Simultaneous electronic recording of video and digital information on the video channel of a VTR or VCR

OWEN BARNES, MARSHALL M. HAITH, and RALPH J. ROBERTS, JR.
University of Denver, Denver, Colorado

A technique is described that permits precise synchronization of video recorded behavior with discrete stimuli and responses. These discrete events are recorded as digital data on the video channel of a video recorder on a horizontal line that lies above the visible video information. These data may be observed in single-frame playback mode by underscanning the playback monitor. A circuit is described for computerized decoding of the digital data. Because each video field is uniquely coded, an updating of the data is possible 60 times each second. The described technique has several advantages over current approaches for synchronizing discrete stimulus and response events with video recorded behavior.

A technique is described for recording data on the video channel of a video tape recorder. The recorded digital data need not be visible to a viewer, because they are recorded on video lines that appear above the displayed picture. This technique solves a variety of problems that occur with available techniques for synchronizing the recording of digital events with video information.

Video recorders are in wide use in psychological laboratories and often require the simultaneous recording of analogue video information and digital events. For example, an experiment may require the video recording of a person's facial gestures and time synchronization of those gestures with the presentation of a stimulus or series of stimuli. Such applications may require precise timing of these stimulus events during video playback, perhaps in a slow-motion mode. A similar need arises when the subject must make discrete responses (e.g., buttonpresses) while other behavior is being videotaped. For example, the experimenter may video record eye movements in a visual search task while the subject presses buttons to report target detections; the time relation between eye movements and buttonpresses must be precisely specifiable. Although no problem arises when there is real-time computer recording of events, a problem does arise if the data are originally recorded on videotape and then the temporal relations among events are reconstructed.

There are at least two ways to ensure playback synchronization of video recorded behavior and the timing of stimuli and responses. Most commonly, the digital information appears on the video display itself, either

directly, through mirrors, or through a video mixer and auxiliary television camera. For example, Markman (1984) video recorded couples as they interacted verbally. Each participant pressed buttons to register a six-level judgment about the quality of each verbal communication that was sent or received. The judgments were represented by six lights that were activated by the buttons and were recorded at the bottom of the video screen. The problem with this approach is that the display of digital events occupies part of the screen and limits the size of the video display that can be recorded. Since the recorded information is not available electronically, the digital information must be transcribed through tedious viewing and recording from the videotape.

In another approach, the digital information is recorded on the audio channel of the video tape recorder. For example, each of six digital events may be represented by one of six tones that is recorded on the audio channel; this procedure provides a degree of synchronization between the information on the audio and that on the video channels. It also solves the two problems associated with recording the information on the video screen, because it uses no space on the visual display. Also, the information can be made available, on playback, through a multi-channel tone detector for computer analysis. However, this approach raises some additional problems. First, it monopolizes an auditory channel (perhaps the only auditory channel) on the recorder. Second, the information on the audio channel can only be decoded in regular-play mode; decoding in a stop-action mode is not possible, and a solution to decoding in a variety of slow-motion speeds is quite complicated. Slow-motion playback might be required for accurate decoding of the conventionally recorded video information. Third, it is difficult to accomplish precise synchronization between the audio information and the video information, because the decoding circuits require several cycles of the audio signal to discriminate one frequency from all other frequencies.

The technique described was developed while Marshall Haith received salary support from NIMH Research Scientist Award MH00367. Ralph Roberts was supported by NIMH Postdoctoral Award MH09392. The development of the technique was supported by two research grants to Marshall Haith: NIMH Grant MH23412 and NICHD Grant HD20026. Please send correspondence to Owen Barnes or Marshall Haith, Department of Psychology, Denver University, Denver, CO 80208.

The number of cycles required depends on the particular frequency that represents the event in question. For example, imagine a situation in which the experimenter video records a person's eye as the person tries to fixate peripherally presented stimuli as rapidly as possible. Recorded on the audio channel are six tones that represent the onset of a sequence of five visual stimuli, in different locations, followed by a response by the subject (to indicate which of the five visual stimuli was the target). If it is important to measure the eye-movement latency to each of the stimulus presentations, as well as the response latency, the audio recording approach may produce serious time distortions because of the time constraints on the tone-recognition circuitry.

