PERSONALCASTING:

INTERACTIVE LOCAL AUGMENTATION

OF TELEVISION PROGRAMMING

By
Victor Michael Bove, Jr.
Bachelor of Science,
Massachusetts Institute of Technology 1983

Submitted to the Department of Architecture
in Partial Fulfillment of the Requirements for the
Degree of Master of Science in Visual Studies
at the
Massachusetts Institute of Technology
September 1985

© **Victor Michael Bove, Jr. 1985**

Signature of Author _____
Ⅴ. Michael Bove, Jr.
Department of Architecture
August 9, 1985

Certified by_____
Andrew B. Lippman
Associate Professor, Media Technology
Thesis Supervisor

Accepted by _____
Nicholas Negroponte
Chairman
Departmental Committee for Graduate Students

CONTENTS

PERSONALCASTING:
INTERACTIVE LOCAL AUGMENTATION
OF TELEVISION PROGRAMMING

By Victor Michael Bove, Jr.

Submitted to the Department of Architecture

in Partial Fulfillment of the Requirements for the

Degree of Master of Science in Visual Studies

September 1985

Thesis Supervisor: Andrew B. Lippman

Associate Professor of Media Technology

ABSTRACT

While a videocassette recorder allows a television viewer to decouple the viewing of a television program from its broadcast, its use would be much more rewarding were it able to "understand" what it had recorded and to utilize this information to vary the presentation of broadcast television programs in a personalized manner. A hardware/software system is developed which uses closed-captioning information as a data input and allows variation in the presentation of television newscasts, with results applicable to the locally-intelligent recording and personalized presentation of other sorts of programming as well.

INTRODUCTION

Personalcasting

The goal of the Personalcasting project is to explore variation of broadcast television programming through a combination of computing and home storage of program material. The recording and playback processes will be changed by the computer's use of descriptive closed-caption information transmitted in combination with programs. Information derived from broadcasts will allow storage and manipulation of program material at a level of granularity smaller than the program level and based upon the content of the broadcast rather than just upon its clock time. The system can then present to the viewer a personalized, processed subset of what it has recorded, and the content information will permit the viewer to vary the playback of the material in a sophisticated manner.

The television news is used as as a test environment for the prototype system for several reasons. The news lends itself to variation, expansion, and compression, its content can bring about changes in the viewer's activities, and it presents some instructive database-creation problems. Also, the news is a good example of an area in which a great deal of useful information is provided by various sources, but without any way for the viewer to relate or utilize it without watching it in its entirety. It is clear that this system could easily expand to create a productive intersection of various sorts of programming including sports, interviews, documentaries, and entertainment.

The Videocassette Recorder

The videocassette recorder, or VCR, has taken the American household by storm. Just three years ago, domestic annual sales of VCR's totaled only a half-million units [Diamond 1982], while this year the industry expects sales to reach twenty-three times the 1982 figure, or 11.5 million [Monitor 1985]. These machines can be used to play prerecorded movies or, when equipped with inexpensive cameras, to make video "home movies," but most units include timers and TV tuners, permitting them to "time-shift," or automatically record broadcast programs to be watched later. Owners of copyrighted programming have attempted to block this practice or to require VCR owners to pay them royalties, but in January 1984 the Supreme Court ruled that taping television programs for later personal viewing is "fair use" of copyright materials and therefore not liable for copyright royalty payment [US 1984][Broadcasting 1984].

The common stereotype of the average American viewer as passively, mindlessly glued to a television receiver was perhaps never entirely true, but given the plethora of channels available on cable systems and the VCR's ability to show both previously-recorded broadcasts and programs rented or purchased from the new flock of "video stores" the viewer now has access to a much greater variety of programming and can watch carefully-selected material at his or her convenience rather than having no choice but the networks' prime-time offerings. The motivation to become an active participant in television watching is not simply that new opportunities exist; there is

also the cost of VCR's, tapes, and monthly cable charges — video information and entertainment is becoming a sizeable enough investment to merit increasingly thoughtful use. That viewers are actively

switching channels can be demonstrated by the fact that top-rated programs vary from network to network throughout the course of a day [Diamond 1982].

A VCR acts as a sort of "asynchronizer," in that it decouples the viewing of a television program from the broadcast. This asynchronization

FIGURE 1
Remote control for a consumer-model videocassette recorder provides many functions, including forward and reverse play, slow and fast forward and reverse, still-frame, one-frame-at-a-time advance, and programmable timer.

takes two forms: first, the VCR's storage capability provides an arbitrary delay between the airing of a show and the viewer's actually watching it; second, its controls permit variation in timing of presentation — using the VCR the viewer can speed up, slow down, skip, or repeat portions of the program, or pause the tape when the telephone rings (Figure 1).
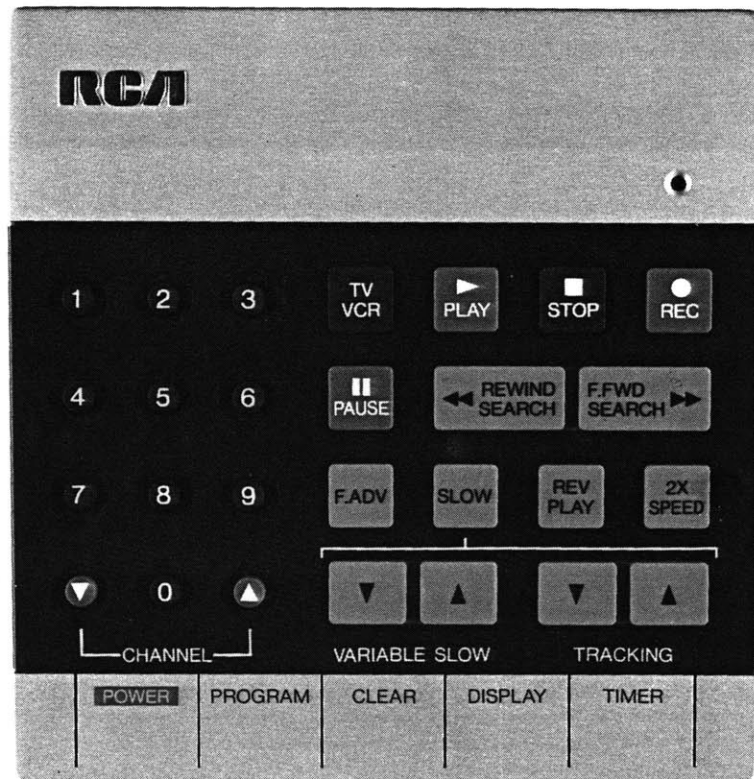
This decoupling of presentation is greatly discussed nowadays, owing to the VCR's capacity for "commercial zapping," or fast-forwarding the tape past commercials, a viewer practice which can considerably reduce the advertising value to a sponsor of a show which viewers are likely to record. Advertisers and networks worry, too, about time-shifting, which not only plays havoc with current methods used to calculate ratings and audience share for programs, but also devalues "prime-time" broadcasting hours when people are likely to be home and watching television [Whiteside 1985b]. Time-shifting appears to be here to stay, however, and at least the Public Broadcasting Service (which, perhaps, worries less than advertiser-supported networks about ratings) has accepted this fact; PBS's programming director recently suggested that the proliferation of VCR's will allow some of their most popular material to be broadcast at non-prime-time hours on the assumption that the audience can record it [Monitor 1985].

Using a VCR's timer, a viewer might record just the sports report from one channel and the weather from another if these segments come on at exactly the same nonoverlapping times every day, but generally a timer will allow only program-level manipulation, and can only assemble programs one after the other in the order in which they were broadcast. Far more useful would be the ability to record broadcast material based on content rather than on clock time, slicing up and rearranging programming from multiple sources to create essentially new programs. Variable playback, too, would be improved

by operating on a content basis. Currently, fast-forwarding through a tape demands as much concentration as playing pinball, requiring the viewer to pay constant attention to the speeded-up pictures. One area in which a content-controlled VCR would be useful is in the personalization of news and information programming — a VCR which actually "understood" what it had recorded could offer features such as a "skip forward to the next item" button, while more complex functionality could effectively de-linearize a program, for example permitting the viewer to stop a news story, watch a previous night's report giving background information on it, and then jump back into the current story.

