1.1 THE REASONS FOR STUDYING BIOMEDICAL SIGNAL PROCESSING

It may seem obvious that signal processing concepts should be part of the core training in biomedical engineering because students in such an applied field must learn to "deal with" signals and systems. The motivation for studying this topic, however, is more profound and can be related to fundamental approaches to conceptualizing and solving biomedical problems. A fundamental construct for interpreting both quantitative and qualitative data in all of biomedical engineering is the conceptualization that measurable, real-world behavior results from interactions among sources of energy and modifiers (or dissipators) of energy. Since a signal is just a record of energy production by a process, this fundamental framework naturally gives rise to questions concerning: (i) the inferences that one can draw from a signal about the properties of the source of the energy, and; (ii) the relationships among simultaneously observed energy records (i.e., signals). These are exactly the questions that signal processing addresses.

The biomedical engineering student should understand that signal processing (and the closely related field of systems analysis) comprises more than just mathematical manipulations that are only useful to control systems engineers. Indeed, these topics provide a fundamental framework for conceptualizing and analyzing physical behavior in a rigorously organized manner, whether that behavior is the output of an electronic control system, the flow of blood through a defective aortic valve, or the reaction of an implant with surrounding tissue (to give a very few examples). Furthermore, while the computational side of signals/systems can produce precise analyses, a qualitative understanding of these subjects is just as important. For example, a student proposing to use wavelets to detect abnormalities in the electrocardiogram signal surely needs to understand the mathematics of wavelet transforms. On the other hand, a student with neurophysiological interests who wishes to study the effects of whole-body vibration on visual function needs to understand the
qualitative concept of resonance (even if that student cannot remember the form of a second-order differential equation to quantify the phenomenon!). Similarly, a student who addresses neural control of heart rate needs to understand the concepts of memory or correlation and the causes of temporal variability in energy records, whether or not the student will utilize methods based on statistics of random processes to describe heart rate. The primary objectives of this textbook are to instill in the student a qualitative understanding of signal processing as an essential component of knowledge for a biomedical engineer and to develop the student's skills in applying these concepts quantitatively to biomedical problems.

A more direct approach to defining these objectives is to ask: "What skills should the student possess after studying this text?" Such skills fall into two broad categories: (1) skills related to the acquisition and processing of biomedical signals for extracting a priori desired information, and (2) skills related to interpretation of the nature of physical processes based either on observations of a signal or on observations of how a process alters the characteristics of a signal. More specifically, the objectives can be summarized by four phrases, which describe what the student should be able to do: (1) measure signals, (2) manipulate (i.e., filter) signals, (3) describe qualities of the source based on those of the signal, and (4) probe the source. Expanding these objectives, the student first should be able to quantify and (if necessary) compensate for the effects of measuring devices on the signal being measured. Second, the student should be able to identify and separate desired and unwanted components of a signal. Third, the student should be able to uncover the nature of phenomena responsible for generating the signal, as well as anticipate its future behavior, based on the identification and interpretation of an appropriate model for the signal. Fourth, the student should be able to relate the properties of a physical system to the differences in characteristics between a signal used to excite (or probe) the system and the measured response of the system.

The didactic framework adopted in this textbook to achieve the goals described above is one in which signal processing is presented as a process for the development and manipulation of a model of an observable variable (the signal). In this framework the "model" becomes a surrogate for the signal source, and one naturally asks to what degree this model is reasonable given one's knowledge of the signal source and one's application. Thus there is a natural connection to real-world problems. This approach is discussed in more detail later in this chapter.

Often in signal processing textbooks there is heavy emphasis on computation and algorithms. For students eager to expand their knowledge in signals and systems, this approach may become a boring repetition of familiar material; for students who see no need for this knowledge, the relevance of the material is not made apparent by such an approach. Yet both groups of students are missing the main message that physical (observable, real-world) behaviors can be described in an organized manner that permits both insights into the underlying physical processes and prediction of unobserved behavior. The "icing on the cake" is that these techniques are independent of the devices used to acquire signals. In other words, the methods are equally applicable to signals from heart rate monitors, signals from strain gauges in an Instron bone test device, spectra from chemical spectroscopy as-
1.2 WHAT IS A SIGNAL?

The electrical current induced in an antenna wire by the electromagnetic field transmitted from your favorite radio station, hourly readings of the atmospheric pressure, and the hum of a mosquito all are examples of signals (see Fig. 1.1). In each case the signal is the output of a sensing device (i.e., antenna, barometer, and ear) and the temporal variations in the signal convey information. To make sense of that information, however, may require some processing of the signal—for example, we learn to recognize the sound signal created by the pressure waves generated by the mosquito and associate its frequency with the insect and its itching bite. How the auditory system discriminates frequency is a complicated story that is not yet fully understood, but the underlying transduction of the pressure wave and manipulation of the resulting neural signals is an example of signal processing by a biological system.

A signal is a single-valued representation of information as a function of an independent variable (e.g., time). The specific type of information being represented may have real or complex values. In the case of physical processes, the information
THE NATURE OF BIOMEDICAL SIGNALS

FIGURE 1.2 (a) Instantaneous mean blood flow velocity in the middle cerebral artery of a human subject obtained from the Doppler shifts of a reflected ultrasound beam. Time marker = 1 s; (b) Electromyogram (EMG) from two wires held firmly under the tongue by a mouthpiece. Subject contracted tongue, then relaxed. Time marker = 0.2 s; (c) Angle of rotation of knee obtained from an angle sensor (Data courtesy of James Abbas); (d) An electrocardiogram (ECG) recording (Data courtesy of Abhijit Patwardhan); (e) Instantaneous heart rate (beats/min, reciprocal of beat-to-beat interval) for 100 consecutive heart beats. Implicit independent variable is
1.3 SOME TYPICAL SOURCES OF BIOMEDICAL SIGNALS

The sources of biomedical signals are infinite in variety and we shall identify only a few as examples of the different types of these signals. A large class of biomedical

---

"heart beat number"; (f) x-axis: Intensity of fluorescence from cells excited by incident laser light beam, y-axis: Number of cells displaying a given intensity. From cells stained with a fluorescent dye that binds to phospholipids (Data courtesy of Margaret Bruce).

