An objective procedure for the hand scoring of scalp average evoked potentials*

DAVID W. SHUCARD† and JOHN L. HORN
University of Denver, Denver, Colorado 80210
and

DAVID METCALF

Department of Psychiatry, University of Colorado Medical Center
Denver, Colorado 80220

A method is described for identifying and scoring average evoked response (AER) components. The method uses background EEG activity as a guide for visual identification of AER components. This procedure provides an advantage in objectivity and precision over most other methods, thus producing an increase in reliability and greater likelihood for interlaboratory agreement in experiments employing measures based upon the AER components.

The purpose of this report is to describe an objective procedure for identification of average evoked response (AER) components derived from scalp-attached electrodes. The method was developed in connection with a broad study of the relationships between several measures based on the AER and on human abilities (Shucard, 1969). The method is straightforward and requires little in the way of complex apparatus or understanding of measurement. Although other methods can be used when more elaborate equipment is available, the technique discussed in this paper is useful when the AERs must be plotted directly from the signal averager and scores arrived at from visual analysis of the plots; moreover, the method is interesting in its own right as a guide to development of more elaborate AER scoring procedures, particularly where the work focuses on component identification.

METHOD

Visual AERs were recorded while the S was in a supine position with head tilted slightly forward. A Grass Model PS-1 photostimulator, set at Intensity 1, delivered the stimuli. The stimulator lamp was located 11 in. from S's eyes. The S was instructed to keep his eyes focused on a circular luminous disk, approximately 1 in. in diam, attached 3 to 3.5 in. in front of the center of the stimulator lamp.

The AER was recorded between F_4-P_4 and F_3-P_3 scalp locations, according to the 10-20 International System. Flashes were presented randomly between 1 and 4 sec. A

Larkins evoked response programmer triggered both the photostimulator and the signal averager.

Two amplifier channels of a Grass Model 78 polygraph were used to record the AER. The frequency and amplification settings were as follows: low frequency—3.0 Hz ($\frac{1}{2}$ amplitude), high frequency—200 Hz ($\frac{1}{2}$ amplitude), sensitivity—7.5 microV/mm. Also, the 60-cycle filters in both amplifiers were operable. These two amplifiers were connected to a Nuclear Data Enhancetron 1024, which in turn was connected to a Tektronix 564 storage oscilloscope for the purpose of monitoring the AER.

Prior to the start of each recording session, a 5-microV calibration signal generated by a Grass Model SWC-1 square-wave calibrator was put through the entire system and averaged over 100 trials. The equipment was calibrated so that the output from the Nuclear Data Enhancetron 1024, as recorded by a Mosley Model 2D-2M X-Y plotter, was scaled such that 1 mm corresponded to 5 msec on the horizontal scale and to 20 microV on the vertical scale when the Enhancetron display setting was at 4.

The entire sweep time was 1 sec. Thus, the output curve traversed over a horizontal distance of 200 mm (20 cm). The stimulus was administered exactly 250 msec after the sweep began. Hence, the first 50 mm on the graphic record corresponded to a period of summated baseline EEG activity preceding onset of the stimulus.

After considering several rather elaborate objective scoring possibilities, it was decided that two kinds of considerations were important for an objective and accurate method of scoring: (1) smoothing curves to reduce the influence of fast activity and other

artifacts, and (2) use of a baseline band, to estimate the bounds of "noise" and thus indicate the amount of change in potential which must occur to indicate "signal."

SMOOTHING CURVES

AER records obtained over 100 trials are jagged—the jaggedness reflecting fast activity artifact in the recording. If a filter to eliminate fast activity above 50 Hz is not used, the jaggedness is more pronounced than it would otherwise be. It can be assumed that jaggedness in the record is error relative to the "true" AER components—i.e., the "true" signal resulting as a direct consequence of the stimulation. Jaggedness introduces a problem in peak (or trough) identification because of the difficulty involved in deciding which of several peaks to identify as signal.

Four possibilities for reducing the jaggedness and easing the task of scoring the records were considered: (1) mathematical transformation—choosing a function to simulate the effects of a high-frequency band-pass filter (this involves rather complex calculations, which are often very cumbersome and time consuming—even on the computer); (2) digital filtering—again involving some rather time-consuming calculations, including conversion of analog records to digital form; (3) conversion to digital form, storing the results on magnetic tape, and playing them back through a filter; and (4) smoothing by graphic procedures. The method employed here was the last mentioned.

According to procedures generally employed in AER component identification, if a change in the sign of the slope of the AER X-Y plot occurs within less than 20 msec, it is too fast to represent a component. That is, 1 to 50 Hz is the frequency range within which workers have generally looked at AERs. The major components (in the first 500 msec) occur between 8 and 20 Hz. In the case of the data considered in this study, 20 msec corresponded to 4 mm on the graphic X-Y record. Therefore, a smoothed curve passing through the average of two points in the same direction (either negative or positive) would not systematically distort the reading of a component, provided that the two points were no more than 4 mm apart. Indeed, this smoothing would roughly approximate the outcome achieved with a low-pass filter set at 50 Hz.

In the graphic process the original records were smoothed in such a way that the smoothed curve passed through the average of all points surrounding the original jagged curve.

*The study and preparation of this manuscript was supported in part by Grant 3-K3-HD-10, 651-02, from the National Institutes of Health, and Mental Health Training Grant 5-TI-MH07082-11.

†Now at Langley Porter Neuropsychiatric Institute, 401 Parnassus Avenue, San Francisco, California 94122.

Fig. 1. Illustration of AER scoring for a S used in this study.