The Television News

While broadcast television became technically feasible in the late 1930s *(Figure 2)*, its development was hindered by World War II and by FCC delays in resolving conflicts over technical standards. After the war, commercial television began in earnest, and the return of many people with electronic, journalistic, and newsreel-filming experience gained during the war helped launch the first television network news operations *[Barnouw 1982]*. Small in budget, these pioneering 15-minute shows depended heavily on film footage purchased from theatrical newsreel production companies or shot on Signal Corps-surplus Mitchell cameras. As television cameras became more portable and less expensive, film footage was supplemented by live cable feeds from correspondents on location; then in 1957 the first Ampex videotape machines, priced at $45,000, hit the market. More

adaptable than film and viewable immediately without processing, tape is a medium perfectly suited for television news production. The communications satellite Telstar I, launched in 1962, permitted events to be televised live from far-flung places. (Telstar, which rapidly traversed continents, was usable only when the signal source and destination were in its line of sight. It was soon replaced by geosynchronous satellites like Early Bird.) Better-budgeted and relying upon increasingly flexible technologies, network news expanded to a half-hour in 1963. Local news programs, too, increased in importance, and news programs began to be supplemented by documentaries

FIGURE 2
John Cameron Swayze reading a Kansas City newspaper on an experimental 1933 telecast.

and "magazine" programs which examined issues in depth.

Nowadays, portable color television cameras with built-in videocassette recorders (for "ENG" or "Electronic News Gathering") and computer-generated graphics both gratuitous and informational are the technological marvels of the TV news world, satellite feeds and "key windows" (which allow one video image to be inserted into part of another) having become so commonplace as to go almost

unremarked *(Figures 3, 4)*.

Technology has unquestionably improved broadcast journalism. Before communications satellites and portable TV cameras, events in remote or underdeveloped portions of the world tended not to be covered by the television news, which emphasized even more than it does now the visual aspects of a story over its importance in a real-world context *[Barnouw 1982]*. Cable television, by increasing the number of channels available, has increased the viewer's choices
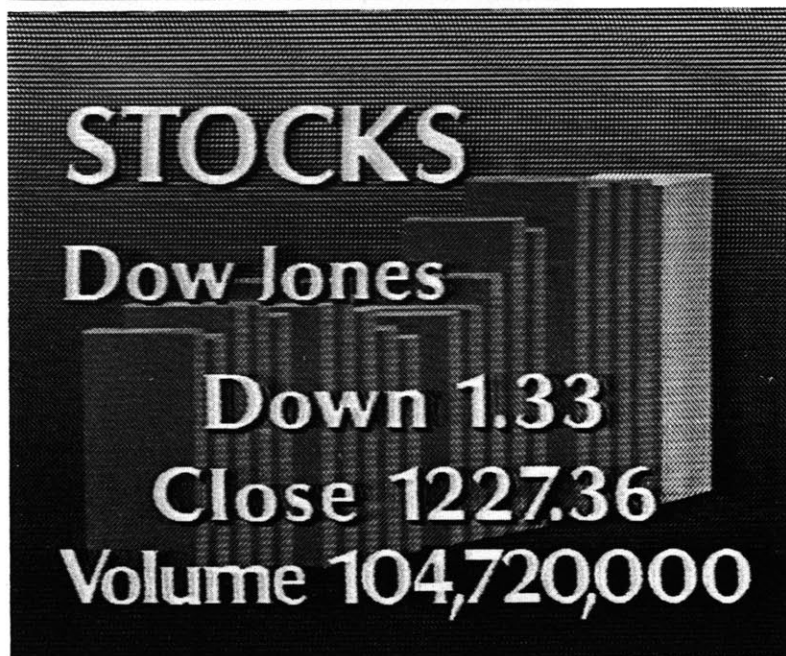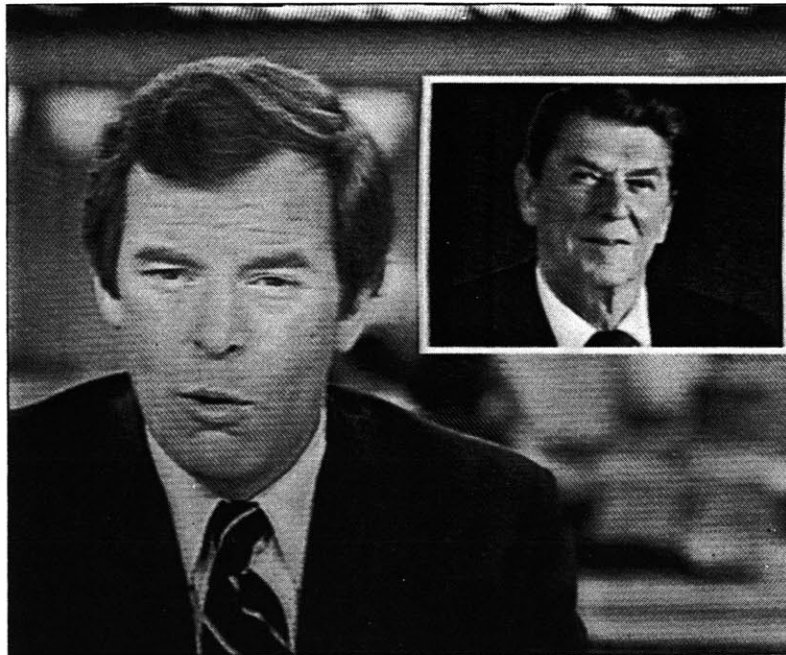
FIGURES 3 & 4
Television news today:
(top) A second image is "keyed" into the first.
(bottom) Computer-generated graphics display numerical information and also make the program more visually interesting.

and has permitted the launching of television entrepreneur Ted Turner's Cable Network News, a 24-hour, news-only channel (though critics charge that always-running CNN, without a strong organizing format, fails to provide a sense of order among "hard news" stories *[Diamond 1982]*). On the other hand, network news has become enough of a prestige vehicle that networks can become wrapped up in presentation over content; the rising importance of the "anchor-as-personality" and the anchor's increasing power in administrative decisions attests to this situation *[Matusow 1983]*.

Despite these technological advances in news gathering and presentation, broadcast television news remains a temporally linear medium. Given only one program at a time, a television network attempts to appeal to the broadest possible audience, and the result is a program which is simply acceptable to all, rather than tailored for a specific segment of the audience. One must watch each news item in its entirety — whether or not it's of interest — before seeing the next, and there is no way of determining what the remaining items are. The latter is a flaw in that a person receiving a thin newspaper will often read an article of marginal interest that he would not read were there a lot of other items competing for his attention.

The variability available by switching the channel knob is at best limited — utopian predictions by cable proponents notwithstanding, television has not broadened to provide a channel for boat-owning pathologists with property holdings in the Midwest, and it is unlikely that it ever will. It does not need to. This thesis explores a

means for providing personalized news broadcasts ("Personalcasts")
for such people — and for anyone else — by varying the presentation
of locally-recorded broadcasts. The lessons learned will have applica-
tion to the broader field of local personalization of other sorts of
broadcast material.

Variable Video through Computing

The desired sort of operation described above requires the ad-
dition of computer control to the VCR, first so that it can make deci-
sions as to what to record and second so that upon playback it can
exploit the potential variability provided by storing away television
programs in rearrangeable fashion.

A number of methods have been explored for providing vari-
able presentation of information through the combination of video
with computing power. A Teletext system turns the home television
receiver into a new information resource; the television broadcaster
cyclically transmits data (either in the vertical blanking interval of ordi-
nary broadcast programs or as full frames of a dedicated channel)
making up a limited number of "pages," and a microprocessor in the
local decoder interprets this information to generate a text and
graphical image of a desired page [Brockhurst 1982]. Viewdata or
Videotex systems such as Great Britain's Prestel [Bright 1979] provide
more flexible, two-way communication via telephone lines, transmit-
ting viewer input to a central computer which returns data for decod-
ing and display. Increasingly complex systems include animated
computer-generated graphics [Pabouctsidis 1984] or real still images

[Clarke 1980], and can support fairly sophisticated applications such as television catalog shopping and remote signature verification.

In the late 1970's, MIT's Architecture Machine Group recognized the possibilities inherent in the laser videodisc, originally developed as a way of merchandising old films in the days before the VCR explosion. One disc provides essentially random-access storage for up to a half-hour of action sequences, 54,000 still images, or a combination of stills and live footage. 54,000 still images is, of course, too many for a user to peruse in any sort of rewarding, useful way without the mediation of a computer — looking through the disc at the rate of one image per second would take fifteen hours! But when the disc player is put under computer control, a the contents of the disc may be accessed and combined in sophisticated ways and a sort of fusion takes place between the computer and the display system.

Through local data storage and computing, the videodisc can become the basis for systems as varied as the interactive Movie-Map *[Lippman 1980][Mohl 1981]* and the electronic book known as the Movie-Manual *[Backer 1981]*. NewsPeek *[Lippman 1984]* uses both local videodiscs and remote databases to create an electronically-delivered video "newspaper."