often is a measure of some form of energy produced by the process. Signals that arise from strictly mathematical processes and have no apparent physical equivalent are commonly considered to represent some form of mathematical energy. A signal may be a function of another variable besides time or even a function of two or more variables, but often we will consider time-dependent functions. Indeed, a great many signals in practice are real-valued, scalar functions of time. Nonetheless, the reader will be reminded throughout that many important biomedical signals are not time-dependent functions—for example, in image processing the intensity values are functions of x and y coordinates—and that the methods being presented apply equally to such signals. For practical purposes, often one can define a signal as the output from a measuring or sensing device (or the result of mathematical operations on such an output), although this definition is too restrictive for some biomedical applications. For example, I can create a function to represent the bases in a DNA molecule by assigning a unique number from 1 to 4 to each of the four bases. Then if I plot the position of each base along the DNA strand as the independent variable and the value of the function for that base as the dependent variable, I have created a signal that is not derived from a sensor or measuring device (nor is it a function of time). As we shall see, if I can represent the signal mathematically, then I can manipulate it using the concepts from this book.

1.3 SOME TYPICAL SOURCES OF BIOMEDICAL SIGNALS

The sources of biomedical signals are infinite in variety and we shall identify only a few as examples of the different types of these signals. A large class of biomedical
signals comprises those which are electrical in nature. The electrocardiogram (ECG), which can be recorded by placing electrodes on the chest, is one such signal. It records the electrical potential at the electrode (or the potential difference between two electrodes) induced by the presence of time-varying electrical activity in cardiac muscle associated with the generation and propagation of action potentials. Each heartbeat produces a sequence of electrical waves (the P, Q, R, S, and T waves) and by examining the shape of the ECG waveforms (Fig. 1.2(d)) a physician can obtain considerable insight about whether the contractions of the heart are occurring normally. In practice, the physician usually compares the electrocardiogram signals from several sites on the body. Although visual inspection of ECG signals is useful clinically, analyses of electrocardiograms can be very sophisticated and involve considerable signal processing. For example, there is substantial interest in detecting small changes in the shape of the ECG waveform occurring after the large peak (the R-wave) during the time that the ventricles are recovering from their contraction. Several approaches being used involve spectral analysis of the ECG signal, a technique which will be discussed in Chapter 9. Another question of clinical relevance is whether the heart is beating regularly or irregularly. (Believe it or not, too much regularity of the heartbeat is considered unhealthy!) Some of the most sophisticated biomedical signal processing in the research literature is aimed at characterizing the type and degree of irregularity of the cardiac rhythm. A portion of the irregularity in heart rate seems to be spontaneous random changes caused partly by random events at the membranes of sinus node pacemaker cells. As we shall study in a later chapter, however, the variations in heart rate at times that are several tens of seconds apart are related. Recently developed signal analysis techniques for characterizing such temporal dependencies are among the most sophisticated and exciting new methods for biomedical signal processing. Chapter 10 will introduce these methods.

Another example of a bioelectrical signal is the electromyogram (EMG). EMG signals are recorded by placing electrodes in, on, or near a muscle and amplifying the electrical potential (or potential difference between two electrodes) that results from the generation and propagation of action potentials along muscle fibers. A multiple-unit EMG (MU EMG) records the potentials from many muscle fibers at once (Fig. 1.2(b)). The presence of MU EMG activity indicates that a muscle is actively contracting and many researchers have attempted to develop techniques for filtering of the MU EMG signal to produce another signal that is proportional to the force that the muscle is generating. This goal is important because it would allow one to estimate muscle force without having to connect a force transducer to the muscle—something that is often impossible to do on human subjects. (How, for example, could one attach a force transducer to the extraocular muscles?) MU EMG signals are utilized also in rehabilitation engineering for controlling a prosthetic device. The subject is trained to contract a muscle that he or she can still control voluntarily and the amplitude of the MU EMG is taken as a measure of the desired degree of activation of the prosthesis—for example, the gripping force to be produced by an artificial hand. But since the MU EMG is inherently noise-like, it must be filtered to extract a smoothed signal which varies with the
1.3 SOME TYPICAL SOURCES OF BIOMEDICAL SIGNALS

The amplitude of the MUEMG. Other examples of bioelectrical signals are the electroretinogram (ERG), electrogastrogram (EGG), and the electroencephalogram (EEG), which is a recording of brain electrical activity from electrodes on the scalp. Processing of EEG signals is an active research area because the EEG is used to determine if a person is asleep and to identify the sleep state. For example, the EEG contains predominantly oscillations in the 8–12 Hz range (alpha rhythm) when the person is awake. In contrast, predominant oscillations in the 1–4 Hz range (delta rhythm) indicate non-dreaming (also called slow-wave or non-REM) sleep.

In Chapters 6 and 8 we will study both analog filters and their digital counterparts that could be used to filter bioelectrical and other signals. We study digital filters because so much of biomedical signal processing is performed on sampled (i.e., digitized) signals. Furthermore, it will become apparent that digital filtering has an important advantage, in that there are many digital filters which have no analog counterpart.

The field of imaging provides many examples of both biomedical signals and biomedical signal processing. In magnetic resonance imaging (MRI) the basic signals are currents induced in a coil caused by the movement of molecular dipoles as the molecules resume a condition of random orientation after having been aligned by the imposed magnetic field. These current signals are complex and substantial. Signal processing is required just to properly detect and decode them, which is done in terms of the spatial locations of the dipoles that caused the currents and the rate of relaxation of the dipoles (which is related to the type of tissue in which they are located). Much of the associated signal processing is based on Fourier transforms, which will be studied in Chapters 5 and 7. Although this book will consider only one-dimensional Fourier transforms and MRI utilizes two-dimensional Fourier transforms, the basic concepts are the same. In addition, once an image is constructed it is desirable to “process” it to enhance the visibility of certain features such as the edges of a tumor. Although there are many advanced methods of image processing involving techniques beyond the scope of this book, the basic techniques are based on concepts derived from digital filtering theory, which we shall encounter in Chapter 8. Indeed, even though other imaging modalities such as positron emission tomography (PET), ultrasound, and x-ray utilize different physical principles for acquisition of the image, the signal processing methods for enhancing images are similar.

