PSYCHOPHYSIOLOGY

Systems, Processes, and Applications

EDITED BY

Michael G. H. Coles
Department of Psychology
University of Illinois
Champaign, Illinois

Emanuel Donchin
Department of Psychology
University of Illinois
Champaign, Illinois

Stephen W. Porges
Department of Human Development
University of Maryland
College Park, Maryland

The Guilford Press New York London

other branches of psychophysiology or not currently in use. In this way, the range of questions that can be answered with respect to a given psychophysiological function can be extended.

Our emphasis on the potential generality of analytical techniques should not be taken to mean that we think that specific measurement techniques are unimportant. Other chapters in this volume discuss the measurement techniques that are typically used in the recording of different psychophysiological functions. Furthermore, for the sake of completeness, we briefly review different approaches to psychophysiological measurement in the next section. However, the bulk of this chapter is devoted to a review of analytic techniques. We present, in detail, two classes of analytic techniques: time domain and frequency domain. Selection between these two classes, and among the different techniques within each class, is dictated by the questions asked by the investigator. Thus, we not only describe the different techniques, but also point to those questions that the techniques are best suited to answer.

DERIVING VOLTAGE × TIME MEASUREMENT FUNCTIONS

In this section, we consider the sequence of events that transpires between variations in the activity of a physiological system in a human subject, and the derivation of the voltage × time functions that represent this activity. This is a brief review. More detailed treatments can be found in other chapters in this volume, and also in Brown (1967), Martin and Venables (1980), Stern, Ray, and Davis (1980), and Venables and Martin (1967a).

Attachments to the Subject

We may distinguish here between two classes of attachments. First, there are those used when the investigator is interested in the activity of a physiological function that manifests itself in variation in electrical activity that can be measured on the surface of the skin. Secondly, there are those used when the activity of the function of interest is manifested in a nonelectrical fashion. We consider these two separately.

Electrodes

Electrodes are used when the activity of the psychophysiological function of interest can be detected in the form of electrical activity at the surface of the skin. Measures of the EEG, the electromyogram (EMG), the electro-oculogram (EOG), the electrocardiogram (ECG or EKG), and EDA all require the use of electrodes. In most cases, the electrodes merely constitute an interface between the subject and amplification equipment (see below), although for some measures of EDA (skin conductance and resistance) the electrodes are used to apply small, constant voltages or currents to the skin in order to quantify properties other than surface voltage.

Electrodes customarily are small metallic discs or disc shapes that are attached to the surface of the subject's skin. Placement depends on the function of interest. Attachment to the subject is generally accomplished through the use of double-adhesive collars that stick to both the electrode and the subject. However, if the electrodes are to be placed on an area that is hairy (e.g., the scalp), then either a glue (e.g., collodion) or a rubber cap may be needed to hold the electrodes in place.

The most critical aspect of the electrode is that it is electrically stable. It should be both inert (i.e., have no inherent electrical activity) and nonpolarizable (i.e., be unaffected by continued exposure to current flow). For all functions, the electrode material of choice is currently silver/silver chloride (silver chloride surface surrounding a solid silver base).

Prior to electrode attachment, the skin is generally cleaned with a mild solvent, such as acetone. With EDA, however, the measure itself can be influenced by the method of cleaning. Venables and Martin (1967b) report that, while acetone, ether, and distilled water do not affect EDA, soap and water lower conductance and raise resistance. To eliminate the possibility of between-subjects variations due to the method of cleaning, these authors advise standardizing procedures across subjects.

Contact between electrodes and skin is maintained by a jelly or paste. For all functions, it is desirable that the jelly be chemically compatible with the skin. For this reason, electrolytes containing sodium chloride (NaCl) or potassium chloride (KCl) are generally used, preferably in concentrations that correspond to those found on the skin. Although commercially available electrolytes do not always satisfy this last requirement, they are usually judged to be acceptable for most recording applications. The primary requirement is that there should be chemical overlap among electrolytes at each interface of material. Silver chloride on the electrode surface plus NaCl in the jelly creates an appropriate sequence of electrolytes between metal electrode and skin. As noted above, measurement of EDA presents a special set of problems, since the behavior of the system itself can be influenced by the electrolyte. Venables and Christie (1980) present a detailed discussion of the problems of electrolyte, with special reference to the measurement of EDA.

The particular characteristics of electrodes, skin preparation, and electrolyte are chosen for one reason—that is, to provide faithful transmission of the electrical activity manifested at the skin to an amplifying system (in a polygraph) where the electrical activity can be magnified. The selection of these characteristics is based on the requirement that whatever reaches the amplifying system should consist of no less and no more than what actually exists at the skin. Note that the activity at the skin may not always represent the activity of interest. Electrodes cannot discriminate among brain electrical activity, muscle electrical activity, or the electrical activity associated with eye movements. For this reason, care must be exercised in ascribing a cause to the electrical activity recorded using electrodes. We consider how this activity is treated by the polygraph after we have discussed the second type of subject attachment.

Transducers

Many physiological functions of interest are not directly manifested in electrical activity at the skin surface. The activity may appear in a number of different ways. First, it may appear directly as mechanical activity. The respiration belt and the strain gauge plethysmograph both rely on the fact that mechanical changes occur with variations in the activity of the function. Respiration may also be measured using a less direct mechanical procedure, the respiratory spirometer, which converts the changes in airflow that occur during respiration into mechanical changes. In other cases, the fact that the function is manifested in changes in the optical quality of tissue is used (e.g., the photoplethysmograph).

The task of the transducer is to convert the mechanical or optical manifestation of the function into an electrical function. With primary and secondary mechanical systems, the conversion can be made to electrical resistance using a strain gauge. The prototypical strain gauge is a plastic tube filled with mercury. Variations in the length and cross-section of the tube, resulting from stretching, are associated with changes in resistance of the tube. Appropriate placement of the strain gauge ensures that variations in the resistance of the strain gauge are due to variations in the function of interest. Using a suitable bridge circuit (see below), these changes in resistance are then converted into changes in voltage.

Other functions that can be monitored using the resistance principle include temperature. In this case, a thermistor is used whose resistance changes with temperature.

With optical systems, the need is to convert variations in the optical properties of tissue that are associated with vascular events into electrical activity (see Jennings, Tahmoush, & Redmond, 1980). In all optical systems, there are two elements, a light source and a receiver. Activity at the receiver depends either on the amount of light transmitted (if source and receiver are on opposite sides of the tissue) or on the amount of light backscattered (if source and receiver are on the same side). Depending on the characteristics of the receiver, variations in the amount of transmitted or backscattered light are converted into variations in electrical current or electrical resistance. In the latter case, a bridge circuit must be used to convert resistance change to voltage change.

In this section, we have considered devices that are used to convert the activity of physiological functions into electrical activity. Note that the transducer can only operate on that aspect of the function it is designed to detect. The function may have many manifestations, only one of which can be detected by the transducer. Furthermore, the transducer will not differentiate between activity that is caused by the function of interest and activity that is caused by extraneous events. For example, respiration strain gauges will be sensitive to all forms of movement—not just those attributable to respiration. Thus, however well a transducer is designed and positioned, it will be blindly faithful in converting what it "sees" into electrical activity. With these caveats in mind, we can now turn to the system that scales these diverse voltage × time functions to a common format.

The Polygraph

"Polygraph" is a generic name for a device that amplifies, shapes, and records psychophysiological functions. Although polygraphs come in different shapes and sizes, they have a number of common features: amplifiers, bridge circuits, integrators, rate devices, analog filters, and a graphic readout facility. The increasing use of computers in psychophysiological research has made the last item redundant for many investigators. The polygraph is interfaced directly with a computer, thus making scoring of polygraph records by hand unnecessary.

The connection between subject and polygraph is achieved via wires or cables (leads). Their function is merely to transmit electrical activity to and from the subject (electrodes) or to and from the transducers. Each psychophysiological measure is processed by a separate channel of the polygraph. Each channel contains a device that is directly connected to the subject or transducer (sometimes called a "coupler") and an amplifying system. The amplifying system is generally the same for all channels. Most manufacturers of polygraphs supply a variety of couplers, each of which is specific for the measurement of a particular psychophysiological function. Below, we review some general characteristics of these couplers and amplifiers.

Amplifiers

The most elementary function of the polygraph is to magnify psychophysiological signals. Amplifiers fulfill this function by increasing the magnitude of the input voltage by a factor of up to 500,000. Following amplification, the signal should have an amplitude on the

er of about ±1 V to be compatible with either the phic readout system of a polygraph or the analog-digital converter of a computer (see below). The size of the amplification factor will depend on : size of the input signal. For example, the magnitle of the ECG signal is about 1 mV, while that of : EEG is about 50 μV. Thus, the amplification :tor for these two measures might be 1000 and),000 times, respectively.

To ensure that the amplifier is performing the appropriate magnification, it is important to pass calibration voltages of known amplitude through the amplification system.

ridge Circuits

ıs we have seen, most transducers represent psychophysiological activity in the form of resistance hanges. For this reason, a critical function of the olygraph is to measure resistance change and to convert it to voltage change. This is accomplished hrough the use of a bridge circuit, which can be as imple as a few resistors arranged in a special way (Malmstadt, Enke, & Crouch, 1974). A bridge circuit provides constant current to the transducer. As the resistance of the transducer changes, so the voltage across the transducer changes. This voltage change is then amplified (see above).

Bridge cricuits are also used in the measurement of two complementary forms of EDA, skin conductance and skin resistance. In this case, either a constant voltage or a constant current is imposed on the subject, and the bridge measures variations in current or voltage that correspond, respectively, to variations in conductance or resistance. Because this procedure involves the imposition of external electrical activity on the subject, safety is a critical factor. However, the procedure is now reasonably standardized (see Fowles, Chapter 4, this volume).

Analog Filtering

As we have mentioned, the task of the electrodes and transducers is to convey to the polygraph a faithful representation of the electrical or other activity associated with a psychophysiological function. In some cases, the signal so conveyed may be filtered by the polygraph, either because it contains artifacts or because it contains aspects of the psychophysiological signal that are of no interest to the investigator.

For the purposes of describing the principles of signal modification or "signal conditioning," the signal is considered as being comprised of different frequencies. Thus, some of these frequencies may be artifactual (due to sources outside the subject or to activity of other, irrelevant functions), while others may simply be of no interest.

For example, a common source of artifact in psychophysiological measurement is 60-Hz (or 50-Hz) activity from standard electrical equipment. This artifact can be minimized by the use of a "notch" filter set at 60 or 50 Hz, which attenuates activity at this frequency while permitting activity at higher or lower frequencies to pass.

Other filters attenuate activity above or below specified frequencies (low-pass and high-pass filters). For example, in EEG recording, the investigator is generally interested only in activity below 40 Hz. Thus, a low-pass filter set at 40 Hz can be used. The ECG consists of frequency components between .05 Hz and 80 Hz (Strong, 1970). If the investigator merely wants to detect the R wave (e.g., to measure interbeat interval), a high-pass filter set at 10 Hz can be used. The high-pass filter attenuates slow shifts in the ECG signal that may be due to EDA or some other unwanted activity.

The various types of electronic circuitry that typically serve as filters can be characterized by their "time constant." The value of the time constant of a high-pass filter is the time for a given sustained input to the circuit to be attenuated to 63% of its original value. While all analog filtering circuits have a time constant characteristic, in practice the concept is associated primarily with that portion of a circuit that serves as the high-pass filter. Some correspondences between time constant and filter cutoff frequency (−3 dB) are as follows ($F = 1/[2*\pi*TC]$, where F = frequency, TC = time constant):

Time Constant (sec)	Frequency (Hz)
10.00	.016
5.00	.032
1.00	.159
.30	.531
.01	15.915

Of course, it is imperative that great care be taken in the use of filters. The investigator does not want to distort the signal of interest (see, for example, Duncan-Johnson & Donchin, 1979). Filters are useful when the characteristics of the unwanted aspects of the signal do not overlap the wanted aspects. The problems that occur when there is overlap, and the solutions to these problems, are discussed below (see "Data Analysis in the Time Domain").

The "analog" filters briefly discussed here are electronic components placed in-line during initial recording of continuous signals, often within the amplifier chassis. Their chief advantages are simplicity and speed. Their disadvantages are that they introduce a phase shift into the signal and that in a particular polygraph they are typically limited to a few settings. Analog filters must be distinguished from digital filters, which are algebraic manipulations of discrete (digitised) signals after recording is complete. Digital filters, discussed below (see pp. 195–196), can be constructed without phase shift and with any desired filter characteristics.

Analog Integration

For some physiological signals, particularly EMG, the investigator is not so much interested in the frequency characteristics of the signal as in the overall amplitude-frequency activity in the signal. Analog integrators provide this measure by first rectifying the signal and then converting the area under the rectified record into a smoothed analog voltage (rectification involves removing or inverting either the positive or negative portion of an alternating current [AC] signal). The resulting voltage × time function will depend on both the amplitude and the frequency of the input signal at any point in time. Because analog integration is normally accomplished with an in-line electronic circuit that is essentially a low-pass filter (smoothing out rapid peaks but preserving average amplitude), different integrators are appropriate for different physiological signals, depending on the frequency characteristics of the signal and the time constant of the integration circuit. Furthermore, the output of such an analog circuit lags the input, again introducing the issue of phase shift. When the frequencies of interest are high in relation to the time resolution needed, as in EMG recording, this lag is inconsequential.

Rate Devices

With some physiological functions, the measure of interest is the rate at which some event occurs, rather than the level of activity. For example, with heart rate (HR), the investigator is concerned with the rate at which R waves are observed in the ECG, rather than with voltage characteristics of the ECG waveform itself.

To accomplish this measurement, most polygraph manufacturers offer rate devices (cardiotachometers), which convert inter-event intervals into an analog signal whose amplitude varies with rate. In some implementations, the conversion is made through a circuit that first detects an R wave, then allows a capacitor to be charged until the next R wave is detected, at which time the capacitor is discharged. The voltage discharged by the capacitor will vary as a function of the duration of the charging period and hence will be proportional to the interbeat interval (and inversely proportional to the rate). Note that the level of the output of the rate device (a voltage × time function) will depend on the previously completed interbeat interval. Thus, the output will lag the input.

Computer Access to Voltage × Time Functions

Digital Input and Analog-to-Digital Conversion

With the development of computers, automatic scoring of physiological data has become a reality. But, before a digital computer can apply the appropriate scoring algorithms, the data must be presented in a palatable form—a set of digitized (i.e., discrete) values. However, the voltage × time functions we have described are inherently analog (i.e., continuous) functions. The requirement, then, is to convert these analog functions into digital representations. Some types of physiological activity are easily represented digitally. For example, while the ECG is a continuous voltage, the occurrence of its R-wave component is easily approximated digitally as a "1" in a series of "0's." Simple electronic circuitry between polygraph and computer, such as Schmitt trigger, readily converts the analog ECG input signal to such a digital output signal. Thus, a continuous voltage × time signal is converted to a discrete voltage × time signal. This method is more accurate than, and obviates the need for, a cardiotachometer rate device, described above.

More elaborate conversion circuitry is required when more information about the continuous input function, besides the mere occurrence of an event, must be represented in the discrete output function. The term "analog-to-digital" (A/D) converter is normally reserved for such circuitry, which produces a series of numerical values that are discrete samples of voltage level from a continuous input. Rather than merely one bit of information (0 or 1), the output has a large number of possible values. For example, a 12-bit A/D converter can output 4096 different values, depending on the voltage input at the time of sampling. Such resolution is essential for measurement of signal amplitude. The sampling intervals used vary as a function of the particular measure. For example, for the auditory brain stem response the intervals are typically 20 μsec (sampling rate of 50 kHz), while for respiration the intervals may be as long as 1 sec (1 Hz). Choice of sampling interval (or sampling rate) is dictated by the expected period or frequency characteristics of the measure in question. The slowest acceptable sampling rate is twice the highest frequency present in the data. A slower sampling rate will provide a distorted digital representation of the analog input (this issue is elaborated further in the section by Porges). A good rule of thumb, then, is to err on the conservative side and sample at least two to five times the highest expected frequency.

The output of the A/D converter, now a discrete voltage × time function, is fed directly to the computer. While logically distinct from the computer itself, pieces of circuitry such as Schmitt triggers, digital input interfaces, and A/D converters are typically integrated electronically into the computer enclosure.

Distributed Processing: Remote Data Acquisition

Given the low price and small size of current microprocessors, laboratory equipment manufacturers have begun to offer "smart" laboratory products that perform the continuous-to-discrete conversion external

to the computer and its associated A/D converters and other circuitry. Data are then passed to the computer in highly palatable form—as the same 8-bit characters that video display terminals send. Thus, the traditional configuration of "dumb" equipment plus a dedicated laboratory computer (with central A/D converter, etc.) can be replaced with "smart" equipment plus a simpler, general, multipurpose computer.

The investigator should, of course, consider the growing variety of configuration options in laboratory equipment when developing a new measurement capability. The point is that across these diverse options all psychophysiological data, whether written on polygraph paper or handled by the most elaborate microprocessor network, can be treated as a voltage × time function, a series of voltage levels in time—a voltage "time series."

DATA ANALYSIS IN THE TIME DOMAIN

Introduction

This section of the chapter distinguishes analytic techniques applied to data in the time domain from those applied in the frequency domain (see the section by Porges). Psychophysiologists intend to monitor the activity of some internal structure manifested as a "signal" conveyed to the body surface by some functional channel. This signal is combined with "noise" coming from other internal and external sources. In many cases, the extraction of the signal from the background noise is a very challenging task.

In the case of data in the time domain, the signal is typically a phasic, nonrepetitive feature of the time series recorded at the surface; this feature is assumed to reflect the activity of a specific internal structure. Important characteristics of the feature commonly include its restriction to a particular time epoch in the record and its variability in latency. Since the signal of interest contributes only part of the variability observed in the time domain, we refer to it as a "component." This component constitutes the target of the signal extraction procedure.

Since signal components are in most cases embedded in noise, the first task for the data analyst is to extract the signal from its background. To accomplish this task, the signal must be defined.

Signal extraction techniques differ in the way in which they define components. The choice of an extraction technique implies a model of the signal, including a specification of its distinctive features and the ways in which these interact to produce the waveforms (time series) that are actually recorded. For instance, a model of event-related potentials (ERPs) could define a component as a deflection of the EEG trace time-locked to a stimulus, with a specific latency and scalp distribution, that "summates" with other components and with noise to produce the waveforms recorded at the scalp. Alternatively, EDA components are deflections of the skin conductance trace, with some shape and latency following an eliciting stimulus. A cardiac cycle can be identified by means of a distinctive feature (R wave), or by its general waveshape, referred to the spatial location used for the recording. Analogous definitions can be given for any component of interest for the psychophysiologist. Specific component models are often highly controversial. Nevertheless, the procedure adopted to extract the signal from the noise in which it is embedded necessarily depends on some kind of model. Therefore, in the present discussion, we pay particular attention to models of signal and noise implicit in different signal extraction techniques.

Once the signal component is defined, the amplitude, latency, or spatial distribution of the raw data can be quantified. These quantification techniques depend on the definition of "components" used for extracting the signal. In many cases, these two stages of data analysis (signal extraction and quantification) constitute a single process. However, the logical distinction between signal extraction and quantification should be kept in mind throughout this section.

Signal Extraction Techniques

The remainder of this section provides a brief sample of the many ways to process the basic voltage × time function. This review is divided into techniques for signal extraction, for data reduction, and for spatial analysis. In fact, since a given technique may serve several such functions, such a division is necessarily somewhat arbitrary.

Signal Averaging

Since the psychophysiological signal is often obscured by noise, many techniques have been proposed to amplify selectively the information of interest for the psychophysiologist. A number of techniques assume that the signal can be differentiated from the noise on the assumption that only the signal is temporally related to an external marker event. Such procedures therefore define the signal as everything in the recording that is time-locked to an external event. All other variability, not time-locked to the external event, is considered noise. This definition is particularly useful when studying perceptual and motor processes. In this case, the relevant external events are readily identifiable, and the temporal relationship of the external event and the internal process is assumed to be constant. The basic procedure consists of the repetition of a large number of essentially identical trials.

Through superimposition or averaging of the single trials, the constant psychophysiological response (signal) to the stimulus remains constant, while variability not consistently related to the external event averages to zero.