SYSTEM PRINCIPLES AND DESIGN GOALS

The central concept behind the Personalcasting project is to store and to play back television programming based upon its content. Programmable, content-based storage in combination with playback controls which utilize the content information possessed by the system will enable the viewer to watch a personalized presentation put together from multiple programs and sources, both programs broadcast on one channel at different times and programs broadcast simultaneously on different channels. The result will be a richer television experience than would have been available without this system.

Merely making recorded information available will not necessarily motivate the viewer to use the added features of a content-based recording / personalized presentation system. The interface to the user will be of great importance in making the use of this system pleasant and rewarding. This concept is illustrated by the large number of people who telephone for directory assistance even though they have a phone directory available and the call nowadays costs them money.

Therefore, several goals will be important throughout this investigation, and will have implications for design of both hardware and software:

• *The system should augment normal television watching, rather than turn it into something totally different.* Just as VCR's and cable provide the viewer with added choices and functionality, so should this system. The viewer's experience, though, should basically

remain that of watching television. Accordingly,

- *The system should endeavor to produce a coherent, interesting presentation whether or not the viewer chooses to interact with it.* The content of the "default case" presentation should be at least as good and probably better than that which the viewer would see watching an ordinary broadcast or tape; it mustn't be necessary to press buttons in order to bring about some result which normally occurs without user intervention. So in order to motivate the viewer to make use of the added features of this system,

- *Possible interactions should be simple to use and sufficiently rewarding,* as is the fast-forward button on a VCR, for example. Finally,

- *Presentation quality should not be forfeited for personalization.* Images and sound should not be appreciably degraded by whatever storage or processing the system will perform upon them.

THE PERSONALCASTING SYSTEM

Parts of the System

In most general form (Figure 5), a system for content-based recording and variable playback will have an input (television broadcasts), a controllable storage and playback device (the recorder), an output (a video monitor and loudspeaker), and a storage and playback controller (the computer). The controller somehow must examine incoming broadcasts and decide whether or not to record them, turning the storage device on and off accordingly, as well as respond to user interactions (alternatively, and perhaps more easily from a software perspective, the controller could record likely programs and then later decide whether or not to "keep" them — if the system has a large amount of storage these two modes of operation are similar in result). It must also create a database describing what it has recorded and later use this information to play back portions of the stored input — not necessarily in the same order in which they were recorded — displaying the resulting presentation on the video monitor. This sys-
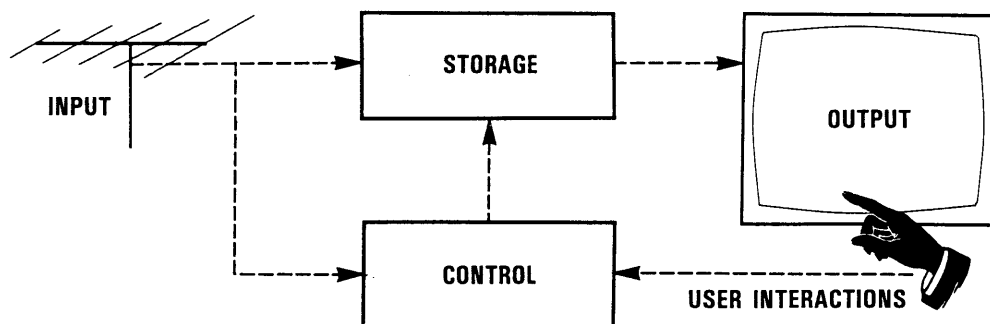

FIGURE 5
A simplified block diagram of a system for content-based recording and variable viewing of television broadcasts.

tem will differ from systems like the Movie-Map in that its repertoire of video sequences will not be fixed. It will constantly be changing the material it has to show and automatically updating the database describing that stored material.

Though the path from the input to the storage device is shown as a single line, and though a television program is generally thought of as a single entity, an instructive way to think of the broadcast channel or stored program material is as several simultaneous but not necessarily related data paths: a high-bandwidth path for the video signal, plus lower-bandwidth paths for audio (nowadays stereo), and perhaps data. In this system, some of these paths carry information which will be presented directly to the viewer, while the content of others may be processed in some way before presentation or may be used for internal "bookkeeping" and essentially invisible except to the computer. Further, this input does not have to be limited to one source — given suitable storage devices the system might record several channels at once if material of interest is being broadcast on all of them.

Storage

The controllability and capacity of the storage medium used will to a large extent affect system characteristics. Currently available disc-type storage, either optical or magnetic, offers fast seek time but limited capacity, while magnetic tape can store substantially more video but being of a linear geometry has a much longer search time which increases with distance.

The storage block in the preceding figure does not necessarily

represent only one physical device. Multiple units may need to be combined in order to fulfill system requirements or to overcome shortcomings of a particular storage medium *(Figure 6)*. When large-capacity, slow-search media like videotape recorders are cascaded in series *(Figure 6b)*, material recorded straight off the air by the first device may be pre-edited onto the second in order to reorganize it for faster access and to eliminate useless material. Also, the system may record onto the first while the second is playing. Storage devices with less capacity may be connected in parallel *(Figure 6c)*, and in the process the system will gain the ability to handle multiple simultaneous inputs. Fairly short tapes (to reduce

a) SINGLE DEVICE

b) DEVICES IN SERIES

c) DEVICES IN PARALLEL
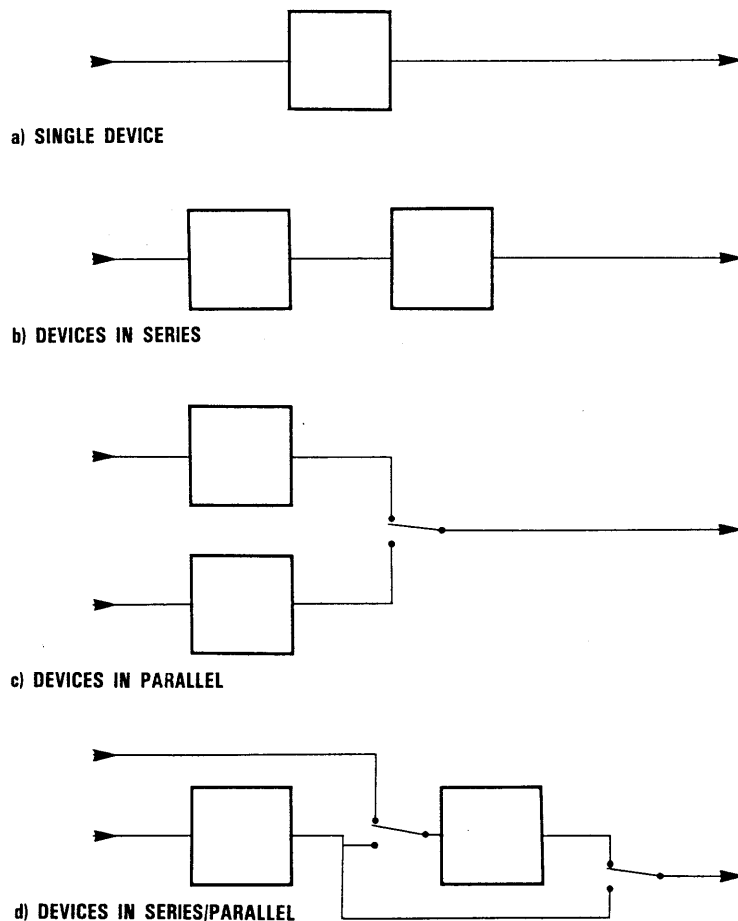
d) DEVICES IN SERIES/PARALLEL

FIGURE 6
Increased flexibility and functionality are provided when a single storage device (a) is replaced by multiple units. These may be connected in serial (b), parallel (c), or serial/parallel (d) form, depending on the capacity and search characteristics of the devices and on the intended applications of the system.

search time) might be used in this arrangement as well as discs. Perhaps the most flexible arrangement is a combination of serial and parallel, as shown in *Figure 6d*. This configuration will allow investigation of both multiple inputs and automatic editing.

