The detection and treatment of artifacts in computer-controlled neurophysiological experiments

JACKSON BEATTY and CARL FIGUEROA
University of California, Los Angeles, California 90024

Pattern recognition algorithms may be employed on time-series data obtained in electrophysiological experiments to detect particular artifacts that would otherwise contaminate the experimental data. Such algorithms may be classified by the statistical properties of the data they test. Four representative algorithms are presented that, on the basis of prior experiments in our laboratory, perform satisfactorily to improve the quality of electrophysiological measurement.

After all normal precautions in experiment preparation and recording methods have been taken, artifacts may still remain in electrophysiological data. Often, computer techniques may be employed to improve the quality of such data. The detection of artifacts in computer-controlled neurophysiological experiments can be conceptualized as a special case of the more general problem of computer pattern recognition. Artifact detection is simply the search for specifiable configurations of data from one or several data sources obtained from the experimental preparation. Thus there can be no generalized algorithm for artifact detection; rather, an appropriate algorithm must be selected for determining the presence or absence of each particular type of artifact, based upon the statistical properties of the pure and the contaminated signal.

The development of computer artifact detection procedure always involves the following stages:

(1) Identification of the artifact or class of artifacts to be detected by computer-based pattern recognition procedures.

(2) Specification of the statistical properties of pure and contaminated signals and the specification of additional sources of experimental data that may be necessary to differentiate the two.

(3) Selection of an appropriate pattern-recognition algorithm to discriminate the presence of the particular artifact based upon statistical properties of the data source or sources.

(4) Determination of criterion parameters based upon the statistical properties of the signals and the mathematical properties of the pattern recognition algorithm. For some algorithms, criterion parameters may be set at theoretically determined values, whereas

for other algorithms, empirical data may be needed to optimize the artifact detection procedure.

(5) Verification of the performance of the computer-controlled artifact detection procedure by other means, such as visual inspection of data, to determine the adequacy of the automatic procedures.

Computer-controlled artifact detection procedures are a part of an automatic artifact management process. Following detection of a specified artifact, experimental data may be edited, adjusted, or rejected, depending upon particular details of the experimental situation.

SOME TYPES OF ARTIFACT DETECTION ALGORITHMS

It is perhaps more useful to classify automatic artifact detection procedures by the statistical properties of the data used in determining the presence or absence of a particular artifact and by the algorithm suitable for making that decision than by a listing of artifact sources and the procedures appropriate to each. Here we present four types of algorithms, each operating on different statistical properties of experimental data, that are capable of reliably discriminating pure and contaminated signals when appropriately employed. For each type of algorithm, an application for a particular common artifact is presented. All procedures described below have been employed in at least one experiment in our laboratory and have performed satisfactorily.

Level Detection Procedures

Level detection procedures discriminate signal values of an experimental variable that exceed a preestablished criterion level. Such algorithms have no memory and may operate on either signed or absolute values of the experimental data source. The Schmitt trigger is an electronic realization of a level detection algorithm.

Application. In high-gain recording of dc cortical potentials, voltage shifts beyond the range of

This research was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Office of Naval Research under Contract N0001-70-C-0350. Additional on-line computer processing for data analysis was provided by the Center for Computer-based Behavioral Studies under Contract F30602-70-00016 with the Advanced Research Projects Agency of the Department of Defense.

Figure 1. An example of a level-detection algorithm.

measurement are sometimes a problem. One alternative is to reduce the system gain, but the resolution of small electrical events is also sacrificed. In discrete trials experiments, it is possible to apply a level detection algorithm to the absolute value of each EEG channel and automatically reject trials on which any channel reaches the limits of the analog/digital converter. Over-ranging is thereby eliminated.

Algorithm. The level detection algorithm takes the following form: As each data point is acquired by the analog/digital converter, it is tested against an upper (LIM 1) and a lower (LIM 2) bound. If either limit is exceeded, data from that trial is discarded.

Variability Detection Procedures

These algorithms compute a measure of the variability of a signal source and compare that quantity with a predefined criterion value. Variability procedures have a finite memory and may be computed over discrete intervals or continuously computed as a running average. Root mean squared voltage measures may be used if dc components are of interest. If a meaningful dc component is not present in the signal, due perhaps to prior high-pass filtering, spurious offsets may be removed by computing the standard deviation of the signal rather than its RMS voltage.

Application. Unwanted variations in arousal state may constitute a significant source of artifact in electrophysiological experiments using chronic recording from freely moving animals. One method of assuring that records are made from a waking animal is to monitor his movement using a "noisy wire" or other movement measurement system, compute the standard deviation or variance of the movement signal over a reasonable period of time, and acquire data if and only if this quantity exceeds a predetermined criterion. Conversely, if immobility is desired, segments in which the variability measure is less than a minimal criterion value would be utilized.

