Reprinted from AMERICAN PSYCHOLOGIST, Vol. 24, No. 3, March 1969
Printed in U. S. A.

A TOOL FOR THE ANALYSIS OF MOTION PICTURE FILM OR VIDEO TAPE[1]

PAUL EKMAN AND WALLACE V. FRIESEN[2]

Langley Porter Neuropsychiatric Institute
University of California Medical Center, San Francisco

Permanent visual records are essential for research on a number of problems in the behavioral sciences, and often data can only be obtained by an observer's inspection of those records. All too often, however, investigators have been seduced by the allure of capturing their phenomena on film or tape and have neglected the question of how to convert the record into data until they have been confronted with an overwhelming mass of motion picture film or video tapes. Such records are, one soon discovers, nearly as complex as the original phenomena recorded. If the investigator could not decide what units of measurement to employ or what to measure when he saw the phenomenon "live," he may well find it just as baffling or complex when viewed on film. Moreover, he may have failed to consider that it takes at least as much time to view a permanent record once as the phenomenon originally required, or to realize that records are not data but must be converted into some digital form in order to be analyzed.

It is no surprise, then, that rather large collections of motion picture films or video tapes of human interactions, stored in laboratories around the country, not only have never been analyzed, but may not even once have been viewed in their entirety by the investigator. While permanent records have the virtue of slow motion and repeated viewing, imagine the time it would take to look at the record of a complete psychoanalysis of 300 hours (even at real time); in addition, include the minimum time required to make occasional notes, sparingly use slow motion, and get the film in and out of the projector!

Commonly, when confronted by such a supply of records, the investigator looks for hardware that might help him wade through what may be an embarrassingly large mass of material. Unfortunately motion picture projectors, film editors, and video tape recorders fail to provide most of the necessary functional capabilities.

Visual Information Display and Retrieval system (VID-R) was designed to rescue the investigator inundated with permanent visual records. It also has applicability to archiving and visual displays in programmed instruction. Before describing VID-R, we will mention some of the circumstances in which permanent visual records are necessary and why their analysis frequently depends upon the observers' inspection of the record.

Permanent visual records are necessary when (*a*) repeated direct observation or the use of multiple observers is not possible, either because the phenomenon is nonrecurrent (e.g., in a primitive society —Sorenson, 1967), or the phenomenon, though recurrent, is inaccessible to observers (e.g., in psychoanalytic sessions); (*b*) accurate or reliable observation requires slowed motion (e.g., of micro facial displays—Ekman & Friesen, 1969; Haggard & Isaacs, 1966) or fast motion (e.g., traffic patterns or very slow, small movements); (*c*) the phenomenon is difficult to describe in words, but can easily be defined with a visual example that allows observers to spot its occurrence in a record; (*d*) the research is exploratory, and the units of measurement or number of events to be measured cannot be specified without inspection of the record. Automated analysis of permanent visual records with an optical scanner is seldom feasible for records taken by behavioral scientists; either the scanner cannot discriminate the event to be measured from the visually complex environment, or the event, while easily recognizable by a human observer, is sufficiently complex in its pattern of variations to be prohibitively expensive to program for recognition by a scanner.

Consequently, if the investigator must look at his record in order to analyze it, he will need the

[1] Research reported in this paper was supported by a research grant from NIMH, MH 11976, and Research Scientist Development Award, 5-K02-MH06092.

[2] Requests for reprints should be sent to Wallace V. Friesen, Studies in Nonverbal Behavior, 1405 Fourth Avenue, San Francisco, California 94122.

following options: to view it at real, slowed, or fast time; automatically to retrieve events he has seen for comparison with new events: to assemble similar events or difficult-to-code events without destroying the original record; to store his measurements or observations in a manner that allows automatic retrieval of the visual phenomena they refer to; and, perhaps, to build a visual dictionary to which he has fast automated access for display of definitions. VID-R performs all of these functions by interfacing two video tape recorders, a video disc recorder, a film-to-video chain, and monitors with the logic of a computer to allow the reading and writing of digital information on the video tape.

Functions

While video tape is the medium used for analysis, the original recording medium may be motion picture film. In such cases a special film-to-video transfer or chain is utilized that has the capacity to enlarge any part of the film frame (6:1 zoom) and to copy the film at real, slowed, or stop frame motion.

The operator utilizes a teletype keyboard to instruct VID-R to perform any of the functions; these include all of the standard video tape recorder functions, additional nonstandard capabilities, and a number of complex operations that combine a sequence of functions.

Viewing speed. The operator can choose to see the video tape on the monitor at real time, in slow motion (with 15 alternate slow speeds), or in fast motion (up to 10 times real time). Not all purposes can be served, however, by this procedure because some loss of definition occurs. In both slow and fast motion, noise appears in the picture in the form of white lines and "wipes," and in slow motion there is also a loss of detail since half of the horizontal lines in the picture are lost. VID-R includes an alternate procedure for slow motion when such loss of detail or noise is not tolerable. Material to be viewed in slow motion may be automatically transferred in up to 30-second units from the video tape to the video disc recorder. There is no loss of resolution and little noise in slow motion when displayed from the disc recorder.