Consumer-model VHS videocassette recorders were used as storage devices in this project, as they not only can store a large amount of material but also are fairly inexpensive (costing a hundredth as much as the first videotape recorders of thirty years ago) and include tuners. Further, it is fairly simple to convert the remote control input to a serial interface, thus obtaining computer control over both the tape transport and the tuner. Most VHS decks can record at three different speeds: standard-play or SP (33.35 mm/sec), long-play or LP (half-speed, or 16.67 mm/sec), and super-long-play or SLP (one-third-speed, 11.12 mm/sec). A standard T-120 videocassette — the size commonly displayed at supermarket checkouts — can hold 2, 4, or 6 hours of recorded material, and the less-common T-160 can record up to 8 hours. On playback, the deck automatically selects the correct speed by synchronizing its transport movement to the control track on the tape, which contains one pulse per video field. Considerable degradation of the video and audio signals takes place at slower recording speeds. The horizontal resolution drops, and the signal-to-noise ratio of the images decreases — a particularly noticeable effect on stills and slow motion. For a typical VHS VCR, from SP to SLP the video signal-to-noise ratio drops from 46 dB to 40 dB and the audio from 43 dB to 40 dB; a far more apparent effect on audio is the reduc-

tion in bandwidth (50 Hz-10 kHz at SP and 100 Hz-6kHz at SLP) [Hitachi 1984]. Cumulative degradation through multiple generations of re-recording is much worse at slower speeds than at standard play speed.

SMPTE audio time code labels each frame on a video tape by encoding frame number information as tones in the tape's audio track. This self-clocking signal is designed so that it may be decoded both forwards and backwards at slower- or faster-than-normal speeds, if the VCR provides audio output while shuttling the tape. Consumer-oriented VCR's (and VHS decks, generally) do not include integral time code hardware, but generators and readers which connect to line-level audio inputs and outputs and provide serial data are easily obtainable. Since VHS decks have stereo sound, one channel can be used for time code and the other for sound recorded from the broadcast. Time code may be placed on a tape while recording video and its corresponding audio, or, if multiple short segments will be recorded onto a tape at different times and contiguous time code from beginning to end is desired, a blank tape is pre-time-coded and recording is disabled on that audio channel. Finding a specific point on the tape involves moving the tape forward and/or backward and repeatedly reading the output of the time code reader until it matches the desired frame number.

While the VHS deck provides inexpensive mass video storage, it is far from a random access device. An optimal search routine using time code is difficult to program, as consumer decks — unlike decks

designed for editing suites — don't output audio during shuttle; there-fore it is necessary to shuttle to the approximate location and play for a few seconds to determine the actual current location. The task of getting to the "approximate location" is made difficult by the fact that the acceleration characteristics of the mechanism vary with the amount of tape on the supply and take-up sides of the cassette and with the length of tape within a cassette. It is possible to modify con-sumer VHS decks to disable the internal audio muting and thus pro-vide constant audio, though the transport mechanism on some mod-els is such that a tape moving faster-than-normal may not always make good contact with the audio head. Even given constant time code and a very good search routine, maximum tape speed greatly limits the system — the VCR still can still require up to several minutes to rewind or fast-forward from one end of a long tape to the other, during which time the screen goes black.

The black-screen problem is a common one in this type of sys-tem. In the case of videodiscs — where, admittedly, the end-to-end search time is measured in seconds, not minutes — the usual solution is to have a second disc cued up to the alternative the viewer is most likely to choose, and to have all related alternatives located near one another on the disc. Searching within a half-hour portion of a videocassette is substantially faster than locating something two hours' of tape away from the current position, but clever reorganiza-tion of material is not available to a one-tape-recorder system whose sequences are recorded directly off the air, and this fact alone argues

strongly for the inclusion of a second VCR. If some other constructive use can be made of the video monitor while a tape is searching, then search speed ceases to be a major problem. This issue will be discussed in more detail later.

Control

The amount of information which the controller can derive from the broadcast input will determine the functionality of the system; the more it can "understand" of the broadcasts it receives, the richer becomes the database and the more complex and useful become the possible interactions the system can support.

For the past several years there has existed in the US a system for the captioning of broadcast television programs for the hearing-impaired. As the digital code inserted during the vertical blanking
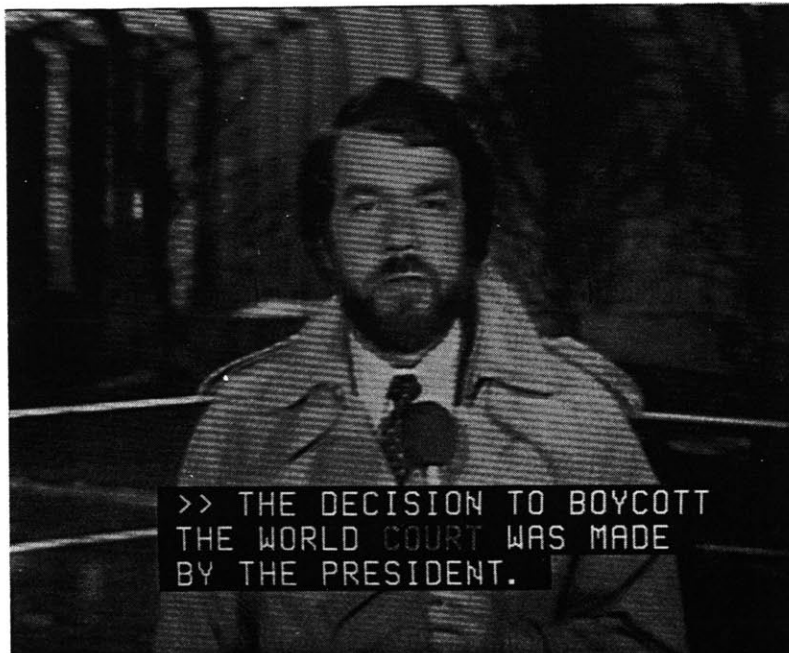

FIGURE 7
Closed-captioning provides a transcript of a network news broadcast.

interval is invisible on television sets not equipped with a decoder — which interprets the transmitted data and superimposes characters onto the picture — the system is known as "closed-captioning" (Fig-

ure 7). Add-on decoders and television sets with integral decoding and display electronics are available from a variety of sources including national chains like Sears, Roebuck and Company *[Sears 1985]*.

An increasing number of television broadcasts are now closed-captioned. The American Broadcasting Company captions its evening news (locally broadcast on Boston's Channel 5), providing a word-for-word transcript of the program. A Sears closed-
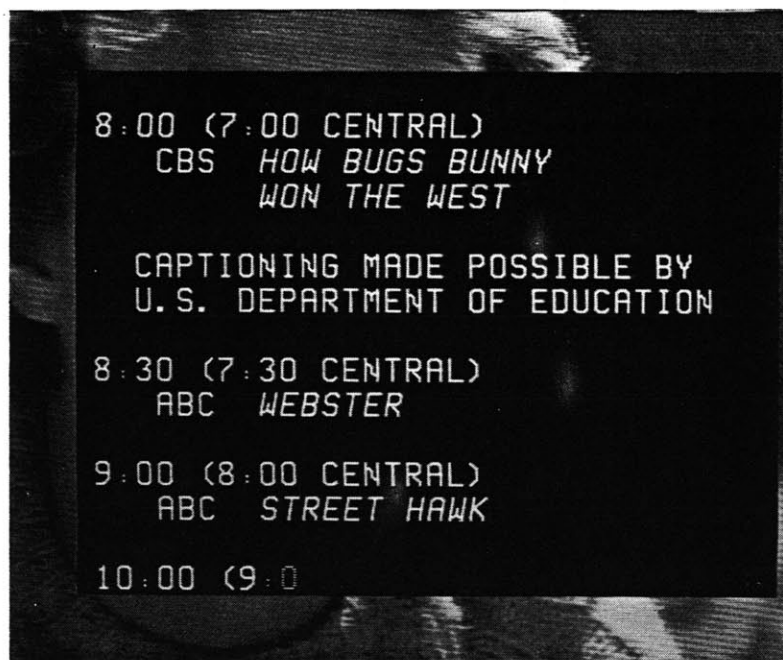


```
8:00 (7:00 CENTRAL)
   CBS   HOW BUGS BUNNY
         WON THE WEST

   CAPTIONING MADE POSSIBLE BY
   U. S. DEPARTMENT OF EDUCATION

8:30 (7:30 CENTRAL)
   ABC   WEBSTER

9:00 (8:00 CENTRAL)
   ABC   STREET HAWK

10:00 (9:0
```

FIGURE 8

Infotext — here, a listing of captioned programs — multiplexed in the same channel as closed-captioning.

caption decoder was modified so that the received data, rather than being displayed on the screen, is buffered and output to a standard RS-232 serial data port for connection to the computer.