The superimposition technique consists simply of overlapping the trace for each of the single trials on a plotter. It can be also obtained with a storage oscilloscope by triggering the display sweep at each presentation of the stimulus. Since superimposition does not require high-speed computing facilities or A/D conversion, it was extensively employed in the 1950s. An advantage of superimposition is that it portrays the range of variability of the single trials. However, it is fairly difficult to detect small potentials, or small differences in amplitude between conditions, by means of this technique. Thus, superimposition is more appropriate when measuring latency than when measuring amplitude. However, in recent years it has been replaced by averaging techniques.

In averaging, the values obtained at each time point are averaged across trials. To employ this algebraic technique, it is, of course, necessary to transform the signal obtained from the amplifier from analog to digital format.

The advantage of averaging over superimposition is the "cleaner" waveform that averaging produces. This expresses the "central tendency" of the sample of trials examined and corresponds to the best statistical estimate of the signal. It is easy to compute the point-by-point standard deviation or range in parallel with the averages, in order to have more complete information about the data.

In principle, averaging can extract an arbitrarily small signal relative to background noise amplitude, if a large number of invariant trials are averaged. The noise will be reduced as a function of the square root of the number of trials. For example, the brain stem auditory ERP, typically less than 1.0 μV, may require several thousand trials.

However, averaging is vulnerable to violations of its assumptions of specifiable external stimulus and invariant response latency and morphology. Particularly when the investigator suspects cross-trial inconsistency in the signal, averages must be interpreted cautiously.

Removing Systematic Noise

Most signal extraction techniques have been developed in order to deal with the problem of *random* noise. Consequently, they are often insufficient in the case of *systematic* noise. In fact, these techniques generally assume that "noise" is that part of the variance that is not systematically related to the experimental variables. Of course, this corresponds to the definition of random noise. However, some of the noise present in the data can be systematically related to the experimental variables. We label this "systematic noise."

In the presence of systematic noise, two important points must be kept in mind. First, the signal must be defined in a more restricted way than simply as "everything related to experimental variables." An example is provided by brain ERPs, where, for a component to be considered a signal, it is not sufficient that it is systematically related to the eliciting event. It is also necessary that it be generated by the brain. Therefore, a systematic ocular potential, recorded at the scalp, does not constitute a brain ERP component, but systematic noise. This kind of systematic noise is commonly called an "artifact."

A second important point concerns the difficulty of dealing with systematic noise by means of traditional signal extraction techniques. A procedure usually adopted to reduce artifact in recording is filtering. There are many ways of filtering data, the most common being frequency filtering. This kind of filtering is discussed elsewhere in this chapter (see pp. 186–187 and pp. 195–196). However, frequency filters are sometimes insufficient for handling artifacts in the data. This is especially the case when signal and artifact have similar frequencies. Eye movement artifact in brain ERPs is an example of this problem.

Fortunately, artifacts are sometimes recognizable by their specific features. These features may be evident in the data themselves or in a recording from electrodes placed near the source of the artifact. In either case, the artifact can be detected (by visual inspection or by some automatic procedure) and the associated record discarded from subsequent analysis. However, although this is a common procedure, such loss of data is not always affordable (Gratton, Coles, & Donchin, 1983).

For this reason, procedures have been developed in order to compensate for artifacts. They are based on the possibility of inferring the effect of an artifact on the records at a certain spatial location from data obtained from a location close to the source of artifact. Data of the latter type may be considered "pure" measures of the activity of the "artifact generator." The remainder of this section describes a recently developed procedure of this type.

This procedure, proposed by Gratton *et al.* (1983), represents an example of an artifact compensation technique. It assumes that the effect of an eye movement on the potential recorded at any scalp location (EEG) can be inferred from activity recorded at a location close to the eyeball (EOG). In order to make this inference, it is sufficient to know how much a signal recorded at the ocular electrode "propagates" to the scalp location under study. Previous researchers (e.g., Corby & Kopell, 1972; Overton & Shagass, 1969; Weerts & Lang, 1973) have demonstrated that not all ocular potentials propagate to the scalp in the same way. In particular, potentials gener-

ated by eyeblinks propagate less than potentials generated by saccadic eye movements.

Accordingly, the proposed eye movement correction procedure (EMCP) distinguishes between time points in the record during which eyeblinks occur (detected by means of a pattern recognition technique; see "Pattern Recognition," below) and time points in which saccadic eye movements occur. Separate propagation factors are then computed for blinks and saccades.

The propagation factors are computed by means of a least-squares-regression technique. However, as noted above, ocular artifacts can be consistently related to some external events. Since brain ERPs can also be elicited consistently by external events, spurious relationships can affect the computation of the correction factors. Therefore, the averaged EOG and EEG traces are subtracted from the single-trial records before the correction factors are computed. In this way, the propagation factors are computed on that portion of the variance of the EOG and EEG recordings that is not related to the external event. The propagation factors are then applied to the original data to correct for the ocular artifact. A schematic representation of EMCP is presented in Figure 10-1.

Although some inaccuracy is present (involving mainly the invariance in time of the EEG and EOG response to the external event, and the difference between the propagation factor for upward and downward eye movements), tests presented by Gratton *et al.* (1983) indicate that EMCP effectively compensates for the ocular artifact.

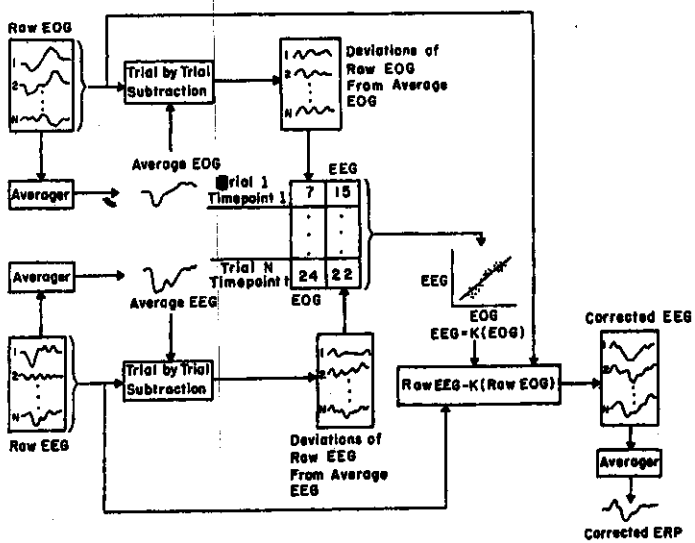

Figure 10-1. Schematic representation of the eye movement correction procedure (EMCP). (From "A New Method for Off-Line Removal of Ocular Artifact" by G. Gratton, M. G. H. Coles, and E. Donchin. *Electroencephalography and Clinical Neurophysiology*, 1983, 55, 468–484. Reprinted by permission.)

Pattern Recognition

INTRODUCTION

Signal-averaging techniques (see pp. 188–189) are particularly useful in separating small signals that are time-locked to an external event from background noise that is not time-locked to the external event. However, in many cases, the assumption of invariance of latency of the signal over trials is untenable, even as a first approximation. In other cases, it is impossible to establish an external event to which the psychophysiological signal can be time-locked. Thus, straightforward signal averaging is not always possible. Pattern recognition techniques can be helpful in these cases. The general assumption underlying such techniques is that the signal is distinguishable from the background noise on the basis of specific features, typically aspects of its waveshape. Two types of pattern recognition techniques may be distinguished: those in which the characterizing features are established *a priori* on the basis of previous data or conceptualizations, and those in which the characterizing features are established *a posteriori* on the basis of characteristics of the data to be processed.

Examples of the first type are the techniques used in psychophysiology to detect the R wave of the ECG, the phasic electrodermal response, or blinks in the EOG trace. Procedures of this type are specific for a particular psychophysiological measure and are not easily generalizable to other measures. Note also that many of these pattern recognition techniques are used to recognize artifacts (see also the discussion of EMCP, above). Although pattern recognition may be performed simply by visual inspection of the records, for reasons of reliability it is preferable to automate the procedure using hardware devices (e.g., Schmitt triggers) or software algorithms.

Pattern recognition techniques based on standard statistical procedures (e.g., cross-correlational techniques, discriminant and canonical analyses, etc.) usually require the use of high-speed computing devices, since they involve large amounts of computation. However, they have the advantage of being generalizable to many different measurement domains; they can also sometimes be applied without any previous knowledge of the "pattern" or "feature" to be recognized. Two examples of this kind of technique (cross-correlation and discriminant analysis) are discussed below in some detail.

CROSS-CORRELATION

A fundamental assumption of the cross-correlational approach (Friedman, 1968) is that the waveshape of the signal component to be detected is constant over trials, while the shape of the noise varies randomly from trial to trial. Thus, that portion of the variance that is constant over trials will contribute to the correlation between trials. Because cross-correlational techniques do not assume invariance of the interval between external event and internal process manifested by the signal component of interest, they are applicable even in the absence of any identified external event and are potentially more versatile than signal averaging, which assumes signal invariance in both morphology and latency.

Cross-correlational techniques involve the computation of a "cross-correlational series" between a "template" (predetermined pattern of consecutive points) and any single trial. A cross-correlational series is an array of correlation values between two time series (or within the same series), where one of the time series is progressively shifted by a certain interval ("lag"). For example, the first correlation index is computed between the elements $a(1)$, $a(2)$, $a(3)$, . . . , $a(n)$ of the template and the elements $b(1)$, $b(2)$, $b(3)$, . . . , $b(n)$ of a given trial. In this case, the lag between the elements of the template and of the trial is 0. Then a second correlation index is computed between the elements $a(1)$, $a(2)$, $a(3)$, . . . , $a(n)$ of the template and the elements $b(1 + \text{lag})$, $b(2 + \text{lag})$, $b(3 + \text{lag})$, . . . , $b(n + \text{lag})$. A series of correlation values is computed by progressively increasing the size of the lag. This procedure is limited only by the number of the elements of the trial array (a correlation involving too small a number of elements would not be reliable). Then, the maximum value in the series of cross-correlations is selected. The lag corresponding to this maximum value is the one at which the trial maximally "looks like" the template. According to the pattern recognition approach, this is the lag at which the signal is "detected." In most cases, if for a given trial some minimal correlation value cannot be reached with any lag, the "signal" is considered to be absent on that trial.

Cross-correlation is vulnerable to two problems. First, the maximum cross-correlation for a particular trial may be unacceptably low. To accept such a trial is to assume that the signal is whatever in the data is least dissimilar to the template. Second, cross-correlation cannot easily handle the presence of multiple components differing in latency. This would constitute a violation of the assumption of invariance of shape of the signal. Cross-correlational techniques should not replace signal-averaging techniques in the case of components with fixed latency, particularly when the signal-to-noise ratio is very small.

Notwithstanding these limitations, cross-correlation techniques have the advantage of utilizing the information provided by the *whole* time series, thereby increasing the power of the analysis. They are particularly useful in identifying components having variable latencies embedded in large amounts of noise.

THE WOODY ADAPTIVE FILTER

The Woody adaptive filter (Woody, 1967) is a particular kind of cross-correlational technique. The term "adaptive" refers to the fact that the template is not established *a priori*, but is extracted by means of an iterative procedure from the data themselves. Each iteration serves to refine the template. This method was originally proposed to identify particular patterns of variation of the EEG recorded in epileptic patients.

Typically, the template used for the initial iteration is the half-cycle of a sine or triangular wave, or the average of the unfiltered single trials. Cross-lagged covariances or correlations are computed between each trial and this template. A new template is obtained by aligning the single trials at the lag that gives the maximum cross-correlation. This procedure is then repeated, using the new average as the template, until the maximal values of cross-correlation become stable. Trials where correlations with the template do not reach a criterion (e.g., .30–.50) at any lag are not used in subsequent template construction and may be discarded entirely from subsequent analysis.

Several studies have been conducted to test the power and reliability of the Woody filter (Nahvi,

Woody, Ungar, & Sharafat, 1975; Wastell, 1977; Woody & Nahvi, 1973). They have concluded that the Woody filter method is often superior to a simple peak detection technique (see pp. 196–198). However, the use of multiple iterations has been questioned (Wastell, 1977). In fact, Wastell reports a decline in validity of the procedure when several iterations are used. Therefore, in contrast with signal averaging, the Woody filter (like most autocorrelation techniques) is not able to improve the signal-to-noise ratio over a definite limit. Therefore, its reliability under conditions of very low signal-to-noise ratio is questionable.

DISCRIMINANT ANALYSIS

Introduction. Discriminant analysis provides a method of discriminating between two or more groups on the basis of systematic differences in the data set. A case classification rule is derived from data whose group membership is known ("training set data"). This rule is then applied to new data of unknown group membership ("test set data"). Thus, discriminant analysis requires that the investigator specify *a priori* the groups into which the data are to be classified. Groups may refer to distinct samples of subjects or to distinct classes of events that vary within subjects.

In addition to providing a method of discriminating among groups, discriminant analysis also provides a means by which to reduce the dimensionality of the data. Such a reduction serves to increase the stability of the discriminant composite. Data reduction is accomplished by selecting a subset of the original variables that best discriminates among the groups. These variables are then used in computing a linear combination of weighting coefficients × variables to produce a discriminant score. The pattern of weighting coefficients provides information concerning the contribution of each variable to the differentiation between the groups. The function employed in the computation of the discriminant score is referred to as the "discriminant function." The purpose of the function is to provide optimal separation between two or more groups by maximizing the between-group variance while minimizing the within-group variance. In the case of psychophysiological data, when the investigator wishes to discriminate between sets of voltage × time functions, the discriminant function consists of a linear combination of time points × weighting coefficients.

As has been discussed above (see pp. 188–189), signal averaging can serve as a relatively simple method of pattern recognition and signal classification. One might doubt the necessity of employing more complex, multivariate techniques, such as discriminant analysis, to accomplish the same goal. In many situations, averaging is adequate. In some situations, however, signal averaging will produce misleading results. For example, averaging is inappropriate when substantial, uncontrolled variation in the amplitude of a component occurs. In such cases, discriminant analysis provides a clear advantage over signal-averaging procedures, since the differential amplitude of the psychophysiological component can become the basis for group classification. In addition to supplementing the signal-averaging procedure, discriminant analysis also provides a technique that can be employed in the analysis of single-trial data. This is clearly advantageous when the investigator is interested in the trial-to-trial variation in both psychophysiological and performance measures. For example, the use of discriminant analysis procedures in the evaluation of single-trial ERPs has had important theoretical implications. Squires, Wickens, Squires, and Donchin (1976) employed it to construct a quantitative expectancy model of the P300 component of the ERP (see pp. 194–195).

Linear Stepwise Discriminant Analysis. The most commonly used discriminant analysis procedure for the assessment of psychophysiological data is linear stepwise discriminant analysis (LSDA) (Donchin & Herning, 1975; Horst & Donchin, 1980; McGillem, Aunon, & Childers, 1981; Squires & Donchin, 1976). The goal of the LSDA procedure is the selection of a subset of variables that maximizes the between-group separation. The process is analogous to stepwise multiple regression, except that in LSDA the predicted criterion can be a multilevel nominal variable.

The first step is to identify the variable that accounts for the largest proportion of between-group variance. A second variable is then selected that accounts for the maximum proportion of between-group variance not already accounted for by the first variable. This successive selection of variables constitutes the stepwise portion of the LSDA procedure. The between-group difference at each step in the procedure is measured by a one-way analysis of variance (ANOVA) F statistic, and the variable with the largest F is chosen. Several LSDA computer programs permit the deletion of variables that no longer provide a substantial contribution to group separation as other variables are added (Dixon, 1979; Jennrich, 1977). These variables may later be re-entered if their F value is again adequate. The process of variable selection is terminated when some specified criterion has been met. Criteria commonly employed include the number of variables already entered, the amount of variance accounted for, or the point at which no further improvement occurs in some criterion (e.g., the U statistic in the BMDP package, see Dixon, 1979).

A separate vector of weighting coefficients will be derived for each of $n - 1$ discriminant functions, n being the number of groups. The discriminant criterion value provides a measure of group differentiation for each discriminant function. The first discriminant function has the largest discriminant criterion value, indicating the dimension of maximal group differentiation. The second discriminant function represents the largest group difference not accounted for by the first dimension. Thus, the discriminant criterion value and hence the group differentiation accounted for by each discriminant function decreases with successive functions.

Although $n - 1$ discriminant functions can be calculated, they might not all contribute significantly to group differentiation. Several procedures are available to test the incremental significance of successive discriminant functions (see Tatsuoka, 1970, 1971). Eliminating discriminant functions that do not contribute significantly to group differentiation serves further to reduce the dimensionality of the data set.

As mentioned above, one of the main functions of discriminant analysis is to provide a classification rule that correctly identifies a high proportion of cases. However, the usefulness of the discriminant function is not determined solely on the basis of its classification accuracy with the original data set (training set). Cross-validation is necessary to establish the validity of the discriminant function. When the investigator has a large number of cases available, the most direct procedure is to divide the data set in half, calculate the discriminant function with one half of the data, and validate it on the other half. This procedure can also be carried out on a new data set collected under the same general experimental paradigm. If, on the other hand, the investigator has an insufficient quantity of data to perform this procedure, there are several alternative techniques. One method, commonly called the "jackknife procedure," removes one case from the training set, computes the discriminant function, and then classifies the case that has been omitted. This procedure is repeated until a discriminant function has been calculated for each of the cases in the data set. Overall classification accuracy is determined by dividing the number of single cases misclassified by the total number of cases contained in the data set. Although the jackknife procedure provides a check on the efficiency of the discriminant function, it does not usually produce results that vary greatly from the original computation. Another cross-validation procedure, the "randomization test," is applied to the entire training set. In this instance, however, the cases in the training set are randomly assigned to two groups. A new discriminant function is then computed for these randomly assigned groups. This process is repeated several times, and a distribution of discriminant functions is compiled. The distribution provides an indication of the classification results that can be expected with random data, thereby providing the investigator with a basis against which to compare the performance of the original discriminant function.

The LSDA procedure outlined above assumes that the covariance across groups is equal and that noise or error in the data conforms to a normal distribution. In cases in which these assumptions are violated, LSDA will provide less than optimal group discrimination performance. A useful alternative in some of these cases is the quadratic discriminant analysis (QDA) technique. The QDA procedure is similar in function to the LSDA technique and has been used successfully in a number of studies (Aunon, McGillem, & O'Donnell, 1982; McGillem, Aunon, & O'Donnell, 1981; Sencaj, Aunon, & McGillem, 1979).

Applications of LSDA. The standardized weighting coefficients obtained in the discriminant analysis procedure can provide valuable information concerning the relative importance of the variables employed in the discriminant function. Examination of the weighting coefficients enables the investigator to assess the contribution of each variable in the discriminant function. Large weights, in either a positive or a negative direction, denote a substantial contribution of their respective variables to group differentiation. In psychophysiological experiments in which the investigator wishes to classify voltage × time functions into two or more groups, the magnitude of the weighting coefficients identifies those features or time points that best differentiate between groups. For example, Horst and Donchin (1980) found that the ERP time points that best differentiated between two pattern reversal conditions were within the region of the voltage × time functions that were predicted to change as a function of experimental manipulations. Furthermore, these time points were consistent with the components derived in a principal-components analysis (PCA) of the data (see pp. 198–202).

Discriminant analysis also provides a classification rule that best differentiates between the training groups. This classification rule can be applied to other data sets collected in similar paradigms. In this case, the investigator is interested in classifying new data according to probability of group membership.

Although the primary purpose of the discriminant function is the correct classification of the greatest possible proportion of cases, useful information may also be obtained from the misclassified cases. This point has been illustrated by several studies that have employed discriminant analysis to assess the group membership of single-trial ERPs. In one such study, subjects were asked to count covertly the total number of high-pitched tones from a Bernoulli series of high- and low-pitched tones. High tones occurred with a probability of .20, while low tones occurred with a probability of .80. The common finding in this general paradigm is that counted, low-probability events produce larger P300 components than un-

counted, high-probability events. Replicating this design, Squires and Donchin (1976) then employed discriminant analysis for the purpose of classifying each single-trial ERP as either a high- or low-probability event. The discriminant function was able to classify 81% of the single-trial ERPs correctly. An examination of the averages of correctly and incorrectly classified ERPs (see Figure 10-2) indicated that the misclassified events resembled the category into which they were classified more closely than they resembled their correct category. These results suggest that some rare stimuli evoked a response characteristic of frequent stimuli, and vice versa. That is, rather than erring in its classification of waveforms, discriminant analysis may have identified trials in which the subject erred in classifying stimuli. This example illustrates the heuristic value of the technique in revealing the fine structure of the subject's behavior, which can be obscured by cross-trial signal-averaging techniques.