Search and retrieval. Fast search and automated retrieval is an essential function, not available in previously existing equipment, for even the most casual inspection of permanent records. Frequently, a record contains surplus information, at least for the purposes of a particular study or inspection. When the observer has viewed the entire record, he needs to reinspect selected portions. Often, the observer will note the occurrence of an event that he wants to compare with another earlier event on the same record; again, rapid access is needed to both events. VID-R searches and retrieves an event when requested either by its time address (location), or by a previously stored observation or measurement. Search and retrieval thus depends upon the capacity of the computer to write and read digital information on the video tape. Before describing search and retrieval procedures we will describe how time addresses and descriptive information is written on the tape.

There are 324,000 video fields on a 90-minute video tape, each being 1/60 of a second in duration. VID-R labels each field by writing a discriminable binary address on the horizontal lines outside the normal viewing area. By writing 12 bits on each line, a total of 20 lines are needed to address each field. This allows space for identifying the tape, writing the address twice, and entering check sums to detect error in reading. If any frames are removed from the original record and placed on a new tape, one address is replaced by its address on the new tape, while the redundant address is returned to identify the original source and location.

Descriptive information is written in the same manner except that all horizontal lines of a field are utilized. A data matrix of 6×120 cells is generated where each cell can be coded from zero to nine; the storage capacity of one data matrix is thus equivalent to nine IBM cards. Although the matrix takes only 1/60 of a second to write, it requires four seconds to tape, for technical reasons. A data matrix is written on 30 consecutive fields to reduce error in reading, and blank space is left before and after the matrix to allow the matrix to be erased and updated. Since such a block erases four seconds of video picture, a new tape must be produced in which the original record is copied with inserted data blocks so as not to lose any of the visual information on the original record. VID-R accomplishes automatically the production of this new second-order tape containing the original picture record with the inserted data blocks. A single data matrix contains considerable space for storing measurements or observations and it is likely that most visual events can be

described before the storage space is exhausted. If more descriptive space is required, additional data blocks can be written prior to copying the visual unit.

Search and retrieval requires that the operator specify information about the event he wants to see, that is, either a time address or descriptive information. Time addresses are obtained during earlier stages of analysis. The operator can, at any time when viewing a tape, request the address of the material he is seeing; this will provide him with at least a rough location since his reaction time will introduce some delay. If a more exact time address or measure of duration is needed, subroutines are available that back up the tape and move it forward in slow motion. In asking for retrieval the operator can request each event separately by typing its time address, or request a series of events by entering their time addresses into case storage. Once instructed, VID-R utilizes the fast forward and reverse capacity of the video tape recorder to search at up to 15 times real time; it locates the exact frame requested, and displays the event on a monitor at the viewing speed requested.

If the operator has previously entered descriptive information into data matrices, he does not need to know the time addresses of the events he wants retrieved but can instruct VID-R in terms of single or multiple descriptive items. Thus, he can ask for "all hand movements," or for "all hand movements of the right hand to the ear when the leg is also moving" (if such descriptions have been made previously), etc. Again, VID-R utilizes fast search to find any event(s) that have the specified descriptions, and then displays them sequentially at the viewing speed requested.

Temporal reorganization (editing). Selecting certain events from a record and collecting them onto a new record for further inspection and analysis has not previously been possible with motion picture film or video tape machinery without either destroying the original record or doing extraordinarily tedious editing of expensive duplicate copies. Yet, the investigator must frequently reorganize a record many times, each reorganization dictated by the needs of a particular study: thus, the original record must be preserved. Various purposes may necessitate temporal reorganizations. Reorganization of a record may be performed to develop a coding scheme for a particular event; for example, all posture shifts from a series of original records may be collected onto a new tape so that the investigator can readily scan the variety that occurred and develop his code for classifying them. Reorganization of records can be crucial in exploring what occurs under specified conditions; for example, the investigator may want to look at the patient's behavior whenever the therapist breaks eye contact and turns his face away.

VID-R automatically performs reorganizations of a single or a series of records, producing new "dubbed" copies of visual events from the original tapes or earlier "dubs." The operator instructs VID-R to perform such reorganizations, on the basis of time addresses or descriptive information contained in the data matrices, providing that information for each item to be reorganized. Once instructed, VID-R uses its fast search and retrieval capacity and an editing subroutine to copy the events into the new ordering. In producing this new tape VID-R can also copy any descriptive information pertaining to the events being dubbed, or leave the space necessary for the later addition of descriptive information.