The smoothed curve thus represented an approximation of a least-squares fit to the jagged curve. A point on the smoothed curve was never allowed to deviate (vertically or horizontally) by more than 3 mm from the corresponding point on the jagged curve. Usually the difference was less than 2 mm. These transformations mean that if the absolute peak was read on the jagged curve as compared to the smoothed curve, the difference in latency score for the two conditions could be no more than 15 msec. It should be emphasized also that any error introduced by the smoothing procedure is unbiased relative to other scores obtained for a S during the study—including scores from other AER recordings. Hence, error due to this procedure can be regarded as random in its influence on correlations and on other statistical analyses, including those upon which interrecord reliabilities are computed. It should be noted that in order to insure that the full voltage range in the prestimulus baseline period be determined for use in subsequent procedures (see Component Identification and Scoring below), this smoothing technique was performed only on the AER waveform (poststimulus activity), not on the prestimulus EEG.

COMPONENT IDENTIFICATION AND SCORING

Having smoothed the AER curve, the next objective must be to identify the components, after which latency and amplitude scores may be obtained. Latency may be defined as the number

of millimeters (converted to milliseconds) from the onset of stimulation to the peak of a component. Amplitude is usually taken as the number of millimeters (converted to microvolts) from trough to peak or from peak to trough of a component.

Components numbered from first to fifth were abbreviated, $N_1, P_1, N_2, P_2, \dots, N_5, P_5$; "N" designates negativity and "P" positivity. Corresponding amplitude scores were identified as $N_1-P_1, P_1-N_2, \dots, N_5-P_5$. These scores were obtained as follows:

Identifying Components (see Fig. 1)

(1) The maximum peak-to-trough biphasic (nonsmoothed) excursion occurring in the prestimulus or baseline period (the first 5 cm or 250 msec of the record) is referred to as the "baseline band." The activity represented within the baseline band range can be assumed to be noise relative to any signal (response to stimulation) which may occur. Moreover, if such noise is assumed to have a random, normal distribution, then, since almost all of a normal distribution can be regarded as contained within plus and minus three standard deviations, the range represented by the baseline band can be assumed to represent a range of about six standard deviations.

(2) Two-thirds of the baseline band will be referred to as the "noise band." A millimeter ruler or millimeter paper can be used to obtain the noise band width. In the scoring operations, this width is used as the measure of nonrandom deflection—i.e., signal.

This range corresponds to about four standard deviations, two on either side of the mean. Thus, if one point (one deflection) differed from another by as much as the noise band interval, the two points would be four standard deviations apart. In order for both of two such points to represent only noise, it would be necessary that they be drawn from extreme, but opposite, sides of the random distribution. For example, if one point was more than three standard deviations above the mean and the second was more than one standard deviation below the mean, they would differ by four standard deviations. In this case, since the probabilities for these two events are independent and thus the product rule holds, the probability of the outcome can be estimated as less than .0001. The maximum probability for such an event (when both points are two sigmas removed from the mean) would be about .0006. Of course, a more rigorous scoring criterion than this could be used—for example, one might use the baseline band or two times the baseline band, etc.—but the one chosen is probably rigorous enough for most applications.

(3) The peaks and troughs for the curve extending beyond the 5-cm point are determined next. If a trough-to-peak distance or peak-to-trough distance exceeds the noise-band width, the peak or trough is identified as a component.¹

(4) The peaks (referring to peaks and troughs) are then labeled in order from first to fifth as $N_1, P_1, N_2, P_2, \dots, N_5, P_5$.

Obtaining Latency Scores

(1) The number of millimeters to each component peak is read along the abscissa to the nearest millimeter and converted to milliseconds by direct transformation.

(2) These are entered as $N_1, P_1, N_2, P_2, \dots, N_5, P_5$ on the score sheet.

Obtaining Amplitude Scores

(1) The number of millimeters from the preceding trough to peak or peak to trough of each component is read along the ordinate and converted to microvolts by direct transformation.

(2) These are entered as $N_1-P_1, P_1-N_2, \dots, N_5-P_5$ on the score sheet.

RESULTS AND DISCUSSION

An estimate of the effectiveness of these scoring methods was made by comparing the latency measures for the same S observed at different times under different conditions of S arousal. The conditions were separated by 5-10 min. The correlations between component measurements obtained across these conditions are shown in Table 1. It can be seen that these

Table 1
Correlations Between AER Latency
Measurements Obtained Under
Different Conditions*

AER Components	Conditions		
	1 and 2	1 and 3	2 and 3
N ₁	.33	.54	.38
N ₂	.70	.69	.74
N ₃	.78	.74	.86
N ₄	.78	.81	.82
N ₅	.78	.78	.82

*Number of Ss = 108.

correlations ranged from .33 for the early N₁ components to .86 for the later components. This finding of higher correlations for the later

components agrees with the findings of Ertl (1968) and Rhodes, Dustman, and Beck (1969). In effect, this means that the reliability of measurement of later components is greater than the reliability for the early two components.

Objective identification of AER components is needed to insure replication across studies. The method outlined here is objective and the results indicate that it leads to reliable measurement.

REFERENCES

- ERTL, J. Evoked potentials, neural efficiency and I.Q. Paper presented at the International Symposium for Biocybernetics, Washington, D.C., 1968.
RHODES, L. E., DUSTMAN, R. E., &

BECK, E. C. The visual evoked response: A comparison of bright and dull children. EEG Clinical Neurophysiology, 1969, 27, 364-372.

SHUCARD, D. W. Relationships among measures of the cortical evoked potential and abilities comprising human intelligence. Unpublished PhD thesis, University of Denver, 1969.

NOTE

1. If there is particular difficulty in identifying a component by this procedure in one record—say that from the left hemisphere—the E may use the information provided by the second record determined under the same conditions. Further, in cases where the record is poor or difficult to score, identification of the second positive component, P₂, can serve as a useful aid in locating the other components, since P₂ is almost always the most prominent and generally occurs around 125 msec after stimulation.