Another example of the use of discriminant analysis in the detailed examination of subjects' behavior is found in tree diagrams of discriminant scores. Figure 10-3 depicts the discriminant scores obtained from ERP waveforms in a two-tone discrimination task (Squires *et al.*, 1976). The investigators calculated discriminant scores for fifth-order sequential stimulus patterns to demonstrate the effect of sequence on the amplitude of several components in the ERP waveform. As can be seen from the figure, the discriminant scores obtained in the experiment closely paralleled the sequential structure of the task. Thus, the discriminant tree diagram provides another means of analyzing the fine structure of subjects' behavior.

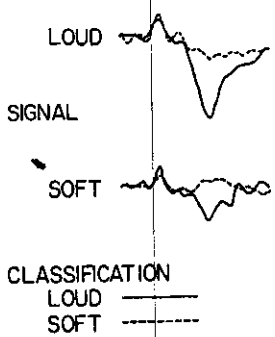

Figure 10-2. Average ERPs sorted by discriminant function classification and type of stimulus. The waveforms represent an average of 16 subjects. (From "Beyond Averaging: The Use of Discriminant Functions to Recognize Event-Related Potentials Elicited by Single Auditory Stimuli" by K. C. Squires and E. Donchin. *Electroencephalography and Clinical Neurophysiology*, 1976, 41, 449–459. Reprinted by permission.)

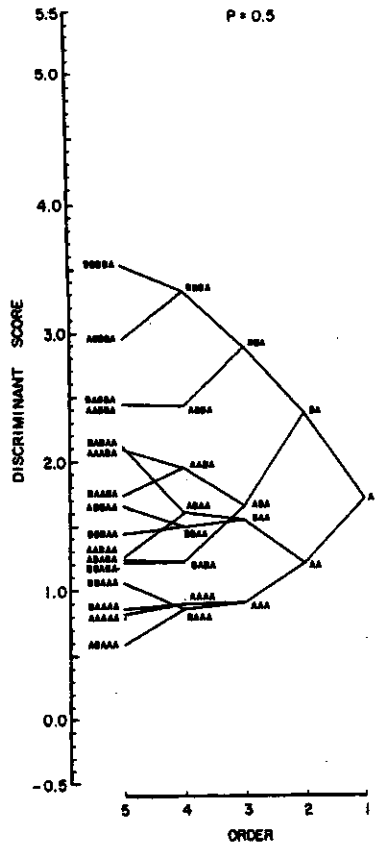

Figure 10-3. Tree diagram of discriminant scores calculated for ERPs elicited by high- and low-pitched tones. The discriminant scores are plotted as a function of stimulus sequence. (From "The Effect of Stimulus Sequence on the Waveform of the Cortical Event-Related Potential" by K. C. Squires, C. D. Wickens, N. C. Squires, and E. Donchin. *Science*, 1976, 193, 1142–1146. Copyright 1976 by the American Association for the Advancement of Science. Reprinted by permission.)

Discriminant analysis may also be employed to evaluate the degree of resemblance of a single trial to the average of one group or another. In the case of voltage × time functions collected in a psychophysiological experiment, the investigator may wish to know how well a single function resembles the average of one of several groups. This information is provided by the discriminant score.

Evaluation of Discriminant Analysis. As with other statistical techniques, there are both advantages and disadvantages associated with using discriminant analysis procedures in the evaluation of psychophysiological data. Discriminant analysis provides an objective, quantifiable method of assessing differences in single voltage × time functions, both within the training set and across other data sets collected under the same general experimental paradigm. In addition to providing classification information, discriminant analysis also provides an alternative to signal averaging in situations in which the amplitude and latency of the voltage × time components vary across trials. The weighting coefficients derived in the process of discriminant analysis provide information that can be interpreted in terms of the voltage × time components. Thus, components derived in the PCA procedure can be compared with the time points selected in the discriminant analysis procedure to give the investigator an indication of the important features in the data set. Although the initial calculation of the discriminant function is computationally costly, its application is relatively simple. In most cases, it requires only the multiplication and summation of a few variables × weighting coefficients.

There are also several disadvantages to discriminant analysis. The need for an independent basis for grouping voltage × time functions can be problematic in some cases, particularly during exploratory data analysis, in which hypotheses are weak or nonspecific. A second problem is that a useful discriminant function can be calculated only if the groups differ significantly. Finally, the need for cross-validation of the discriminant function imposes additional requirements on the investigator. It is preferable that sufficient data be collected so that the discriminant function can be computed with one set of data and validated on another.

As with other analytic techniques discussed in the present chapter, discriminant analysis cannot be profitably employed without consideration of its limitations and assumptions. However, correct application of the discriminant analysis procedure can produce valuable information for the psychophysiologist.

Digital Filtering

Digital filters have been little used explicitly in psychophysiology. They constitute an interesting contrast with analog filters (see pp. 186–187). A digital filter is most easily described by example. Conceptually, perhaps the simplest digital filter consists of replacing each value in a time series with the average of that number, the number preceding it, and the number following it. Such a common smoothing operation is a rudimentary low-pass filter, in that high-frequency components are reduced.

Specific digital filters vary along several dimensions, which determine the bandpass characteristics and computational speed of each filter. In the example above, three weights are used, each having a value of ⅓. Somewhat less smoothing is accomplished if a different set of weights is used: ¼, ½, ¼. Alternatively, smoothing is also altered if the number of weights (the "window width") is changed to 5, each weight perhaps being ⅕. If the number and values of the weights are held constant but the time interval between data points is changed, the filter will again have different characteristics. A final choice is whether to apply the weights recursively—that is, after applying the filter at point T, does the filter applied to point $T + 1$ employ the unfiltered T (nonrecursive) or the filtered T (recursive) in computing the filtered $T + 1$?

Clearly, psychophysiologists routinely manipulate their data algebraically in ways that constitute digital filtering. Even the computation of a mean of n values can be seen as (1) assigning each value a weight of $1/n$, (2) applying the filter to the midpoint value in the time series by summing the weighted values, and (3) discarding all but the "filtered" midpoint value in the time series. What are typically not discussed when simple digital filters are used are the bandpass characteristics of the filter procedure. Ruchkin and Glaser (1978) describe simple digital filters and present their characteristics. More generally, Cook (1981) has developed a Fortran program, based on the methods of Ackroyd (1973), which determines the optimal values for a set of weights for a nonrecursive filter, given sampling interval, bandwidth, and number of weights desired. Glaser and Ruchkin (1976) present a mathematical discussion of digital filters oriented to the psychophysiologist.

A more elaborate digital method for filtering voltage × time function is Wiener filtering (Walter, 1968; Wiener, 1964). Naitoh and Sunderman (1978) outline the application of this method to ERP data. As they describe it, an estimate of the frequency characteristics of background noise is made from a comparison of the spectra of the average ERP with the average of the spectra of single-trial ERPs, the spectra being obtained via Fourier analysis. This noise estimate is then used to correct the single-trial spectra. Finally, the original ERPs are regenerated via inverse Fourier transforms of the corrected spectra. Naitoh and Sunderman (1978) review evidence that Wiener filtering does not adequately preserve high-frequency information. Furthermore, they suggest that as a technique for general use the slight improvement in signal-to-noise ratio is not worth the trouble (see also Carlton & Katz, 1980; Ungar & Basar, 1976). However, they describe special circumstances for which it might be very appropriate.

Other than for simple smoothing, digital filters are perhaps most commonly employed prior to a pattern recognition procedure such as the Woody filter (see pp. 191–192). However, they are potentially appropriate for any voltage × time function. They deserve serious consideration in the laboratory, particularly given the continually decreasing cost of additional computation.

Data Reduction Techniques

Introduction

Although the headings "Signal Extraction Techniques" and "Data Reduction Techniques" serve to illustrate the fact that these are distinct processes to which psychophysiological signals are subjected, they are not mutually exclusive. One technique included under the heading of "Signal Extraction Techniques" that would also fit under the present heading is LSDA. Discriminant analysis techniques serve both to provide a method of signal extraction and pattern recognition and, at the same time, to reduce the magnitude of the data set to a much smaller subset of variables. Another technique, PCA, which is discussed under the present heading, could have been included under "Signal Extraction Techniques." As with discriminant analysis, the PCA procedure serves to reduce the size of the data base from numerous dimensions to a relatively few "components." In addition, the PCA technique does not require the restrictive, *a priori* assumptions of group membership that characterize the discriminant analysis procedure. Thus, we do not wish to assert that any of the techniques illustrated in this chapter fit into a single category, but instead that there are distinct stages in the process of data analysis.

A major problem in the analysis and interpretation of psychophysiological data is the determination of the specific criteria by which a signal is defined. For example, if one averages single-trial data, one makes certain assumptions about the signal and noise distributions that underlie the data: That portion of the voltage × time function that is temporally invariant over repeated presentations of a stimulus is defined as the "signal," while other, randomly varying portions of the epoch that are reduced as a result of averaging are defined as the "noise." Even if we adopt the signal–noise model implied by the averaging procedure, the problem of determining the important features of this "signal" remain. One commonly employed procedure for subdividing the average signal is to define its features on the basis of their relationship to the experimentally induced variance. In this case, the important features of the signal become identical with the components of variance in the data set. This type of definition of features or components of the voltage × time function requires not only the proper use of signal extraction and data reduction techniques, but also the exercise of tight experimental control. Since components of the signal are defined in terms of their relationship to experimentally manipulated variance, poor experimental design can lead to spurious components. Thus, another point to be emphasized is that the proper use of methods of analysis can provide the investigator with useful information only within the framework of good experimental design.

As mentioned above, the signal extracted from the raw or average voltage × time function is typically subdivided into features or components that are related to the experimentally induced variance. Each of these derived components can be thought of as a linear combination of weighting coefficients × time points. The problem with this approach, however, lies in the fact that there are an infinite number of possible linear representations for a vector of voltage × time values. Therefore, criteria must be adopted to aid in the selection of a subset of possible linear combinations. The determination of these weighting coefficients and their application to the voltage × time signal constitute the primary topic of the remainder of this section.

Peak Measurement

The identification of a peak in a voltage × time vector is perhaps one of the oldest measurement procedures in psychophysiology. The procedure is relatively simple, and it provides both amplitude and latency information. Although there are several methods of defining the peak of a component, they all involve a simple linear combination rule that assumes a weighting coefficient for each time point. This rule typically involves setting all of the weighting coefficients to zero, except for the one weighting coefficient $a(x)$ that corresponds to the time point $t(x)$ at which either the largest or smallest voltage is observed within a prespecified temporal window. This coefficient is set to one. Thus, in the case of peak measurement, the component derived from the voltage × time vector is defined as a single point. The principal advantages of this measurement procedure are its intuitive appeal and computational simplicity. Peak measurement algorithms represent a direct analogue of the visual inspection of voltage × time data, with the added advantage of an easily standardized selection procedure.

A few representative procedures for peak measurement are presented here. The identification of "peaks" (single-point events) as zero crossings along the voltage × time function was proposed in the mid-1960s (Ertl, 1965; Ertl & Schafer, 1969). The method was suggested for peak identification in average ERPs and provides a reliable means for determining latency information. However, amplitude information is not available, since the peak has been defined as the zero point. Another method of peak identification that has been widely employed involves selecting either the largest or smallest voltage within a prespecified temporal window and defining it as the peak. Amplitude information can be obtained from a base-to-peak difference, with the baseline usually being defined as some relatively inactive portion of the voltage × time function, such as that for some period prior to stimulus presentation. Alternatively, a peak-to-peak difference can be derived. In both cases, latency information is provided by the time point $t(x)$ at which the largest or smallest voltage is obtained. The peak or peaks of the voltage × time function can also be defined in terms of the intersection of the tangents of their positive and negative slopes. Amplitude and latency information is also provided by this method.

All of the peak measurement techniques outlined above provide latency information, and, with the exception of the zero-crossing technique, all give amplitude information. Note that each of the procedures makes the assumption that the psychophysiological component of interest can be defined as a single point in the voltage × time vector. When measuring a well-delineated peak with a large signal-to-noise ratio (either the single trial or the average), this assumption would appear to be appropriate. Examples of psychophysiological signals that would meet these criteria include the cardiac R wave, the skin conductance response, and the systolic and diastolic peaks in the blood pressure cycle. However, even peaks that normally are sharply defined can easily become obscured by nonsystematic variance, producing spurious measurements. Another disadvantage of defining a component in terms of a single point is the loss of information concerning the morphology of the voltage × time function. This information, which may be of benefit to the investigator, is discarded prior to analysis of the peak measurement. In effect, all information other than a single point in the voltage × time function is defined as noise in the peak measurement procedure.

Other signal extraction and data reduction techniques also make assumptions about the nature of the signal and noise distributions. However, a subset of these provide information that is similar to that given by the peak measurement techniques, while also retaining some morphological information. For example, the polarity histogram is one measurement technique that provides amplitude information in the form of probability instead of voltage (Callaway & Halliday, 1973; Kubayashi & Yaguchi, 1981). The procedure is performed by incrementing a frequency count whenever an individual time point in the voltage × time function is above or below a zero baseline. A component is then defined whenever the time × probability histogram exceeds some criterion value. The advantages of the technique include its computational simplicity and relative insensitivity to random fluctuations in the voltage × time signal. Some morphological information is also retained in the form of probability values.

Another procedure that provides amplitude as well as morphological information (symmetry and peakedness) has been proposed by Callaway, Halliday, and Herning (1983). In this procedure, called PEAK, a grand average template is computed. The important features (peaks and troughs) of the template are defined by means of a standard algorithm. Lagged correlations are then computed between the template and the individual voltage × time vectors. Components in the voltage × time vectors are defined as the maximum lagged correlations between the template and the individual vectors. A series of measurements are then made on the features, such as amplitude, latency, peakedness, and symmetry. Other component measurement techniques, such as area measurement and PCA, that also provide alternatives to traditional peak measurement procedures are discussed below.

Another disadvantage of the peak measurement methodology is the difficulty encountered in defining the peak of a relatively slow component. Can a single time point accurately represent a slow component—and, even if it could, which point would be selected? Several psychophysiological signals would qualify as slow components (e.g., respiration, skin conductance response, contingent negative variation [CNV]). Such techniques as area measurement and PCA may provide a more appropriate representation of these components.

In addition to the limitations mentioned above, peak measurement techniques also fail to provide information concerning component overlap. The measurement of a single point does not permit the assessment of the actual number of temporally overlapping components that may jointly be responsible for the voltage recorded at the specific time point. Several examples of this particular problem have been addressed in the ERP literature (Donchin, Tueting, Ritter, Kutas, & Heffley, 1975; Squires, Donchin, Herning, & McCarthy, 1977). While carefully designed factorial experiments can alleviate this problem to some degree, a better solution lies in the application of a procedure that will permit a direct evaluation of the overlapping components.

A final problem concerns independence among peaks when several peaks are measured in the voltage × time function. This is particularly important if statistical inference techniques are to be applied to the data, since most of these techniques assume independence among measures.

In summary, peak measurement procedures must be applied with caution when they are used to define a psychophysiological component. It must be realized that data reduction may result in the loss or distortion

of relevant information. However, the techniques outlined above can provide useful information in situations in which the psychophysiological signals are relatively fast, are well delineated, and possess a high signal-to-noise ratio.

Area Measurement

Like peak measurement, the measurement of the area of a psychophysiological component can also be conceptualized in terms of a linear combination of n time points. In this case, however, the weighting coefficients that correspond to the temporal region of the component are set to $1/n$, while the rest of the weighting coefficients are set to zero. Thus, unlike the peak measurement procedures, area measurement defines the component of interest in terms of a range of contiguous time points. These points are then integrated relative to a baseline to produce the area measurement of the component. The assumption underlying the use of area measurement is that the psychophysiological component is most accurately represented by the area of some specific epoch along the voltage × time function. This appears most reasonable in the case of slow components, such as the skin conductance response, respiration, and CNV.

The measurement of the area or amplitude of a component is performed relative to some baseline. In most cases the baseline is defined as that portion of the voltage × time vector that occurs prior to stimulus presentation. It is assumed that the baseline represents an inactive portion of the vector. However, this is not always the case. In some situations, anticipatory activity is present (e.g., CNV). In this case another method of defining an inactive baseline is required. One such method is the use of "trimmed" averages, which are relatively insensitive to extreme deviations in the data (see Donchin & Heffley, 1978).

Like peak identification, area measurement also possesses a good deal of face validity, since many psychophysiological signals extend over more than a few time points. The need for elaborate computational algorithms is also minimized by area measurement. Furthermore, area measurements are less susceptible to modest amounts of latency jitter in the component, as well as less sensitive to random amplitude variations in a few time points, than is peak measurement. The degree of insensitivity to random fluctuations is a function of both the number of points included in the area and the temporal range of the latency variability.

Although area measurement presents a distinct advantage over peak identification in some cases, it still fails to deal adequately with several measurement issues. The determination of integration limits is often difficult and/or arbitrary, due to the poor resolution of component limits in the raw or average voltage × time function. The issue of the establishment of reliable integration limits becomes less of a problem with components that are easily recognized. The issue of component overlap is also not addressed by the area measurement procedures: It is difficult to assess the relative contribution of overlapping components to the voltage measured at either one or several time points. As has been mentioned above, one way to lessen this problem is to control the experimental variables that are known to affect the amplitude and latency of the overlapping components. Finally, as with peak measurement, area measurement techniques may fail to provide the investigator with a clear, detailed picture of the morphology of the voltage × time function.

In summary, although area measurement procedures alleviate some of the problems encountered with peak identification techniques, there still remain unresolved issues. Area measurement would appear to be most appropriate when nonoverlapping, slow components are evaluated.

PCA

INTRODUCTION

Unlike discriminant analysis, PCA does not require that the subclasses be known *a priori*. Thus PCA makes less restrictive assumptions about the number of relevant categories into which the data will be subdivided. This is particularly useful to the investigator when the nature and number of subclasses are unknown prior to the analysis. In addition to the pattern recognition information garnered from PCA, the technique also provides a means by which a huge data base is reduced to a few components that most parsimoniously describe the experimental variance. Although the PCA procedure has been employed most frequently in the analysis of ERP data, it is clearly relevant to the analysis of other psychophysiological signals.

Like peak and area measurement techniques, PCA can also be conceptualized in terms of a linear combination of time points. To reiterate, the peak measurement procedure defines the psychophysiological component as a single time point in the voltage × time function. The other time points are discarded prior to analysis. In the case of the area measurement, the psychophysiological component is defined as the integration of equally weighted values at several time points. Area measurement represents a distinct improvement over peak measurement procedures in the assessment of slow components. However, neither procedure addresses the issues of the selection of optimal weighting coefficients or the effects of component overlap on the observed voltages.

Unlike the peak and area measurement techniques, the PCA procedure employs the complete voltage × time data matrix to determine the weighting coefficients. In the present case, we describe the R-PCA procedure, which involves the computation of a time point × time point input matrix. Other investigators (John, Ruchkin, & Villegas, 1964; John, Ruchkin, & Vidal, 1978) have suggested the usefulness of the Q-PCA procedure, which involves the computation of a waveform × waveform input matrix. In the former case, the interest is in the relationship among time points across the voltage × time function. In the latter case, the analysis provides information concerning the relationship among individual waveforms in the data matrix. Although the present discussion is concerned with the R-PCA procedure, its general points also apply to the Q-PCA technique.

In terms of providing optimal weighting coefficients for the determination of components, PCA is clearly preferable to the methods employed in peak and area measurement. In the case of the PCA procedure, the weighting coefficients (component loadings) represent the contribution of the derived component to the variance at each time point in the voltage × time function. Another advantage of the component extraction procedure employed in PCA is that the weighting coefficients associated with each component are uncorrelated with the weighting coefficients associated with each of the other components. Thus, the component scores computed from the linear combination of time points × weighting coefficients are orthogonal. Therefore, in contrast to peak and area measurements, PCA permits the investigator to assess the independent effects of the experimental manipulations on temporally overlapping components (Donchin *et al.*, 1975; Glaser & Ruchkin, 1976).

