of the journalists had also taken part in the Vietnam war: David Greenway was a reporter at the time and Walter Robinson was in the army Intelligence. The third, Jonathan Wiggs, was a photographer in the Gulf pools set up by the Pentagon.

While working for the same newspaper, the three of them had different perspectives. Walter Robinson was extremely critical of the Pentagon rules. He also provided examples regarding the differences between

television and the printed press in terms of mobility. David Greenway, while criticizing some extreme cases, expressed some understanding of the Pentagon's arguments. Jonathan Wiggs provided his view as a photo-journalist who experienced the pools. Paul Strassmann didn't really represent the point of view of the Pentagon. He contributed more as an expert in information warfare. He did, however, provide reasons for supporting the rules imposed by the Pentagon.

Time and money not with-standing, this database would take even more advantage of the technology if the following viewpoints could be included. An interview of a Pentagon representative whose opinion would have better counter-balanced the many answers I gathered from the journalists. Another missing voice is CNN's with in particular Peter Arnett who turned into a controversial media hero during the Gulf War. Finally I would have liked to add an interview of Alvin & Heidi Toffler.

Adding the Pentagon and CNN's voices would also reinforce the value of the software developed. They would have made more salient the changes in viewpoint when the viewer shifts focus. For instance, *"Story Hunting under Pentagon Rules"* only slightly changes when switched from David Greenway to Walter Robinson.

Adding Alvin & Heidi Toffler's interview would further demonstrate that a point of view can be supported by various individual views. The advantage would have been to explore another example of how to combine in the database's representation individual views that compose a same perspective. In this case it is the point of view of those who believe that the future of war is all about information. This means that if the viewer were to focus on Paul Strassmann s/he could listen as well to arguments brought in by Alvin & Heidi Toffler. They would in fact make Paul Strassmann's position even stronger.

"We could see, hear and talk all through the war. After a few hours he could not" Air Force official Martin Faga

"An ounce of silicon may have been worth more than a ton of depleted uranium" Col. Alan D. Campen, USAF (Ret.)

"CNN went off the air at the very moment that the war called for a bomb from an F-117 stealth fighter to penetrate the city's main telecommunication center" Col. Alan D. Campen, USAF (Ret.)

"Frequency mangement challenges were enormous" Ernest May, "Intelligence Backing into the Future", Foreign Affairs Summer 1992

"We will never know how well Iraq's military might have done because the initial attacks took down its control network and, with it, any hope for Iraq to know what had happened, what was about to happen and what it might do about it". Col. Alan D. Campen, USAF (Ret.)

"Knowledge came to rival weapons and tactics in importance, giving credence to the notion that an ennemy might be brought to its knee principally through destruction and disruption of the means for command and control". Col. Alan D. Campen, USAF (Ret.)

6 Related Work

The purpose of this chapter is to draw a map of where "Viewpoints on Demand" stands in relation to the various research areas it overlaps with: personalizing and presenting information in electronic publishing, interactive storytelling, representing the content of video.

1. Customizing & Presenting Digital Information

W.Bender at al., 1991
Newspace: Mass Media and Personal Computing

The early nineties are marked by increasing interest in the digital delivery of information and entertainment, the so-called "information highway". By now it is almost impossible to find a magazine, newspaper or TV channel that has *not* approached the topic. Since the 70's the Architecture Machine group, which in 1985 became part of the MIT Media Laboratory, conducted research in the area of electronic publishing, digital television, interactive media and personalizing content. The initial concept that triggered the creation of the Media Lab was the convergence of television, publishing and computers. In terms of news and information, the central idea has always been that - as readers or viewers - in order to cope with the deluge of information, we need to get *personalized* information. For many years this made the Media Lab research very unique as no other lab was so invested in personalization.

J.Orwant,1991,
Doppelganger Goes to
School: Machine Learning
for User Modelling, MIT
M.S. in Media Arts and
Sciences

Various projects -- now grouped under the News In the Future consortium (NIF) -- have approached the problem of personalization. Jon Orwant began research on **User Modeling** with the Doppelgänger system: this software is designed to sense, gather and process personal information about an individual and from this information, Doppelganger builds an agent to select information, advertizing or entertainment that is of interest to the user. In a commercial environment this raises legal issues about privacy.

Several Media Lab projects dealing with customizing text news have used early versions of Doppelgänger. Other projects have dealt with personalizing media by inventing ways in which content can adapt to user input without accessing a system expert in providing personal information on a given individual like Doppelgänger. "Viewpoints on Demand" belongs to this second category. The first category uses *"passive sensors"* as it does require any input from the user. The second category uses *"active sensors"*.