When an ultrasound beam strikes a moving object, the frequency of the reflected beam differs from that of the incident beam in proportion to the velocity of the object (Doppler shift). Because high-frequency ultrasound signals can penetrate hard biological tissues (such as thin bones), this property provides a means of measuring velocities of inaccessible tissues such as blood cells (Fig. 1.2(a)). Although this measurement is not a direct assessment of the bulk flow of blood, it is used in humans to identify vessels in the brain in which flow is sluggish.

Sensors that transduce mechanical actions into electrical signals are common in biomedical engineering applications. In studies of whole-body biomechanics, accelerometers record the acceleration of limbs or joints (Fig. 1.2(c)) and force plates
use strain gauge sensors to measure the reaction forces where the feet contact the ground. Although the outputs of such sensors are electrical signals, it is usual to calibrate the electrical signals in terms of the mechanical signal being measured and to think of the electrical signal as the mechanical signal. That is, one does not describe the accelerometer output as “the electrical equivalent of the acceleration signal” but as the acceleration signal itself. Since accelerometers are sensitive to noisy vibrations, it is usually necessary to remove the noise on (computer-sampled) accelerometer signals using digital filters, such as those described in Chapter 8.

Force transducers, displacement transducers, and pressure transducers are additional types of sensors which produce electrical signals that correspond to mechanical actions. Some of these transducers can be made quite small—for example, to fit inside an artery for measuring blood pressure. Some can be miniaturized for measurements, for example, from single excised muscle fibers or from cells grown in culture. In some of these situations the transducer modifies the measured behavior because its mechanical characteristics are not negligible compared to the system under study. In such cases it is necessary to correct the measured signal for this “loading” effect. Often this correction comprises designing a digital filter based on the frequency response of the transducer. The concept of frequency response will be examined in several chapters, beginning in Chapter 4, and the difficulties in designing these inverse filters will be discussed in Chapter 8.

As biomedical engineers become more involved in probing behavior at the cellular and molecular levels, sensors for events occurring at the microscopic level have assumed greater importance. Often these sensors directly or indirectly measure products of biochemical reactions. For example, the concentration of calcium ion in a cell can be measured indirectly by introducing into the cell a chemical (Fluro2) which fluoresces with an intensity that is a function of the calcium concentration. One then focuses a microscope on the cell in question and directs the light at the eyepiece to a photocell through an optical system that is tuned to the frequency of the fluorescent light. The electrical output from the photocell is proportional to the calcium concentration. A similar approach using a voltage-sensitive dye permits one to measure intracellular potentials by optical means. Other types of optical biosensors measure the absorption (or reflectance) of an incident light beam containing one or a few frequencies of light. For example, the hydrogen ion concentration in a thin tissue can be determined by staining the tissue with a particular red dye and measuring the relative transmission of light through the tissue at two different frequencies. Some newer approaches to measuring the concentration of protein in a sample utilize a bioactive electrode that reacts with the desired protein. For certain reactions, the reflectance of an imposed laser light having a specific frequency is proportional to the concentration of the reactant. Again the signal of interest is derived by leading the reflected beam to a photocell or photosensitive solid-state device (i.e., CCD). Often in such sensor systems the analysis of the signal involves taking a ratio or some nonlinear processing such as logarithmic scaling. For ratios or nonlinear scaling the noise might be amplified disproportionately, so it becomes important to filter the signal from the photocell to reduce this effect. Furthermore,
there may be contamination by background light which should be removed by filtering. The digital filters we will analyze in Chapter 8 could be used for these purposes. Finally, if there is motion of the environment, movement of such small sensors might introduce a time-varying artifact. In some cases one can reduce this latter type of artifact using adaptive filtering techniques, but these methods are beyond the scope of this book.

Up to this point the discussion has assumed implicitly that biomedical signals comprise a combination of a desired component which represents a true measure of a physical behavior and an undesired component categorized generally as noise. We need to recognize that certain biomedical signals which visually resemble noise actually represent true physical behavior. If, for example, one uses a fine-tipped glass microelectrode to measure the current flowing through a tiny (i.e., a few microns square) piece of a cell membrane, the fluctuations of the current with time appear to be rather random but this apparent randomness can be explained by applying certain simple rules which determine the conformation of proteins that form the channels for conducting ions through the membrane. The consequence of these rules is that the openings and closings of channels for ionic current flow exhibit a self-similarity across many scales of time. This self-similarity causes the appearance of rapid fluctuations in the signal, thus creating the illusion of unpredictableness (i.e., randomness) in a signal that is obeying simple rules. Such signals are said to be fractal. Chapter 10 will introduce the analysis of fractal signals.

1.4 CONTINUOUS-TIME AND DISCRETE-TIME

Many biomedical measurements such as arterial blood pressure or the torque at a joint are inherently defined at all instants of time. The signals resulting from these measurements are continuous-time signals and it is traditional to represent them as explicit functions of time in the form $x(t)$, $p(t)$, and so forth. On the other hand, we could sample the values of a continuous-time signal at integer multiples of some fundamental time increment and obtain a signal that consists of a sequence of numbers, each corresponding to one instant of time. Often we ignore the time increment and represent this discrete-time signal as a function of the integer multiple of the time increment, saying that the argument of the function can assume only integer values—that is, the first entry in the sequence corresponds to zero times the time increment and is called $x[0]$, the second entry corresponds to one times the time increment and corresponds to $x[1]$, and so on. Notice that in order to distinguish continuous-time and discrete-time signals, their arguments are enclosed in different types of brackets.

**Example 1.1** An example of obtaining a discrete-time signal by sampling a continuous-time signal at a constant rate is shown in Fig. 1.3. Note the different ways of graphically representing continuous-time (CT) and discrete-time (DT) signals.
Discrete-time signals can arise from inherently discrete processes as well as from sampling of continuous-time signals. Within the biomedical realm one can consider the amount of blood ejected from the heart with each beat as a discrete-time variable, and its representation as a function of a time variable (which assumes only integer values and increments by one with each heartbeat) would constitute a discrete-time signal. If one were counting the number of cells in a culture dish at hourly intervals, then a plot of this count versus an integer index of the time would be a discrete-time signal. Figures 1.2(e) and 1.2(f) show “discrete-time” signals which, however, are displayed as if they were continuous-time by connecting consecutive points with interpolating straight lines. Often we aggregate data by counting events over a fixed time interval—for example, the number of breaths taken by a subject in each minute, the number of action potentials generated by a neuron in each second—and thus create a discrete-time signal from an inherently continuous-time signal (e.g., Example 1.2 and Fig. 1.4).