As with the peak and area measurement procedures, the method of determining the weighting coefficients in the PCA procedure implies a particular definition of the psychophysiological component. The PCA procedure defines a component in terms of the covariation between time points in the voltage × time function. A pattern of high covariation among time points implies that a specific component (source of variance) can be assumed to be influencing them jointly. These derived components are represented in terms of the variance in the data. The component score produced by the linear combination of the time points × weighting coefficients provides a measure of the magnitude of a specific component in a specific voltage × time function. Thus, for each PCA component a separate weighting coefficient is obtained for each of the time points, and a separate component score is derived for each voltage × time function in the data matrix. An example of a component loading plot is presented in Figure 10-4. This figure displays four sets of component loadings and the grand mean waveform from an ERP experiment. There are 128 component loadings, which correspond to the 128 time points in the waveform. A separate set of loadings is calculated for each of the four components.

Several assumptions underlie the PCA model. It is a linear model, and thus assumes that the derived components simply sum together to produce the voltage × time function without interaction. A second assumption concerns the sources of variability in the data. It is assumed that the sources of variance in the data are orthogonal. Although there is no foolproof method of assuring that this assumption is met, good experimental design (in terms of the factorial manipulation of experimental variables that are believed to influence the major sources of variance) is one way to minimize intercomponent correlation (Donchin & Heffley, 1978). Techniques are also available for testing the assumption of orthogonality (Harman, 1967). In cases in which two or more sources of variance are highly correlated across voltage × time functions, PCA will yield a set of weighting coefficients and a single component score, which represent a composite of these correlated components. The interpretation of this composite component in terms of the voltage × time data set will be misleading (Roessler & Manzey, 1981; Wastell, 1981a). A third assumption of the PCA model concerns the domain of component variability. PCA can reliably and efficiently handle variability in the amplitude of the component. On the other hand, variability in the latency of the component over voltage × time functions can cause substantial problems in the interpretation of the derived components. The PCA procedure does not discriminate between variance in the data that is due to variations in amplitude of the underlying component and variance that is due to variations in latency of the underlying component. Therefore, if both the amplitude and latency of a component are changing over trials, PCA will not be able to distinguish the two dimensions. In the case of latency variability, some attempt needs to be made to decrease the variability over trials prior to the use of the PCA technique (Picton & Stuss, 1980). One such procedure, which is described in the present chapter (see pp. 191–192), is the adaptive filter for the analysis of variable-latency neuroelectric signals (Nahvi *et al.*, 1975; Woody, 1967;

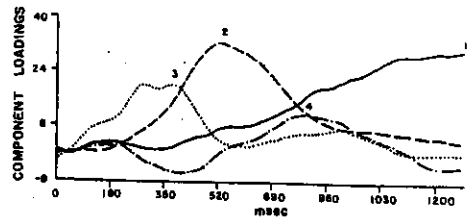

Figure 10-4. Plot of four sets of component loadings derived from a principal-components analysis (PCA) of an ERP data set. Each of the component loading vectors is composed of 128 points corresponding to 128 time points (100-Hz digitizing rate) in the waveforms.

Woody & Nahvi, 1973). Callaway *et al.* (1983) demonstrate the improvement in PCA results that latency correction can provide.

APPROPRIATE EXPERIMENTAL DESIGN

The assumptions of the PCA model that are outlined above, and the ease with which they can be violated, suggest that PCA cannot be blindly employed in the analysis of psychophysiological data. The exercise of good experimental design, as well as sensitivity to the assumptions of the PCA model, is of paramount importance if the technique is to provide valid information. The PCA technique represents a multistep procedure for the analysis and interpretation of psychophysiological data. Each step in the procedure requires forethought about the assumptions of the model and the design of the experiment. The initial step, and perhaps the most important, concerns the design of the experiment. There are several issues that must be considered prior to the design of an experiment destined for PCA.

The first issue concerns the second assumption of the PCA model mentioned above—that the major sources of variance are orthogonal. One method to minimize intercomponent correlation is the factorial manipulation of the major sources of variance. A second issue to be considered during the design of the experiment is that the number of cases (typically, subjects × conditions) should exceed by a factor of 10 or more the number of variables (typically, number of time points) in the voltage × time function (Picton & Stuss, 1980). As the number of cases decreases relative to the number of variables, the stability of the component structure will also decrease. In terms of a practical example, this means that a voltage × time vector with 60 time points (variables) would require 600 separate cases to insure stability. A third issue to be considered during the design of the experiment concerns the requirement of the PCA model that the number of variables be of sufficient quantity to determine a stable component structure. For many psychophysiological data sets, this mathematical precondition is usually not a problem, since a large number of variables produce relatively high loadings on each component. However, if underdetermination of the component structure is suspected (too few variables having high loadings on a specific component), there are several techniques that permit the investigator to assess the resulting instability (Mulaik, 1972; Thurstone, 1935; Tucker, 1973).

SELECTION AND COMPUTATION OF THE INPUT MATRIX

Once the investigator has designed the experiment and collected the data, the next step is to decide on the type of input matrix to be employed in the PCA solution. This selection has important implications for the interpretation of the resulting component structure. The input matrices that are accepted by most PCA programs include mean cross-products, covariance, and correlation matrices.

Calculation of the mean cross-products matrix involves summing the products of cross-multiplication of the voltage values for all time points with all other time points. Note that in this case all of the experimental variance is analyzed, since neither the mean nor the variance at each time point is removed from the data set before analysis. The fact that the mean is not subtracted from the cross-products matrix has certain implications for the component structure. To begin with, the loadings of the first derived component usually duplicate the grand average voltage × time function. Second, large base-to-peak deflections in the voltage × time function will produce components even when they are not influenced by the experimental manipulations (Donchin & Heffley, 1978). The use of the cross-products matrix would appear to be most appropriate when the investigator wishes to retain information about the absolute variations in amplitude, as well as the polarity of the corresponding component in the raw data (see Ruchkin, Sutton, & Stega, 1980; Squires *et al.*, 1977).

The calculation of the covariance matrix is similar to that of the cross-products matrix, with the exception that the grand average voltage × time vector is subtracted from the individual voltage × time functions prior to the computation of the cross-products. Thus, the portion of variance that is contributed by the differences between the variable (time point) means is removed in the process of calculating the covariance matrix. In terms of the component structure that will be derived from the covariance matrix, the important issue will be the degree to which the individual voltage × time functions differ from the grand average, not the absolute amplitude or polarity, as is the case for the cross-products matrix. Thus, component scores will reveal relative rather than absolute differences in the component. The covariance matrix has been used most frequently in the analysis of ERPs, since the differences among ERPs relative to the grand mean waveform are usually of primary importance (see Isreal, Chesney, Wickens, & Donchin, 1980; Ruchkin *et al.*, 1980).

The correlation matrix is another option in the selection of the input matrices for the extraction of principal components. The calculation of the correlation matrix requires that the mean of each variable be subtracted (as in covariance), and, additionally, that the difference be divided by the variable's standard deviation. Thus, in the case of the correlation matrix, the variance attributed to the differences between the time point means, as well as the variance due to differences in time point variability, is removed during the process of calculation of the matrix. Essentially, each time point value is converted prior to PCA to a standard *z* score, based on that point's mean and variance across voltage × time functions. The components extracted from the correlation matrix will be similar to those derived from the covariance matrix, with the exception that the loadings will be more uniform across the length of the component, due to the standardization of the variables (Donchin & Heffley, 1978). Thus, in the case of the correlation matrix, the loadings will not reflect the component morphology as well as when the covariance matrix is employed. This standardization also serves to obscure the magnitude of differences in variance across time points. This may result in the assignment of relatively high loadings to time points at which differences are small.

As can be seen from the preceding discussion, the choice of an input matrix for the PCA procedure constrains the conclusions that can be drawn from the derived component structure. Thus, the investigator must take a careful look at the specific questions that are to be addressed with the PCA, prior to the selection of the input matrix.

EXTRACTION OF PRINCIPAL COMPONENTS

The third step in the PCA process involves the extraction of the weighting coefficients to be used in the linear combination of time points. The extraction procedure, consisting of a sequence of standard matrix manipulations normally performed by packaged statistical software, produces one vector of weighting coefficients for each of the derived components. A separate weighting coefficient is derived for each of the time points in the voltage × time function. Thus, if six components are extracted from a series of voltage × time functions, each composed of 60 time points, there would be six sets of 60 weighting coefficients derived in the PCA procedure. As has been mentioned above, a vector of weighting coefficients represents the contribution of the derived component to the variance at each time point in the voltage × time function. The weighting coefficients associated with each component are uncorrelated with the weighting coefficients associated with each of the other components.

The orthogonality of the components produced by the PCA technique represents a distinct advantage over the peak and area measurement procedures in terms of later inference testing. Univariate ANOVAs can be performed on the component scores for each of the components. On the other hand, computation of separate ANOVAs for each peak or area measurement is of doubtful validity, due to the possible correlation between measures in different parts of the voltage × time function.

The first component extracted in the PCA accounts for the largest proportion of systematic variance in the data matrix. The second derived component accounts for the largest possible percentage of residual variance and is orthogonal to the first component. This process of component extraction continues until all possible components have been derived.

It must be noted that the components derived via PCA need not reflect the physiological generators underlying the recorded voltage changes in a one-to-one fashion. Instead, the components represent merely one summary of the systematic variance present in the data. Theoretical inferences and converging measurement operations are required to verify the relationship of PCA components and physiological components.

One of the goals of the PCA technique is the reduction of the data base to a subset of meaningful components. That is, the hope is that a few orthogonal dimensions (components) will be able to account for most of the variability in the raw data, or that most of the information in the raw data can be more simply represented. Intuitively, this is possible to the extent that the original observation time points are redundant. Determination of the number of components to retain is usually based on such criteria as the amount of variance accounted for and the parsimony of interpretation of the component structure. Several statistical methods have been suggested to assess the number of components to retain (Cattell, 1966; Humphreys & Montanelli, 1975; Kaiser, 1960; Montanelli & Humphreys, 1976; Tucker, 1973).

One point that is specifically relevant to component extraction with psychophysiological data concerns the temporal range of the components in the voltage × time function (Wastell, 1981b). The PCA procedure initially selects components associated with relatively slowly varying regions of the voltage × time function, since these components typically encompass a large amount of the variance. Somewhat faster components, such as the P300, are then selected. Components that extend over a relatively limited temporal range will be extracted much later in the PCA procedure. Therefore, by virtue of the component extraction procedure employed in PCA, some fast components will not constitute a sufficient amount of variance to produce a component that will meet the selection criteria. This point is especially important if the voltage × time functions consist of both slowly and quickly varying components.

ROTATION OF COMPONENT LOADINGS

Once the desired number of components has been extracted from the input matrix, the next step usually involves trying to simplify the component structure. In most cases the component loadings for each derived component vary across the entire voltage × time function, because of a nonzero correlation among time points. The purpose of the rotation procedure is to simplify the pattern of loadings so as to localize each component to a portion of the voltage × time function.

The varimax rotation procedure has been frequently used with ERP data and provides one method by which the interpretability of the component struc-

re can be enhanced. The procedure retains an or-
ogonal component space while maximizing the var-
ce of the component loadings by attempting to
ve the high loadings to unity and the low loadings
zero. Thus, the varimax rotation maximizes the
ociation between each component and a few time
ints and minimizes the association at all other time
ints for each component. The rotation redistributes
e component variance among the time points, but
es not alter the goodness of fit of the component
odel. Note that the PCA extraction procedure pro-
des the component structure while the rotation tem-
orally localizes the components, thereby permitting
e evaluation of the components in terms of the
iginal voltage × time functions. In terms of psycho-
ysiological data, the varimax procedure emphasizes
e peak of the signal and is therefore analogous to a
se-to-peak measurement.

Following the completion of the rotation proce-
ure, the next step is to compute the linear combina-
on of time points × weighting coefficients for each
mponent. This transformation will produce a sepa-
te component score for each voltage × time func-
on in the input matrix. The component score repre-
nts a measure of the magnitude of a specific
mponent in a specific voltage × time function.

The majority of psychometric studies that employ
e PCA technique terminate prior to the calculation
f component scores. In many cases, investigators are
nly interested in the association between the princi-
al components and the observed data. Component
adings are sufficient to provide this information.
he psychophysiologist, on the other hand, is also
ncerned with the effect of experimental manipula-
ons on the components derived from the voltage ×
me data. In this case, the component scores as well as
e loadings are of interest. Calculation of the compo-
nt scores permits the investigator to locate the ob-
rved voltage × time functions in a simpler and
esumably more meaningful component space. Dif-
rences among the component scores reflect the ef-
ct of experimental manipulations on the principal
mponents and may be subjected to inference-testing
ocedures (see "Inference Testing," below).

mmary Comparison of Data Reduction Techniques

t this point it is appropriate to summarize the advan-
ges and disadvantages of PCA in the analysis of
sychophysiological data, relative to simpler data re-
uction techniques. To begin with, PCA provides an
bjective and statistically based method for identify-
g and computing linear combinations of time points
weighting coefficients. This serves to reduce experi-
enter bias in the selection and definition of the
sychophysiological components. Another advantage
the PCA procedure concerns the method of calcu-
ting the weighting coefficients. In the peak and area
easurement procedures, weighting coefficients are

set to either zero or one. The PCA technique permits the assignment of graded weighting coefficients on the basis of the contribution of the derived component to the variance at each time point in the voltage × time function. Thus, the entire data set is employed in the calculation of component scores, rather than a few time points. This serves to increase the sensitivity of the experimental procedures, as it attenuates the effects of noise and sampling fluctuations on the components. Furthermore, unlike the peak and area measurement techniques, PCA provides information about both the amplitude variability and the morphology of the voltage × time functions. Amplitude information is available in the form of component scores. The morphological characteristics of the component are provided by the weighting coefficients. PCA also gives the investigator information about the degree of component overlap, provided that the underlying components are not highly correlated. Since the components derived from the PCA are orthogonal, univariate tests of significance may be appropriately applied to the component scores. Finally, PCA provides an efficient summary of a very large data base by providing a simpler and therefore more readily interpretable data structure.

Although PCA presents numerous advantages over some traditionally employed psychophysiological analysis techniques, it has some limitations that should be mentioned. For example, the PCA model assumes that the components embedded in the voltage × time function are temporally invariant over trials. In cases in which this assumption is not met, PCA confounds the amplitude and latency variability of the components and provides a component structure that is difficult to interpret. There are, however, several techniques that can be employed as preprocessors to reduce the latency variability prior to employing the PCA technique (e.g., the Woody filter or other autocorrelation measures). The transformation process employed in the PCA is certainly not as intuitively clear as that used in peak or area measurements. This may sometimes lead to confusion when raw voltage × time functions are compared with the reduced component structure. Another point to consider is that components in the voltage × time functions that span a relatively few time points may not constitute sufficient variance to meet the component selection criteria prior to rotation. Finally, since PCA does in fact employ the entire time point × time point data matrix, substantial computing power is required to carry out the transformations.

Spatial Analysis

Introduction

Although some psychophysiological signals can be treated as reflecting the activity of a single structure

(as in the case of heart rate), in other cases the signal reflects the activity of what are functionally multiple generators (e.g., EEG). Furthermore, the signal produced by these generators, propagated through space to the body surface, can vary as a function of the spatial characteristics of the generators and the conductivity characteristics of the structures interposed between the operators and the skin. As a result, the signal recorded at the surface will depend on the location of the electrode or other transducer.

In some cases the variability due to electrode location is not of interest to the psychophysiologist but constitutes merely a source of error to be eliminated. For example, ECG morphology depends greatly on electrode location. However, the psychophysiologist might only be interested in interbeat interval. Thus, variation in the morphology of the ECG waveform with electrode position can be ignored. Of course, when variation due to location is ignored in this way, the psychophysiologist is assuming that a single "channel" or "generator" is of interest and that the variability observed at different locations on the body surface is irrelevant. This model is more often adopted for measures of autonomic activity than for EEG or EMG. However, by ignoring the spatial distribution over the body surface of the psychophysiological signal, we may miss a relevant part of the information provided by the signal.

Although there are serious problems in making inferences about location of the ERP component generators from the scalp distribution, measures derived from multielectrode recordings can still be very useful as an empirical method for defining components (see Donchin, 1978). In fact, if a component recorded at the scalp represents the sum of many fields generated by the activity of neurons functionally linked together (although not necessarily localized in a specific brain structure), the scalp distribution of a component will reflect its spatial properties. If we accept this basic model, and if we use scalp topographical information merely to infer functional, not physical, generators, the actual relationship between anatomical generators and their scalp manifestations need not be known. To this end, it is only necessary to record from those locations that allow us to discriminate among functional systems.

The remainder of this section is concerned with a brief description of some procedures devised to study the spatial distribution of psychophysiological measures. Although these procedures have been devised for analyzing brain ERPs, they can be applied to any other measure that can be recorded simultaneously from multiple locations.

Isopotential Maps

Isopotential maps are one way of expressing the values of a psychophysiological variable at different locations on the body surface. They involve recording at a large number of locations in order to obtain an accurate description of the similarities in voltage between different points of the body surface at a particular time point. Isopotential maps have been most frequently used for the EEG.

In an isopotential map (e.g., Ragot & Remond, 1978), the body surface is schematically represented on paper in the same way that terrain is represented in topographical maps. Voltage values observed at any location on the body are presented at the corresponding points of the map. Values of the intervening points are extrapolated by means of algorithms that typically rely on values at adjacent points. Points with equal values are then connected by lines, and a convention is adopted to distinguish positive and negative values.

Isopotential maps constitute only a graphical representation of the data and do not therefore imply any particular assumptions (beyond those concerning interpolation). However, they do not simplify the structure of the data, and therefore they do not qualify as signal extraction techniques. Rather, they are a preliminary tool for investigating the spatial distribution of the psychophysiological variable, where no assumptions about signal and noise are made.

A particular kind of isopotential map is the spatiotemporal map (Remond, 1962). In this map, one of the axes is given by time. Therefore, the spatial information is restricted to a line, but information about the variation over time of the spatial distribution is included. As indicated above, this kind of map is more a data description technique than a signal extraction procedure. The problem of defining the signal remains unsolved.

Another kind of spatial map is the significant probability map (Duffy, Bartels, & Burchfiel, 1981). This kind of map plots z or t statistics obtained by the comparison of pairs of values from two data sets. Maps for different time points are compared. A signal is defined as those aspects of the distribution which differentiate significantly between two sets of data. Note that this kind of definition yields a signal that is specific to the data sets used, and comparisons between data obtained in different experiments are problematic.

Univariate and Multivariate Approaches to Spatial Analysis

An isopotential map is essentially a graphical way of representing the information obtained with a multiple-electrode recording. Because it does not make any distinction between signal and noise, it does not qualify as a signal extraction technique. However, signal extraction from an isopotential map can be accomplished in at least two ways. A peak detection algorithm (see pp. 196–198) can define a signal. Alternatively, the signal may be defined on the basis of a pattern in point-by-point t tests between what are understood to be signal-present and signal-absent con-

ditions (see the discussion of significant probability mapping, above).

However, the use of typical univariate techniques to test inferences about data from multiple-electrode recordings is unsatisfactory for two reasons. First, the large number of resulting significance tests greatly inflates experiment-wise error rate. Standard adjustment of the alpha level is likely to undercorrect for this problem, because error variance is likely to be correlated across recording sites. Second, univariate analysis provides little information about effects or patterning at different sites. A further limitation of standard methods of signal extraction and inference testing with isopotential maps is the inability to distinguish between overlapping sources of activity at each time point.

The remainder of this section is concerned with a description of vector analysis (VA), a multivariate approach to representing the spatial distribution of a psychophysiological variable. The procedure is both powerful (in that all the available information is used) and efficient (in that signal and noise are clearly distinguished).