News

H.W. Lie, 1990
The Electronic Broadsheet,
All hre news that fit the
display M.S. MIT Media Lab

J.S. Donath, 1986
The Electronic Newstand:
Desing of an Intelligent
Interface to a variety of
News Sources in several
media, MSVS MIT Media
Lab

A.W. Blount, 1993
Self-Organizing New
M.S..MIT Media Labs

B. Sheth & P.Maes
Evolving Agents for
Personalized Information
Filtering, CAIA'93

The "Newspeek" and "Newspace" projects were the precursors of NIF. As part of Newspace, the Electronic Broadsheet by Hakon Lie focused on the **automatic layout** of custom news. The Electronic Newstand was a first attempt to collect information from various sources for an individual viewer.

Later on, Alan Blount implemented a system that organizes news articles with regard to their content and style, acting as an intermediary between the article database/user modeling system and the graphic designer. Research on **agents** that can collect personalized information is another major area of research .

Presentation of Information

Michelle Fineblum developed a system that automatically generates hypermedia presentations at run-time based on the user's style selections . The content's subject matter was research at the Visible Language Work-

shop (a Media Lab group that works on dynamic design). The user would specify whether s/he is a researcher or not, and whether s/he is technical or not. The system used a knowledge-based representation of presentation styles linked to these user types, along with a database of richly described media. The output was a presentation based on an ad hoc branching structure.

M.A. Fineblum, 1991
Adaptive Presentation
Styles for Dynamic
Hypermedia Scripts, MSVS
MIT Media Lab

"Viewpoints on Demand" has in common with this system the idea that different users benefit from different presentation of the information. It has also in common the goal of customizing the presentation of information based on some initial user input parameter. A major difference is that "Viewpoints on Demand" is not based on a branching model (even though it can incorporate branching) [see chapter 7]. It is based on constraint based editing to generate a linear video documentary. In terms of content, another difference is that "Viewpoints on Demand" informs by using viewpoints and opinions. Michelle Fineblum's project is not meant to introduce people and their opinions on a topic. It is meant to present the results of research.

Graphical design

The research direction taken by "Viewpoints on Demand" is not only to present viewpoints based on initial input from the viewer but also to suggest opinions *related* to the sequence played on the screen at a given moment. In the context of graphical design, the Media Laboratory's Visual Language Workshop addresses similar issues. Its research exploiting its **Typographic Space** concept explores the graphical presentation of text information in a three dimensional space and deals in particular with *point of view in space*: how to avoid the distorsion of the typographic form caused by perspective and an arbitrary viewpoint, but also how to communicate visually shifts in viewpoint that indicate a change in focus to a new topic and its related information. Ultimately what is needed is an active sensor system and the dynamic display of related stories

**D. Small, S.Ishizaki,
M.Cooper,**
Typographic Space, MIT
Media Lab, Internal Report

E. Rennison, 1994
Galaxy of News, An
Approach to understanding
Expansive News Landscapes

Galaxy of News by Earl Rennison shows relationships within large quantities of information (news stories) by automatically building and visualizing links between related articles. It explores the semantic relation between graphical design *and content* in a 3D space. With semantic zooming and panning, the user is able to select a topic and explore it in more detail: as s/he zooms in, headlines to appear first and then the ar-

ticles body begins to drop in. As more time is spent in an area, it becomes related visually to other areas.

2. Interactive Narrative Research

B.Rubin, 1989
Constraint-Based Cinematic Editing. MSVS MIT Media Lab

Ben Rubin's constraint-based cinematic editing system is probably closest to "Viewpoints on Demand". Rubin used a different set of constraints on the content: instead of viewpoint and opinion, he used plot. The system had knowledge of flow. Events in the plot were weighted according to importance. Potential uses listed are: assisting an editor, doing queries to video databases or interactive systems. No specific interface was designed for the story "Rowe's Warf".

The results were mixed. Some sequences produced by the system were reasonable. Others showed that the system had trouble with spatial continuity and story coherence. Plot actions are classified according to chronology and subjective importance.

R.Schank et al., 1992
Agents in the Story Archive, TR 27, Northwestern University

Roger Schank's team at the Institute for Learning Science also uses stories as a retrieving mechanism for *mediated* presentation [Laurel 90]. The goal is to teach by re-telling stories that answer questions from the user suggesting them under appropriate circumstances. According to Schank, this implies that the computer doesn't need to "know what's in the stories or be able to converse with the user about the knowledge they contain. It simply helps users find and access the expertise they need".