Captions are encoded into line 21, field 1 of each frame of an NTSC broadcast signal in the form of two 7-bit ASCII characters with odd parity added, preceded by a data clock synchronizing "run-in" signal and a start bit. Bits are sent according to a standard non-return-to-zero (NRZ) code. As the frame rate of NTSC television is 30 frames/second, the net data rate is 480 baud, fairly low-speed communica-

tion when compared with typical computer data transfer rates but sufficiently faster than typical speech to allow multiple captioning channels to be multiplexed onto one signal. Sequences of non-printing control characters exist for positioning caption windows (so that individual characters may have "voice balloons" like those in comic strips), as well as scrolling, underlining, italicizing, and coloring the text. Other control sequences allow the multiplexing of two languages of captioning and the further multiplexing of the captioning signal with an "Infotext" channel which provides screenfuls of information not necessarily related to the program being broadcast (Figure 8). A typical closed-caption decoder is equipped with mode switches allowing either of the languages or text mode to be selected for display [Lentz 1980].

The Database

Software was written in the C programming language on a SUN Microsystems workstation, a minicomputer based on the Motorola 68000 microprocessor and running the UNIX operating system. The amount of storage and speed of processing required for database handling and device control are well within the capabilities of even smaller personal computers.

The development of schemes for the organization of knowledge, and further for the organization of knowledge about knowledge, proved a frustrating task for ancient philosophers and has more recently proven almost no less baffling to modern library scientists. The task may be made easier by limiting the scope of the field of

knowledge being cataloged, and by dividing the knowledge into well-defined types of facts placed within a simple structure. It is not necessary here to classify everything that might be put onto a videotape but rather to limit the contents to items gathered from television news broadcasts, which are already fairly consistent in form. When a set of facts the system needs to know about a given item is established, designing a database for the system becomes not at all formidable.

FIGURE 9

The contents of a videotape are described in a hierarchical manner: a tape is made up of multiple programs, each containing multiple items of multiple scenes. Specific pieces of information are associated with each tape, program, item, and scene.

The form of the database is based upon the hierarchical nature of a news broadcast (Figure 9). A TAPE, which corresponds to a physical cassette, has an associated "label", and is in turn made up of some number of PROGRAMs. Each of these has a name, a date, a time, and a channel, and contains ITEMs ("stories"). Each item has a headline, a location (a dateline-type location, not the physical location on the tape), a category (local, national, international, financial, health, sports, weather, features, commentary), a "major story" (if any)
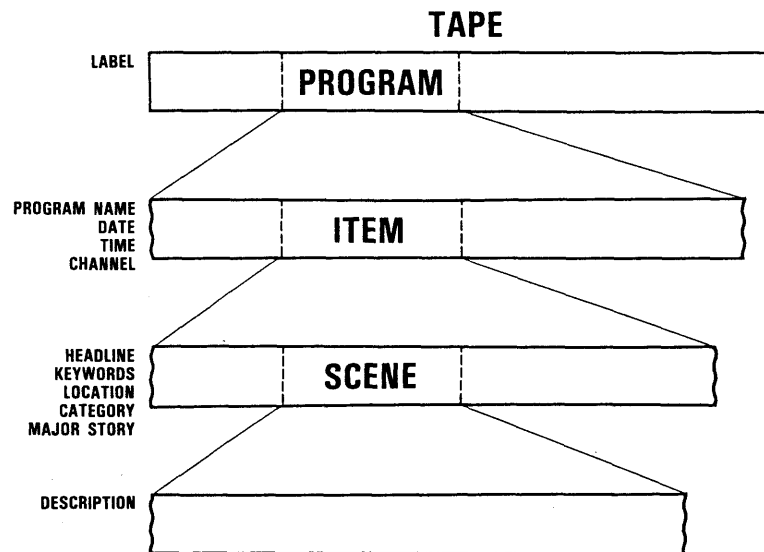
of which this report may be a part or update, and a group of associated keywords. At the lowest level are SCENEs, each with its description (like "anchorman" or "Kremlin").

C data structures containing the appropriate pieces of information for each level are defined for tapes, programs, items, and scenes. Most of the low-level software functions do not directly handle these structures in a hierarchical manner, however, but for more efficient processing deal with pointer unions called DETAILs:

```
typedef union {
        TAPE *tape;
        PROG *prog;
        ITEM *item;
        SCENE *scene;
        } DETAIL;
```

A linked list of DETAILs and their timecodes is created at each level in the hierarchy (Figure 10), where each DETAIL is given pointers to the next one and preceding one on its level, as well as to its constituent parts (if any) on the next lower level and to the DETAIL of which it in turn is a constituent part:

```
typedef struct encode {
        struct encode *parent;
        struct encode *parts[MAXPARTS];
        struct encode *back;
        struct encode *front;
        int number_of_parts;
        int start_timecode, end_timecode;
        int flags;
        DETAIL detail;
        } ENCODE;
```

These linked lists are the actual database describing the tape, with the DETAILs arranged in increasing timecode order. Scenes are therefore treated as the smallest rearrangeable elements of the re-

corded material, though it will always be possible to stop the tape at any point within a scene and to skip to the beginning of another one (it is easy to see how sentences, words, or even frames might be added to this database at increasingly lower levels, should an application require it).

A printout of the information regarding the contents of a tape might look like the following:

tape.0: 3765 - 208645
Michael and Judith's current news tape
    program.0: 3884 - 54673
    program: ABC News; **date:** 8 April 1985; **time:** 19:00; **channel:** 5
        item.0: 3884 - 10704
        headline: Soviets propose missile freeze
        keywords: Reagan, Gorbachev, missiles, arms, weapons, disarmament, defense,
            US, USSR
        location: Moscow; **category:** international; **majorstory:** Disarmament
            scene.0: 3884 - 4533
            description: anchorman
            scene.1: 4533 - 6606
            description: Kremlin
            scene.2: 6606 - 10704
            description: White House press conference
        item.1: 10704 - 11276
        headline: British and West Germans Protest US Missiles

. . .and so on. Portions of the tape not of use to the system are not explicitly labeled as skipped, but rather the ending time code of the preceding entry in the database differs from the beginning time code of the following one. For simplicity, commercials are not currently included in the database (the system "commercial-zaps" automatically), but the system could be modified to pick out commercials that are of specific interest to the viewer.

A linked list of *DETAIL*s will be followed by the tape-playing software when the viewer watches recorded broadcasts. Therefore a

different, rearranged linked list of *DETAILs* can be made to describe an alternative "trajectory" or presentation order for material contained on a tape or group of tapes, and the list might be further reordered by the viewer's input to the system while watching.

It must be noted that this database is "ideal" in the sense that its development depended in no way upon the information that might easily be obtained from the incoming video signal and caption data, but rather upon the sorts of information that might be useful to a system handling news videotapes. Still, a fair portion of the blanks in this database may be filled in quickly and automatically by programs which process the data in the closed-caption channel.

The closed-caption decoder sends ASCII null characters when no caption information is coming in, and thus if the output of this
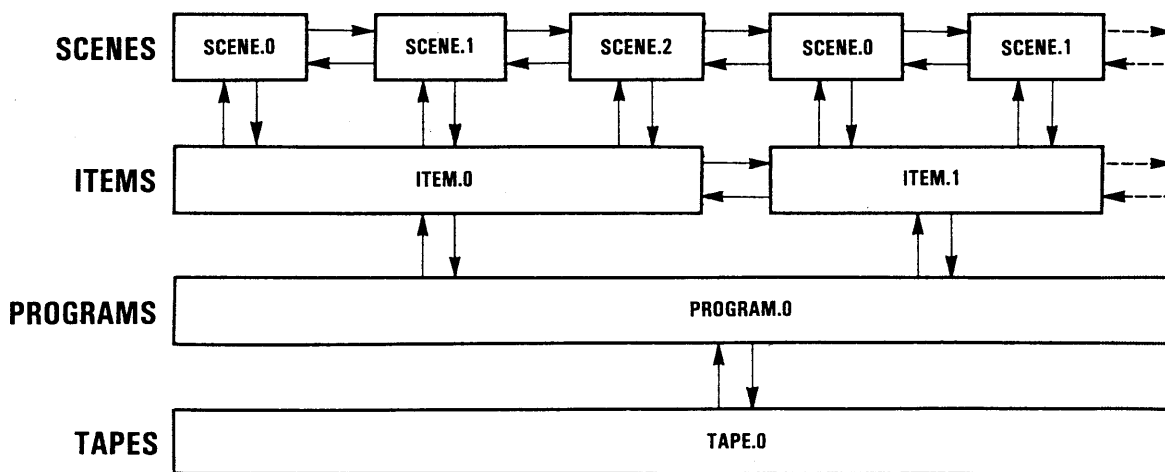
FIGURE 10
The database contains linked lists of scenes, items, programs, and tapes, and elements at each level point to their "parents" and constituent parts.

device is stored on a computer the resulting file will be padded out with null characters so that there will be exactly twice as many characters as frames. It is therefore necessary to note the time code only at the beginning of a video recording — the tape location of any caption character can then be determined by simple character counting.