**Example 1.2** Figure 1.4 demonstrates a common method for obtaining a DT signal from biomedical data. The top tracing is a representative extracellular recording from a neuron, showing its action potentials during a 10-second time interval. The number of action potentials which occur each second are counted and graphed in two different ways. In the middle tracing a CT signal is constructed by holding a constant value during each one-second period. This constant value equals the action potential count during that time period. In the bottom tracing a DT signal is constructed from these same data by plotting single values, representing the number of action potentials which occurred during the previous second, at one-second intervals.
1.5 ASSESSING THE RELATIONSHIPS BETWEEN TWO SIGNALS

The above examples have the common theme of extracting information from observations (i.e., measurements) of spontaneous behavior. Another broad area of signal processing involves assessing the relationship between two signals. By discussing such a relationship, we are implying that there exists some interconnection between the sources of these signals. Figure 1.5 indicates the two general possibilities for this interconnection: (1) one signal directly activates the source of the second signal; (2) a third signal activates the separate sources of the initial two signals. To continue this discussion we need to formally define what the boxes in Fig. 1.5 represent. In the most general sense these boxes, which will be called systems, represent any mechanism through which one signal depends on (or is derived from) another signal, and we shall define a system as any physical device or set of rules which transforms one variable (i.e., signal) into another variable (or signal). A typical system has an input variable and an output variable, although some systems have an output with no explicit input. These latter systems are called autonomous systems, whereas those having an explicit input are called non-autonomous systems. It should be noted that a system may be a physical device (or collection of devices which interact with one another) or may simply be a set of rules which can be written down. Because this book is a signal processing text, we shall place considerable emphasis on the viewpoint that a system transforms one signal into another signal and we shall devote some effort to formal descriptions of various classes of trans-

**FIGURE 1.4** Constructing CT and DT representations of the firing rate (number of action potentials per second) of a neuronal spike train. See Example 1.2 for further explanation.
formations. Transformations will be symbolized as $T[.]$, as shown in Fig. 1.5. Often we will utilize the attributes of linear transformations, which obey certain properties that are discussed in Chapter 2. An example of a physical device that can be considered a system is an electronic amplifier into which an electrical signal can be led, and which produces an output that is a scaled version of the input signal. A pressure transducer can be described as a system whose input and output are the pressure on the sensor and the electrical voltage at its output terminals, respectively. Analysis of the blood flow through an artery may be cast in the framework of a system in which the input is blood pressure, the system comprises a description of the mechanical properties of the arterial wall and the blood, and the output is the blood flow rate. Rule-based systems are common in the fields of artificial intelligence and fuzzy logic. In biomedical applications a simple model of information processing by the nervous system may be based on a set of rules since detailed neurophysiological information may be lacking. A system may have more than one input. Consider, for example, the system comprising the extraocular muscles and the eyeball. The inputs to this system are the neural signals that activate the muscles and the output is the position of the eyeball in its orbit. If one were to consider the control of both eyes simultaneously, then the previous multiple-input, single-output system would become a multiple-input, multiple-output system.

**Example 1.3** The systems framework is very important in many applications of biomedical signal processing. In developing drug therapy regimens, for example, one needs to know how the target behavior will respond to a given amount of drug. Thus, when a drug such as sodium nitroprusside is given to maintain blood pressure, it is desirable to be able to predict the amount of drug needed to raise the blood pressure by the desired amount. Conceptually, one could inject several different levels of the drug, measure the blood pressure responses, and establish a look-up table for determining the drug dose. In some applications this approach may be viable,
but a better alternative will be developed in Chapter 3. For many systems one can calculate the response of the output variable of a system (e.g., blood pressure) to any input signal (e.g., amount of drug injected into the blood) if one knows the response of the output to certain specific input signals. Even when this method is not directly applicable, it is often possible to approximate the behavior of a system over restricted ranges of its input and output signals.

Even in the case of observing spontaneous behavior (when there is no observed input signal) it is possible to apply signal processing methods by pretending that the signal being observed is the output of a system. The choice of a specific signal processing method will dictate what the fictitious "input" signal is assumed to be. This conceptual framework, which is developed later in this chapter, provides a common basis for comparing signal processing methods and explicitly highlights implicit assumptions underlying the use of each method.

Obviously the above illustrations are only a few of the many examples of biomedical signals and signal processing and many more will be introduced in the discussions and homework problems of this book. The student surely will encounter many others in his or her own work. The purpose of presenting these examples is in part to illustrate the breadth of biomedical engineering. Another purpose is to demonstrate that signal processing is a critical element of biomedical engineering activities, whether they address problems at the whole body, organ system, tissue, cell, or molecular level.

1.6 WHY SIGNALS ARE "PROCESSED"

Signal processing can be defined as the manipulation of a signal for the purpose of either extracting information from the signal, extracting information about the relationships of two (or more) signals, or producing an alternative representation of the signal. Most commonly the manipulation process is specified by (a set of) mathematical equations, although qualitative or "fuzzy" rules are equally valid. There are numerous specific motivations for signal processing but many fall into the following categories: (1) to remove unwanted signal components that are corrupting the signal of interest; (2) to extract information by rendering it in a more obvious or more useful form; and (3) to predict future values of the signal in order to anticipate the behavior of its source. The first motivation clearly comprises the process of filtering to remove noise and this motivation will be encountered in almost every chapter, as most methods of signal processing implicitly provide some basis for discriminating desired from undesired signal components. (An important issue is the basis on which the user decides what is desired signal and what is noise!) The general problem of discriminating noise from signal is discussed later in this chapter and the major focus of Chapter 8 will be the use of filters for noise removal.