"Viewpoints on Demand" takes a similar approach in the sense that it does not represent all the details of a given story in order to customize its presentation. It represents only what relates to the *opinion* expressed. It also uses a representation of the story's *structure* in order to generate a custom version of it. Schank's approach is to retrieve a *whole* story given a certain *context.* At some point in the dialog a story on Scuds might be relevant and one of the agents will retrieve it. Finally Schank does not recompose and customize stories according to a given viewpoint. Knowledge about viewpoint and opinion is not the focus: *there is no indexing based on viewpoint but indexing that shows relations between stories:* "Our approach concentrates on indexing stories rather than fully representing them. Before we can represent all the information in a given story we must consider the connectivity of one story to another and the conditions under which a giv-

en story ought to be accessed and then told".

T. Oren, G. Salomom, K. Kreitmen, A. Don, 1990 Guides: Characterizing the Interface, in The Art of Human-Computer Interface Design, Addison Wesley

The **Guides project** developed at Apple shows in particular how narrative and information retrieval can be combined. The major difference is that Viewpoints on Demand presents views (opinions) in relation to each other instead of using viewpoints as guides to navigate through the material.

K.E.Steiner, T.G.Moher, 1992 Graphic Story Write: An Interactive Environment for Emergeent Storytelling, CHI'92

Graphic Story Writer uses graphics and text with no video but it does include knowlege about story structure which is precisely what it wants to teach to children in addition to helping them learn to read. With Graphic Story Writer they create interactively a story by dragging characters and objects in a set. Selecting initially a characteristic for a character like "greedy" helps define his/her goal. The system progressively generates the text of the story as more objects are added. It tries to guide periodically the user in order to work toward the resolution of the central conflict. It would insert for example: "the girl still looked for something fun to do" if her goal is indeed to have fun.

The elements of story are simple and help generate simple stories that children can manage. The structure of the story itself is not pre-determined as it is in "Viewpoints on Demand". It is created by interaction between the various elements added to the set at a given time. This direction is worth pursuing in fiction in order to give the user freedom of creativity. As I said in a previous chapter, with information the goal is different and so the influence of the user on the content must be of different nature.

Ryan Evans, 1994, LogBoy Meets FilterGirl, MS MIT Media Lab

Mark Halliday, 1993 Digital Cinema: An Environment for Multi-threaded Stories. MS, MIT Media Lab

Train of Thought and **Just One Catch** are two examples of interactive movies in the fiction domain. Just One Catch plays back differently if the viewer asks for more dialog or more action. This could be applied in the information domain: a viewer might want to see more interviews about an event than descriptive reports of the event itself.

3. Point of View and Opinions

W.Sack, 1994,
Indexing Multimedia by
Ideology, AAAI-94
Workshop on Indexing and
Reuse in Multimedia
Systems

Future News: Constructing
the Audience Constructing
the News, Proceedings of
WRITE'94

Coding News and Popular
Culture, IJCAI'93

Warren Sack's project "Machine-Mediated Ideological Analysis of News" explores also viewpoint by trying to *analyse* it and *recognize* it in text based news. "Viewpoints on Demand" focuses on the *editing* and *presentation* of point of view instead. They relate to each other in the same way speech recognition relates to speech synthesis. W.Sacks' software, SpinDoctor, looks for textual clues that give away the point of view of the author(s) of a story. In order to do so it incorporates a representation of ideology and viewpoint. SpinDoctor is able to do a search in a database of articles by matching a query including both a pattern and a viewpoint to an article.

SpinDoctor and Viewpoints on Demand could in fact complement each other if SpinDoctor could be applied to video as well as text. The first one would search for viewpoints on a topic based on input from the second one. The second would then edit a video sequence with the material brought in by SpinDoctor.

4. Video Logging & Video Representation

**Semantic nets and
information retrieval**

Anil S. Chakravarthy's work addresses the issue of incorporating background knowledge of semantic relations into picture and video retrieval systems. His system attempts to match the semantic relations describing an object in the database with the relations extracted between the words of the query. This work treats video sequences *individually and not in relation to each others.*

A.S. Chakravarthy, 1994,
Information Access with
Semantic Background
Knowledge, PhD Thesis
Proposal

However, one could imagine that this work could be easily extended to include the semantic relations about how video sequences relate to each other along with the inferences that can be made.

7 Future work: a Navigational Paradigm and a Storyteller Agent

BUT, AND, OR, BECAUSE...are meaningful navigational text connectors to explore opinions in video.

Techniques that enable dynamic reactions to changes in the environment pave the way toward a storyteller agent.

