Principles of Neurobiological Signal Analysis

Edmund M. Glaser
Daniel S. Ruchkin

Principles of Neurobiological Signal Analysis

Edmund M. Glaser
Daniel S. Ruchkin

School of Medicine
University of Maryland
Baltimore, Maryland

ACADEMIC PRESS New York San Francisco London 1976
A Subsidiary of Harcourt Brace Jovanovich, Publishers

ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
24/28 Oval Road, London NW1

Library of Congress Cataloging in Publication Data

Glaser, Edmund M
Principles of neurobiological signal analysis.

Bibliography: p.
Includes index.
1. Neural analyzers—Mathematical models. I. Ruchkin, Daniel S., joint author. II. Title. III. Title: Signal analysis.
QP363.G58 591.1'88 76-42267
ISBN 0-12-285950-2

PRINTED IN THE UNITED STATES OF AMERICA

PREFACE

More years ago than we care to think of or mention, we convinced ourselves of the need for a monograph on the principles of signal analysis as applied to the electrical activity of the nervous system. This book is the result. Our premise in organizing it has been simple: that neurobiologists are generally uneasy in their use of signal analysis simply because they have had little formal training in the mathematics underlying its framework and that therefore they have little intuitive feel for what signal analysis procedures mean. Our goal, consequently, is to provide neurobiologists with a reasonably detailed discussion of signal analysis as it has been variously applied to neuronal signals. We wish to make them more aware of what these analyses can and cannot do, their implications, and limitations. We have used mathematics where it is essential, but in doing so we have tried to avoid unnecessary rigor. We have assumed that mathematically the reader is equipped with a hazy recollection of calculus. Our hope is that we can dispel most of this haze in the early going. On another front, we have consciously refrained from treating the cuisine of signal analysis. Recipes or programs for signal analysis are readily available for a variety of computers. We do not feel they provide much elucidation of the basic issues.

The first three chapters establish the theoretical groundwork of signal analysis. Chapter 1 presents an introductory discussion of the properties of signal and noise, especially as they apply to the nervous system. It reflects our judgment that the essential ingredients of neurobiological signal analysis are the related concepts of signal spectra and covariance functions. They are likely to remain so even as the present, predominantly linear methods of signal analysis are broadened to encompass nonlinear techniques. Chapter 2 discusses the methods of sampling and converting biological signals into sequences of digital numbers readily digestible by a computer. Chapter 3 then develops more thoroughly the concepts of spectrum and covariance analysis. This chapter is mathematically somewhat more demanding than the first two. Those who find it too trying should not feel distressed since much of what appears subsequently will still be comprehensible. The loss is in the appreciation of some of the analytic details.

Chapters 4 and 5 deal with techniques for extracting evoked responses from background noise and with multivariate statistical procedures for treating evoked response waveshapes as variables dependent upon the experimental manipulations performed upon a subject. Chapters 6 and 7 deal with the analysis of spike (action potential) activity generated by individual neurons and small groups of neurons. Chapter 8 presents methods for studying how such spike activity may be related to

the concurrently observed slow wave (EEG-like) activity of the nervous system.

A number of individuals have contributed to the completion of this work. It was Dr. José del Castillo who provided us with facilities at the Laboratory of Neurobiology of the University of Puerto Rico. It was there that this book had its inception. Drs. Donald Childers, Emanuel Donchin, and George Gerstein reviewed various chapters and provided much helpful criticism. A special note of thanks goes to Drs. José Negrete and Guillermina Yankelevich de Negrete who lent much encouragement during the initial tribulations of writing. Finally, we would like to express our special appreciation to Mrs. Frances Pridgen who, equipped with an extensive background as a legal secretary, typed the manuscript and suffered with us in guiding it to completion. In a moment of relaxation, when all was done, we asked her opinion of the work. She flipped slowly through its pages, smiled and said, "Naturally, this is taxable." We wonder.