These problems can be resolved by recording the digital information on the video channel. The procedure to be described records the digital information on a horizontal line that appears immediately after the vertical synchronization pulse. This line appears above the visible information (with a conventional monitor). The audio channel remains free for other uses, and there is no time lag between the video and the digital information because all events are recorded simultaneously on the same video frame. The digital information can be decoded electronically. Finally, stop-image and slow-motion video play-

back do not affect the integrity of the digital information since the video-scan rate remains constant for all playback modes. (In these modes, the digital information can be read visually on monitors that permit underscanning of the video image.) One problem with this approach is that distortion of the digital information can occur when there are imperfections in the videotape in the region of the recorded data or the vertical sync pulse. However, this problem is rare and can be circumvented by dedicating some of the recorded data bits to an error-checking function.

This technique was designed to study how people learn to play video games. The situation required simultaneous video recording of the game screen and recording of buttonpresses, which the player utilized to fire shots, turn (left or right), or thrust a space ship. Later, we needed to synchronize, precisely, events on the game screen with subject responses. Since the decoding technique required stop-image analysis every $\frac{1}{10}$ sec, we needed a recording technique that was insensitive to tape-speed playback.

Functionally, the digital recording circuitry scans the state of eight digital switches for each video field (60 times per sec). The state of each switch is represented at specific locations on the eighth horizontal scan line by a white (0) or black (1) video level (see Figure 1). On playback, the

Figure 1. Illustration of the video encoding technique. The state of each of eight switches is encoded on the eighth horizontal line following the vertical sync pulse. Each bit is encoded as a white or black level (0 or 1) 2–4 μ sec after the beginning of each 6- μ sec window.

decoding timer (initiated at the onset of the horizontal sync pulse for the eighth line), samples the intensity level (black or white) at the location of each of the 8 encoded bits. The state of each bit is held in an 8-bit buffer for the remainder of the video field and is available for transfer to a parallel computer port for permanent storage. Of course, this scheme can be generalized for use of several television lines if recording of additional digital information is necessary. The description that follows and the circuit schematics shown in Figures 2 and 3 represent the case for single-line recording only.

Line 8 Encoder

Figure 2 is a schematic of the encoding circuitry.¹ The first task is to separate the vertical and horizontal sync pulses generated by the recording camera. Circuitry associated with integrated circuits (ICs) U1 through U4 constitutes the sync separator.²

The next step is to identify the vertical sync pulse, which signals the beginning of a new video field and then the

eighth horizontal line following the vertical sync pulse. The divide-by-eight counter, U5, locates horizontal line 8 as follows. First, the reset signal on U5 is released at the trailing edge of the vertical sync when U7B-4 sets flip-flop U6A, removing the NOT-Q output on pin 6. U5 then counts horizontal sync pulses and, after the eighth count, generates an inverted output at U5-11 which clears U6A. A positive pulse appears at U6A-6 which sets U6B, whose Q output on pin 9 generates "Line 8 Gate."

Now the digital representation of the state of eight switches is formatted in black and white on line 8. The state of the switches affects the outputs of U12, U13, and U14. The 8 bits are represented across line 8 at 6- μ sec time intervals for each bit. A 0 bit is coded as a 2- μ sec black level, and a 1 bit is coded as a 4- μ sec black level.

When "Line 8 Gate" goes high, it removes the reset level from U17A and B. U17A is also triggered on pin 10. R18 and C10 provide a short delay in removal of the reset level from U17B-3, which ensures that U17A always fires first. When U17A times out, U17B is triggered. U17B-13

Figure 2. Schematic of the video encoding circuitry.

pulses gate U9B-4 and triggers one-shots U16A and U16B. The data timing role of U16A and U16B will be described later.

The first clock pulse from U17B-13 does not reach divide-by-eight counter U10, because inhibit one-shot U18A was fired by "Line 8 Gate," and its NOT-Q output disables gate U9B-5. As a result, counter U10 contains 0 counts and keeps the 0 output of 3-line to 8-line decoder U11 (at pin 15) low. This output, inverted by U9D, is applied to 3-input gate U12A-2. The third input to gate U12A is the external switch data on pin 13. Assume that the switch input to U12A-13 is a 0 (switch input grounded). U12A-12 will remain high, and 8-input NOR gate U15-8 will remain low, holding a reset level on U16B-3. Now, consider the timing roles of U16A and B.