This chapter describes possible extensions. There are two major directions: navigating opinions and developing a storyteller agent. The last section positions this research in relation to research in automatic logging.

1. Navigating opinions

Navigational problems are a difficult but essential issue in interactive media. A major problem is to offer the user a navigational tool that will not prevent immersion in the content or encourage meaningless exploration.

What I foresee is the ability to interact with opinions while you are watching the story develop. This suggests viewer intervention at a clip level in addition to intervention at a global story level. Interaction with opinion requires *structure* but how much structure is there in the raw content?

An opinion, argument or debate expresses views about the world. Believability depends on how well the statements are supported, how much a given view takes into account other views, what emotions are conjured up and how much.

In its most structured form, an opinion is made up of statements that are supported by facts with lines of arguments. However, an opinion may not be *"substantiated by positive knowledge or proof"*. It can be the position of an expert: *"a judgment based on **special knowledge** and given by an expert"*.

In news reports, journalists collect opinions that may not be well supported by facts and arguments when they are recorded. Their role is to find and verify the facts that may lead to such opinions. Journalists also collect viewpoints of experts. In fact most of the time opinions collected by a journalist are not very likely to be as structured unless an expert or a politician is interviewed in a talk show with enough time to go in depth. This, of course, does not mean that there are not holes in the discourse of politicians. It only means that they are trained to be articulate and convincing. A political speech is more likely to be a form of preaching, aimed at seducing a crowd, relying on dramatic statements that might only be loosely supported by facts. It is very likely to ignore arguments that it cannot counter.

Research papers are also opinions of experts in some form which are very structured, because they want to demonstrate a point and need to be extremely logical.

Opinions on video

Documentary or news footage in its raw form *captures bits of reality in which arguments are not always well structured.* Some people verbally articulate arguments better than others. The more the interview is structured by the reporter, the more likely the answers will be structured but with the risk of preventing surprise and discovery. Cinéma Vérité in particular [ref.] is a form of reporting that does not impose a structure on the content when it is recorded. It is purely observational, and that's the source of its power. Structure comes after, in the editing room and eventually in the narration.

Most of the argumentation in TV and radio reporting appears in the voice-over that articulates events with pieces of interviews and comments on them. The journalist writes and reads a script that describes events, makes some statements or comments and then illustrates them with quotes and images of events.

Opinions in text

Print journalism collects opinions in a descriptive form, including occasional quotes. As with video, the journalists build relations between vari-

ous opinions, events,and facts.

In academic discussions structuring arguments is particularly important. Arguments are presented with *text*. A text-based system called Euclid lets users state a position and defend it using a hypertext type of branching. Other users can then attack arguments by extending this net. This provides an extremely rigorous framework for debate.

This need for structuring appears also in email discussions. In fact, Euclid was created to make text debates more efficient. Text turns out to be a very convenient, flexible and malleable medium for this type of interaction. Even if asynchronous videoconferencing were widely available - allowing you to react when convenient to you - text is easier to edit, making it more flexible when you are shaping an argument that requires preparation.

Opinions: Text vs Video

TV and film brings people to the screen. *"Camera allow us to get to know someone,"* comments Glorianna Davenport. Even if video is a representation of reality, and not reality itself, it can create a very strong sense of "being there" and bring to the viewers people with their sensitivity, their tone of voice and gestures. *Only with the moving image is it possible to show a debate to an audience.* Debates have taken place and are broadcast to an audience on television, not in newspapers. Newspapers on the other hand provide the context and the background information to fully understand the debate. They are more interactive: the reader is guided and gets oriented with front pages, headlines, tables of content, sections, page numbers.

"Literature often has the problem of making the significant somehow visible, while film often finds itself trying to make the visible

The major differences are that 1) TV is extremely limited in time and so has to reduce explanation and supporting facts to a minimum, 2) TV can show people, processes, places, events whereas newspapers describe them, 3) TV can report in real time. Newspapers need more time for distribution but this is changing with electronic delivery. They provide more in-depth analysis. Despite these three main differences, pros and cons are presented in a similar manner.

However, there are similarities. A TV news report has a *structure* that is

somehow very close to a newspaper report. *First,* the written script is an important component of a news report. The voice-over exists in *text* before it is recorded as a narration and combined with images [refs.]. *Second,* interview bites serve a similar purpose as quotes do to support or illustrate a point.