Multivariate Approach to Spatial Analysis

Vector Analysis (VA) is a multivariate procedure proposed by Gratton, Coles, and Donchin (1985) to quantify information about the spatial distribution of a psychophysiological variable. "Spatial distribution" is here defined as the polarity and relative amount of activity observed at any number of electrode sites, independent of the absolute size of this activity. VA estimates the portion of the activity recorded at several different electrode locations that can be attributed to one or more components, defined in terms of spatial distribution. Therefore, VA defines the signal as one or more components characterized by a specific spatial distribution, and the noise as the remaining variance.

VA treats the voltage values of the electrode locations at a given time point as being the elements of a vector (the data vector). Thus, there is one vector for each time point, and within each vector there is a value for each recording site. This voltage × electrode arrangement contrasts with the usual voltage × time representation. The data vector can be represented geometrically in a space (the vector space) having one dimension for each recording site. Any vector may be characterized by its length and its orientation when plotted in the Euclidean space defined by the dimensions.

VA uses a specific multivariate approach to data reduction, such that a univariate approach to inference testing may be employed. The length of the data vector is a measure of the total activity recorded at all the electrode locations, independent of their relative value or sign. The orientation of the data vector relative to the dimension axes (recording sites) is determined by the relative amplitude and polarity at the different electrode sites, independent of the total activity recorded. Therefore, when a polar notation is adopted to describe the data vector, the information concerning the spatial distribution at each time point can be isolated and expressed by a series of angles between the vector and arbitrary reference axes. Figure 10-5 shows an example of this notation.

This approach yields two important benefits: Information about the spatial distribution at any given time point can be quantified, and any imaginable spatial distribution can be represented by an orientation in the vector space. In other words, a combination of angles (with the reference axes) in the vector space defines a given spatial distribution.

It is therefore possible to measure the degree to which the spatial distribution observed at a given time point compares with a distribution defined *a priori*. This relationship is described by the cosine of the angle between the observed data vector and the vector representing the hypothesized distribution. A first application of VA to the analysis of spatial distribution consists of establishing the vectors of interest in the vector space, computing the angle with the observed data vectors, and testing the differences. In such an analysis, the same time point from different trials could enter as a replication factor into a one-sample significance test. The difference between two or more observed spatial distributions can also be tested.

Given the usual rules of vector arithmetic, an observed data vector can be viewed as the sum of two or

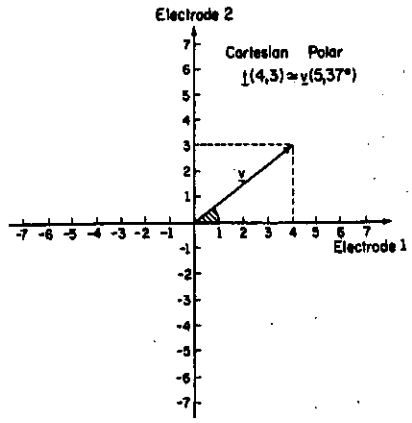

Figure 10-5. Vector analysis: Geometrical representation of a two-element vector (v). Values of the corresponding Cartesian and polar coordinates are also shown.

more component vectors, one of which can be considered as an error vector. Each component vector will be characterized by its own orientation in the space (corresponding, as shown above, to a specific spatial distribution) and its own length (which is a measure of the weight of each component vector in determining the data vector). The "contribution" of each component to the data vector is equal to the length of the component vector. Several different procedures to estimate either the length of the component vectors, their orientation, or both, are available (Gratton et al., 1985). These procedures can be labeled as "vector decomposition." In the simplest case (the vector filter, or VF), a single component, with known spatial distribution, is considered responsible for the spatial distribution observed at a given time point, and discrepancies between this expected distribution and the observed distributions are attributed to sampling error or noise. A brief description of this procedure is given here.

The purpose of the VF technique (Gratton *et al.*, 1985) is to determine the amount of the activity recorded with a multiple-electrode montage at a given time point that can be attributed to a particular target component, defined *a priori* by the investigator. The target component is defined in terms of a spatial distribution that can be represented as a vector in a multidimensional space. All the activity that cannot be attributed to the target component is defined as error. Orthogonality between the orientations of the target vector and of the error vector is postulated (i.e., at each time point, signal and error are uncorrelated). Therefore, the model adopted by VF assumes that the data vector at each time point is given by the sum of a target vector, with given orientation and unknown length, and of an error vector, with orientation orthogonal to the signal vector and unknown length. A statistical test such as Hotelling's one-sample T-square can be used to test the hypothesis that the discrepancies between the observed vector (equal to the mean vector of a sample of vectors) and the theoretical vector may or may not be attributed to chance.

The task of the VF procedure is to estimate the length of the target vector, which, as shown above, corresponds to its contribution to the observed distribution. This can be accomplished by projecting the data vector onto the target vector. This operation is equivalent to rotating the vector space to align one of the axes with the target vector, thus projecting the data vector onto the new axis (see Figure 10-6). The length of the target vector is equal to the length of the data vector multiplied by the cosine of the angle between the observed and the target vectors. Therefore, the length of the target vector will depend on its orientation. This orientation can be chosen *a priori*, on the basis of knowledge of the spatial distribution of the target component, or of some standard experimental procedure known to elicit the target component. An alternative procedure is to select the orientation of the signal vector on the basis of some post hoc statistical procedure (e.g., discriminant analysis, PCA, etc.).

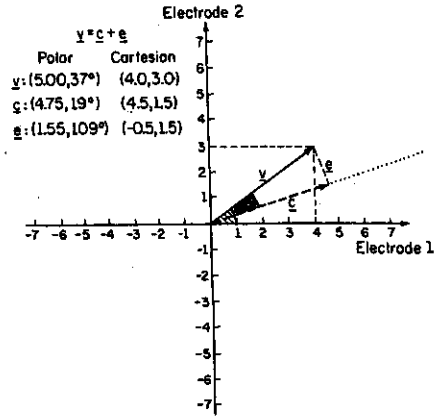

Figure 10-6. Vector filter: Projection of the observed vector (v) on the target component vector (c). Values of the Cartesian and polar coordinates for observed, target, and error (e) vector are also shown.

The lengths of the target vectors obtained at each time point can be directly submitted to standard inferential procedures. The practical result of VF is to "filter" the data for the component of interest (defined by the spatial distribution expressed in the target vector), with the filter output proportional to the goodness of fit between data vector and target vector. Therefore, VF qualifies as a signal extraction technique. A series of filter output values, one for each time point, constitutes a time series, whose values refer to the estimated contribution of the target component to each time point. This time series can be submitted to the analytical and inferential procedures described elsewhere in this chapter.

VF has several advantages in comparison with the traditional procedures of spatial analysis. First, it involves a small amount of computation. Second, it makes use of all the information available at any given time point. Third, it provides a tool for testing hypotheses concerning spatial distribution.

Since the distribution of the target component is established *a priori*, VF needs no cross-validation. Actually, VF itself can be considered as a test for the distribution of the target component. VF does not make use of the information obtained at different time points in determining the target component. While this can be, in some cases disadvantageous (as

the analysis is conducted separately for each time point), it is useful when the latency of the components is variable.

Another limitation is that VF is not able to distinguish between the overlapping contribution of several components to the same time point, unless their distributions correspond to orthogonal vectors in the vector space. In the latter case, for each time point, the independent contributions of as many components as the number of recording sites can be assessed. Note that, since the vector space is defined by the recording sites used, an appropriate choice of the electrode sites can greatly improve the resolution of the effects of overlapping components by the VF technique. However, components with different spatial distribution will be differently amplified, or filtered out, by a VF. On the other hand, the general approach of VA allows the investigator to distinguish between overlapping components. Procedures particularly devised to solve this problem are presented in Gratton *et al.* (1985).

DATA ANALYSIS IN THE FREQUENCY DOMAIN (by Stephen W. Porges)

The Description and Partitioning of Variance

The description of physiological activity in terms of both dependent and independent variables is difficult. Physiological activity is seldom in a "binary" state that can be described as either being "on" or "off." Moreover, changes in level or frequency seldom are complete descriptors of physiological activity. The physiological systems of interest to psychophysiologists are continuously changing, reflecting the dynamic regulatory function of the nervous system. It would, of course, be naive to believe that these systems are sensitive solely to those variables we choose to manipulate in our experiments. Thus, we are faced with a series of paradoxical problems. For example, we may be interested in monitoring the central nervous system during manipulations of "mental effort" or "information processing." However, the dimensions of physiological activity that may be the most sensitive to the "neural mediation" of information processing may also be the most sensitive to the "neural mediation" of basic homeostatic function.

In the psychophysiological literature, two methodological procedures have been employed to deal with the problems of partitioning the impact of "stimulus processing" from the background "neurophysiological regulation." Both procedures reside within the time domain (i.e., the stimulus and the physiological activity are indexed by time). The first procedure is characterized by indexing the changes in mean level or variance produced by an experimental event. Implicit in this type of procedure is the notion of a "statistically significant" response. This notion is based upon a model that assumes that the variance of the physiological process associated with the experimental manipulation is conveyed by the mean level of the physiological process.

The second procedure is characterized by averaging across repeated trials. This procedure is based upon the view that physiological signals reflect neurally mediated responses to stimuli and are superimposed on the background regulatory neurophysiological function. The averaging model assumes that the statistical distribution of the background activity is not influenced by the stimulus and that the background activity is identically and independently distributed. This means that the background activity is assumed to be distributed randomly. Thus, the two prevalent quantitative strategies of decomposing physiological response variance are insensitive both to the possibility that the "signal" is encoded in a parameter other than level and to the possibility that the "signal" is encoded in the background "noise."

Although all voltage × time functions are "time series," few psychophysiologists have used time series statistics (applied to the frequency domain) as analytic tools. Instead, most psychophysiological researchers have attempted to describe responses via more traditional descriptive statistics. For example, the experimental designs that have been prevalent in psychophysiology have involved traditional, repeated-measures ANOVAs, which test effects of stimulus manipulations on the descriptive statistic of mean level. In a few instances, the pattern of the physiological response as a voltage × time function has been estimated with measures of variability such as the standard deviation (e.g., heart period variability). However, in most cases, pattern is described by directional or polarity shifts in the voltage × time function (e.g., HR deceleration or P300).

The repeated-measures ANOVA design tends to evaluate "pre-," "during," and "post-" stimulation periods. By partitioning the variance in this manner, the variance is divided into "treatment" or "time" (repeated measures) and error effects. The error tends to include the variance associated with individual differences among the subjects. To reduce the variance associated with individual differences (i.e., error variance in the analysis of variance design), potent manipulations are used. The objective of this strategy is to enhance the signal-to-noise ratio by maximizing the difference between the "baseline" spontaneous activity and the "response" elicited by the stimulus manipulations.

Ironically, massive treatments often violate the homogeneity-of-variance assumption of ANOVAs. Although the ANOVA is viewed as a "robust" test and is relatively insensitive to violations of the homogeneity of variance assumption in between-groups designs, slight variations in the variance between repeated measures in the repeated-measures design will produce difficulties in interpretation (see page 214; see also Porges, 1979).

An alternative method of describing voltage × time functions is to incorporate "time series statistics" into the experimental and quantitative strategies. Time series methods may be used to detect changes in the voltage × time functions in response to an event by describing the pattern of the function during baseline or stimulus conditions. Time series methods may be classified into two broad categories: time domain and frequency domain. As a general rule, all time series may be represented in either domain. However, certain data may be more easily or more appropriately described in one domain than in the other. One domain may lead to a more natural interpretation. For example, the frequency domain is often used to describe the periodic characteristics of spontaneous EEG and fits nicely into our conceptualization of rhythmic generators in the central nervous system.

Time Series Analysis: Definitions and Methods

The Definition of a Time Series

Although most psychophysiological data are presented in terms of mean levels within or across subjects, the sequential pattern, on which the mean is based, may contribute important information. Time series statistics provide methods to describe and evaluate these patterns. A set of sequential observations, such as the circumference of the chest sampled every second or the time intervals between sequential heart beats, constitutes a time series. Mathematically, a time series may be described as a string of variables that are sequentially indexed—for example, X_t, X_{t+1}, X_{t+2}, ..., X_{t+n}. In this example, the index t represents time.

Time Domain and Frequency Domain Methods: An Overview

There are two basic approaches that may be used to describe and analyze a time series. The series may be represented and analyzed in the time domain or in the frequency domain. Time domain representations plot data as a function of time (see the "Data Analysis in the Time Domain" section). Those time domain methods that are most closely related to the frequency domain are based on autocorrelation and cross-correlation measures. As their names imply, the techniques are mathematical extensions of traditional correlational techniques. An autocorrelation is the correlation of one time series with a time-shifted version of itself. If the time series is periodic, the plot of the autocorrelations (the autocorrelogram) at different time lags will be periodic. Similarly, a cross-correlation is the correlation of one time series with a time-shifted version of a second time series. The cross-correlation function provides information regarding the statistical dependence of one series on another. If the two time series are identical, the peak value of the cross-correlation function will be unity at the lag that makes the two series identical and less than unity at all other lags. In most cases, since the second series is not simply a time-shifted version of the first series, the peak value of the cross-correlation will be less than unity.

Autocorrelation techniques are effective in detecting periodicities only when the series are characterized by a relatively pure sinusoid, uncontaminated by other influences. Cross-correlation techniques lose their effectiveness and sensitivity to assess the communality between two series when the difference between the series is more than a temporal displacement.

Autoregression techniques are more commonly used in developing models of baseline activity and using the model to forecast into the future. These techniques consist of predicting the value of a time series function at a particular time on the basis of previous values of that function. In a multiple-regression sense, each previous time point serves as an independent predictor variable to which a weight is assigned. Stock market forecasting "systems" are dependent upon this type of modeling. Once the model is generated, confidence intervals can be calculated for the forecasted values. In the case of psychophysiological research, one can define a significant response in any physiological system, on any trial, for any subject, by evaluating whether the stimulus manipulation produces a physiological response that occurs outside the confidence intervals of the forecasted values. Autoregression models may be as simple as a linear forecast (i.e., projecting best linear fit from the baseline) or may involve higher-order models. Individuals interested in applying time domain forecasting and prediction models to detect the impact of an intervention are encouraged to study the Box and Jenkins models (Box & Jenkins, 1976) and to be familiar with the interrupted time series model described by Campbell and Stanley (1966).

If the goal is to describe a periodic signal that represents only a small percentage of the total variance of the series, then the successful application of time domain techniques will be limited to the experimenter's ability to filter the data by removing trend and periodicities other than the one of interest (see pp. 186–187 and pp. 195–196). This requires *a priori* knowledge of the underlying periodic structure of the process.

In contrast to time domain techniques, frequency domain techniques are those based upon the spectral density function, which describes how the periodic

variation in a time series may be accounted for by cyclic components at different frequencies. The procedure estimates the spectral densities at various frequencies and is called "spectral analysis." For bivariate series, the "cross-spectral" density function measures the covariances between the two series at different frequencies.

Spectral technology decomposes the variance of a time series into constituent frequencies or periodicities. There is a mathematical relationship between the time domain correlation procedures and spectral analysis. The spectral density function is the Fourier transform of the autocovariance (unstandardized correlation) function, and the cross-spectral density function is the Fourier transform of the cross-covariance function. (The Fourier transform is an algebraic method of decomposing any time series into a set of pure sine wave of different frequencies, with a particular amplitude and phase angle for each frequency.)

Other Frequency Domain Methods

There are other frequency domain techniques. A simple and often visually appealing method is "zero crossing." This method quantifies the frequency with which a waveform crosses an arbitrary baseline. It provides a relatively accurate estimate of the frequency of the process if, and only if, the process contains only one periodic component and is not contaminated by background noise. The periodogram is effective at finding periodic components and may be efficiently calculated using the fast Fourier transform. However, the periodogram has poor statistical characteristics and should not be used without appropriate frequency domain smoothing (see Bohrer & Porges, 1982).

Periodic covariation may be described with "cross-spectral analysis." Cross-spectral analysis generates a coherence function, which is a measure of the best linear association of each observed rhythm in one variable with the same rhythm in a second variable. The coherence is the square of the correlation between the sinusoidal components of the two processes at a specific frequency. The coherence at any specific frequency is the square of the cross-spectral density divided by the product of the spectral densities of each series at the specified frequency. Note the similarity of this equation with the calculation of a squared correlation coefficient; the cross-spectral density parallels the squared cross-products, and the spectral density parallels the variances. Conceptually, the coherence may be thought of as a time series analogue of Ω^2 (see Hays, 1981), or as the proportion of variance accounted for by the influence of one series on the other at each specific frequency. Since physiological processes are not perfect sinusoids, but occur over a band of frequencies, we have developed a summary statistic that describes the proportion of shared variance between two systems over a band of frequencies (see Porges, Bohrer, Cheung, Drasgow, McCabe, & Keren, 1980). We have labeled this statistic the "weighted coherence" (C_w). In our laboratory, C_w has been used primarily to describe the relationship between heart period and respiration. However, the application of C_w is not limited to the assessment of the coupling between respiration and heart period activity, but may also be used to determine the proportion of shared variance between any two processes that fit the statistical assumptions for spectral analysis.

Spectral analysis is based upon a model that assumes that the constituent periodic components of a time series are statistically "independent" and linearly additive. There are situations in which one frequency component in a system could trigger a faster frequency. For example, consider a physiological system in which four breaths occur before there is a general shift in blood pressure. Both frequencies will be manifested in the spectrum of blood pressure. Using traditional spectral analysis, one would assume that the periodic components are independent. However, by using a spectral technology called "polyspectral" (see Brillinger, 1975), it is possible to identify potential "coherences" between two frequency components within one physiological process, or between two different frequency components represented in two different physiological processes.

Time Series Statistics: Methods to Partition Variance

By viewing psychophysiological variables as a time series (i.e., voltage × time functions), and by viewing experimental procedures as a method of partitioning variance, we may arrive at two insights into the construct of "variance." First, the variance associated with the "treatment" must be partitioned from the variance associated with the background physiological activity. This procedure is necessary, since physiological activity is omnipresent and physiological responses must be evaluated against a varying, rather than a constant, baseline. Second, the variance of any physiological process is not uniquely determined by any one specific physiological mechanism. Virtually all physiological response systems represent the actions of antagonistic mediators that reflect the organism's quest to maintain dynamic homeostasis. Therefore, the variance of the physiological process contains "component" variances representing potentially independent mechanisms. Thus, time series methods may be useful in partitioning the variance of the complex physiological response patterns into components. Moreover, it is possible that the statistical behavior of the components will be different; that is, different components will be differentially sensitive to various manipulations.

The discussion above leads to a revised conceptualization of the physiological response pattern in the psychophysiological experiment. Most physiological response patterns may be conceptualized as the sum of two uncorrelated processes: a baseline trend, and an ensemble of rhythmic influences that are superimposed on the baseline trend. The impact of a stimulus or psychological state may reliably influence either or both "component" physiological processes. To complicate matters, the constituent rhythmic components may be manifestations of different underlying neurophysiological processes. For example, in HR there are two obvious rhythms: One is modulated at the respiratory frequency (i.e., respiratory sinus arrhythmia); the second, an oscillation at a slower frequency, appears to represent the influence of the rhythmic oscillation of blood and cerebrospinal fluid, since the same rhythm is observed in vasomotor activity, blood pressure, and cerebrospinal fluid.

Time domain approaches focus on evaluating changes in trend as an indicator of the impact of the stimulus manipulation. These methods tend to remove the background rhythmic activity by averaging across trials. The averaging method assumes that the phase relationship between the underlying rhythmic background activity and the stimulus is identically and independently distributed. Thus, when the data are averaged across trials, the rhythmic background activity will average to zero. This assumption, of course, is only tenable in experiments in which the timing of stimulus presentation is independent of the physiological process. In self-paced experiments, it is highly unlikely that a rhythmic component of the background physiological activity is not phase-related to the self-initiated trial onset, since behavior is neurophysiologically mediated.

In contrast, frequency domain approaches tend to focus on describing the rhythmic components of the background physiological activity that are superimposed on the trend. Thus, it appears that the frequency domain approach tends to evaluate the component of variance that is treated as "error" variance in the time domain approach. Moreover, appropriate implementation of many frequency domain techniques requires that the trend be removed prior to partitioning of the variance into frequency-specific components. Frequency domain approaches tend to be associated with spectral analysis technology. The theories underlying the spectral technology have been, for the most part, developed for "stationary" data sets (Chatfield, 1975). (A time series is said to be stationary when the mean, variance and autocovariance function are independent of time.) Application of the spectral technology to nonstationary data will result in potentially unreliable and uninterpretable spectral density estimates.