The idea of applying signal processing to extract information is pervasive in biomedical applications. Often the objective is to discriminate abnormal from nor-
mal signals and on this basis diagnose the presence of disease. Just as the physician utilizes auditory discrimination to detect abnormal heart or lung sounds, biomedical engineers often separate a signal into a sum of basic signal types in order to detect the presence of abnormal signals that are suggestive of disease. Many of these approaches involve searching for unusual features in the Fourier transform of a signal. For example, in several cardiorespiratory diseases, such as congestive heart failure, blood pressure oscillates with a period of a few or several tens of seconds. To detect this oscillation one can examine the power spectrum (which is proportional to the squared magnitude of the Fourier transform) of blood pressure. On the other hand, we shall find that many biomedical signals do not adhere well (in a practical sense, not in a mathematical sense) to the basic premise of the Fourier transform: that the signal can be expressed as a sum of sinusoids. Consequently, an active and important area of research is to develop alternative methods for decomposing a signal into basic signal types that better represent the properties of biomedical signal sources.

The need to predict future values of a signal arises in two different situations. Most commonly this need occurs when one is interested in controlling a behavior—for example, regulating the blood glucose level by periodic injections of insulin. Because the result of any control action requires some finite time before its effect is evident, it is useful if one can predict the behavior of interest a short time in the future. In this way one can apply a control action based on the anticipated behavior at the time that the control action will be effective. The concepts of control theory underlying such applications are not discussed in this textbook, but in Chapters 3 and 9 we will address the question of how one processes a signal to predict its future values. The other biomedical situation in which the prediction of future values is important is the early detection of the onset of a disease process. Often in this case the problem is to predict the limits of future normal behavior so that small deviations from normalcy which might signal the onset of disease may be identified. This type of application has been a "Holy Grail" of biomedical signal processing since its inception. It is exemplified today, for example, by attempts to correlate indices of fractal behavior of heart rate with the presence or absence of disease. This area is an active research field whose theoretical advances are driven, in part, by the need for better methods for applications involving biomedical problems.

### 1.7 TYPES OF SIGNALS: DETERMINISTIC, STOCHASTIC, FRACTAL AND CHAOTIC

As one examines the examples of signals in Fig. 1.2, there is the impression of qualitative differences among them. Some appear to be smooth and one has the impression that it is possible to predict how these signals would behave beyond the period of measurement. In contrast, other signals are quite variable and give the impression that it would be difficult to predict their exact future behaviors. Such impressions are correct and reflect a fundamental problem for biomedical signal processing: There are different types (or classes) of signals. The class to which a signal belongs
strongly influences the inferences one can draw from the signal regarding the source of the signal and determines, to a certain extent, the applicable signal processing methods.

We shall recognize four classes of signals: (1) deterministic, (2) stochastic, (3) fractal, and (4) chaotic. Examples of these types of signals are shown in Fig. 1.6. Deterministic signals are encountered commonly in textbook examples but less frequently in the real world. A deterministic signal is one whose values in the future can be predicted exactly with absolute confidence if enough information about its past is available. Often one only requires a small amount of past information. For example, once I know one cycle of a sine wave, I can predict its value at any time in the future. All of the common textbook signals such as impulses, steps, and exponential functions are deterministic. In fact, any signal which can be expressed exactly in closed mathematical form as a function of time is deterministic.

Stochastic signals are signals for which it is impossible to predict an exact future value even if one knows its entire past history. That is, there is some aspect of the signal that is random, and the name “random signal” is often used for these signals. This book will follow this practice but it should be noted that some random signals are completely unpredictable (i.e., uncorrelated), whereas others can be predicted with greater (but not absolute) confidence. For example, predictions of the direction of change of these latter signals may be right more often than not, even though predictions of their exact value will almost certainly be incorrect every time.

Random signals are abundant in physical processes. Noise generated by electronic components in instrumentation is a common type of random signal that is present in much biomedical data. Although contemporary electronic design mini-

(a) Deterministic

(b) Filtered white noise

(c) Fractal noise

(d) Chaotic

FIGURE 1.6 Examples of the four types of signals. Note the visual similarities. The reader should explore other examples of these signal types using the m-file sigtype.m. See text or m-file for description of how these signals were generated.
mizes this noise, it can be significant compared to signals from microminiature biomedical sensors such as extracellular potentials recorded with glass microelectrodes. Almost all bioelectrical measurements contain random noise caused by random fluctuations of ionic currents and by the stochastic nature of action potential generation. In addition, because most biomedical systems are complex, it is not possible to appreciate all of the factors that influence a measurement. Often we classify those parts of a signal which are not understood as “noise.” Because of our ignorance, these signal components may appear to vary randomly relative to known mechanisms, thereby reinforcing the impression of stochastic behavior. Furthermore, it is often valid to treat these signals as stochastic.

The presence of fractal signals in biomedical processes has become widely recognized in the last decade. Fractal signals (Fig. 1.6(c)) have the interesting property that they look very similar at all levels of magnification, a property referred to as scale-invariance. For example, if I create a new signal from the one in Fig. 1.6(c) having one-fourth of the time resolution (i.e., by averaging the first four consecutive points, then the next four points, etc.), I cannot discriminate between that and the initial one-fourth of the original signal using many of the usual measures for quantifying a signal (Fig. 1.7). Visually they are not the same signal but they look very much alike. You might expect that random signals would exhibit this property also, but there are important quantitative differences between the scaling properties of fractal signals and of random signals. These differences will be discussed in Chapter 10. There is very good evidence that a part of the beat-to-beat heart rate signal (e.g., Fig. 1.2(e)) is fractal, as well as the signal representing current through a single ionic channel of a cell membrane. It is likely that many other biomedical signals are fractal. Furthermore, the concept of fractals can be applied to spatial variations, such as the branchings of blood vessels or airways or inhomogeneities in an electrode, as well as to temporal variations. More applications of spatial fractals in biomedicine are likely to appear. If a signal is shown to be fractal, then the challenge is to understand how the structure and properties of its source could produce the scale

![Image](image.png)

**FIGURE 1.7** The similarity of fractal signals observed at different scales is seen by comparing (top) 50 points from the signal of Fig. 1.4(c) with (bottom) a 50-point signal constructed by averaging groups of four consecutive data points of the same signal.
invariance described above. Many of the simple physical processes that are often invoked as sources of biomedical signals cannot reproduce this property. Directly proving that a signal is fractal is difficult at this time, although there are fairly reproducible (but complex) techniques for determining indices of scale-invariance. One’s confidence in a specific calculated value for any of these indices, however, may not be strong.