The following section suggests how to apply a structure to a content composed of opinions in order to help the viewer *visualize on the fly a pseudo debate.*

Navigating opinions with text connectors

A promising improvement is the integration in the selection and scheduling mechanism of a specific type of branching to follow lines of argument. A viewer can become an active participant if provided with a means of browsing related opinions and arguments on demand. The interface has to be much more simple than, for instance, what Euclid offers to the academic world, yet it has to guarantee as much rigor as the material allows.

My solution is to use the common *conjunctions of coordination* to link related opinions and show the arguments that are behind them: if a BUT button pops up on the screen while someone is talking, the viewer can choose to see an opposite view. If an OR pops up, s/he could get a similar view. An AND would bring more arguments supporting an opinion or more details. This gives an option to watch additional clips on demand.

These clips need to be integrated with the initial selection. In particular it is necessary to make sure that the fast forward and rewind functions take them into account. Suggestions to the viewer should change accordingly: the same suggestions should not be made twice if the viewer already checked them. This requires *additional adaptability* from the system to changes over time.

It is a combination of video and words: BUT, AND, OR, BECAUSE, SO connectors appear as powerful yet intuitive navigational devices to wander through arguments and opinions. They can pop up on the screen when relevant as you watch a story unroll. They can link you to material that was not selected initially by the story generator.

They provide a branching mechanism that allows meaningful choices instead of blind choices [ref. to limited look ahead]. Navigation through content happens in logical steps under the control of the viewer.

Example

What follows describes an example of interaction using the database built for this thesis.

High level meaning: [fig.1] shows the net of the relations between all the elements of answers currently in the database that support (or explain) the statement "The Gulf War Ended in the Blink of an Eye". A sharp contrast appears between Paul Strassmann's set of reasons and the other interviewees of the database.

Terminology: an object of the net can be a link, a connector or an "opinion item" (or argument). Basically, an opinion item is a video sequence that shows the smallest element that can be cut supporting an opinion.

"The Gulf War Ended in the Blink of an Eye"

BECAUSE

BUT

AND.4

Paul Strassmann

Destruction of command and control happened very fast: first the radars were destroyed, then telephone switches, bridges with cables in them.

Iraq completely lost control. Its air force fled to Iran. This hadn't happened before, even in WWII Germany had some fighting force left at the end of the war

The winning strategy was: we will make the enemy deaf and blind. That had never happened before in the history of mankind.

OR

AND.1

AND.2

AND.3

The impact of air power was completely underestimated. Desert with no trees.

We never knew how low the morale was

The right tactics and right strategy in the right place

Col. Powell's doctrine of overwhelming force

David Greenway

Carpet bombing, as in Vietnam which is what brought the Viet Cong to the peace table

Jonathan Wiggs

Iraq had a strong army but Americans had technology for night vision which made a big difference

We overestimated their equipment

Low morale and bad training

Clearly the bombing had an impact, maybe more by its effect on casualties

Walter Robinson

"Why was the Gulf War so short ?"

P. Strassmann

1

The winning strategy was: we will make the enemy deaf and blind. That had never happened before in the history of mankind.

BUT
AND

Narration

Information as a key element changing the nature of warfare and the outcome of the Gulf War is known only to some information experts. There are alternative explanations to the success of the war.

D. Greenway

2

The right tactics and right strategy in the right place

BUT
AND
OR

3

The impact of air power was completely underestimated. Desert with no trees.

BUT
AND
OR

Figure 1 describes the internal representation of the relations between clips of explanation. Objects like the large gray OR box can be complex and contain recursively sets of connections. *Hierarchies* of logical connections need to be created. The BUT node shows an opposition between two sets of answers.

P. Strassmann
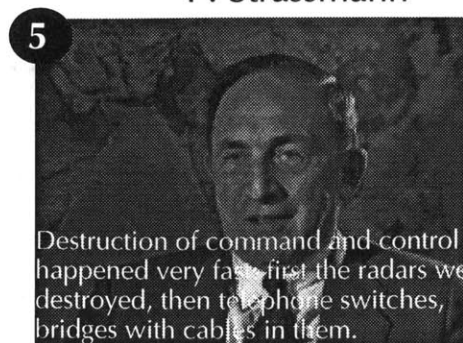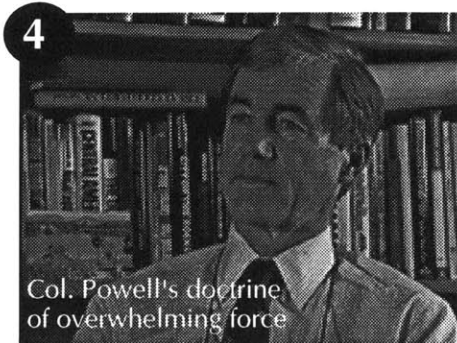
Subsets of arguments produced by various speakers can be in opposition with sets of arguments from other speakers. What does it imply at the level of the user interface and playback of the story?
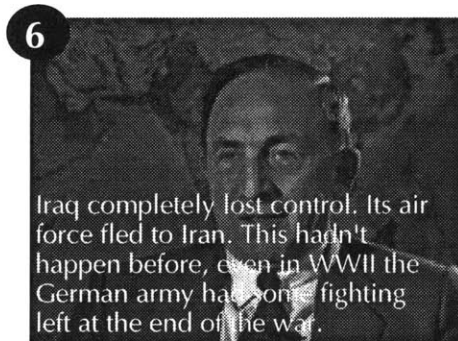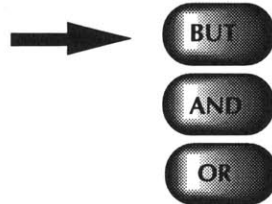
In particular, due to the limited length of the story, the story generator may have selected only some of the arguments related to a given statement like "The Gulf War ended in the blink of an eye". Here is a typical case (read page6):

