

SCORIT: A computer subroutine for scoring electrodermal responses*

WILLIAM F. PROKASY

University of Utah, Salt Lake City, Utah 84112

Given digitized electrodermal response records sampled at a rate of 20/sec and at a sensitivity level between 140 and 170 ohms, a computer subroutine for scoring skin resistance responses is described. It is designed to score multiple responses on a trial, which makes it particularly suitable for classical conditioning experiments involving long interstimulus intervals. The subroutine returns to the main program response latency, base resistance, peak resistance, and time from base to peak. Since SCORIT is written in FORTRAN IV, it can be employed in virtually any modern central computing facility at costs substantially below those involved in hand scoring.

Typically, skin resistance response (SRR) protocols are scored by hand from a hard-copy recording. If more than simple response frequency is measured, hand scoring is expensive and time consuming. In addition, the criteria employed for identifying responses are usually limited to specifying a minimum resistance change occurring within a defined interval following signal onset. Omitted, and largely subjective within each laboratory, are criteria for the change in resistance per unit time required to define a response, for changes in rate of change required to define a second response superimposed on one in progress, and for minimal response durations. The increased availability of equipment that permits A/D conversion and digitalization of skin resistance levels (SRLs) makes it feasible to consider using a computer program rather than a human reader to score SRR protocols. SCORIT is a FORTRAN IV subroutine designed for that purpose. It was written with the specific objective of eliminating hand scoring but has, in addition, the advantage of specifying in programming language a kind of response model. In this sense, SCORIT constitutes a step toward standardization of scoring in an area in which intuitive judgments specific to each laboratory have predominated.

INITIAL ASSUMPTIONS

SCORIT has been written as a kind of "black box" subroutine. That is, it can be appended to any program that makes SRR protocols available and will score the protocols given a minimal amount of information. There are two important constraints on the initial digitizing of data which the scoring generality imposes: The sampling rate must be 20/sec, and each unit of change in the

digitized sample should be equated to a resistance change in the range of 140 to 170 ohms.

The sampling rate was chosen as one that provides sufficient precision together with low cost. Latency measurements in the SRR conditioning rarely exceed .1 sec in accuracy, which means that the .05-sec value assumed by SCORIT provides more precision than typically is available. The sampling rate of 20 Hz was also chosen to be divisible into line frequency. This reduces the likelihood of 60-cycle interference during the on-line recording. A higher sampling rate could have been chosen, but each increase in sampling rate results in a corresponding proportional increase in scoring costs.

SCORIT was developed based on dependent variable units of 156.25 ohms. This value happened to be convenient, although variations from 140 to 170 ohms should constitute no problem. The value is such that at a tonic level of 38,000 ohms a change of unit (i.e., 156.25 ohms) is approximately .1 μ ohms conductance, a satisfactory minimum for most experimentation.

The main program designed by the user that calls the SCORIT subroutine must provide for all I/O. SCORIT assumes that a single protocol (e.g., a single trial of 200 sample points) is in core in an array common to both the user's program and SCORIT. The array is also assumed to be coded in positive integer format. The main program should be designed to assign absolute resistance values and perform any desired conversions. Given positive integer data, the only argument necessary for the execution of SCORIT is the total number of sample points in the array, and this can be as large as desired. SCORIT returns the following information: response latency (measured from the first point in the protocol), time from base to peak, base and peak values, and the scoring criterion employed for each identified response.

CURVE SMOOTHING

In the course of A/D conversion or of original analog recording, occasional noise effects are observed. At a sampling rate of 20 Hz, for example, single points may deviate from an otherwise smooth function. Incorporated in SCORIT are two sets of curve-smoothing rules that eliminate some noise. The first set simply examines all points and redefines as part of a smooth function all one-point deviations, assigning to them the value of the point just preceding or following it, depending upon which is the closer. The second set repeats the process but adjusts instead all two-point deviations from an otherwise smooth function.

Since the need to adjust for two-point deviations may

*This research was supported by NIMH Grant MH-15353. I wish to thank William C. Williams, Craig Clark, and William Yu-Ming Lee for serving as judges of the response protocols.