Chaotic signals are deterministic signals that cannot be predicted exactly in the future. The apparent contradiction in this definition is explainable on the basis of the “sensitive dependence on initial conditions” of chaotic signals. For some deterministic signals their trajectories in the future are so sensitive to their past values that it is impossible to specify those past values with sufficient accuracy that we can predict ahead in time with certainty. Thus, in theory, these signals are deterministic, but beyond a short time into the future the error of the prediction becomes very large. Because a chaotic signal is not fully predictable, visually its behavior has some characteristics of a random signal (but random signals are not chaotic and chaotic signals are not random!). As was the case with fractal signals, it is only in the last decade or so that researchers have recognized that biomedical systems could generate chaotic behavior that is reflected as chaotic signals. Again, the challenge is to understand the underlying structure and properties of a physical process which could produce this sensitive dependence on initial conditions and the unpredictable fluctuations in the signal.

Proving that a signal is chaotic is difficult at this time and developing new methods for such proof is an active research topic. One of the difficulties is the essentially universal presence of stochastic signal components, which seriously corrupt the analyses for chaotic signals. Another difficulty is that a process which is chaotic under some circumstances may not be chaotic under a different set of conditions. There is evidence, however, that some immunological and biochemical regulatory processes can exhibit chaotic behavior, and that EEG activity and breathing may have chaotic characteristics. Several neurophysiological systems, ranging from multicellular oscillators to single neurons, have been reported to exhibit chaos and the spread of disease in recurring epidemics has been assessed from this perspective also. Although evidence of chaotic behavior in real-world signals has been circumstantial, often it is possible to demonstrate the potential for chaotic behavior in a mathematical model of a biomedical process and then infer that the real-world process might also exhibit this behavior.

Example 1.4 Figure 1.6 displays signals of the four types discussed above which were generated using the file sigtype.m. The signals were constructed as follows: The deterministic signal is a sum of five sine waves having randomly selected values of amplitude and frequency. The stochastic signal is the output of a linear, lowpass filter whose input is uncorrelated white noise. The fractal Brownian noise signal was constructed using the spectral synthesis method with 300 frequency components and \( H = 0.8 \). The chaotic signal is the solution for the Henon map with parameters \((1.4, 0.3)\) and randomly chosen initial conditions.
You may create other examples by running the file `sigtype.m`, which will generate other examples of these signal types. It is strongly recommended that you examine other examples of these types of signals because different realizations from these examples can be visually quite different. Use the MATLAB command `help sigtype` to retrieve information about the parameter values in `sigtype`.

1.8 SIGNAL MODELING AS A FRAMEWORK FOR SIGNAL PROCESSING

Signal processing was defined above as the manipulation of a signal for the purpose of either extracting information from the signal (or information about two or more signals) or producing an alternative representation of the signal. The parallels between this definition and the previous discussion of a system should be obvious. In signal processing the original signal is analogous to the input signal of a system, the rules specifying the manipulation process are analogous to the system itself, and the information (in the form of a derived signal) or the alternative representation of the original signal is analogous to the system output (Fig. 1.8). “Manipulation process” is simply another name for the transformation that specifies the input–output relationship of a system. The viewpoint of this book is that signal processing can be formulated as a process of applying a transformation to a signal. In this case, however, one does not necessarily describe the transformation in the same terms that would be used to describe an electronic amplifier, for example. A transformation is just a compact way to describe how to combine current and past (and perhaps future) values of the input signal in order to determine the current value of the output signal. The algorithms for signal processing often are not mathematically simple, but they can be expressed in a form that by analogy accomplishes this same end. An advantage of this framework is that signal processing algorithms can be grouped according to their similarities, and generalizations about the group can be derived. This structure will be called the analysis framework for signal processing. As an example, consider the construction of a Fourier series representation of a periodic signal. As shown in Fig. 1.9(a), the input to the “signal processing system” is the periodic signal, the system in this case is the set of equations for calculating the Fourier se-

![FIGURE 1.8 System model of signal processing.](image)
Fourier Series

**Equations**

\[ \sin(\omega_0 t) \]

\[ \cos(\omega_0 t) \]

\[ \sin(n \omega_0 t) \]

\[ \cos(n \omega_0 t) \]

\[ a_0 \]

\[ b_0 \]

\[ \cdot \]

\[ \cdot \]

\[ a_n \]

\[ b_n \]

\[ \cdot \]

**FIGURE 1.9** (a) Signal processing analysis model of Fourier series calculations; (b) Signal processing synthesis model of a periodic signal based on Fourier series summation of sinusoidal inputs.

Another way to view this process is shown in Fig. 1.9(b). Here we represent not the analysis process but the model of the signal that is implied by the Fourier series analysis. This “signal modeling” structure will be called the *synthesis* framework for signal processing. This approach explicitly presumes that the observed periodic signal can be represented as a specific transformation of a large number of harmonically related sine waves. The transformation \( T[\cdot] \) is exactly the usual equation for a Fourier series expansion with the appropriate values for the coefficients. It should be apparent that knowledge of the input signals and the transformation equation(s) is equivalent to knowledge of the output signal. Both the analysis and the synthesis frameworks will be important for developing signal processing methods and the reader is strongly urged to clarify in his own mind the distinctions between these two approaches. (This specific example will be developed in detail in Chapter 5.)

**Example 1.5** The synthesis framework for Fourier series analysis can be carried a step further by utilizing a normalized time scale—that is, let the duration of one cycle be normalized to one and the frequencies of the input sine waves be integer multiples of unity. Then, in order to construct the output signal another parameter is needed to complete the description of the input signals, the fundamental frequency.
to be used. Now this model structure for Fourier series analysis is analogous to the ARMA model structure for random signals, which is studied in Chapter 9. In the latter case the input signals are sample functions of a white noise process and the additional parameter that is needed is the variance of the white noise.

A differential equation that expresses a relationship between a function, say \( y(t) \), and its derivatives and a forcing function, say \( x(t) \), and its derivatives may be considered to be a signal processing model in the synthesis form. That is, the differential equation relates how one “constructs” the current value of \( y(t) \) by adding (or subtracting) current and past values of \( x(t) \) and past values of \( y(t) \). To visualize this result, think of the approximation

\[
\frac{dy(t)}{dt} \approx \frac{y(t) - y(t - \Delta t)}{\Delta t}.
\]

Consequently, any situation that can be described by a differential equation also can be considered as an example of signal processing.