Fig.2: While Strassmann is talking, the viewer has the choice between getting more arguments supporting his position or getting alternative views

from Greenway, Wiggs or Robinson. In the first case, his options remain the same until the system runs out of clips for Strassmann. In the second case, s/he has a choice between getting more of Greenway (for instance) or switching to Wiggs or back to Strassmann or Greenway. The system keeps track of what has been shown and what remains relevant.

2. Towards a Storyteller Agent

Variations on a theme: "My storyteller knows me"

Under the concept "My storyteller knows me" [MIT Media Lab 5th Anniversary talk], Glorianna Davenport envisions a partnership between a viewer and a smart computer program. This is part of the more global picture in which intelligent agents who can learn about us will help us cope with the deluge of information we are more and more exposed to. As defined in the Society of Mind by Marvin Minsky [ref], multi-agent systems use individual agents which each have a specific expertise. Solutions to problems emerge from the interaction among those agents. We can imagine agents expert in video editing and storytelling that would collect pieces of information in various places in order to assemble a customized documentary.

Why and how would an agent-like structure work

A very promising improvement would be to develop further the engine's ability to adapt by changing on the fly its plans about the selection and scheduling of the clips. Agent-based technology would be appropriate: adaptability to a changing environment is what makes a real difference between an ad hoc program understanding only a fixed data structure and an *agent* having some meta-knowledge about its environment and its task.

"Intelligence is determined by the dynamics of interaction with the world" Rodney Brooks[1]

The traditional A.I. approach has been to decompose into competence modules that are centralized and hierarchical. It does not work well for uncertain, unpredictable and changing environments. Under this model, a robot cannot find its way in a room if it is not lit in a certain way and

1. Rodney Brooks is professor at the MIT Artificial Intelligence laboratory. He has been advocating the use of more distributed knowledge representation models for robots to better adapt to their environment.

if it contains moving objects of shapes which are not known in advance. The key idea introduced by Marvin Minsky in his Society of Mind theory [ref.] is that the competences need to be decentralized and distributed.

This system's environment *evolves* and gradually lets *the viewer extend what s/he already knows* of the content. The narration in particular should adapt to the variation in the settings. A story should be introduced and eventually structured differently when the viewpoint changes. In the current version, as the parameters change, the narration remains the same. It is rigid whereas the rest of the content selected changes. If the viewer pauses in the middle of the story and changes its settings, the story should adapt without going back to the beginning. In the current version, a story is replayed from the beginning with a new selection to fill its template.

Appendix A gives a rough description of the various modules that could interact in order to achieve the desired goal.

Including more sophisticated rules

Editing creates meaning by juxtaposing images. It can be used to reinforce a viewpoint. The viewer can infer this or that, feel one way or another, depending only on how a story is edited.

Juxtaposition can, for instance, create a strong impact on the viewer by showing contrast and conflict. Michael Rabiger [ref.] takes the example of a college student confessing being nervous about a graduation speech. It contrasts with her mother saying how confident she usually is. Juxtaposing both will trigger reactions from the viewer.

A storyteller agent typically would embed rules of behavior inherited from human experience in editing. A very simple rule would be to use a chronology of events instead of a topical approach. More sophisticated rules can be imagined.