Although any data set that is described in the frequency domain may be represented in the time domain, or vice versa, the two approaches do not provide identical information. For example, in the discussion above, I have described the primary emphases of time domain (i.e., the description of trend) and frequency domain (i.e., the description of rhythmic activity) approaches. In both approaches, the data set typically is modified prior to analysis. In the time domain approach, the data have been "smoothed" to remove the variance associated with background activity. In the frequency domain approach, the data have been "detrended" to provide a stationary data set with a constant baseline. However, any time series, which in the examples given here would be voltage × time functions, could be described via frequency domain spectral technology in terms of the sum of spectral density estimates and could be "reconstituted" into the original time series with knowledge of the spectral density estimates and the phase relationships among the constituent frequency components. The time domain autocorrelation approach and the frequency domain spectral approach are merely transformations of each other, although time domain models are more likely to be used to describe changes in trend and frequency domain models to describe changes in the constituent frequency components.

The discussion above is relevant, since most physiological systems monitored by psychophysiologists tend to have both aperiodic (i.e., trend) and periodic (i.e., rhythmic) components. For example, with HR, we have the basic problem of the directional HR responses associated with motor and cognitive function being superimposed on the naturally occurring respiratory sinus arrhythmia. In the case of HR, most psychophysiological investigations attempt to maximize the impact of the stimulus or psychological state on the trend. This is done by averaging and treating the rhythmic oscillations as background "error." Similarly, averaging across trials minimizes the background oscillations in EDA. However, in the case of EEG recordings, it is the periodic characteristics that are emphasized, and it is the trend that is filtered from the data set and treated as "error." In both situations, the assumption is made that the physiological response "component" (i.e., trend or periodic) is a sensitive index of the psychological process being monitored. However, it is conceivable that there may be situations in which the "level" of the output of the physiological system manifested in a change in "trend" may be unresponsive to the manipulation, while the treatment effect may be easily observed in a change in the pattern—or vice versa.

Constraints and Limitations of Sampling Procedures

Physiological Activity: Continuous Processes

Sensitive evaluations of physiological activity must necessarily include sophisticated techniques to evaluate pattern and change. The quantification strategy

that the researcher employs in psychophysiological research must rely on an *a priori* definition of the response parameters being investigated. In most psychological research, background spontaneous activity is considered unimportant. Meaningful responses can be easily identified as a discrete change in the ongoing activity of the system. However, in the investigation of physiological processes, it is clear that most physiological systems function continuously. Although we can easily identify the occurrence of many discrete behavioral responses, meaningful physiological responses are often much more difficult to define and isolate. One must assume that virtually every physiological system is continuous, even though the measurable datum is manifested at discrete times (e.g., heart beats).

Physiological Activity: Discrete Processes

Although the underlying physiological processes are assumed to be continuous, the prevalent quantification strategies necessitate estimates of the physiological activity at discrete points in time. There are two reasons for this procedure: First, most analytic methods are based upon statistical models in which the continuous process is sampled at sequential points in time; and second, the prevalent quantification techniques associated with digital computers necessitate time-dependent sampling. Thus, although many physiological processes are continuous, the statistical and computer technologies generally force the researcher into quantifying and analyzing the voltage × time functions as discrete processes sampled at sequential points in time.

How fast should one sample continuous processes? The sampling rate or "time window" must be fast enough to describe the variance of the process accurately. The decision regarding sampling rate requires an *a priori* understanding of the physiological response system being monitored. If relevant information is encoded in a periodic component of the physiological process with a duration shorter than twice the sampling interval, then the sampled data set will not convey the relevant information. For example, if peripheral vasomotor activity is being sampled from a finger at a rate slower than the HR, the variance in vasomotor activity associated with the beating of the heart will be "aliased" or "folded back" on a slower periodicity. The fastest frequency about which we can derive meaningful information from a data set is called the "Nyquist" frequency. The Nyquist frequency is one-half the sampling frequency.

To illustrate the impact of sampling too slowly, consider sampling a 60-Hz pure sine wave 30 times/sec. Because the signal would always be at the same point in the cycle when sampled, the samples would all have the same value, implying that no signal is present. Sampling this signal 60 times/sec would still yield a flat line. Sampling slightly faster than that would mean measuring successive portions of the cycle, implying a very slowly changing sine wave. For example, sampling a 60-Hz signal 70 times/sec would yield a time series of discrete values resembling a pure 10-Hz signal. Indeed, the investigator could not distinguish true 10-Hz activity from "aliased" 60-Hz activity. Only if the 60-Hz signal were sampled 120 or more times/sec would the 60-Hz signal not be distorted.

As a more complex example, imagine that three physiological variables (i.e., HR, respiration rate, and finger vasomotor activity) are being sampled at a rate of 1 time/sec. In this example, the heart is beating at 90 beats/min (i.e., 1.5 beats/sec, or 1.5 Hz), and the breathing frequency is 15 times/min (i.e., one breath every 4 sec, or .25 Hz). If each variable is sampled 1 time/sec, the fastest periodic process we can evaluate in each of the variables is a process that is slower than one oscillation every 2 sec, or .5 Hz. This does not cause any serious problems with the respiration series, since the breathing is slower than the Nyquist frequency of .5 Hz. Similarly, in the cardiac system, the fastest periodic activity is the respiratory sinus arrhythmia at the frequency of breathing. However, although vasomotor activity exhibits rhythmic processes at the respiratory frequency and at even slower frequencies, it also oscillates at the frequency of the heart beat, since the flow of blood to the periphery is changing on each systole and diastole. Therefore, the peripheral vasomotor activity should exhibit a rhythm of approximately 1.5 Hz (i.e., 90/min), concordant with the average HR. However, if the vasomotor activity is sampled only 1 time/sec, what happens to the variance associated with this fast oscillation? The variance associated with the fast oscillation will be "folded back" and added to the variance of frequencies slower than the Nyquist frequency (which in this case is one-half the 1-Hz sampling rate, or .5 Hz). These lower frequencies are said to be "aliased." The same problem will exist if these variables are sampled every 500 msec when the average HR is about 90 beats/min. In this example, the frequency decomposition (spectrum) of the vasomotor time series will result in a periodic component at a frequency slower than breathing, a second "peak" at the breathing frequency, and a third "peak" at a frequency faster than breathing. This faster frequency does not represent a true neurophysiological process, but rather the impact of an inappropriate sampling rate. In this example it would be necessary to sample at 3 Hz or faster (at least twice the 1.5-Hz HR) to prevent aliasing. To decompose the rapidly changing vasomotor waveform into frequency components accurately, or to be sensitive to short-latency changes in amplitude, it would be preferable to sample more than 3 times/sec.

The dangers of inappropriate sampling rates are clear, but how does one avoid these problems? If one were interested in the relationship among various physiological variables, such as HR and respiration, it would be necessary to sample the activity of all variables at a frequency that is at least twice the frequency of the fastest variable. Note that the problem of aliasing is not problematic solely in the frequency domain: One can see the inappropriate interpretations or loss of relevant information in the time domain, if slow sampling results in not detecting the response component that is sensitive to the stimulus. Fundamentally, sampling a "continuous" process necessitates an understanding of the periodic components and response latencies of the physiological system being studied.

Physiological Activity: Point Processes

Some physiological processes are, by their nature, events that may be characterized as "binary"—categorized as "occurring" or "not occurring." These processes are called "point processes." For example, the beating of the heart may be operationalized as a "binary" event indicated by the occurrence of the R wave. Similarly, single-unit activity in the central nervous system is characterized by "spikes" and "interspike intervals." Point processes pose special statistical problems. The primary problem arises when attempting to sample a point process at equal intervals in time (e.g., second by second). Time series texts (e.g., Gottman, 1981) deal primarily with equal time sampling of continuous processes. Fortunately, this is not problematic with many physiological processes, since they may be represented as continuous voltage × time functions. However, how does one deal with such processes as heart period and the ensemble of processes temporally determined by the beating of the heart? Although blood pressure changes are time locked to the beating of the heart, is it legitimate to view blood pressure as a continuous process and to sample it at equal time intervals? Moreover, how would one estimate the duration of any specific cardiac cycle component (e.g., the P-R interval) across time? These questions have never been adequately discussed in the psychophysiological literature and can be reduced to two points: First, how does one sample event-related physiological data in equal time intervals; second, how frequently must one sample event-related physiological data?

Although Bartlett (1963) provides a method for performing spectral analysis on the "interval" characteristic of binary data, it is of little use to the psychophysiologist. The reasons are self-evident, since the data are assumed to be stationary for this analysis. Recall the arguments given above that the spectral analysis of "nonstationary" time series provides uninterpretable estimates of the spectral densities and that physiological processes tend to be "nonstationary" time series. It is, therefore, necessary to "detrend" the interval time series to generate a data set that is at least weakly stationary. (A process is called "weakly stationary" if its mean is constant and its autocovariance function depends only on lag; see Chatfield, 1975.) Moreover, even in the time domain, equal time interval estimates are necessary for assessing trends. Since most methods of detrending data to produce "stationary" time series for frequency domain analyses are actually time domain methods that have been developed for equal time sampling of continuous process, it is necessary to generate an estimate of the point process at equal points in time.

There are a variety of methods that may be used to generate an estimate of a point process at equal points in time, such as interpolation, weighting, and sampling. Each method has its own unique characteristics. An important requirement is to make the "time window" short enough to map into the temporal variability of the process. If the time window is longer than twice the shortest inter-event interval, then the time window may smooth or alias a component of the variance of the process. In the case of heart period, it is necessary to estimate the heart period in sequential intervals of approximately one-half the duration of the fastest heart period. By estimating the heart period process at sequential intervals that are shorter than half the duration of the fastest heart period, the variance of the heart period process will be preserved in the transformed data set. Moreover, the transformed data set will now be amenable to time domain detrending and filtering techniques, as well as spectral analysis techniques.

Conclusion

The investigator should consider the relative assumptions, advantages, and disadvantages of time domain and frequency domain techniques. Attention to periodicities in a time series, rather than to trends alone, can enhance our understanding of psychophysiological processes.

INFERENCE TESTING

Introduction

"Inference testing" involves procedures that evaluate the probable validity of statements about one set of phenomena, where those statements are based on knowledge about a second set of phenomena. The inference being tested may be "inductive" (one knows real-world event X, which appears to be generalizable to principle Y), "deductive" (one entertains theory Y, which predicts real-world event X), or some elaborate combination of these.

Both inductive and deductive inferences typically contribute to a psychophysiological experiment. First, a general concern or hypothesis is stated, and a highly specific instance of it is studied (deduction). Straightforward algebraic manipulations might then be performed on the resulting voltage × time functions to evaluate whether the claim of the hypothesis was manifested in the data obtained. These manipulations produce "descriptive statistics," merely summarizing the data in some highly specified way. Within the confines of a particular experiment, the validity of a hypothesis is tested by inspection of the data; inferential statistical tests are unnecessary. The discussions of time domain and frequency domain data analysis in this chapter catalog such algebraic procedures, ranging from the computation of a sample mean to PCA.

The investigator is rarely content merely to evaluate the validity of the hypothesis in the specific case alone, however, because the purpose is to confirm the original generalization or to derive new generalizations from initial ones. Thus, the experimenter is likely to attempt to apply one set of concepts to a real-world procedure (deduction), the results of which can then be used to infer new concepts (induction). "Inferential statistics" are those used in this way to evaluate the generalizability of the findings of a particular experiment. Such statistics address the extent to which findings in a specific case can be expected to hold for some superset of similar cases, versus the alternative that specific results are merely the result of variations in the phenomena not accounted for by the theory under consideration.

As noted earlier in this chapter, the algebraic manipulations that raw data often endure are not easily localized in a single stage of analysis. Just as it is artificial to distinguish between signal extraction and data reduction, so it is artificial to segregate data analysis and inference testing. However, while particular techniques straddle such boundaries, the logical distinction between description and inference is essential. The investigator's statistical options become severely curtailed when moving from descriptive and/or exploratory analysis to inferential analysis.

There are numerous texts on the general use of inferential statistics (e.g., Hays, 1981; Myers, 1979; Winer, 1971). Rather than a complete user's guide to statistical inference, the remainder of this section provides a sampling of issues of particular relevance to the psychophysiologist, highlighting assumptions of statistical tests, common violations of those assumptions, and remedial solutions.

Univariate ANOVA

Two issues arise when the traditional ANOVA as an inferential process is applied to psychophysiological data. One issue concerns the need to study phenomena independently of pre-existing basal levels. The use of analysis of covariance (ANCOVA) and of change scores has been particularly controversial. The other issue, the assumption of homogeneity of covariance, derives from the special constraints on ANOVA when repeated measures are used, either alone or crossed with between-subjects variables (mixed-model ANOVA).

Both issues arise because the psychophysiologist typically studies the time course of voltage × time functions in the context of changing inputs from independent variables. If one wishes to measure acute, event-related responses, variations in average or basal level may contribute a major source of statistical noise (variance not controlled by factors in the ANOVA table appearing in the error term). Alternatively, such variance may constitute the main phenomenon of interest if one studies slower, homeostatic actions.

Clearly, it is valuable (though not always possible) for the investigator to determine, *a priori*, the likely sources of variance in the dependent measure. Gaining experimental control over variance is inherently preferable to attempting to assert post hoc statistical control.

ANCOVA

The ANCOVA is commonly the method of choice for post hoc removal of undesired sources of variance. Unfortunately, it is often difficult to achieve such statistical control over undesired sources of variance without systematically distorting the data of interest. For example, it has been argued that ANCOVA is not valid in the very situation for which it is intuitively most appealing (see Chapman & Chapman, 1973, pp. 82–83; Lord, 1967). These authors claim that ANCOVA is legitimate only if two requirements are met: The regression slopes of the dependent variable on the covariate must be the same for each level of the independent variable, and the mean value of the covariate must be the same for each level of the independent variable. As an illustration, assume an experiment in which HR during imagery is believed to vary systematically as a function of imagery ability, a between-subjects factor. To complicate matters, however, some of the subjects are athletes having resting HR levels as much as 40% below those of other subjects. Relative to the hypothesis of the experiment, this source of variance in HR is merely statistical noise. The investigator wishes to employ resting HR as the covariate in an ANCOVA. In order to permit the use of ANCOVA in this case, the investigator would have to show that, for each level of imagery ability in the design, (1) the regression slopes of imagery HR on resting HR are equal and (2) the mean resting HR levels are equal. It is the second requirement that is most likely to disappoint the investigator, since it is often such hypothetically irrelevant, *a priori* group differences that tempt the use of ANCOVA. The former requirement, on the other hand, is less constraining: Differences in regression slopes amount to an interaction of experimental variables with the covariate, which may be a meaningful, if unanticipated, result (see pp. 215–216).

Opinion on the second requirement is not uniform (see Benjamin, 1967; Cohen & Cohen, 1975; Lubin, 1965; Overall & Woodward, 1977). It has been argued that ANCOVA may be permissible despite group differences on the covariate, depending on the reason for the difference. Overall and Woodward (1977) have argued in favor of ANCOVA in the case where experimental treatments do not affect the covariate and subjects are assigned to experimental groups either randomly, or nonrandomly but on the basis of scores on the covariate.

In the case of nonrandom group assignment, Cohen and Cohen (1975) suggest that the issue be considered in terms of causality. When it is believed (for theoretical—not statistical—reasons) that the covariate is causally dependent on the independent variable, ANCOVA would not be legitimate. Specifically, if covariate C shares variance with independent variable X because X affects C, then removing their shared variance from X unfairly robs X of variance with which X may actually affect dependent variable Y. Thus, ANCOVA in this case would distort the experimental effect of X on Y. On the other hand, if C is causally prior to X, then any shared variance with which they jointly affect Y properly belongs to C, not X. Y would be affected by that source of variance whether or not X were present, so X should not be credited with that variance. In the example above of imagery and HR, in which group assignment is nonrandom (based on imagery ability), it can be argued that there is no theoretical basis for basal HR determining imagery ability. Thus, the use of ANCOVA in the face of group differences on the covariate can be defended.

In the case of random assignment to groups, Cohen and Cohen (1975) are quite comfortable with ANCOVA. They reason that, although two random samples may by chance differ on the covariate, what matters is the population they represent, about which inferences will be drawn. Randomization assures that the expected (population) differences between samples will be zero, regardless of actual sample differences. Thus, C and X share no variance in the population, even though they may do so by chance in the samples selected. Although the groups differ on C, they will tend to regress toward their population mean (i.e., no difference on C) when they are observed in order to measure Y. Of course, this tendency will be realized only for large samples.

We do not attempt here to advocate one or the other of these positions on the validity of ANCOVA in the face of group differences on the covariate. Clearly, the investigator should evaluate whether the two assumptions are met in a given data set and should consider whether violation of the assumptions will pose an interpretative problem. Perhaps the important lesson to be drawn from such debates is that what constitutes a proper use of inferential statistics is partially a function of the experimental purpose and the conceptual framework. Just as a given statistical model is developed under certain assumptions, the importance of a violation of those assumptions rests on the use made of the statistic. Furthermore, particular assumptions differ with respect to the consequences of violation. However, the confidence intervals of traditional statistics are not the only means of testing inferences. Repeated sampling from the superset of cases to which one's first experiment belongs—known as "replication" and "cross-validation"—is a respectable alternative.

Change Scores and the Law of Initial Values

Wilder (1957) first formulated the "law of initial values" (LIV), which states that response amplitude is a function of prestimulus level. Assuming an implicit ceiling effect, the prediction is that higher prestimulus levels will be associated with smaller responses. A number of early papers report confirmatory data (e.g., Hord, Johnson, & Lubin, 1964; Lacey, 1956; Sternbach, 1960). The LIV has serious implications for the most popular metric in psychophysiology, the change score. Specifically, the size of the change score may be partially a function of initial level. Clearly, the LIV phenomenon can affect change score data in ways that are unrelated to the experimental manipulation, generally reducing statistical power. To deal with this problem with autonomic measurs, where the LIV problem is most widely acknowledged, Lacey (1956) proposed the autonomic lability score (ALS) transformation. While the ALS was intended to take account of the LIV by standardizing the effect of the homeostatic influence on response amplitude scores, it was later shown (Benjamin, 1963) that the ALS actually removes LIV effects completely. Essentially, the ALS method is a customized elaboration of ANCOVA. As such, it is vulnerable to the same constraints and debates as ANCOVA (see the discussion of ANCOVA, above). When the investigator is satisfied that the ALS method is statistically acceptable in a particular data set, it can be useful for standardizing data that consist of multiple dependent measures having different measurement scales, variances, and the like. It is particularly appropriate for a close examination of response pattern across situations and measures (e.g., Lacey, 1956; cf. the discussion of coefficient of concordance, page 217). However, the ALS is rarely used in current research.

The validity of change scores is still arguable (e.g., Benjamin, 1973; Etaugh & Etaugh, 1972; Harris, 1963; Lubin, 1965). We are in sympathy with Benjamin (1973), who has argued that one's metric achieves validity because of its theoretical appropriateness, not its statistical purity. Thus, if one's theory

actually makes predictions about change scores, one should measure change scores.

The Assumption of Homogeneity of Covariance

"Homogeneity of covariance" means that the covariance between each pair of repeated-measures factor levels is constant. The assumption of homogeneity of covariance in ANOVA is perhaps less controversial than the requirement for ANCOVA, but it is still frequently violated. For example, Jennings and Wood (1976) reported that 84% of the articles using repeated-measures ANOVA designs in Volume 12 of *Psychophysiology* (1975) appeared to have ignored the assumption. Violations of the assumption generally bias the *F* test toward a Type I error (Myers, 1979; Winer, 1971).

Violations of the assumption are possible in any repeated-measures design, but almost inevitable in psychophysiological studies analyzing voltage × time functions. Neighboring time points tend to be highly correlated, but samples that are more widely spaced in time will generally be less tightly coupled. The problem is exacerbated when periodicities exist in the signal, such as sinus arrhythmia in HR, producing systematic irregularities in the covariance among sample points. Thus, the covariances among pairs of points in a time series will vary as a function of their temporal separation. This problem is clearly larger when "time" is measured in milliseconds between digitized samples than in minutes between trial blocks or days between sessions. Of course, the critical issue is not the absolute time scale but the stability of the psychophysiological function relative to the sampling interval and the total sample epoch. Neighboring 10/sec samples of skin conductance level are likely to be much more closely intercorrelated than 10/sec samples of EEG.