1.9 WHAT IS NOISE?

To a considerable degree the direct answer to this question does not lie in the realm of signal processing. Rather, the ultimate basis for deciding what constitutes noise should derive from considerations about the experimental or clinical measurements and the known properties of the source of a signal. Ideally the biomedical engineer decides a priori the criteria for judging whether certain components of a signal represent the desired measurement or not. Then the signal processing method is chosen to enhance the desired signal and reduce any undesired signal components. That is not to say that the user should be able to visually identify the noise in the signal before any signal processing occurs; but he or she ought to be able to specify properties of the noise which will permit specification of signal processing parameters that will reduce it. In some cases this information may not be known a priori and it may be necessary to examine the results of the signal processing steps to assess whether the output signal exhibits some apparent separation into desired and noise components. The user is strongly cautioned, however, to try to understand this separation relative to the original signal through a process of “reversing” the signal processing steps, if at all possible. Any determination that a signal component is unwanted noise ideally should be a reflection of knowledge of the process under study. Signal processing should implement that knowledge rather than being the sole basis of its determination.

Example 1.6 In many cases it will be obvious that certain components of a signal are noise, and one will not even contemplate that this result is based on knowledge of the properties of the signal source. Consider, for example, a recording of
EMG activity from an intercostal muscle of the chest in Fig. 1.10. The ECG signal invariably is present within a chest muscle EMG recording, but we recognize from the shape of the waveform, from its regular timing that is asynchronous with muscle contraction, and from its amplitude, that the ECG is not part of the EMG. Note that those discrimination criteria are independent of the signal processing steps. (The reverse is not true, of course. The signal processing steps are likely to be designed to reduce the ECG signal components based on their waveshape and timing.)

It is important to reiterate two related points made earlier: First, just because a signal "looks like" noise, it may not be noise. An obvious possibility is that it has a fractal character or that it arises from a chaotic system. Second, by deciding what will be considered noise, the user may well have limited the choice of signal models that can reasonably be applied to the non-noise signal.

Example 1.7 Assume I have a measurement that appears to be a sine wave with a lot of added noise (Fig. 1.11(a)). With this model in mind I can devise a method of filtering that removes as much of the "noise" as I can. Probably the resultant filtered signal could reasonably be analyzed only as a sinusoid. On the other hand, I might assume that the original signal represents a filtered random signal that has a strong component at one frequency (Fig. 1.11(b)); therefore both the desired and noise components are random signals. Unfortunately, there are no easy guidelines to resolve dilemmas like this. One must incorporate other information about the process under study and even then it may be necessary to test several models of the signal.

![Figure 1.10 Example of a chest wall EMG during two breaths, showing contamination by the ECG signal (top). Signal processing can detect and remove the ECG signal (middle). After rectifying and lowpass filtering, one obtains a signal representing "amplitude" of the EMG (bottom). Short line (lower right) represents one second.](image)
and try to judge which one is best. Once again, that type of assessment is an area of very active research.

Several examples of the need for noise removal by filtering were mentioned in the previous discussion of the sources of biomedical signals, although the causes of noise were not always discussed. Noise from electronic instrumentation is invariably present in biomedical signals, although the art of instrumentation design is such that this noise source may be negligible. Sometimes signals of interest are contaminated by signals of a similar type from another source. One example was given above: records of EMGs from muscles of the chest invariably contain an ECG signal which one wishes to minimize, usually by filtering. An extreme example of this type of noise contamination occurs during the recording of potentials from the scalp that are evoked by a brief sensory test stimulus such as a flash of light. Often the evoked potential is not even apparent in the recording because of the background EEG activity and a great deal of signal processing is necessary to permit visualization of it. Sometimes the motion of recording devices cannot be escaped and this motion adds a contaminating component to the signal being recorded. Other examples of unwanted signal components in biomedical applications are ubiquitous.

1.10 SUMMARY

This chapter has discussed the reasons for studying biomedical signal processing, presented some examples of biomedical signals, explained the different types of
signals which appear in biomedical applications, established some basic definitions, and described a framework for the introduction and analysis of signal processing techniques. A signal is a representation of information as a function of an independent variable, which is often time. Signals may be either continuous-time or discrete-time. From the signals they acquire, biomedical engineers want either to extract particular information or to draw inferences about the properties of the sources of the signals. In other applications they need to determine the relationships between two signals. Signal processing is the manipulation of a signal for obtaining information from the signal, deriving an alternative representation of the signal, or ascertaining the relationships of two or more signals. Often the intent is to extract information by suppressing unwanted signal components, but the determination of what constitutes noise should be based on knowledge of the signal source, if possible. In some applications it is desired to predict future values of the signal.

There are four types of signals that might be encountered—(1) deterministic, (2) stochastic, (3) fractal, and (4) chaotic—and it is important to determine to which class a signal of interest should be assigned. Often, however, this task is very difficult because the methods for identification of fractal and chaotic signals are evolving.

Signal processing can be viewed in the same framework as systems analysis—the signal is the input to a “black box” which contains the rules for processing the signal, and the output is the desired information or derived signal representation. This framework would be considered the analysis model of signal processing. In addition, each signal processing method can be placed into a “signal modeling” or synthesis framework. In this framework the signal under study is the output and the inputs are the signal waveforms which, according to the rules of the signal processing method, are to be used to construct the original signal. The “black box” then contains the rules for this construction. We will utilize both frameworks in the ensuing chapters.

**EXERCISES**

1.1 Consider how each of the following situations can be put into a systems framework. Specify the input and output signals, and describe the contents of the system “box” (i.e., the transformation process):

a. A miniature, solid-state pH sensor is placed at the end of a catheter which is then inserted into an artery. Wires running through the catheter connect the sensor to an external amplifier to permit recording of the pH.

b. To determine sleep state a neurologist examines a 12-lead EEG recording to determine the predominant rhythms (called alpha, beta, gamma, and delta). She also looks at a chin EMG recording to see if continuous muscle activity is present and at EMG recordings from the extraocular muscles to detect rapid eye movements. From these observations, every 30 seconds she classifies the sleep state into one of six possible states which are referred to as: awake; stage I, II, II, or IV of nonREM; REM.
c. Nerve cells are grown in a culture dish. Every day the culture dish is exposed to an electromagnetic field of a particular frequency and intensity for four hours. Afterwards the lengths of any axonal processes are measured microscopically to see if the electromagnetic field promotes axonal growth.