Sample of possible high-level rules using editing:

■ IF *two sequences are contrapuntal,* THEN *tie them together*

■ IF *ratio interview/action is high* THEN *intercut more action*

■ IF *rhythm needs to be fast* THEN *make shorter cuts (and vice versa)*

■ IF *exciting scene* THEN *find follow up scene and a closure*

■ IF *goal is to make a strong point* THEN *juxtapose sequences that make a similar point (and vice versa)*

■ IF *expectation is created by a current sequence* THEN *check if new sequence meets this expectation (and act upon the result)*

Some rules can focus more specifically on enhancing a viewpoint:

■ IF *one main character is picked* THEN *maintain it as a strong character*

■ IF *character is afraid of X* THEN *make X look scary*

■ IF *character hates Y* THEN *show Y in a negative light*

■ IF *Y is main speaker* THEN *make Y talk last*

■ IF *Y is main speaker* THEN *insert a silence at end of clip* (to reinforce its impact)

Stylistic "tricks" that can reinforce a viewpoint:

- Audio from an "illustrative" clip starts as a person is still speaking, in a way that reinforces the meaning of what is said ("I hate street noice, I like to live in a farm" followed by a view of the farm with birds singing).

- Achieving the effect visually by using camera framing, camera position or lighting

Note 1: Rules involving cinematography instead of editing are much more difficult to achieve, because the technology for manipulating the content of a frame is still in its very early stages.

Note 2: There is a risk of bias if a speaker looks as though s/he is supporting another viewpoint when in fact s/he is not.

3. Automatic Logging of Video Databases

H.J. Zhang at al., 1994,
A Video Database System
for Digital Libraries, Institue
of Systems Science National
University of Singapore

A question that often comes to mind with video databases is how to deal with large archives of footage. Currently news footage is archived on tapes and labeled using keywords. Keyword-based search applied to digital media is only a very limited retrieval mechanism. It also still implies manual labelling of the content which is a tedious process as video takes time to watch.

There is on-going research aiming at recognizing automatically moving and static objects in a sequence of images. Other projects are also attempting to do speaker recognition and natural language understanding.

M. Hawley, 1993, Structure out of Sound, PhD MIT Media Lab

B. Arons, 1993, Speech Skimmer: Interactively Skimming Recorded Speech, Proc. ACM User Interface Software and technology Conference

C. D. Horner NewsTime: A Graphical Interface to Audio News, MS MIT Media Lab

Parallel to these efforts, it is essential to understand the representation or annotation that should be generated automatically in order to be used by a storyteller system. My project has focused on the presentation of opinions. Future work should involve connecting what has been understood about representation to produce intelligent behavior with results obtained in the areas of Natural Language retrieval, speech recognition (for the audio track) and vision (video track).

It is already possible to pre-log digitized interviews by doing speaker recognition and by locating pauses in the audio track [ref.]. This would also allow the system the footage to precut. A finer level of detail could then be done manually.

8 Conclusion

Thinking about the future too often distracts us from the past as if everything was novelty. Personalization is hardly a new idea. In some cases, it is the presence of mass production that tends to make us overlook the human need for personalization.

Even before the industrial era, one can find examples off suppressed individuality. During his reign, Louis XIV consolidated French power and caused French craft to dominate Europe. However, in the second half of the 18th century, it was Paris, not Versaille that became the center of culture and fashion. The lesser nobility and the bourgeoisie were taking over in Parisian salons. As a result, urban taste developed, in opposition to official court style, a more intimate, *personal* style. Art was funded by private connoiseurs following their own taste instead of the academic style decided by the court (which had brought Molière and the Comédie Française). Individuality had become the rule.

A couple of centuries before that, books were usually sold unbound and most people liked to have them bound to their own personalized requirements.

One might see the automatic generation of personal news and information as the demonstration of automation pushed to an extreme, casting a

negative light and showing the further loss of craftmanship. It used to be that book collectors would look down on printed books, some refusing to have anything but manuscripts in their libraries. Nevertheless, books produced during the first century of printing are now regarded as priceless treasures by book collectors today. One might see a trend: there was more craftmanship in them than in today's books. Personal news, selected and assembled by a machine would only be the next step along that curve.

Or personal news could be seen as a return to the "Siècle des Lumières", to individual values. Automation, instead of resulting in mass production, could result in automated craftmanship. What makes it possible is the multi-agent model that suports the autonomy of the individual as opposed to the centralized model that engenders mass-production.

Il suffisait que Mme Swann n'arrivât pas toute pareille au même moment pour que l'Avenue fût autre. Les lieux que nous avons connus n'appartiennent pas qu'au monde de l'espace où nous les situons pour plus de facilité. Ils n'étaient qu'une mince tranche au milieu d'impressions contigües qui formaient notre vie d'alors; le souvenir d'une certaine image n'est que le regret d'un instant et les maisons, les routes, les avenues, sont fugitives, hélas, comme les années.