Two ways of coping with heterogeneity of covariance have been proposed. Jennings and Wood (1976) apprised psychophysiologists of Box's (1954) solution as developed by Geisser and Greenhouse (1958; see also Games, 1975, 1976; Keselman & Rogan, 1980; Keselman, Rogan, Mendoza, & Breen, 1980; McCall & Appelbaum, 1973; Richards, 1980; R. S. Wilson, 1974). This method reduces the degrees of freedom used for evaluation of the significance of the *F* statistic. The reduction is proportional to the amount of heterogeneity among the covariances. Assuming K treatment levels and n subjects, the maximum possible reduction is from $(K-1)$ and $(K-1)\times(n-1)$ to (1) and $(n-1)$ degrees of freedom. Jennings and Wood (1976) and Myers (1979) provide a formula for calculating the correction factor, known as ϵ or λ.[1] These writers point out that the *F* can first be evaluated using the most conservative [(1), $(n-1)$] and the most liberal [$(K-1)$, $(K-1)\times(n-1)$] limits; if the *F* is significant in the first case or nonsignificant in the second case, there is no need to calculate the correction factor.

To illustrate the potential impact of the Geisser-Greenhouse correction factor, consider a standard CNV paradigm (Simons, Ohman, & Lang, 1979). The EEG was digitized at 30 Hz, and sets of 15 consecutive points were averaged to one value every .5 sec. One of us computed the correction factor for this data set to be .19. Thus, using the correction factor would have cost the investigators 80% of their degrees of freedom in analyses using the .5-sec data points. Using median HR data obtained during sequential 30-sec periods of an imagery experiment (Lang, Kozak, Miller, Levin, & McLean, 1980), the correction factor was computed to be .35. This higher value (i.e., less heterogeneity) reflects the much longer interobservation interval in this study than in the CNV example, though about two-thirds of the degrees of freedom would still have been lost had the correction been made.

In sum, the Geisser–Greenhouse correction factor can take a very serious toll on the apparent statistical power of a psychophysiological experiment (though, of course, it merely reclaims the inflated "power" caused by heterogeneity of covariance). In addition, Davidson (1972) has pointed out that with small sample sizes there may be inadequate statistical power in the procedure to detect a violation of the homogeneity assumption. Finally, it has been suggested that the product of the correction factors for the separate main effects be used for testing interactions of repeated-measures factors, but this application has not been fully validated (Jennings & Wood, 1976). These factors together undoubtedly explain how rarely the correction factor is used in published research.

The second way to cope with violations of the homogeneity-of-variance assumption, advocated by Richards (1980), is to do multivariate analysis of variance (MANOVA), rather than the usual repeated-measures, univariate ANOVA. The different levels of the repeated-measures factor in ANOVA become separate dependent variables analyzed simultaneously in MANOVA (see pp. 216–217).

Power of the F Test in ANOVA

Statistical power is generally well understood conceptually but is rarely considered quantitatively in the design or evaluation of psychophysiological studies. "Power" may be defined as the ability to find an effect that actually exists. More formally, if β is the probability of a Type II error (failure to reject a false null hypothesis), then power is $1-\beta$.

Cohen and Cohen (1975) discuss statistical power as a joint function of three other parameters, such that power is increased when any of the following is increased: sample size, effect size, and α (the probability of a Type I error, rejection of the null hypothesis when it is true). Fixing the values of any three of these parameters determines the fourth. Conversely, the appropriate value for any of them, such as sample size, cannot be determined without knowing the other three. Effect size is often most difficult to deal with in experimental design. The magnitude of an experimental effect can intuitively be understood as the value of an equivalent correlation coefficient. Although the investigator's hypotheses may not specify the effect size anticipated, prior experience with a given paradigm often implicitly guides the decision, with sample size and α level set accordingly.

However, in our opinion, statistical power is frequently too low in many psychophysiological studies. While it may be argued that power is low only for weak effects and that perhaps we should normally confine ourselves to seeking strong effects, the issue is rather that investigators too rarely consider the questions of effect size and power explicitly when designing experiments.

As an illustration of low power levels prevalent in psychophysiological research, effect size and power were calculated for a subset of data published in Lang *et al.* (1980). For a simple between-subjects main effect, with 16 subjects in each of two groups—a relatively large sample size for a psychophysiological study—the effect sizes actually obtained for two dependent measures were equivalent to correlations of .43 and .35. These were conceptually important effects and were statistically significant. Assuming an α level of .05 and an effect size of .40 (Cohen & Cohen, 1975, suggest .30 as "moderate" and .50 as "large" effect sizes when one has no basis for estimate), a total sample size of 32 provides a power level of .64. In other words, Lang *et al.* could expect to find such an effect, upon replication, in only two experiments out of every three attempts. If sample size were reduced to 24, power would fall to .51, or only an even chance of replication.

The point of this illustration is that statistical power is surprisingly low in many psychophysiological studies. Cohen and Cohen (1975) provide an enlightening discussion of power and a straightforward method for its computation. They recommend .80 as a reasonable target value for power in many situations. Koele (1982) discusses the relative power of different types of ANOVA designs and emphasizes the simplicity of determining power. A more detailed treatment is available in Cohen (1977).

Multiple Regression/Correlation

Traditionally, the correlational approach in psychology has been associated with psychometric test construction and with studies of individual differences and clinical phenomena. In these cases, independent variables generally vary between subjects and are typically difficult to manipulate experimentally, due to logical, practical, and ethical constraints. Although causality is more difficult to demonstrate with correlational than with experimental designs, there is much to recommend multiple regression/correlation (MRC) as a general data-analytic strategy that subsumes ANOVA and ANCOVA (Cohen, 1968; Cohen & Cohen, 1975, especially Section 8.7 and Chapters 9 and 10).

In fact, correlational methods are increasingly used in the processing of psychophysiological data prior to the stage of inference testing. EMCP, Woody filtering, discriminant analysis, PCA, and a number of techniques in the frequency domain employ correlational calculations (see the discussions of these techniques earlier in this chapter).

Several factors have probably contributed to the underutilization of MRC for inference testing in psychophysiology. First, investigators are usually interested in establishing whether a "significant" relationship exists between two variables, rather than in estimating the strength of the relationship, or predicting the exact value of one variable given the other. More conceptually, the specific prediction of a physiological variable on the basis of a psychological variable would be meaningful only given a level of theorizing beyond what is often available. Third, the traditional association of correlational techniques with correlational designs (and their limitations) unnecessarily constrains appreciation of the MRC approach. Fourth, when used for inference testing, MRC is not conveniently adapted for use in repeated-measures or mixed-model designs. Fifth, most computer packages do not include in their MRC routine an easy facility for accomplishing an ANOVA-type analysis, especially with nominal independent variables. Finally, MRC normally requires the investigator to make explicit choices about error terms, interactions, and the priority among independent variables. ANOVA programs typically deny the investigator these often difficult choices, with the benefit of convenience but at the cost of flexibility and, sometimes, statistical power.

The last point bears expansion. It must be understood that, both conceptually and algebraically, MRC is a superset of ANOVA. Indeed, such ANOVA complexities as trend analysis, ANCOVA, planned comparisons, continuous independent variables and their interactions, and unequal cell sizes can be handled very routinely within MRC. For example, nonparallel regression slopes that preclude ANCOVA (see pp. 212–213) amount to an analyzable interaction effect in MRC—that is, a problem becomes a significant finding. Similarly, complex planned comparisons are easily tested using appropriate "dummy coding" of the levels of the independent variable(s).

As one gains experience in the use of MRC to accomplish inference testing of data collected in a traditional ANOVA-type experimental design, one realizes the extent to which most ANOVA packages

1. There is a typographical error in the formula given in Jennings and Wood (1976). The penultimate parenthesis should be deleted.

encourage the investigator to ignore certain basic statistical questions. Specifically, the inclusion or exclusion of particular interaction terms in the statistical model, and the pairing of independent variables and error terms, involve difficult scientific questions with which the investigator should struggle. Typically, in ANOVA, all possible interactions are included in the statistical model and the source table. However, in principle the investigator is free to select, on conceptual grounds, which interactions should be included and which ones should be left in the error term. Use of an incomplete design is particularly appropriate when an interaction term has little theoretical meaning and when its associated degrees of freedom could be put to better use reducing the mean square error. Similarly, ANOVA generally forces the testing of the significance of each factor against an error term, which is the residual after all available sources of variance have been removed (i.e., with a minimum of both error sum of squares and error degrees of freedom—a mixed blessing); MRC permits one to use any of several estimates of error, with potentially greater statistical power—consistent with the original Fisherian emphasis on hierarchical (sequential) rather than simultaneous analysis. Again, the choice among these options should be consciously made, not delegated to an ANOVA program. As an example, consider an experiment in which subjects' autonomic response to snake exposure is measured. Should the effect of subject gender be tested before or after the snake fear questionnaire score has been partialed out of the autonomic measure, and/or partialed out of the gender variable? The converse question can also be raised. ANOVA normally tests each effect after all other effects have been removed from the dependent variable. In other words, each factor is evaluated with all other factors treated as covariates. The investigator may not always find this to be theoretically appropriate.

In sum, MRC is potentially of great use to the psychophysiologist as a general, conceptually stimulating method of inference testing. The highly readable text by Cohen and Cohen (1975) is recommended to the ANOVA-oriented investigator seeking to employ the more general methods of MRC, particularly when considering ANCOVA.[2]

Multivariate Techniques

Multivariate Analysis of Variance

Although most inferential statistics used in psychophysiology involve multiple variables, multivariate analysis of variance (MANOVA) is a term normally reserved for a technique that is the extension of the typical univariate ANOVA (multiple independent variables but a single dependent variable) to the simultaneous analysis of multiple *dependent* variables. MANOVA appears to be highly appropriate for inference testing in psychophysiological research, because measurement of multiple dependent measures is routine. MANOVA has been especially advocated as an alternative to univariate ANOVA when repeated measures are involved (Davidson, 1972; Richards, 1980). In the MANOVA approach to such a design, the levels of the repeated-measures variable(s) in the ANOVA become separate dependent variables.

A particular advantage of MANOVA over ANOVA is that while both assume homogeneity of covariance (see page 214), studies have shown MANOVA to be very robust to violations of this assumption, especially if cell sizes are equal (Hakstian, Roed, & Lind, 1979; see Richards, 1980). Furthermore, MANOVA is more sensitive than ANOVA to certain types of small but reliable effects (Davidson, 1972). MANOVA is clearly in a better position to control experiment-wise error rate and to test hypotheses involving several response systems.

Despite these advantages, MANOVA is rarely used in published psychophysiological research. Besides the lack of familiarity most investigators have with the technique, two obstacles probably account for this neglect. A practical obstacle is the greater difficulty of computation and statistical interpretation of MANOVA than of ANOVA, including continuing disputes over choice of test statistic (e.g., Olson, 1976, 1979; Stevens, 1979). Standard statistical packages appear to be improving in this regard. However, a conceptual obstacle is the lack of theory to specify the relationship among multiple physiological dependent variables. On what common scale should HR and skin conductance be quantified? How readily can one interpret a multivariate F indicating significant systematic variability somewhere among 100 digitized samples from each of eight EEG sites? Thus, even though the basic phenomena of interest are fundamentally multivariate, psychophysiologists have preferred the narrower, univariate ANOVA approach. It is difficult to evaluate how this understandable restriction of vision constrains the hypotheses proposed and the inferences made. The interested reader may consult Cooley and Lohnes (1971), Press (1972), Tatsuoka (1971), Van Egeren (1973), R. S. Wilson (1974), Winer (1971), or Woodward and Overall (1975) for basic discussions of MANOVA.

Canonical Correlation Analysis

The traditional Pearson product-moment correlation coefficient, a measure of linear association between variables x and y, may be generalized in two stages. If variable x is replaced by set X consisting of several variables, MRC can evaluate the association between y and set X. If variable y is then replaced by set Y of several variables, it is canonical correlation that measures the association between set X and set Y. Specifically, canonical correlation analysis seeks a linear combination of the variables in set X and a linear combination of the variables in set Y such that a maximum correlation between these two linear combinations is achieved—in other words, such that set X controls a maximum amount of the variance in set Y. Cohen and Cohen (1975, Chapter 11) and Knapp (1978) demonstrate that canonical correlation subsumes a wide variety of common univariate and multivariate parametric methods of inference testing, including ANOVA, ANCOVA, MRC, MANOVA, MANCOVA (MANOVA with covariance), and discriminant analysis. Given the common practice of quantifying several types of physiological phenomena from multiple recording sites, this technique appears highly appropriate for psychophysiological inference testing. It combines the benefits of MRC and MANOVA (see above) over traditional ANOVA. However, it faces the same interpretative difficulties described for MANOVA.

In general, multivariate statistics have not been widely adopted in psychophysiology, probably because investigators have not felt the need to go beyond what ANOVA will do for them. When the questions asked and the hypotheses tested no longer fit within such strictures, multivariate methods will have to be dealt with. Conversely, once they become routine, they will undoubtedly influence the questions that are asked.

Nonparametric Tests

Given the frequency with which psychophysiological data violate the assumptions of parametric statistics, nonparametric statistics would seem a highly appropriate alternative. They typically require fewer assumptions, though they do so at some cost of statistical power. However, nonparametric approaches have not been developed adequately to accommodate the complex experimental designs often used in psychophysiology. One-way and two-way analogues of standard parametric ANOVA have been described and occasionally appear in the literature (Kruskal–Wallis one-way ANOVA by ranks and Friedman two-way ANOVA by ranks; Siegel, 1956). Methods for testing post hoc comparisons following these analyses exist (Levy, 1979; Marascuilo & McSweeney, 1967). K. V. Wilson (1956) offered a more general nonparametric ANOVA analogue, which is computationally more cumbersome and has not generally been used.

Several other nonparametric statistics deserve consideration when the design is appropriate. The well-known Spearman rank-order correlation (Rs) has been generalized to the coefficient-of-concordance statistic (W; Siegel, 1956). Where Rs reflects the agreement between two sets of rankings, W reflects the agreement among multiple sets of rankings. For example, W can reflect the degree of agreement among judges ranking a series of responses. Intuitively, W is the average of the pairwise Rs values in the data set. Siegel (1956) presents a test of significance for W. This statistic is highly suitable as a summary statistic computed for individual subjects across multiple physiological dependent measures in multiple situations, providing a measure of response stereotypy. Thus, it is potentially useful as a means of classifying subjects. However, it is not so readily applicable to hypothesis testing about relatively homogeneous populations of subjects, the more common goal in experimental design. Some early studies of response patterning did use W (e.g., Schnore, 1959), but it has received little attention since then.

Dependent variables in psychophysiological studies are usually quantified as continuous variables. However, when the dependent variable in a repeated-measures design is dichotomous (e.g., responses might be scored as present or absent, as is sometimes done for skin conductance responses), the usual parametric ANOVA is not in general appropriate (Marascuilo & McSweeney, 1977; Winer, 1971). The nonparametric Q statistic (Cochran, 1950) is recommended instead, although under some conditions (including a sufficiently large number of observations), the F of ANOVA approximates Q (D'Agostino, 1971; Lunny, 1970). Q is easily computed and follows a chi-square distribution. Q is vulnerable to heterogeneity of covariance in a manner analogous to the F statistic, and the Box/Geisser–Greenhouse correction factor for degrees of freedom in the F test is appropriate for Q (Bhapkar & Somes, 1977; Myers, DiCecco, White, & Borden, 1982). Methods for testing post hoc comparisons following the Q analysis exist (Levy, 1979; Marascuilo & McSweeney, 1977). Although developed for a simple subjects × conditions design, Q has been extended to certain cases of interaction (Marascuilo & Serlin, 1977). However, Q has not been extended adequately to cover the complex designs typical of much of psychophysiology.

A Few More Caveats

The use of change scores pervades data analysis in psychophysiology. Nevertheless, change scores are notorious for having low reliability. Furthermore, if treatment means and variance in true scores are held constant, statistical power is directly related to the reliability of the dependent variable (i.e., inversely related to variance in error scores). Thus, change scores seem poor candidates for inference testing. However, Nicewander and Price (1978) have demon-

2. The content of this section owes much to Cohen and Cohen (1975). As the present chapter is being written, a new edition of their book is in press.

strated that high reliability is not necessarily optimal for inference testing, largely because variance in true scores is often *not* held constant in experiments. They show, in fact, that under certain conditions statistical power is paradoxically maximized when the reliability of dependent measures is minimized. They conclude that the optimal level of reliability depends on the nature of the hypothesis being tested. One caveat would be that the investigator should consider carefully whether, on conceptual grounds, change scores are appropriate in a given study.

A related issue includes the direction of statistical inference and the relative reliability of different measurements. Though more often discussed in the clinical realm, this issue arises in psychophysiological research as well. Chapman and Chapman (1973) provide a clear example: "If schizophrenics are as inferior to normal subjects on one ability as on another, but the test that is used to measure one of the abilities is more reliable than the test for the other, a greater deficit will be found on the more reliable measure" (p. 67). In psychophysiology, a number of examples can be given. If two conditions produce equal real changes from a third condition, the measured change will be larger for the condition measured with greater reliability. Similarly, if one quantifies HR with greater reliability than finger pulse volume, genuinely equal changes in both will yield data indicating a larger change in HR than in finger pulse volume. As a final example, if P300 in the brain ERP can be measured more reliably (in a statistical sense) at the parietal than at the frontal electrode placement, it will be easier to find same-size effects at the parietal than at the frontal location. A second caveat, then, would be that investigators should evaluate the statistical reliability of psychophysiological measures and, in making statistical inferences, should be careful not to confound differences in genuine effects with differences in reliability of measurement.

CONCLUSION

In this chapter, we have reviewed techniques that can be used to analyze psychophysiological functions result in a voltage × time function. For this reason, all analytic techniques can, at least in principle, be applied to any psychophysiological function. Selection of which technique to use must be guided in part by the particular question the investigator seeks to answer and in part by the nature of the underlying physiological system. We have attempted to indicate the advantages and disadvantages of each technique. By doing this, we hope that investigators will look beyond those techniques that have traditionally been associated with a particular function. We also hope that, in spite of space limitations, we have given enough guidance to enable the interested researcher to make an intelligent selection of a technique. The references we have provided should ensure that users go beyond a "cookbook" approach.

We should emphasize that analytic techniques are, in some sense, only as good as the data to which they are applied. There is clearly no substitute for careful recording procedures and appropriate experimental design.

ACKNOWLEDGMENTS

The preparation of this chapter was supported in part by the Air Force Office of Scientific Research (Contract No. F49620-79-C-0233); the Environmental Protection Agency (Contract No. EPA CR 808974-02); the School of Aerospace Medicine, Brooks Air Force Base (Contract No. F33615-82-C-0609); and the Defense Advanced Research Projects Agency (Contract No. MDA903-83-C-0017).

The preparation of Porges's section was supported, in part, by Research Scientist Development Award K02-MH-0054 from the National Institute of Mental Health and by Grant No. HD 15968 from the National Institutes of Health.

REFERENCES

Ackroyd, M. *Digital filtering*. London: Butterworth, 1973.

Aunon, J. I., McGillem, C. D., & O'Donnell, R. D. Comparison of linear and quadratic classification of event-related potentials on the basis of their exogenous or endogenous components. *Psychophysiology*, 1982, 19, 531–537.

Bartlett, M. S. The spectral analysis of point processes. *Journal of the Royal Statistical Society* (Series B), 1963, 25, 264–280.

Benjamin, L. S. Statistical treatment of the law of initial values (LIV) in autonomic research: A review and recommendation. *Psychosomatic Medicine*, 1963, 25, 556–566.

Benjamin, L. S. Facts and artifacts in using analysis of covariance to "undo" the law of initial values. *Psychophysiology*, 1967, 4, 187–206.

Benjamin, L. S. Remarks on behalf of change scores and associated correlational statistics: A response to the Etaughs. *Developmental Psychology*, 1973, 8, 180–183.