Further processing of the closed caption file finds words (by looking for space characters) and sentences (by looking for punctuation), and the location of each of these syntactical elements is noted. In the case of the ABC News, two greater-than signs or right arrows (►►) in the caption introduce a new speaker and three (►►►) a new story. A simple heuristic model of this program based on observation therefore enables the system quickly to find the beginnings and ends of news stories and insert these into the database. More sophisticated (though not always accurate) processing to find certain key phrases or elements can make guesses as to when the studio anchorman is on screen versus when remote footage is being shown, or to the location of commercials (this latter case involves more than waiting for gaps in the captioning, as many television commercials are now captioned — indeed, advertising agencies tend to make much more effective use of the position, scrolling, and other display capabilities of closed-captioning than do most program producers). Analysis of the video signal is probably necessary for reliable detection of scenes within a story, but equipment is now on the market which automatically compares sequences of frames in order to find changes of scene [Dubner 1984]. At this point scene-detector boxes are costly

enough to be practical only for television stations and large production houses, but there is every reason to believe such devices can be mass-produced fairly inexpensively if the demand warrants. The database-filling software for now creates one scene per story. Ideally, of course, the broadcaster could transmit data in the second caption channel which would tag the beginnings and ends of items and include the rest of the necessary information as well.

The easiest way of automatically filling in the keywords involves having a list of known keywords and searching the text of each story for these, recording those which are found. This method is obviously far from foolproof, as is attempting to assign a category based on what keywords are found. Headline-writing for newspapers is a difficult skill to master, and no attempt is made at automatic generation of real headlines, though the first sentence of the text might be used. Similarly, the major story is not currently entered by the computer. Since it is not possible to fill in the entire database automatically and since entries occasionally require correction, a simple database editing program was written for testing and demonstration purposes. This program permits printing out the database in "outline" form, as above, and adding or changing entries by hand.

True computer understanding of the transcript of a news broadcast is a topic right on the outer limits of artificial intelligence research, and falls far outside the scope of this thesis as well as beyond the capabilities of current small computers. However, certain elements of "expert" or rule-based systems principles could be applied to the

database-creation problem *[Hayes-Roth 1983]*. Rather than having a hard-coded parser which derives desired information from the caption input, a set of rules ("On the ABC News, three arrows means start-of-story," "The financial segment of the ABC News is always introduced 'On tonight's Money Matters' ". . .) could be established separate from the database-accessing software. On playing the tape, a different set of rules would then interpret a fairly general viewer request and find the most likely item from the database. At the level of research described in this thesis, though, the important issue is demonstrating the basic principles and functions of this system, and it does not bear on the immediate problem whether the database is created by a quick-and-dirty parser or an elegant expert system, or whether it comes in pre-fabricated via the second caption channel.

A Simple, One-VCR System

A simple setup for a news-viewing system, and the one initially used in this project for testing basic ideas, is shown in *Figure 11*. In practice, the computer records the captioned news program (or whatever programming is defined in a schedule file) every evening onto a "permanently time coded" tape, simultaneously placing the output of the closed-caption reader into a file, which is post-processed to create the database describing this program. Because of the desire for fairly clean stills, and also the potential desire to re- record the output of this system, the tape was run at standard-play speed. A T-120 videocassette at standard-play permits the system to hold on line the preceding three nights' news as well as the current program — each

night the tape is re-recorded over .the four-day-old broadcast.

For viewing, the video from the VCR connects directly to a monitor, with the viewer interaction taking place via a computer terminal. A user-interface program handles terminal input and output, accessing the database through a set of functions which return pointers to desired information such as the location of the beginning of the next story. The interface program also sends commands to a tape-playing routine, which in turn calls a search function which reads output from the time code reader. Note that the closed-caption reader is used only when recording and not while viewing — this fact suggests that with another VCR and tuner programming could be recorded even while the viewer was using the system for viewing.
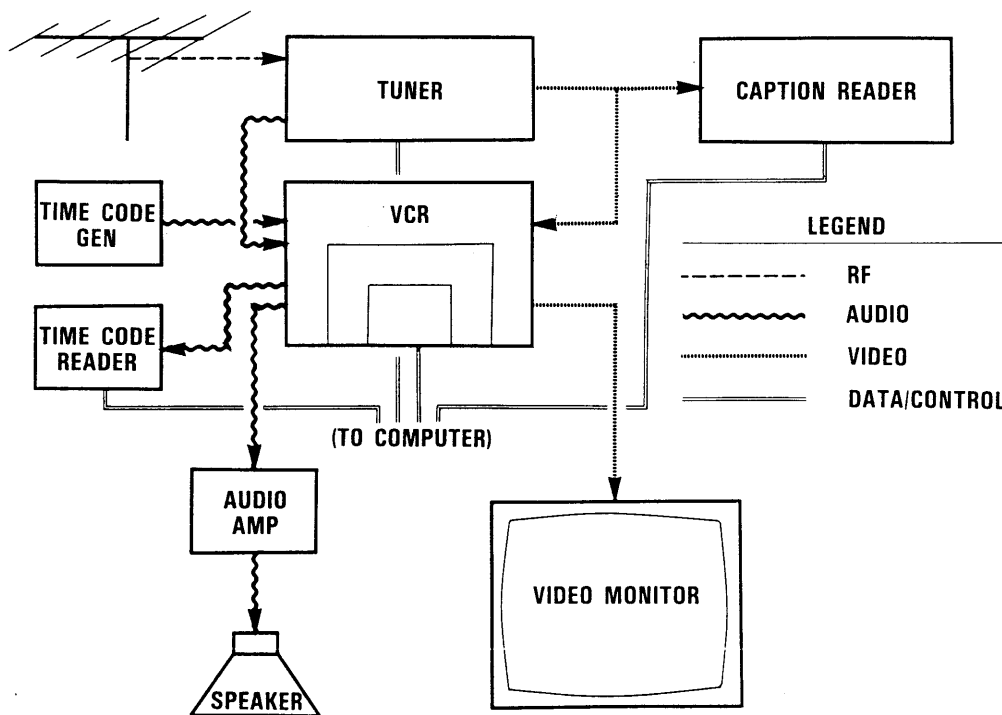
FIGURE 11

A simple, one-VCR setup for computer-assisted recording and viewing of videotapes.

This version of the system presents the news sequentially, printing out the headline of the upcoming story on the terminal and then, if the viewer doesn't press a key requesting a skip to the next story, playing this portion of the tape on the monitor. After showing a story, the tape-playing routine indexes the tape to the beginning of the next item as indicated in the database; generally the search routine which it calls will return instantly, as the tape will already be in the right place, but portions of the tape without database entries (like commercials) will be automatically skipped. The tape may also be interrupted

FIGURE 12
A demonstration of the need for a second VCR in the system.

while playing and skipped to the next item. Other options include listing and showing stories (both current and from past broadcasts) with corresponding keywords, majorstory, or category, or searching the database for items containing a typed-in keyword. The system remembers which items have been viewed, and no longer presents them as options in response to later requests. Also, if the tape is interrupted and a related story requested, the tape will return to the same point after showing the requested item.

The simple system demonstrated the basic principles, but it proved unsatisfactory for several reasons: it requires watching both the terminal and the monitor, it creates distracting black-screen pauses whenever the tape is searching, and, as *Figure 12* suggests, it cannot be used to view yesterday's news while recording today's. More significantly, a typewriter-type keyboard is too complex a control for use while watching television. Further, automatic editing of videotapes was considered a potentially desirable feature meriting exploration.