Marcel Proust, "A la Recherche du Temps Perdu", "Du Côté de chez Swann".

Bibliography

B. Arons, 1993, Speech Skimmer: Interactively Skimming Recorded Speech, Proc. ACM User Interface Software and technology Conference

W.Bender at al., 1991 Newspace: Mass Media and Personal Computing, USENIX, Nashville

A.W. Blount, 1993 Self-Organizing New M.S..MIT Media Lab

A.D. Campen (ed.), 1992, The First Information War, AFCEA

A.S. Chakravarthy, 1994, Information Access with Semantic Background Knowledge, PhD Thesis Proposal, MIT Media Lab

Jeff Conklin, 1987"Hypertext: an introduction and Survey", IEEE Computer Vol.20 No 9

M. Davis, 1991, The Director's Workshop: Semantic Video Logging with Intelligent Icons. AAAI-91 Workshop on Intelligent Multimedia Interfaces. Anaheim, CA

J.S. Donath, 1986, The Electronic Newstand: Desing of an Intelligent Interface to a variety of News Sources in several media., MSVS MIT Media Lab

Ryan Evans, 1994 "LogBoy Meets FilterGirl", MS MIT Media Lab

M.A. Fineblum, 1991 Adaptive Presentation Styles for Dynamic Hypermedia Scripts, MSVS MIT Media Lab

M. Halliday 1993 Digital Cinema: An Environment for Multi-threaded Stories. MS, MIT Media Lab

M. Hawley, 1993, Structure out of Sound, PhD MIT Media Lab

C. D. Horner 1994 NewsTime: A Graphical Interface to Audio News, MS MIT Media Lab

Brenda Laurel 1989, A Taxonomy of Interactive Movies, New Media News, Vol.3 issue 1

B. Laurel, 1990, Interface Agents: Metaphors with character, in The Art of Human-Computer Interface Design, B.Laurel (Ed.), Addison Wesley

M. Murtaugh, 1994, Con Artist: A System for Graphically Representing Story Knowledge, BS MIT

H.W. Lie, 1990 The Electronic Broadsheet, All hre news that fit the display M.S. MIT Media Lab

J.Orwant, 1991, Doppelganger Goes to School: Machine Learning for User Modelling, M.S. in Media Arts and Sciences

Michael Rabiger, 1987, "Directing the Documentary", Focal Press

E. Rennison, 1994 Galaxy of News, An Approach to understanding Expansive News Landscapes, submitted to UIST'1994

B.Rubin, 1989 Constraint-Based Cinematic Editing. MSVS. MIT Media Lab

Aguierre Smith, T. G., & Davenport, G. (1992). The Stratification System: A Design Environment for Random Access Video. Workshop on Networking and Operating System Support for Digital Audio and Video. San Diego, CA: ACM.

W.Sack, 1994, Indexing Multimedia by Ideology, AAAI-94 Workshop on Indexing and Reuse in Multimedia Systems, Seattle Washington

W.Sack, 1994, Future News: Constructing the Audience Constructing the News, Proceedings of WRITE'94

W.Sack, 1993 Coding News and Popular Culture, IJCAI'93 Proceedings, Chambéry France

R.Schank et al., 1992 Agents in the Story Archive, TR 27, Northwestern University

D. Small, S.Ishizaki, M.Cooper, Typographic Space, MIT Media Lab, Internal Report

K.E.Steiner, T.G.Moher, 1992 Graphic Story Writer: An Interactive Environment for Emergeent Storytelling, CHI'92

Paul Saffo 1992 Consumer Technology Purchase Behavior in the 1990's - IFTF

B. Sheth & P.Maes, 1993, Evolving Agents for Personalized Information Filtering, Conference on Artificial Intelligence Applications, Orlando Florida

A.&H. Toffler, 1993 War and Anti War, Little Brown

Sun-Tzu, the Art of Warfare, c. 403-221 B.C. translation and comments by Roger Ames, Ballantine Books, 1993

Upside, April 1994 An interview with Eric Nee of Scott Cook, CEO of Intuit

P.H.Winston, 1992, Artificial Intelligence, 3rd Ed. Addison Wesley

Nigel Woodhead 1991 Hypertext & hypermedia, Addison Wesley

Barbie Zelizer, 1992, CNN, the Gulf War, and Journalistic Practise, Journal of Communication, Vol. 42 No1

H.J. Zhang at al., 1994, A Video Database System for Digital Libraries, Institue of Systems Science National University of Singapore

2809-44