Visual dictionary. Often the coding of body movements or facial expressions involves the development of classes of behavior or units that, while easy to see, are difficult to describe adequately in words or to remember solely through a verbal description. A visual dictionary can contain examples of each classification item so that an observer can refresh his memory or make a paired visual comparison between the dictionary definition and the event to be classified. The construction of such dictionaries with previously existing equipment was not feasible because of the cumbersome task of updating definitions, and the lack of any means for quickly retrieving any definition.

VID-R builds a visual dictionary by utilizing the procedures described for temporal reorganization. A descriptive data block containing a descriptive matrix is inserted before each visual definition, or before a number of definitions. These data matrices contain descriptions that characterize different aspects of the definition, and that are used by the operator to determine whether such a definition is contained in the dictionary and, if so, to retrieve it. For example, if the operator is attempting to code leg movements from an original record, when he sees a particular leg movement he would first describe it on a series of standard log descriptors, entering these into the computer through the teletype, for example, "unilateral, foot only, repetitive, lateral swing." The dictionary tape would

then be searched automatically and any items that had that same set of descriptors would be retrieved and displayed on a monitor for the observer. If the dictionary did not contain any such definitions, VID-R would produce a copy of all such undefined items for the investigator to consider for entry into the dictionary. Updating or making new entries into the dictionary are also accomplished through fast search and retrieval procedures.

Components

VID-R was designed to incorporate as much off-the-shelf hardware as possible, and to contain maximum flexibility in logic. Both of these objectives led to choosing a small computer to provide the logic for the system rather than constructing specialized logic hardware.

Off-the-shelf hardware includes: two Sony PV 120U video tape recorders (the essential requirement being that all functions of the recorders be electronically controlled); one MVR video disc recorder; three high-resolution television monitors; one teletype, ASR 33, keyboard printer with papertape punch and reader; one Digital Equipment Corporation PDP-8 computer that provides the logic. Specially built equipment includes: a video and audio interface to perform data transfers between the computer and recorders; video tape recorder controllers capable of performing the instructions from the computer; one film-to-television chain that allows the transfer of 16-millimeter optical sound movies to video at various speeds and with enlargement capability. There is also a fairly extensive programming package. More technical detail on the choice of components, interface, controllers, and programming is given in ~~Haggard and Isaacs (1966)~~ Ekman, et al, 1969.

Applications

There are three main applications of VID-R, outlined only briefly here since they are described fully elsewhere (Ekman, Friesen, & Taussig, 1969) —analysis of behavior, archives, and programmed instruction. VID-R can facilitate the analysis of behavior from permanent records regardless of how fine the units of analysis may be, or how casual the use of the permanent record. If the investigator must look at his record more than once, he will want to return to specific portions; the capacity to store his observations on-line and later to search, retrieve, and either view them or produce a reorganized selected copy will be of great use.

Film will probably remain the first choice for permanent archives, because little is known about the long-term life of video tape, and because film provides both better resolution, if taken under optimal lighting conditions, and superior color to television. However, indexed video tape copies of film archives could provide a feasible means for different potential users of an archive to preview and determine which portions of the archive might be relevant to their interest.

In programmed instruction, VID-R offers the possibility to store both single-frame visual material and sequential, moving visual events, which, when indexed, are retrievable at whatever point they might be necessary for presentation to the student. The maximum search time of a 90-minute tape might reach six minutes, but this could be reduced if the material were stored in other than a random fashion.

These three examples of the potential utility of the VID-R system suggest only a few of the possible usages. Specialized hardware has been used only to gain control over the video recorders, while the specific capabilities are determined by the computer programs. Furthermore, the basic function controls of the computer program have been written in the most general terms so that variations in the order or manner in which several functions are performed require no changes in programming but merely the substitution of subroutines. Any task requiring the capacity to locate, retrieve, and reorganize preselected portions of visual material, or which could utilize the capacity to write descriptions, codes, or instructions with the visual material on the video tape, could be simplified or even made possible for the first time through the use of VID-R.

REFERENCES

EKMAN, P., & FRIESEN, W. V. Nonverbal leakage and clues to deception. *Psychiatry*, 1969, in press.

EKMAN, P., FRIESEN, W. V., & TAUSSIG, T. VID-R and SCAN: Tools and methods in the analysis of facial expression and body movement. In G. Gerbner, O. Holsti, K. Krippendorff, W. Paisley, & P. Stone (Eds.), *Content analysis*. New York: Wiley, 1969.

HAGGARD, E. A., & ISAACS, K. S. Micro-momentary facial expressions as indicators of ego mechanisms in psychotherapy. In L. A. Gottschalk & A. H. Auerbach (Eds.), *Methods of research in psychotherapy*. New York: Appleton-Century-Crofts, 1966.

SORENSEN, E. R. A research film program in the study of changing man. *Current Anthropology*, 1967, 8, 443–469.