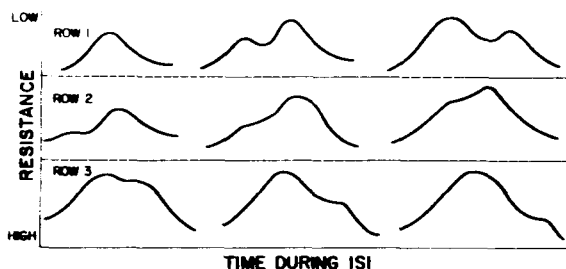


Fig. 1. Examples of three response categories. See text for fuller discussion.

not be great, the user may find it wiser to save about 10% of the scoring line by eliminating that section of SCORIT. Similarly, if one-point deviations are no problem, the relevant section of SCORIT can be eliminated with an additional savings of approximately 10% of the scoring time. However, SCORIT assumes each change in value to be reliable, in the sense that it is examined against scoring rules. Therefore, one- and two-point variations can, depending upon their precise location, identify responses when none exist.

RESPONSE FORM

In recent years many writers have been concerned with the fact that multiple responses occur during SRR conditioning when the interstimulus interval is at least 4 or 5 sec in duration (Grings, Lockhart, & Dameron, 1962; Dengerink & Taylor, 1971; Lockhart, 1966; Prokasy & Ebel, 1967; Prokasy & Kumpfer, in press; Stewart, Stern, Winokur, & Fredman, 1961). To illustrate the kinds of response identified by SCORIT, some prototypes are illustrated in Fig. 1. Row 1 in Fig. 1 includes three examples of responses that are easily identified. In each instance, a baseline, peak, and return toward baseline are clear. The latter two examples include two responses each, and the baseline for the second is defined as the resistance value at the point at which resistance begins to decrease rather than the point of initiation of the first of the two responses.

The three examples provided in each of Rows 2 and 3 are less common response forms but are particularly important in view of the existence of multiple responses. In Row 2 each example is of two responses, but there does not exist (as in Row 1) a return toward baseline prior to a clear increase in the rate of resistance drop. SCORIT identifies both responses in each of the three examples by seeking a point of inflection. The peak of the first is identified as the resistance at the point of inflection, and the latency of the second is also defined at that point. The amplitude of the second is assessed as the difference between resistance values at the point of inflection and the subsequent peak (i.e., minimum resistance value reached). Row 3 is similar to Row 2 in providing three examples of double responses, although in these cases the point of inflection is identified after a maximum deflection is obtained for the first response. The latency of the second response, located on the

return limb of the first response, is identified at the point of inflection, i.e., at the point in which there is a clear decrease in the rate of return toward baseline. Since the second response is identified strictly through a retardation in return to baseline, a measure of amplitude is not possible.

There are several marginal rules for identifying a response. A one-unit change (i.e., a 156.25-ohm resistance drop) is defined as a response only if it lasts at least .5 sec and no longer than 4 sec and if it is terminated with a return to baseline. A two-unit change in response is defined if the second unit change occurs within .5 sec of the first unit shift. A two-unit shift reached within .25 sec of response onset is also defined as a response, regardless of subsequent duration. Examination for changes of three units or more are made only if there is initially at least a two-unit shift within a .5-sec period.

Since some investigators may wish to employ fewer criteria for responses, SCORIT allows the user to pass arguments that can eliminate any one of, or combination of, the routes for response definition. This is done by specifying the type to be eliminated. Thus, if the investigator does not wish to incorporate one-unit changes as responses, this can be indicated by passing an argument to SCORIT. Similarly, classes of responses, such as those that would otherwise be defined on the return limb, can be eliminated. The broad category changes reduce scoring time, but elimination of individual response types does not save time.

TONIC CHANGES

Changes in tonic SRL are exhibited either by gradual increases or decreases in resistance. Increases in resistance pose little problem for SCORIT, since a response is defined by a drop in resistance. It is possible, of course, for the increase to be at such a rate as to obscure a response by altering its slope in a way that SCORIT cannot detect. We have not found this to be a problem.

A potentially troublesome sort of tonic change exists when S_s ' resistance decreases throughout the session. SCORIT is designed to ignore constant rate changes during a protocol, although it can be fooled by local rate fluctuations in the midst of an overall decrease in tonic SRL. In the event that at no point during a trial does resistance increase (i.e., there are only decreases), response definition is somewhat more stringent. For example, SCORIT classifies all resistance decreases requiring more than 4.5 sec from base to peak as tonic rather than phasic changes. This risks eliminating responses, but the likelihood of this sort of loss is no more than .002 of all responses scored.