CONTENTS

CHAPTER 3 POWER SPECTRA AND COVARIANCE FUNCTIONS

CHAPTER 6 SPONTANEOUS AND DRIVEN SINGLE UNIT ACTIVITY

Chapter 1

SOME PROPERTIES OF BIOLOGICAL SIGNALS

1.1. INTRODUCTION

Speaking in a somewhat general way, we say that all biological data can be considered to be signals. Obviously, however, some data are more signallike than others. The dividing line between data that can be profitably considered to be signallike and data that cannot depends upon both the origin of the data and how we propose to process it and analyze it conceptually. A discussion of the many facets of this idea in the light of modern computer data processing methods is one of the major purposes of this book. Embarking in this direction requires that we first establish some of the major concepts and properties of signals insofar as they relate to biological processes. The properties of these signals influence, guide, and sometimes determine the ways in which computer programs are developed to perform signal analysis.

Signal: A variation in the amplitude and polarity of an observed physical quantity produced by a process whose mechanisms we desire to understand by experimental investigation.

The requirement that the variation be produced by a mechanism we are interested in is of basic importance and brings us to consider at once, noise, the inseparable companion of signal.

Noise: A variation in the size of an observed physical quantity we are investigating produced by a process or an aspect of a process that we have no present interest in.

Data: Some combination, often additive, of signal and noise. The additive situations are easiest to deal with in terms of analysis and interpretation of results. In much of what follows we will assume it applies. In general, however, additivity should not be taken for granted.

The errant course of scientific progress is such that often what is considered to be a signal in one investigation turns out to be noise in another. Or more colloquially, one man's signal is another man's noise.

The variations in the size of a physical quantity are often time-dependent. When they are, the data is said to be a function of time and written $x(t)$. Temporal data variation is most convenient for us to consider and also most appropriate since a real-time computer generally accepts data in time sequential form. However, we may also profitably consider data which are functions of such variables as distances or angle, for it is usually a simple matter to convert them into functions of time by a signal transducer. As an example, a scanning densitometer converts the spatially varying density of a translucent object into a function of time as the densitometer is moved over the scanned object. An oscilloscope screen is an example of the process in reverse for there the time-varying data is converted into a function of distance along the horizontal axis of the oscilloscope screen. Hereafter, when we mention data signals and noise, we will consider them to be temporally varying.

We are interested in establishing the basic principles of a wide assortment of procedures by which we analyze the signallike data of neurobiological investigations. Temporally generated signals and noises exhibit a wide variety of waveform features or parameters, and it is essential to classify them according to such features, for the validity of much of the subsequent data processing depends upon the presence or magnitude of these features. The following pages contain a discussion of some of the properties of signals to serve as the basis of understanding the signal analysis procedures and techniques to be described in later chapters.

1.2. CONTINUOUS SIGNALS AND THEIR DISCRETE COUNTERPARTS

Let us begin with data which consist only of signals. A

signal is said to be continuous if it is defined at all instants of time during which it occurs. A continuous signal may, however, possess discontinuities or sudden changes in amplitude at certain instants of time. At these instants the slope of the signal is infinite. At other times the signal amplitude changes gradually so that by choosing an interval short enough, the corresponding change in amplitude can be made as small as we like. While continuous signals without discontinuities are the rule in such biological phenomena as the EEG, deliberately generated discontinuous signals may be generated by the instrumentation associated with neurobiological investigations. As an example, the signal produced by a rat when it pushes a switch to obtain food is discontinuous. This type of signal is referred to as a step function. Illustrations of continuous and discontinuous signals are shown in Fig. 1.1. It is also to be noted that whether continuous or not, the signals are always single valued: they have only one value at any particular instant in time. A particularly interesting and important discontinuous signal is the unit step signal of Fig. 1.1(c):

$$u(t) = \begin{cases} 0 \text{ when } t \leq t_d, \\ 1 \text{ when } t > t_d \end{cases} \tag{1.1}$$

t_d is the instant of discontinuity. The equation indicates that the signal jumps to 1 as soon as t becomes greater than t_d. The unit step is used, among other purposes, to describe a stimulus that has a sudden onset.