A Multi-VCR System

The result of the preceding observations is the substantially more complicated hardware setup shown in *Figure 13* (this figure is considerably simplified for clarity — pathways of sync signals are not shown, nor is the encoder which takes the R/G/B component outputs of the frame buffer and converts them to composite NTSC). In order for control to take place via the television screen, a computer frame buffer was added to the system. A second VCR takes its audio and video inputs from the outputs of the first; editing can take place from the first VCR to the second, or either can be played to the monitor. A Grass Valley computer-controlled video switcher is used to switch among these inputs and to select one to go to the monitor through its "program" output (as it can key, mix, and wipe, it actually has the capability to create a combination of the video inputs — locally-generated graphics superimposed over tape output, for example). A second independently-selected output ("preset") goes to the digitizing in-

put of the frame buffer, allowing it to freeze frames either for immedi-
ate display or for storage on the computer. An audio switcher would
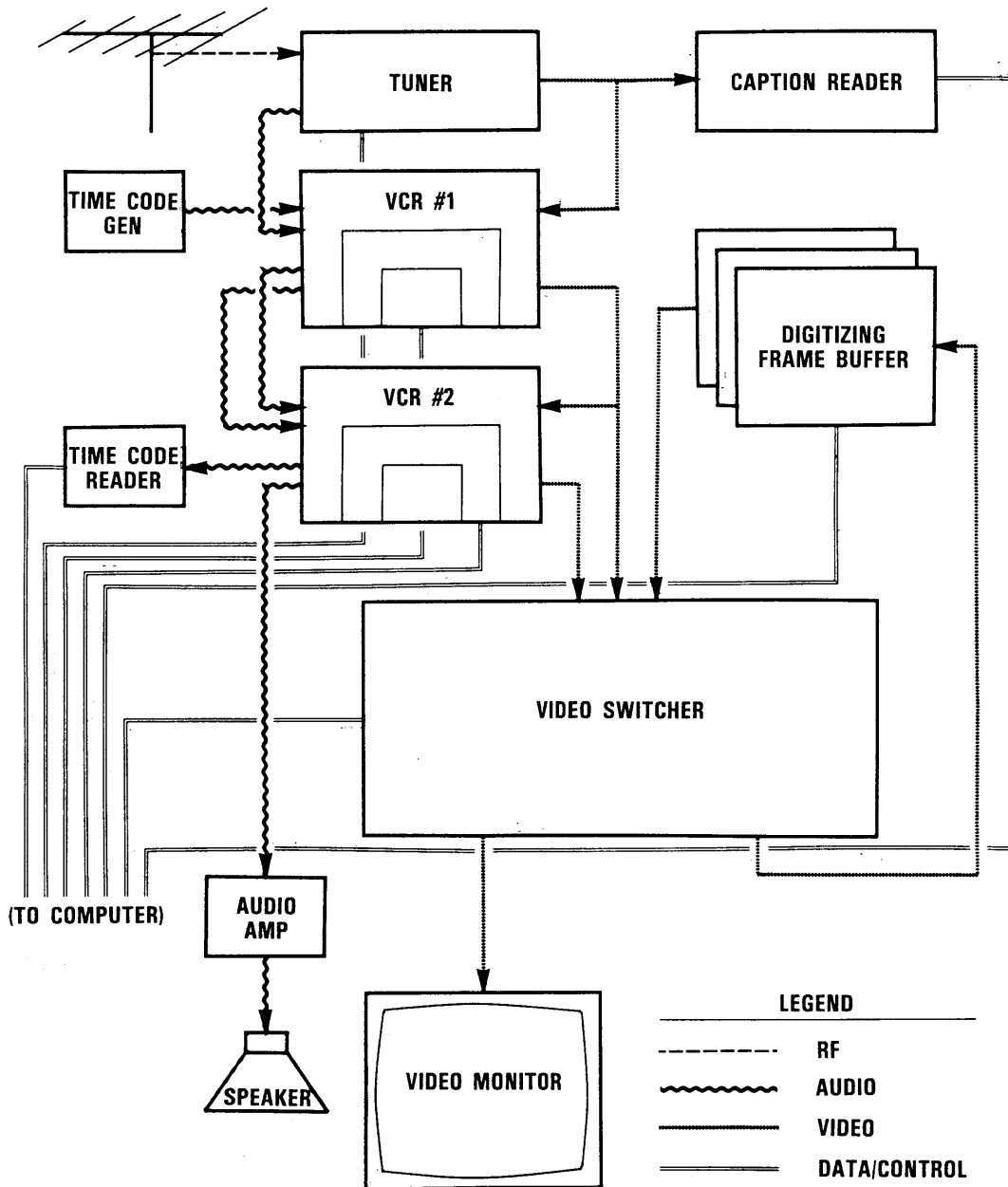seem to be a necessity here, but since a VCR passes its audio input

FIGURE 13

Adding a second VCR to the previous arrangement permits automatic tape editing, while
the use of a frame buffer permits computer-generated text and graphics to be mixed with
the tape output.

through except when in play mode, the audio can be "daisy-chained" through the two VCR's to the amplifier and the time code reader — this configuration generally works but does preclude the possibility of searching both tapes simultaneously or of combining the outputs of both tapes on the screen and taking the audio from VCR #1. A multiple-output audio crosspoint switcher and a second time code reader would take care of this shortcoming and simplify operation greatly.

In Moscow today the Soviet leader Mikhail Gorbachev rolled out the red carpet for a bipartisan congressional delegation. They spent nearly four hours together, and most of the Americans came away impressed.

■ Soviets propose missile freeze

■ Shultz and Gromyko to meet in Geneva

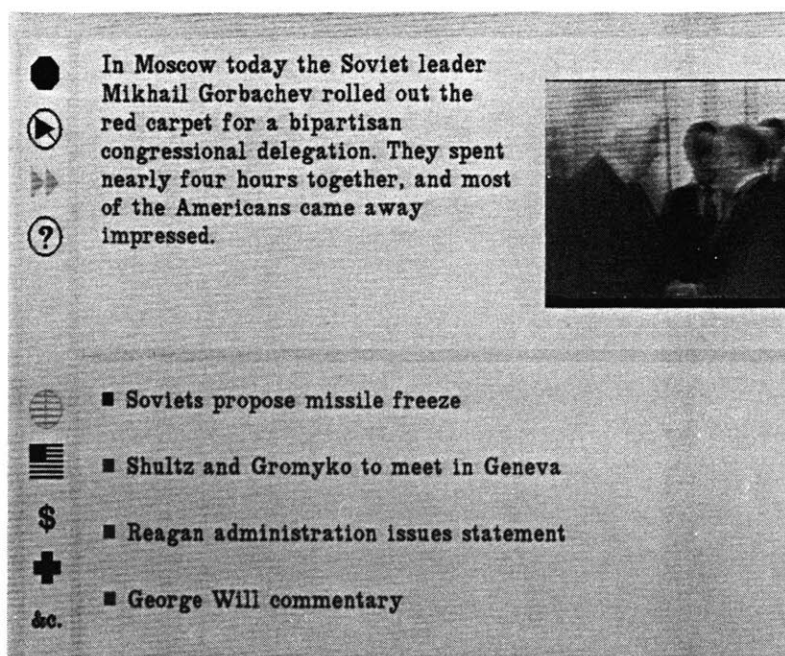■ Reagan administration issues statement

■ George Will commentary

FIGURE 14

A graphics display for interactive news viewing. The top half of the screen is the current story, and the bottom half contains headlines of upcoming items. "Buttons" in the upper left permit exiting the program, pausing the tape, skipping to the next item, or receiving a directory of current news items available. The lower left contains light-up symbols indicating the category of the current item.

With the increased flexibility of the new hardware setup, much better software could be created (after seeing the preceding version in action, an observer remarked that it was probably very good for exercising the low-level software but that one certainly wouldn't want to have to watch every night's news that way!). The first task undertaken was to graft a real user interface onto the already-described functionality.

Before each story is shown, the image from the frame buffer is routed to the monitor and the viewer sees a screen display similar to *Figure 14*. A reduced still-frame representative of the upcoming story, digitized by the frame buffer and stored on the computer, appears in the upper right corner of the screen.

Deciding which frame of the tape to digitize for each story is a problem, and in the absence of a better solution (like markings in the caption channel) the system just takes a still from half-way through the story. Accumulating these still frames requires the computer for the first time to rewind and play through a tape of a news broadcast; previously all post-processing was done strictly on the database. Since the computer has only to index the tape to the middle of each story, frame-grab and shrink, this added step takes only a few minutes and represents neither a major time delay nor a difficult processing task.

Alongside the picture are the first few sentences of the story displayed using "dejaggied" two-bit-per-point proportionally-spaced fonts; these were designed based upon signal characteristics of an NTSC display and allow a substantial amount of text to be shown without edge ringing or luminance-chrominance crosstalk *[Schmandt 1983]*.

Graphical symbols (a globe for "international," a US flag for "national," a dollar sign for "financial," and so forth) light up to indicate the category of the current story, while headlines of upcoming stories are displayed in the lower portion of the screen. "Buttons" in the upper left are activated by pushing the buttons on the SUN

mouse, which is used only as a remote-control keypad and not for its position-sensing abilities. These controls permit the viewer to exit the program, pause the program at this point, skip to the next story, and receive a listing of all the stories not yet viewed. This screen display is intended to give the viewer sufficient information to decide whether or not to view this story; if the viewer pushes none of the buttons, after a delay of a few seconds the tape begins playing and the display cuts to it. While the story is in progress, hitting any button stops the tape and returns the same frame-buffer display and the same choices as before. When the complete headline listing is on-screen, the viewer may select any story from the listing and see it next, or may delete any from those to be viewed. It is also possible as before to view all of the stories in a given category or those related to the current one. The graphics display is on the television screen at all times the tape is not actually playing; the eventual intent is to display additional information useful to the viewer (weather maps and so forth) should a request for a story involve a search time of more than a few seconds.