Until SCORIT is employed over a wide range of laboratory conditions and in different laboratories, the extent to which tonic decreases in SRL will cause difficulties cannot be known. The estimate available from present data indicates that, with a conditioning

preparation in which there is a general, overall decrease in SRL throughout the session, the likelihood of a classified response being attributable to tonic shifts is less than .005.

VALIDITY

In the absence of a formal model of response form, the criterion of validity is the degree to which SCORIT is able to simulate the judgments of human observers. Three judges, all of whom were familiar with SRR conditioning, scored the records of 20 Ss who had received 49 CS-UCS trials. For each S there were 49 protocols of 11-sec duration (i.e., 220 data points beginning at CS onset) obtained with a CS-UCS interval of 6 sec and an average intertrial interval of 45 sec.

The records were scored independently by each judge. The protocols were displayed on a cathode-ray tube, and for each protocol the judge indicated the sequence and type of responses observed. The outcome of the scoring of the records of the first 10 Ss was compared with the outcome of a preliminary version of SCORIT. Based on the kinds of discrepancy obtained between the judges and SCORIT, the subroutine was modified. These initial data will not be reported here, but the initial discrepancy rates were less than 50% greater than those to be reported on the final 10 Ss.

A total of 939 responses was identified by SCORIT over the 490 (10 Ss, 49 trials each) 11-sec protocols, and this base figure will be used in identifying relative discrepancy rates. Table 1 is an interscorer discrepancy matrix: The entries are the number of times any one scorer differed in judgment from another scorer. In all but a few instances, these discrepancies concerned the presence or absence of a single response. Thus, the discrepancy rate, based on total number of responses, varied between .06 and .1, with the average for each scorer varying from .065 to .085. The average discrepancy rate for SCORIT is .07.

The discrepancy matrix is not an error matrix. The majority of the entries in Table 1 were of equivocal sorts, e.g., whether or not an observed rate change was sufficient to identify a response of the type illustrated in either Row 2 or Row 3 of Fig. 1.¹ One indication of error rate is the number of times each scorer differed from all of the others, given that all of the others agreed. For humans, such a number would reflect the number of times there was either an error in perceptual judgment or an individual scoring style. For SCORIT, this number would reflect idiosyncratic scoring rules. Table 2

Table 1
Number of Times a Discrepancy Existed
Between Each Pair of Scorers

	S*	1	2
1	69		
2	82	86	
3	68	56	67

*S stands for SCORIT, numbers for each human scorer.

Table 2
Errors of Omission and Commission for
Each Response Category

Row†	S*		1		2		3		Base Total**
	O	C	O	C	O	C	O	C	
1	2	2	5	3	2	5	1	1	811
2	4	9	8	3	3	7	1	1	104
3	5	2	2	0	0	13	2	4	19

*S refers to SCORIT, numbers to human scorers.

**Total number of responses identified by SCORIT in each category.

†Rows 1, 2, and 3 refer to response categories in Fig. 1.

presents these data, partitioned according to response type and to whether the error was one of commission or omission. Error rates varied from approximately 1% to 3%, with SCORIT at approximately 2.7%. The table shows that one of the judges made very few errors, that one tended to be conservative in identifying responses, and that the third tended to be liberal in identifying responses. SCORIT was approximately balanced across errors of omission and commission, though it can be seen that it tended to be sensitive to rate changes on the rise limb of an initial response and less sensitive to rate changes on the return limb.

Since in many, if not most, laboratories response identification would be confined to responses exhibited in Row 1 of Fig. 1, SCORIT's 2.7% error rate is inflated. For example, on only 4 instances (2 omissions and 2 commissions) out of 811 identified responses of Row 1 type did SCORIT make an error. The two omissions were resistance changes that required more than 4.5 sec between base and peak, and the two commissions were identifications as responses of shifts that the human scorers identified as changes in tonic level. Regardless, the relative error rate for SCORIT on the Row 1 response type was less than .005.