With a reset level on U16B-3, U16B will not be fired by the clock pulse from U17B-13. However, the reset level at U16A-11 was removed by "Line 8 Gate," so U16A (a 2- μ sec one-shot) does fire, and its output at pin 12 is routed through NOR gate U14C, gate U9A, and inverter U7E to the video out terminal. In this case, a short (2- μ sec) black-level pulse occurs, representing a 0, and is gated to the video out. If the data level at U12A-13 had been a 1 (high level), U12A-12 would have been low,

U15-8 would have been high, and both U16A and U16B (a 4- μ sec one-shot) would have fired. Since both U16A and U16B feed NOR gate U14C, a long (4- μ sec) black-level pulse, representing a 1, would have been gated onto the video. The second clock pulse from U6B-9 increments divide-by-eight counter U10; U11-14 goes low, enabling 3-input gate U12B. The switch data at U12B-5 determines whether a 0 or 1 is gated to the video. This sequence continues for the remaining seven switch inputs. The leading edge of the eighth clock pulse to reach U10 clears U6B, terminating "Line 8 Gate," which now inhibits any further video data during the current field. Line 8 will now contain 8 data bits. Each bit is a black level of either 2- μ sec or 4- μ sec duration, followed in each 6- μ sec bit time interval by a 4- μ sec or 2- μ sec white level, respectively. The timing values for clock oscillator U17A and U17B were chosen so that the data string occupies most of line 8 (about 48 μ sec of the 50 μ sec available).

This scheme could be expanded to include more data bits per line, or to utilize more lines.

Line 8 Decoder

For decoding, the first task is to separate the vertical and horizontal sync pulses generated by the videotape playback of the recorded information (see Figure 3). Cir-

Figure 3. Schematic of video decoding circuitry.

cuitry associated with ICs U1 through U6 duplicate the sync separator portion of the encoder circuit, except that the ninth horizontal sync pulse after the vertical sync clears U6B.

The next task is to sample line 8 at eight points, 3 μ sec after the beginning of each 6- μ sec bit window, for a black or white video level. When U6B generates "Line 8 Gate," Q3 is switched off, enabling comparator U8. The threshold of U8 is adjusted by R24 to pass only the eight data time windows during line 8. Each data level is inverted by U7F and applied to the serial input of shift register U10, and also triggers 3- μ sec one-shot U9A. When U9A times out, the data present at U10-1 will be clocked into the register. If the data level is a 0 (2 μ sec), its level will return to low before U9A times out, so a 0 is shifted into U10. If the data present at U10-1 is a 1 (4 μ sec) when U9A times out, a 1 is shifted into U10. After the 8 data bits during line 8 have been strobed, the data are available in parallel form at U10 for a parallel input buffer of a computer. These data remain available for about 16 msec, until the next vertical sync pulse appears.

Other Applications

Although the described scheme will have greatest application to the temporal coordination of videotaped behavior and discrete events, there are other situations for which it may be applicable.

Research on infant visual expectancy (Haith, Hazan, & Goodman, in press) requires the presentation to infants of a sequence of video-graphic displays. A video record of eye movements is analyzed in slow motion for calculation of reaction times. For a precise calculation of response latency, there must be a recorded indication of stimulus onset. When a computer generates a display, an indication is easily generated by switching on a digit of a time-date display (when the stimuli appear) that is superimposed on the eye recording.

There is a problem for researchers who want to use our visual displays but do not want to use a computer; this is often the case, for example, for infant research done in the home. The displays are easily recorded on videotape, which can then be shown to the infant as the infant's eye is recorded. But how can the temporal occurrence of each visual event be recorded on the video recording of the infant's eye?

The technique described here can be used for this purpose by encoding digital information on the videotape presented to the baby. The baby would not see the information, which would lie above the visible portion of the video display. However, decoding circuitry can be attached to the video output which switches on with stimulus onset and controls, for instance, digits on a time-date display that appears on the video recording of the eye.

REFERENCES

- ALBING, B. (1982, January). Sync separator provides speed, accuracy. *Electronic Design News*, 27, 207-208.
- HAITH, M. M., HAZAN, C., & GOODMAN, G. (in press). Infants' expectation and anticipation of future events, *Child Development*.
- MARKMAN, H. J. (1984). The longitudinal study of couples' interactions: Implications for understanding and predicting the development of marital distress. In K. Hahlweg & N. S. Jacobson (Eds.), *Marital interaction: Analysis and modification* (pp. 253-281). New York: Guilford Press.

NOTES

1. The prototype model of this encoder was built with parts in stock, which partially accounts for the mixture of IC families and for the total IC count being higher than necessary.
2. This circuit was described in Albining (1982) and has been found to be very reliable; the only drawback is a requirement for three supply voltages. The original circuit was modified for our application by the addition of U4B, because we wanted to eliminate equalization pulses from the horizontal sync, if present.

(Manuscript received February 9, 1987;
revision accepted for publication October 9, 1987.)