When the second VCR is used in an automatic editing application, the VCR with the tuner becomes the "input" VCR and the second the "presentation" VCR, or the one whose output the viewer actually sees on the monitor. As one demonstration of the possibility of automatic editing using the two-VCR system, the system was set up to condense each evening's news broadcast down to a headline service by assembling just the anchorman lead-ins from each story. Viewers

agreed that the resulting short tape conveyed most of the important information in the broadcast, while a tape of the opposite case (just the non-studio footage) was found to be of little use unless one had some background on the stories beforehand.

CONCLUSIONS AND FURTHER DIRECTIONS

The work described in this thesis has been only the beginning of a new project, and a large number of areas remain for exploration:

• *Multiple tuners* — Since there is only one closed-captioned newscast available locally and entering databases by hand for non-captioned programs is very time-consuming, only one tuner was used for input even though the second VCR also has one. Perhaps the system should watch a number of channels simultaneously, each having its own VCR. Being able to pick among several versions of a story would be an attractive option, and intelligently combining them together into a single more informative presentation even better. Given more than one input source, the system would need an internal model for each program in order to be able to create its tape-description database.

• *Constant recording* — Which programs to record shouldn't be hard-coded into the software, but rather the computer should look at the television schedule from the Infotext channel. Also, it might watch and record constantly, which would allow it to find unscheduled news bulletins; upon running out of tape it would transfer anything interesting it had run across to another tape (the one the viewer would actually see when using the system) and then rewind.

• *User profile and auto-editing* — If the system maintained a profile on the likes and dislikes of the viewer, it could automatically show the news in a personalized order, like financial news first. More efficient presentation without search delays in this case might be ac-

complished by re-recording the news from the "input" VCR to the "presentation" VCR in this rearranged order. This profile might also contain information like "show me international news from NBC and financial from ABC."

• *Interface* — The best way to interact with the television screen in this application has yet to be determined. Perhaps the remote control should be a small touch-sensitive device which maps to the screen, various regions becoming virtual buttons.

• *Database and archiving* — If a closed-caption encoder existed locally, the database (and even the user profile) could be recorded onto the beginning of a tape instead of being stored on the computer. Tapes could then be moved from VCR to VCR without causing confusion, and tapes of archived items could be stored on the bookshelf — the system would know what was on a tape as soon as it was loaded. Moreover, the viewer could take tapes recorded on one such system and view them on another.

• *Expansion to non-news programs* — In general, viewers probably won't want to carve up non-news programs in the same manner (documentaries and interviews might be indexed into the news system, though). Still, intelligent timeshifting — looking at on-line TV listings and automatically recording all Celtics games or flims starring Lauren Bacall — is quite within the capabilities of such a setup.

• *Presentation* — Many fascinating possibilites exist for presenting the news, for instance a computer-generated world map with headlines pointing to locations of major stories. Also, the computer

could create a set of local informational graphics (personalized financial data or local weather) for display when the tape is searching for an especially distant item. A question that often comes up in discussions of proposed high-definition television (HDTV) systems is, "But what are we going to do with all that additional bandwidth?" One idea is to place the 525-line NTSC broadcast images into a full-resolution window or windows on the bigger raster, which would be generated by a local frame buffer. Several channels could be shown at once without degradation (since they really wouldn't need to be "shrunk," just read out faster from a standards converter), or text, graphics, and controls could be placed in the additional space.

This investigation has shown the feasibility and desirability of varying television programming through sophisticated, content-sensitive control of storage and playback. The lessons learned indicate several fruitful directions in which further research should proceed, and the hardware/software system created provides a flexible environment in which this work may take place.

REFERENCES

[Backer 1981] Backer, D. S., "Personalized Visualization: Interactive Cinema for Training," *Proceedings of 1981 Computer Graphics '81 Conference,* Online Conferences, London, England, 1981.

[Barnouw 1982] Barnouw, E., *Tube of Plenty: The Evolution of American Television,* Oxford University Press, Oxford, England, 1982.

[Bright 1979] Bright, R. D., "Prestel — The World's First Public Viewdata Service," *IEEE Transactions on Consumer Electronics,* vol. CE-25, July 1979.

[Broadcasting 1984] "Sony Betamax Victory Likely to Ignite Hill Copyright Activity," *Broadcasting,* vol. 106, no. 4, January 28, 1984.

[Brockhurst 1982] Brockhurst, D. *et al.,* "Progress in the Development of UK Level Four Teletext," *Television: The Journal of the Royal Television Society,* vol. 19, no. 5, September/October 1982.

[Clarke 1980] Clarke, K. E., "The Application of Picture Coding Techniques to Viewdata," *IEEE Transactions on Consumer Electronics,* vol. CE-26, August 1980.

[Diamond 1982] Diamond, E., *Sign Off: The Last Days of Television,* The MIT Press, Cambridge MA, 1982.

[Dubner 1984] "Put Dubner's Scene Detector on the Case," (pamphlet), Dubner Computer Systems Inc., Fort Lee NJ, 1984.

[Hayes-Roth 1983] Hayes-Roth, F. *et al., Building Expert Systems,* Addison-Wesley Publishing Company, Reading MA, 1983.

[Hitachi 1984] *Hitachi Portable VTR Model VT-7A Instruction Manual,* Hitachi Sales Corporation of America, Lindhurst NJ, 1984.

[Lentz 1980] Lentz, J. et al., *Television Captioning for the Deaf: Signal and Display Specifications,* Report no. E-7709-C, Engineering and Technical Operations Department, Public Broadcasting Service, Washington DC, May, 1980.

[Lippman 1980] Lippman, A. B., "Movie-Maps: An Application of the Optical Videodisc to Computer Graphics," *Proceedings of the 1980 SIGGRAPH Conference,* Association for Computing Machinery, 1980.

[Lippman 1984] Lippman, A. B., "Imaging and Interactivity," *Proceedings of the 15th Joint Conference on Image Technology,* Tokyo, Japan, 1984.

[Matusow 1983] Matusow, B., *The Evening Stars: The Making of the Network News Anchor,* Houghton Mifflin Company, Boston, 1983.

[Mohl 1981] Mohl, R. F., *Cognitive Space in the Interactive Movie Map: An Investigation of Spatial Learning in Virtual Environments,* doctor's thesis, Massachusetts Institute of Technology, February, 1981.

[Monitor 1985] "Videoculture 5," *The Christian Science Monitor,* June 14, 1985.

[Pabouctsidis 1984] Pabouctsidis, C., "The Coding of Graphics Animation in a Videotex Terminal," *IEEE Transactions on Consumer Electronics,* vol. CE-30, August 1984.

[Schmandt 1983] Schmandt, C. M., "Fuzzy Fonts: Analog Models Improve Digital Text Quality," *Proceedings of 1983 National Computer Graphics Association Conference,* National Computer Graphics Association, 1983.

[Sears 1985] *1985 Spring/Summer Catalog,* Sears, Roebuck and Company, Chicago IL, 1985.

[US 1984] *Sony Corp of America v. Universal Studios, Inc.,* -- U.S. -- No. 81-1687 (1984).

[Whiteside 1985a] Whiteside, T., "Onward and Upward with the Arts: Cable—I," *The New Yorker,* May 20, 1985.

[Whiteside 1985b] Whiteside, T., "Onward and Upward with the Arts: Cable—III," *The New Yorker,* June 3, 1985.

ACKNOWLEDGEMENTS

The Media Laboratory has been an exciting and rewarding place in which to work, and I would like to thank everyone there for the help and support I received.

Special thanks go to:

Nicholas Negroponte, for the opportunity to work in advanced areas of media technology;

Andy Lippman, for letting me take time to finish writing this thesis when he would have preferred my moving on to the next part of this research;

Walter Bender, for always-thoughtful advice, always-available assistance, always-astute criticism, and always-welcome encouragement;

Judith Donath, for writing the database-handling software, thus providing me more time to expand the definition of "non-destructive hardware modification," as well as for fascinating conversation;

Kyra Lowther, for acknowledging me in *her* thesis (it's only fair); and

Ellen Robinson, for being Ellen Robinson.

This thesis was designed on the Tech Electronic Newsroom system and typeset in Compugraphic Oracle, a not-too-terrible facsimile of Hermann Zapf's Linotype Optima.