Besides speaking of a continuous signal, $x(t)$, we will also have occasion to speak of its time derivatives, the first derivative being written $dx(t)/dt$ or, alternatively, $x'(t)$. The first derivative is, of course, the time rate of change of the variable. When it is zero, the variable itself is at a local maximum or minimum value or, less frequently, at an inflection point. (The derivative of a constant signal is always zero.) This property is often used in determining when a spikelike waveform reaches a maximum or minimum. A peak detection device which essentially

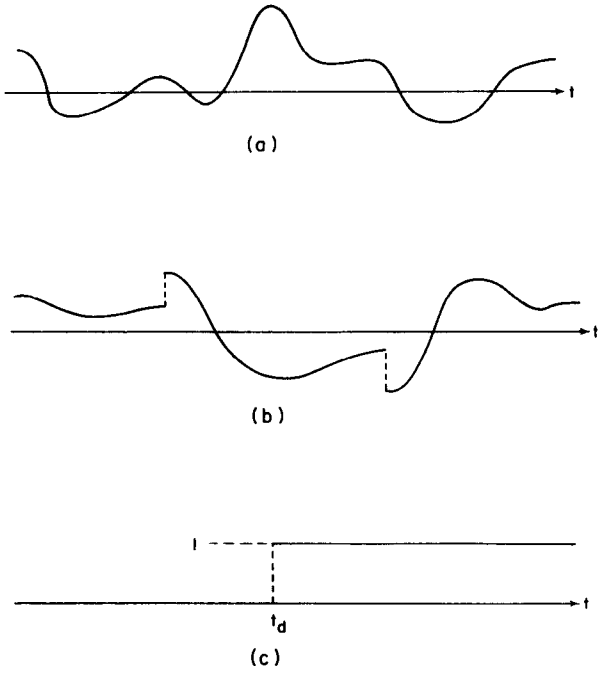

Fig. 1.1. (a) A continuous signal; (b) a discontinuous signal; (c) the unit step $\mu(t)$, showing step onset at $t = t_d$.

takes the time derivative of the waveform is commonly employed for this. When its output, the waveform time derivative, goes through zero in a negative direction, a positive maximum has occurred; when it goes through zero in a positive direction, a negative maximum has occurred. Figure 1.2(a) illustrates the situation for the former case. The first derivative is also important in indicating when the signal is changing most rapidly because it has its greatest value at that time. A positive maximum in the first derivative indicates the time when the signal is increasing most rapidly; a negative maximum, when it is decreasing most rapidly. Just as a continuous signal may exhibit discontinuities, so may its derivatives. A discontinuity in the first derivative occurs when there is a cusp in the original signal. An example is the sawtooth signal of Fig. 1.2(b). When it is at its maximum and

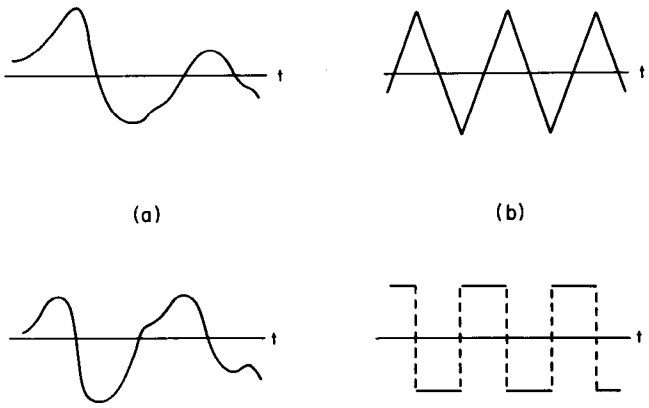

Fig. 1.2. (a) Above, a continuous signal; below, its time derivative. The negative and positive going zero crossings of the derivative correspond to positive and negative peaks in the signal. (b) Above, a periodic sawtooth signal; below, its time derivative which is a periodic discontinuous square wave.

minimum values, discontinuities occur in its first derivative, a square wave.

The derivative operation is not without practical difficulties since noise contributions tend to corrupt the derivative measurement. In computer analysis of data, the derivative operation is approximated by comparing successive sampled values of the signal with one another to see when maximum and minimum rates of change occur. Although this is an approximation, the results are often more than adequate. It is worth noting here that approximation is different from estimation, the latter being a statistical procedure whose meaning will be made clear in the subsequent pages.