Bhapkar, V. P., & Somes, G. W. Distribution of Q when testing equality of matched proportions. *Journal of the American Statistical Association*, 1977, 72, 658–661.

Bohrer, R. E., & Porges, S. W. The application of time-series statistics to psychological research: An introduction. In G. Keren (Ed.), *Statistical and methodological issues in psychology and social sciences research*. Hillsdale, N.J.: Erlbaum, 1982.

Box, G. E. P. Some theorems on quadratic forms applied in the study of analysis of variance problems: II. Effects of inequality of variance and covariance between errors in the two-way classification. *Annals of Mathematical Statistics*, 1954, 25, 484–494.

Box, G. E. P., & Jenkins, G. M. *Time series analysis, forecasting and control*. San Francisco: Holden Day, 1976.

Brillinger, D. R. *Time series: Data analysis and theory*. New York: Holt, Rinehart & Winston, 1975.

Brown, C. C. (Ed.). *Methods in psychophysiology*. Baltimore: Williams & Wilkins, 1967.

Callaway, E., & Halliday, R. A. Evoked potential variability: Effects of age, amplitude and methods of measurement. *Electroencephalography and Clinical Neurophysiology*, 1973, 34, 125–133.

Callaway, E., Halliday, R. A., & Herning, R. I. A comparison of

methods for measuring event-related potentials. *Electroencephalography and Clinical Neurophysiology*, 1983, 55, 227–232.

Campbell, D. T., & Stanley, J. C. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally, 1966.

Carlton, E. H., & Katz, S. Is Wiener filtering an effective method of improving evoked potential estimation? *IEEE Transactions in Biomedical Engineering*, 1980, 27, 187–192.

Cattell, R. B. The scree test for the number of factors. *Multivariate Behavioral Research*, 1966, 1, 245–276.

Chapman, L. J., & Chapman, J. P. *Disordered thought in schizophrenia*. New York: Appleton-Century-Crofts, 1973.

Chatfield, C. *The analysis of time series: Theory and practice*. London: Chapman & Hall, 1975.

Cochran, W. G. The comparison of percentages in matched samples. *Biometrika*, 1950, 37, 256–266.

Cohen, J. Multiple regression as a general data-analytic system. *Psychological Bulletin*, 1968, 70, 426–443.

Cohen, J. *Statistical power analysis for the behavioral sciences*. New York: Academic Press, 1977.

Cohen, J., & Cohen, P. *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, N.J.: Erlbaum, 1975.

Cooley, W. W., & Lohnes, P. R. *Multivariate data analysis*. New York: Wiley, 1971.

Cook, E. C. FWTGEN—An interactive FORTRAN II/IV program for calculating weights for a non-recursive digital filter. *Psychophysiology*, 1981, 18, 489–490.

Corby, J. C., & Kopell, B. S. Differential contribution of blinks and vertical eye movements as artifacts in EEG recording. *Psychophysiology*, 1972, 9, 640–644.

D'Agostino, R. B. A second look at analysis of variance on dichotomous data. *Journal of Educational Measurement*, 1971, 8, 327–333.

Davidson, M. L. Univariate versus multivariate tests in repeated-measures experiments. *Psychological Bulletin*, 1972, 77, 446–452.

Dixon, W. T. *BMDP biomedical computer programs*. Los Angeles: University of California at Los Angeles, 1979.

Donchin, E. Use of scalp distribution as a dependent variable in event-related potentials studies: Excerpts of preconference correspondence. In D. Otto (Ed.), *Multidisciplinary perspectives in event-related brain potential research* (EPA-600/9-77-043). Washington, D.C.: U.S. Government Printing Office, 1978.

Donchin, E., & Heffley, E. Multivariate analysis of event-related potential data: A tutorial review. In D. Otto (Ed.), *Multidisciplinary perspectives in event-related brain potential research* (EPA-600/9-77-043). Washington, D.C.: U.S. Government Printing Office, 1978.

Donchin, E., & Herning, R. I. A simulation study of the efficacy of stepwise discriminant analysis in the detection and comparison of event-related potentials. *Electroencephalography and Clinical Neurophysiology*, 1975, 38, 51–68.

Donchin, E., Tueting, P., Ritter, W., Kutas, M., & Heffley, E. On the independence of the CNV and the P300 components of the human average evoked potential. *Electroencephalography and Clinical Neurophysiology*, 1975, 38, 449–461.

Duffy, F. H., Bartels, P. H., & Burchfiel, J. L. Significance probability mapping: An aid in the topographical analysis of brain electrical activity. *Electroencephalography and Clinical Neurophysiology*, 1981, 51, 455–462.

Duncan-Johnson, C. C., & Donchin, E. The time constant in P300 recording. *Psychophysiology*, 1979, 16, 53–55.

Ertl, J. P. Detection of evoked potentials by zero crossing analysis. *Electroencephalography and Clinical Neurophysiology*, 1965, 18, 630–631.

Ertl, J. P., & Schafer, E. W. P. Brain response correlates of psychometric intelligence. *Nature*, 1969, 223, 421–422.

Etaugh, E. F., & Etaugh, C. F. Overlap: hypothesis or tautology? *Developmental Psychology*, 1972, 6, 340–342.

Friedman, D. H. *Detection of signal by template matching*. Baltimore: Johns Hopkins University Press, 1968.

Games, P. A. Computer programs for robust analyses in multifactor analysis of variance designs. *Educational and Psychological Measurement*, 1975, 35, 147–152.

Games, P. A. Programs for robust analyses of ANOVA's with repeated measures. *Psychophysiology*, 1976, 13, 603.

Geisser, S., & Greenhouse, S. W. An extension of Box's results on the use of the F distribution in multivariate analysis. *Annals of Mathematical Statistics*, 1958, 29, 885–891.

Glaser, E. M., & Ruchkin, D. S. *Principles of neurobiological signal analysis*. New York: Academic Press, 1976.

Gottman, J. M. *Time-series analysis*. New York: Cambridge University Press, 1981.

Graham, F. K. Constraints on measuring heart rate sequentially through real and cardiac time. *Psychophysiology*, 1978, 15, 492–495.

Gratton, G., Coles, M. G. H., & Donchin, E. A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, 1983, 55, 468–484.

Gratton, G., Coles, M. G. H., & Donchin, E. *A vector analysis of ERPs*. Manuscript in preparation, 1985.

Hakstian, A. R., Roed, J. C., & Lind, J. C. Two-sample T2 procedure and the assumption of homogeneous covariance matrices. *Psychological Bulletin*, 1979, 86, 1255–1263.

Harman, H. H. *Modern factor analysis* (Rev. ed.). Chicago: University of Chicago Press, 1967.

Harris, C. W. (Ed.). *Problems in measuring change*. Madison: University of Wisconsin Press, 1963.

Hays, W. L. *Statistics*. New York: Holt, Rinehart & Winston, 1981.

Hord, D. J., Johnson, L. C., & Lubin, A. Differential effect of the law of initial value (LIV) on autonomic variables. *Psychophysiology*, 1964, 1, 79–87.

Horst, R. L., & Donchin, E. Beyond averaging II: Single trial classification of exogenous event-related potentials using stepwise discriminant analysis. *Electroencephalography and Clinical Neurophysiology*, 1980, 48, 113–126.

Humphreys, L. G., & Montanelli, R. G. An investigation of the parallel analysis criterion for determining the number of common factors. *Multivariate Behavioral Research*, 1975, 10, 193–206.

Isreal, J. B., Chesney, G. L., Wickens, C. D., & Donchin, E. P300 and tracking difficulty: Evidence for multiple resources in dual-task performance. *Psychophysiology*, 1980, 17, 259–273.

Jennings, J. R., Tahmoush, A. J., & Redmond, D. P. Non-invasive measurement of peripheral vascular activity. In I. Martin & P. H. Venables (Eds.), *Techniques in psychophysiology*. Chichester, England: Wiley, 1980.

Jennings, J. R., & Wood, C. C. The epsilon-adjustment procedures for repeated-measures analyses of variance. *Psychophysiology*, 1976, 13, 277–278.

Jennrich, R. I. Stepwise discriminant analysis. In K. Enslin, A. Ralston, & H. S. Wief (Eds.), *Statistical methods for digital computers*. New York: Academic Press, 1977.

John, E. R., Ruchkin, D. S., & Villegas, J. Experimental background: Signal analysis and behavioral correlates of evoked potential configurations in cats. *Annals of the New York Academy of Sciences*, 1964, 112, 362–420.

John, E. R., Ruchkin, D. S., & Vidal, J. J. Measurement of event-related potentials. In E. Callaway, P. Tueting, & S. H. Koslow (Eds.), *Event-related brain potentials in man*. New York: Academic Press, 1978.

Kaiser, H. F. The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 1960, 20, 141–151.

Keselman, H. J., & Rogan, J. C. Repeated measures F tests and psychophysiological research: Controlling the number of false positives. *Psychophysiology*, 1980, 17, 499–503.

Keselman, H. J., Rogan, J. C., Mendoza, J. L., & Breen, L. J. Testing the validity conditions of repeated measures F tests. *Psychophysiological Bulletin*, 1980, 87, 479–481.

Knapp, T. R. Canonical correlation analysis: A general parametric significance-testing system. *Psychological Bulletin*, 1978, 85, 410–416.

:, P. Calculating power in analysis of variance. *Psychological ulletin*, 1982, 92, 513–516.
yashi, H., & Yaguchi, K. A statistical method of component lentification of average evoked potentials. *Electroencephalo-aphy and Clinical Neurophysiology*, 1981, 51, 213–214.
, J. L. The evaluation of autonomic responses: Toward a neral solution. *Annals of the New York Academy of Sciences*, 956, 67, 123–164.
P. J., Kozak, M. J., Miller, G. A., Levin, D. N., & lcLean, A. Emotional imagery: Conceptual structure and pat-rn of somatovisceral response. *Psychophysiology*, 1980, 17, 79–190.
K. J. Nonparametric large-sample pairwise comparisons. *Psychological Bulletin*, 1979, 86, 371–375.
F. M. A paradox in the interpretation of group comparisons. *ychological Bulletin*, 1967, 68, 304–305.
, A. [Book review of *Problems in measuring change*, C. Harris, d.]. *American Journal of Psychology*, 1965, 78, 324–327.
, G. H. Using analysis of variance with a dichotomous pendent variable: An empirical study. *Journal of Educational easurement*, 1970, 4, 263–269.
stadt, H. V., Enke, C. G., & Crouch, S. R. *Electronic measure-nts for scientists*. Menlo Park, Calif.: W. A. Benjamin, 1974.
icuilo, L. A., & McSweeney, M. Nonparametric post hoc mparisons for trend. *Psychological Bulletin*, 1967, 67, 401–2.
icuilo, L. A., & McSweeney, M. *Nonparametric and distribu-n free methods for the social sciences*. Monterey, Calif.: Brooks/ le, 1977.
icuilo, L. A., & Serlin, R. Interactions for dichotomous varia-s in repeated measures design. *Psychological Bulletin*, 1977, , 1002–1007.
n, I., & Venables, P. H. (Eds.). *Techniques in psychophysiology*. ichester, England: Wiley, 1980.
ll, R. B., & Appelbaum, M. I. Bias in the analysis of repeated-easures designs: Some alternative approaches. *Child Develop-nt*, 1973, 44, 401–415.
lem, C. D., Aunon, J. I., & Childers, D. G. Signal processing evoked potential research: Applications of filtering and pat-n recognition. *Critical Reviews in Bioengineering*, 1981, 6, 5–265.
lem, C. D., Aunon, J. I., & O'Donnell, R. D. Computer ssification of single event-related potentials. *Psychophysiology*, 31, 18, 192. (Abstract)
nelli, R. G., & Humphreys, L. G. Latent roots of random a correlation matrices with squared multiple correlations on diagonal: A Monte Carlo study. *Psychometrika*, 1976, 41, 1–347.
:, S. A. *The foundations of factor analysis*. New York: McGraw-l, 1972.
J. L. *Fundamentals of experimental design* (3rd ed.). Boston: yn & Bacon, 1979.
J. L., DiCecco, J. V., White, J. B., & Borden, V. M. Re-ted measures on dichotomous variables: Q and F tests. *Psy-logical Bulletin*, 1982, 92, 517–525.
M. J., Woody, C. D., Ungar, R., & Sharafat, A. R. Detec-n of neuroelectric signals from multiple data channels by imum linear filter method. *Electroencephalography and Clini-Neurophysiology*, 1975, 38, 191–198.
. P., & Sunderman, S. Before averaging: Preprocessing slow ential data with a Wiener filter. In D. Otto (Ed.), *Multidisci-ary perspectives in event-related brain potential research* (EPA-/9-77-043). Washington, D.C.: U.S. Government Printing ce, 1978.
nder, W. A., & Price, J. M. Dependent variable reliability the power of significance tests. *Psychological Bulletin*, 1978, 405–409.
C. L. On choosing a test statistic in multivariate analysis of nce. *Psychological Bulletin*, 1976, 83, 579–586.
C. L. Practical considerations in choosing a MANOVA test statistic: A rejoinder to Stevens. *Psychological Bulletin*, 1979, 86, 1350–1352.
Overall, J. E., & Woodward, J. A. Nonrandom assignment and the analysis of covariance. *Psychological Bulletin*, 1977, 84, 588–594.
Overton, D. A., & Shagass, C. Distribution of eye movement and eye blink potentials over the scalp. *Electroencephalography and Clinical Neurophysiology*, 1969, 27, 544–549.
Picton, T. W., & Stuss, D. T. The component structure of human event-related potentials. In H. H. Kornhuber & L. Deecke (Eds.), *Motivation, motor and sensory processes of the brain: Electrical potentials, behavioral and clinical use*. Amsterdam: North-Holland Biomedical Press, 1980.
Porges, S. W. Developmental designs for infancy research. In J. D. Osofsky (Ed.), *Handbook of infant development*. New York: Wiley, 1979.
Porges, S. W., Bohrer, R. E., Cheung, M. N., Drasgow, F., McCabe, P. M., & Keren, G. New time-series statistic for detecting rhythmic co-occurrence in the frequency domain: The weighted coherence and its application to psychophysiological research. *Psychological Bulletin*, 1980, 88, 580–587.
Press, S. J. *Applied multivariate analysis*. New York: Holt, Rinehart & Winston, 1972.
Ragot, R. A., & Remond, A. EEG field mapping. *Electroencephalography and Clinical Neurophysiology*, 1978, 45, 417–421.
Remond, A. Construction et ajustment des enregistrement cartographiques et spatio-temporals des EEG. *Revue Neurologique*, 1962, 106, 135–136.
Richards, J. E. *Multivariate analysis of variance of repeated physiological measures*. Paper presented at the annual convention of the Society for Psychophysiological Research, Vancouver, British Columbia, October 1980.
Roessler, F., & Mansey, D. Principal components and varimax-rotated components in event-related potential research: Some remarks on their interpretation. *Biological Psychology*, 1981, 13, 3–26.
Ruchkin, D. S., & Glaser, E. M. Simple digital filters for examining CNV and P300 on a single-trial basis. In D. Otto (Ed.), *Multidisciplinary perspectives in event-related potential research* (EPA-600/9-77-043). Washington, D.C.: U.S. Government Printing Office, 1978.
Ruchkin, D. S., Sutton, S., & Stega, M. Emitted P300 and slow wave event-related potentials in guessing and detection tasks. *Electroencephalography and Clinical Neurophysiology*, 1980, 49, 1–14.
Schnore, M. M. Individual patterns of physiological activity as a function of task differences and degree of arousal. *Journal of Experimental Psychology*, 1959, 58, 117–128.
Sencaj, R. W., Aunon, J. I., & McGillem, C. D. Discrimination among visual stimuli by classifications of their single evoked potentials. *Medical and Biological Engineering and Computing*, 1979, 17, 391–396.
Siegel, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.
Simons, R. F., Ohman, A., & Lang, P. J. Anticipation and response set: Cortical, cardiac, and electrodermal correlates. *Psychophysiology*, 1979, 16, 222–233.
Squires, K. C., & Donchin, E. Beyond averaging: The use of discriminant functions to recognize event-related potentials elicited by single auditory stimuli. *Electroencephalography and Clinical Neurophysiology*, 1976, 41, 449–459.
Squires, K. C., Donchin, E., Herning, R. I., & McCarthy, G. On the influence of task relevance and stimulus probability on event-related potential components. *Electroencephalography and Clinical Neurophysiology*, 1977, 42, 1–14.
Squires, K. C., Wickens, C. D., Squires, N. K., & Donchin, E. The effect of stimulus sequence on the waveform of the cortical event-related potential. *Science*, 1976, 193, 1142–1146.
Stern, R. M., Ray, W. J., & Davis, C. M. *Psychophysiological recording*. New York: Oxford University Press, 1980.

Sternbach, R. A. Some relationships among various "dimensions" of autonomic activity. *Psychosomatic Medicine*, 1960, 22, 430–434.
Stevens, J. P. Comment on Olson: Choosing a test statistic in multivariate analysis of variance. *Psychological Bulletin*, 1979, 86, 355–360.
Strong, P. *Biophysical measurements*. Beaverton, Oregon: Tektronix, 1970.
Tatsuoka, M. M. *Discriminant analysis: The study of group differences*. Champaign, Ill.: Institute for Personality and Ability Testing, 1970.
Tatsuoka, M. M. *Multivariate analysis: Techniques for educational and psychological research*. New York: Wiley, 1971.
Thurstone, L. L. *The vectors of mind*. Chicago: University of Chicago Press, 1935.
Tucker, L. R. *Note on number of attributes in a battery to support a given number of common factors*. Unpublished manuscript, 1973.
Ungar, P., & Basar, E. Comparison of Wiener filtering and selective averaging of evoked potentials. *Electroencephalography and Clinical Neurophysiology*, 1976, 40, 516–520.
Van Egeren, L. F. Multivariate statistical analysis. *Psychophysiology*, 1973, 10, 517–532.
Venables, P. H., & Christie, M. J. Electrodermal activity. In I. Martin & P. H. Venables (Eds.), *Techniques in psychophysiology*. Chichester, England: Wiley, 1980.
Venables, P. H., & Martin, I. (Eds.). *A manual of psychophysiological methods*. Amsterdam: North-Holland, 1967. (a)
Venables, P. H., & Martin, I. Skin resistance and skin potential. In P. H. Venables & I. Martin (Eds.), *A manual of psychophysiological methods*. Amsterdam: North-Holland, 1967. (b)
Walter, D. O. A posteriori "Wiener filtering" of average evoked responses. *Electroencephalography and Clinical Neurophysiology*, 1968, (Suppl. 27), 59–70.
Wastell, D. G. Statistical detection of individual evoked responses: An evaluation of Woody's adaptive filter. *Electroencephalography and Clinical Neurophysiology*, 1977, 42, 835–839.
Wastell, D. G. On the correlated nature of evoked brain activity: Biophysical and statistical considerations. *Biological Psychology*, 1981, 13, 51–69. (a)
Wastell, D. G. PCA and varimax rotation: Some comments on Roessler and Mansey. *Biological Psychology*, 1981, 13, 27–29. (b)
Weerts, T. C., & Lang, P. J. The effects of eye fixation and stimulus and response location on the contingent negative variation (CNV). *Biological Psychology*, 1973, 1, 1–19.
Wiener, N. *Extrapolation, interpolation, and smoothing of stationary time series*. Cambridge, Mass.: MIT Press, 1964.
Wilder, J. The law of initial values in neurology and psychiatry: Facts and problems. *Journal of Nervous and Mental Disease*, 1957, 125, 73–86.
Wilson, K. V. A distribution-free test of analysis of variance hypotheses. *Psychological Bulletin*, 1956, 53, 96–101.
Wilson, R. S. CARDIVAR: The statistical analysis of heart rate. *Psychophysiology*, 1974, 11, 76–85.
Winer, B. J. *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill, 1971.
Woodward, J. A., & Overall, J. E. Multivariate analysis of variance by multiple regression methods. *Psychological Bulletin*, 1975, 82, 21–32.
Woody, C. D. Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. *Medical and Biological Engineering*, 1967, 5, 539–553.
Woody, C. D., & Nahvi, M. J. Application of optimum linear filter theory to the detection of cortical signals preceding facial movement in the cat. *Experimental Brain Research*, 1973, 16, 455–465.