Algorithm. The variability detection procedure may be represented in the following manner: Using a memory of length L , the variance of the time-series is computed and compared with a criterion, CRIT 1. When the criterion is satisfied, an experimental trial is begun. To facilitate computation, the mean (M) and mean sum of squares (MS) are computed for the first L data points. If the criterion is not satisfied, the oldest

data point is replaced and the variance is recomputed. This updating process continues until the criterion is met, at which time the experimental trial may proceed.

Procedures for Waveshape Detection With Amplitude Specified, Phase Unspecified

These algorithms are well suited for the identification of events of specified waveshape and amplitude that occur in the signal at unspecified points in time. The procedures have memory and utilize the serial properties of the digitized data.

Application. In an experiment on evoked response morphology in a developing focal epilepsy in the rat, it is necessary to discriminate epileptiform spiking from other high-voltage events in the electroencephalographic record. This may be accomplished by comparing, point by point, a template of predetermined length of an isolated epileptiform spike sampled at a suitable rate with the ongoing EEG sampled at the same rate. If the difference between the sampled wave and the template is less than a set criterion at every point, then the EEG segment may be considered to represent an epileptiform spike, with a precisely defined onset, and not a high-voltage transient or other artifact. It can be seen that this method is, in fact, an application of the level detection procedure applied to a set of temporally related points with the final decision equal to the logical and of the separate decisions.

Algorithm. This waveshape identification procedure may be represented as follows: A template ($T_1, T_2, \dots, T_L$) of the idealized event is first defined as a set of L points, and the maximum permissible absolute deviation from these points (CRIT 1, $\dots$ CRIT L) is set at reasonable values given the signal-to-noise ratio of

Figure 2. An example of a variability-detection algorithm.

Figure 3. An example of a waveshape detection algorithm with amplitude specified and phase unspecified.

the data. The template may be set by averaging several visually identified occurrences of the desired waveform. Automatic identification is then accomplished by obtaining a set of L data points and serially comparing each point with the corresponding point of the template. If any point differs from the template by more than the criterion value, a new data point is entered into the sample array at the next clock period. Identification is made when all sampled points match the template.

Procedures for Waveshape Detection With Phase Specified, Amplitude Unspecified

These procedures involve computing the correlation of a specified segment of time series data with a predefined template which is compared with a predefined criterion.

Application. In the measurement of peripheral pulse waveforms, it is necessary to exclude movement artifact which appears as large transient shifts in potential in the output of the pressure transducer. Similar problems occur in blood velocity measurements using an ultrasonic Doppler flowmeter. In both cases, the robust QRS complex of the electrocardiogram may be used to signal the beginning of the epoch in which the pulse pressure wave is to appear, thereby specifying phase of the signal. A low correlation coefficient indicates the presence of a movement artifact. Such data is discarded and the subject is signaled that unwanted movement has occurred. Slow changes in pulse pressure or velocity waveform may be tracked by periodic replacement of the template by the average of several recently obtained artifact-free waveforms.

Algorithm. This waveshape identification algorithm can be expressed in the following way: The EKG is sampled rapidly and tested against a level detection criterion to determine the onset of the R wave which serves as a phase-locking signal for the peripheral pulse pressure wave. Upon detection, L data points are

Figure 4. An example of a waveshape detection algorithm with phase specified and amplitude unspecified. (A) Template formation. (B) Decision process.

acquired from the peripheral transducer at a fixed sampling rate. N such samples are obtained and are averaged together. If visual inspection reveals no movement artifacts, this average is used as a template and the mean together with L times the variance of these points are computed and saved.

In data acquisition, the EKG is again scanned for the R wave. Upon detection, L samples of the peripheral pulse measure are obtained. Between samples, the sums of X_i , X_i^2 , and X_iT_i are computed. When L points have been obtained, these sums are combined with the previously stored values to compute the Pearson product moment coefficient of correlation, which is then compared with the criterion. If the obtained correlation exceeds the criterion, the data are accepted.

CONCLUSION

The use of statistical algorithms for the further purification of time series data obtained in neurophysiological experiments can improve the quality of experimental procedures. Careful attention must be paid to the statistical features that discriminate proper from artifactual data in a particular situation. Once identified, automatic tests of these features may be devised to permit the rejection, editing, or correction of artifactual data within the confines of the experiment. These tests may be considered as pattern recognition algorithms applied to time-series data.