In contrast to temporally continuous (T-continuous) signals are the temporally discrete (T-discrete) signals. These are signals which exist only at discrete instants in time. For our purposes the most important discrete signals are those which occur when a continuous signal has its amplitude measured or sampled at discrete instants of time that are usually equally spaced. A T-discrete signal is thus a sequence of measurements $x_1, x_2, \ldots,$

x_T lasting for the duration of the time the signal is observed. In digital data processing it is furthermore usually quantized in amplitude by an analog-to-digital converter. This gives it the property of being amplitude discrete (*A*-discrete). The result is a signal, *T*- and *A*-discrete, which provides the basic data thereafter for all subsequent computer analyses of the original signal.

Having introduced the continuous signal and its sampled *T*-discrete representation, it is useful to establish here a form of notation which permits us to distinguish between them with a minimum amount of confusion. We will use the symbol ° to distinguish a sampled *T*-discrete signal from its continuous source signal. We will drop the ° when no confusion seems possible. Similarly, we will use t to represent continuous time and $t°\Delta$ to represent those instants that a signal is sampled at a uniform rate. Δ is the interval between neighboring samples, and $t°$ is an integer-valued index: 1, 2, 3, ..., etc. Signal analyses are often most easy to describe when $\Delta = 1$. This results in no loss of generality. When there is no possibility of confusion, the Δ will be dropped.

The signals or data handled by a digital computer are discrete not only in time but also in amplitude. This arises from the fact that the amplitude of a signal at a particular sampling instant is represented as a number within the computer, a number containing a limited number of digits or bits depending upon the computer's structure. To arrive at this numeric representation a continuous signal is first transformed into its *A*-discrete amplitude version by quantization in an analog-to-digital (A-D) converter. At each sampling time the quantization procedure assigns to the signal amplitude one of a finite number of levels. This level has a numeric value which represents the sample in subsequent data analysis computations. The subject of A-D conversion, or quantization, is discussed more thoroughly in Chapter 2.

Perhaps the simplest way of reconstructing a continuous signal from a set of its samples is shown in Fig. 1.3. Here the signal is assumed to remain constant at its sampled value for the

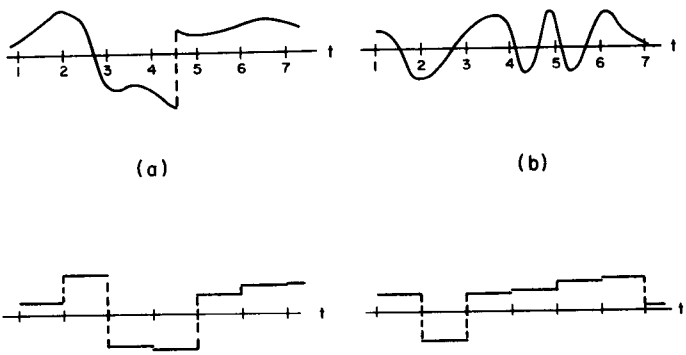

Fig. 1.3. (a) Above, a signal with a discontinuity between t = 4 and 5; below, a reconstruction of that signal by interpolation with a constant value between sampling instants. (b) Above, another continuous signal fluctuating rapidly between the 4th and 5th sampling instants. Note how the same type of sampling reconstruction totally lacks evidence of the rapid fluctuation of the original.

time interval between the present and the next sample time. It is important to recognize that the sampling and interpolation process can produce severe alterations of the signal depending upon the interrelationships between signal and sampling parameters. Two of the simplest errors are seen in Fig. 1.3 where in (a) a discontinuity is lost and in (b) a rapidly fluctuating component is suppressed because the sampling rate is too low. This type of error occurs regardless of how the interpolation between sampling instants is performed. A more thorough discussion of sampling problems is also presented in Chapter 3.

In some cases a signal is intrinsically *T*-discrete as for example is the count of the number of events occurring within an interval of time, such as the number of times an EEG waveform has a zero-crossing (a transition through zero amplitude) in one second. A second example is a list of measurements characterizing the structure of an object. It is important to note, however, that in the latter example the order in which the measurements are placed into a sequence may be of little or no importance. In temporal measurements or in measurements that are functions of a scanning