On balance, then, SCORIT's error and discrepancy rates fall well within the bounds that characterize human scorers. Further improvement might be made in SCORIT by programming around unusual protocols, but program complexity and processing costs would increase markedly with very little scoring gain.

INFORMATION PROVIDED BY SCORIT

For each identified response, SCORIT returns five pieces of information to the main program. The first is helpful for diagnostic purposes: It indicates the route (out of 22 different routes) through which a response was defined. The ability to identify these routes permits the investigator to eliminate as needed any specific route in subsequent scoring. The second piece of information is response latency, measured from the beginning of the data array. Adjustments for a "true" latency would have to be made by the main program. The base and peak values are the third and fourth pieces of information, and the amount of time elapsing from base to peak is the fifth. When measures cannot be defined, the program

arbitrarily assigns a value of -2. For example, in Row 3 it would be possible to define a base level and latency for a response occurring on the return limb of a prior response, but the peak value and the time to reach peak cannot be assessed. In these instances, the number returned is -2. Resistance values are returned to the main program with the values found in the data protocol, and both time measures are returned in units defined by the ordinal position. At 20 samples per second, if a response begins with the 32nd point, the number returned would be 32, which would be converted by the main program to 1.6 sec.

COST

Since the kinds of equipment employed and their maintenance costs vary from laboratory to laboratory, assessing the cost of scoring is difficult. Our experience is that SCORIT reduces costs on the order of at least 20 to 1, while providing the additional benefit of identifying responses of the sort depicted in Rows 2 and 3 of Fig. 1. Some idea of cost can be obtained from the central processor time required to score the protocols. The data of 104 Ss were scored. Each S was administered 49 trials, and sampling was taken for 11 sec on each trial. The average amount of time per S was slightly under 1 sec, which, for most computing centers, means a cost of no more than a maximum of 10-15 cents per S. Given that data are on magnetic tape in digitized form, our processing costs for scoring the records, converting to conductance measures, and storing the results on tape in a form ready for statistical analysis are clearly less than 50 cents per S. This contrasts, for example, with a minimum cost of \$3 per S simply to hand score the records, to say nothing of key punching and related costs. SCORIT, when compiled by our UNIVAC 1108 compiling system, occupies approximately 6,000 words of core, including system programs. Core needs, then, are not large, and in some

computer centers this means a reduced unit cost for processing.

CONCLUDING COMMENTS

For those interested in SCORIT, a FORTRAN program deck and documentation can be provided without charge upon request. The program contains comment cards that should simplify the work of a programmer who might wish to trace or alter its features. SCORIT will be under continuous review for improvements, and users are invited to make suggestions for useful modifications.

REFERENCES

- Dengerink, H. A., & Taylor, S. P. Multiple responses with differential properties in delayed galvanic skin response conditioning: A review. *Psychophysiology*, 1971, 8, 348-360.
- Grings, W. W., Lockhart, R. A., & Dameron, L. E. Conditioning autonomic responses of mentally subnormal individuals. *Psychological Monographs*, 1962, 76, 39(Whole No. 58).
- Lockhart, R. A. Comments regarding multiple response phenomena in long interstimulus interval conditioning. *Psychophysiology*, 1966, 3, 108-114.
- Prokasy, W. F., & Ebel, H. C. Three components of the classically conditioned GSR in human subjects. *Journal of Experimental Psychology*, 1967, 73, 247-256.
- Prokasy, W. F., & Kumpfer, K. L. Classical conditioning. In W. F. Prokasy and D. C. Raskin (Eds.), *Electrodermal activity in psychological research*. New York: Academic Press, in press.
- Stewart, M. A., Stern, J. A., Winokur, G., & Fredman, S. An analysis of GSR conditioning. *Psychological Review*, 1961, 68, 60-67.

NOTE

1. A number of the entries in the discrepancy matrix were a direct result of sampling beginning in the midst of a response. Though of little practical consequence, were sampling to begin, say, at least 1.0 sec prior to signal onset the percentage of discrepancies would be reduced somewhat by ignoring all responses identified prior to signal onset. Similarly, if the duration of the interval over which samples were taken was extended 1 sec beyond that needed, fewer discrepancies would result, provided that responses initiated during the final second were ignored.

(Received for publication October 5, 1973;
revision received November 28, 1973.)