1.2 As traffic passes by a research building, the floor vibrates. To minimize vibrations of a microscope sitting on it, a table in the building has each of its four legs sitting in a tub of sand. The viscous damping provided by the sand should reduce the vibration of the table top relative to that of the floor. To assess the degree of damping to be expected, a researcher wants to relate the displacement of the table top to that of the floor. Formulate this problem in a systems framework, identify the input and output signals, and discuss the system which relates these signals.

1.3 Draw a block diagram for both the analysis and synthesis signal processing models for each of the following:

a. Rectal temperature of a subject is measured continuously for four days and these data are approximated using four cycles of a sine wave having a period of 24 hours.

b. The floor vibrations in Exercise 1.2 are recorded with a sensitive accelerometer and found to have five main frequencies whose amplitudes each vary with time during the day. To model this signal it is represented as a summation of five sinusoids whose amplitudes may change every hour.

c. Instantaneous heart rate (IHR) is defined as the reciprocal of the duration of a beat. The heart rate of a subject in the Intensive Care Unit (ICU) is being monitored by averaging the IHR for all beats during each 5-minute time interval to obtain an index of the mean heart rate.

1.4 Consider the following signals and decide which components of each are information-bearing biomedical signals or noise. State your reasons for your choices:

a. A new biomedical sensor continuously measures the glucose level in the blood.

b. A set of ECG electrodes records the electrical activity of the heart and the electrical activity of respiratory muscles of the chest.

c. An ultrasound beam is detected by an ultrasound microphone after the beam reflects off a solid tumor.

d. A microelectrode implanted in the motor cortex of a monkey records action potentials from many neurons, but especially from one neuron that becomes active during a reaching movement with the right hand.

1.5 Classify each of the following signals as CT or DT signals and specify either an appropriate unit of time for each signal or the independent variable (if it is not time):

a. The instantaneous velocity of the left heel during a long jump.

b. The concentration of calcium inside a muscle cell.

c. The amount of blood ejected from the left ventricle with each heartbeat.

d. The number of red blood cells passing through a pulmonary capillary each second.

e. The average velocity of red blood cells in a pulmonary capillary.
f. The concentration of oxytocin in 5 ml samples of arterial blood taken every hour.
g. Pressure inside the eyeball.
h. The number of nerve cells in a thin slice of the brainstem, where each slice is taken from a different experimental animal (but from the same location).
i. The brightness of a light that is supplied with a current proportional to the work done by a subject peddling an exercise bicycle.

1.6 Classify each of these signals as deterministic, stochastic, fractal, or chaotic, and explain your reasoning. (There may be more than one correct answer for some.)
   a. The signal from a blood glucose sensor after it is inserted into an artery.
   b. The signal from a blood glucose sensor before it is inserted into an artery.
   c. The signal from a pH meter whose electrode is in contact with gastric contents.
   d. The heart rate signal of Fig. 1.2(e).
   e. The intensity signal of Fig. 1.2(f).
   f. The EMG signal of Fig. 1.2(b).
   g. The three-dimensional coordinates of the path traced by the movements of a molecule of oxygen deep in the lung.
   h. \( x(t) = 0.5 \cos(6\Omega t) + 14 t u(t) \).
   i. The voltage across the membrane of a muscle cell of the heart.

1.7 Consider the set of all functions which can be constructed by linear summations of scaled and delayed versions of the three basis functions shown in Fig. 1.12. Any such function can be represented in the form
   \[ f(t) = a \, x(t - \tau_1) + b \, y(t - \tau_2) + c \, z(t - \tau_3) + d, \]
   where \( a, b, c, \) and \( d \) are constants. Evaluate these parameters \( (a, b, c, d, \tau_1, \tau_2, \tau_3) \) for each of the three functions graphed in Fig. 1.12.

FIGURE 1.12 See Exercise 1.7.
1.8 In the steady state the liters per minute that a subject inhales during breathing (i.e., ventilation, symbolized as \( V_I \)) is linearly related to the partial pressure of carbon dioxide in the arterial blood (symbolized as \( \text{PaCO}_2 \)). To calculate this relationship a physician measures ventilation and \( \text{PaCO}_2 \) from a subject and plots the former versus the latter (Fig. 1.13). Because these data are noisy and do not fall on a perfectly straight line, he uses linear regression to fit a line through these data points. Draw an analysis model of this signal processing step, identifying the input signal and output(s), and describe the transformation processes which are occurring in the system. If you know about the procedure of linear regression, you should be able to give the equations of the transformation.

1.9 One tool for assessing the presence of lung disease is to measure the maximum expiratory airflow effort that a subject can make, starting from his maximum inspired volume. Fig. 1.14(a,b) shows simulated data from such a maximal expiratory flow maneuver. Theoretically the maximum flow is a function of the fluid mechanical properties of the lung and it is a decreasing function of lung volume. Therefore it is usual to plot airflow versus expired volume, as shown in Fig. 1.14(c), and to compare such flow–volume curves to typical curves from normal subjects. A common problem is the noise near the peak of the flow signal (Fig. 1.14(a)), which is “spread out” over a large initial part of the flow–volume curve. Furthermore, it is often difficult to filter this noise from the flow signal without also removing too much of the desired flow signal. Let’s consider whether it is possible to filter the flow–volume curve directly.

a. If I digitize the flow and volume signals during the measurement (in this case, at 25 samples per second) and make a “discrete-time” plot from these data, I obtain the graph in Fig. 1.14(d). Explain why the data samples are not spaced uniformly along the volume axis.

![Graph](image)

**FIGURE 1.13** Steady-state ventilation of a human subject vs. partial pressure of CO\(_2\) in arterial blood (\( \text{PaCO}_2 \)). Solid circles: measured data. Dashed line: best fit line obtained from linear regression.
b. In theory, I could use interpolation of the data of Fig. 1.14(d) to generate samples of flow at uniformly spaced increments along the volume axis. Roughly sketch the “discrete” flow–volume plot obtained by interpolating at every 0.1-liter increment. Have I obtained a digitized signal that could be processed like any other “discrete-time” signal? Explain your answer.