Chapter 1

Introduction

The function of the human body is frequently associated with signals of electrical, chemical, or acoustic origin. Such signals convey information which may not be immediately perceived but which is hidden in the signal's structure. This information has to be "decoded" or extracted in some way before the signals can be given meaningful interpretations. The signals reflect properties of their associated underlying biological systems, and their decoding has been found very helpful in explaining and identifying various pathological conditions. The decoding process is sometimes straightforward and may only involve very limited, manual effort such as visual inspection of the signal on a paper print-out or computer screen. However, the complexity of a signal is often quite considerable, and, therefore, biomedical signal processing has become an indispensable tool for extracting clinically significant information hidden in the signal.

Biomedical signal processing represents an interdisciplinary topic. Knowledge of the physiology of the human body is crucial to avoid the risk of designing an analysis method which distorts, or even removes, significant information. It is also valuable to have a sound knowledge of other topics such as anatomy, linear algebra, calculus, statistics, and circuit design.

Biomedical signal processing has, by some, been viewed as a steppingstone for developing diagnostic systems which offer fully automated analysis. Some decades ago when computers first arrived in the area of medicine, automation was the overriding goal. However, this goal has been considerably modified over the years, not only because of the inherent difficulties in developing such systems, but equally so because the physician must be ultimately responsible for the diagnostic decisions taken. While fully automated analysis may be warranted in a few situations, today's goal is rather to develop computer systems which offer advanced aid to the physician in making well-

founded decisions. In these systems biomedical signal processing has come to play a very important role.

Research in biomedical signal processing has so far mainly been concerned with the analysis of one particular signal type at a time ("unimodal signal analysis"); a fact, which to a large extent, influences the content of the present textbook. However, the emerging interest in multimodal signal analysis will definitely help to explain, in more detail, how different physiological subsystems interact with each other, such as the interaction between blood pressure and heart rate in the cardiovascular system. By exploring the mutual information contained in different signals, more qualified diagnostic decisions can be made. The increased algorithmic complexity associated with multimodal analysis is not a serious limitation since it will be met by the rapid advancement of computer technology and the ever-increasing computational speed.

1.1 Biomedical Signal Processing: Objectives and Contexts

1.1.1 Objectives

Biomedical signal processing has many objectives, and some of the most important ones are presented below. We also describe the main contexts in which biomedical signal processing is applied. Other challenging objectives and contexts can certainly be defined by those interested in pursuing a career in this fascinating, interdisciplinary field.

Historically, biomedical signals have often been assessed visually, and manual ruler-based procedures were developed to make sure that measurements could be obtained in a standardized manner. However, it is well-known that there is relatively poor concordance between manually obtained measurements, and this may lead to unreliable diagnostic conclusions. A fundamental objective of biomedical signal processing is therefore to *reduce the subjectivity* of manual measurements. The introduction of computer-based methods for the purpose of objectively quantifying different signal characteristics is the result of a desire to improve measurement accuracy as well as reproducibility.

In addition to reducing measurement subjectivity, biomedical signal processing is used in its own right for developing methods that *extract features* to help characterize and understand the information contained in a signal. Such feature extraction methods can be designed to mimic manual measurements, but are equally often designed to extract information which is not readily available from the signal through visual assessment. For example,

small variations in heart rate that cannot be perceived by the human eye have been found to contain very valuable clinical information when quantified in detail using a suitable signal processing technique; see Chapter 8 for more details on this particular topic. Although it is certainly desirable to extract features that have an intuitive meaning to the physician, it is not necessarily those features which yield the best performance in clinical terms.

In many situations, the recorded signal is corrupted by different types of noise and interference, sometimes originating from another physiological process of the body. For example, situations may arise when ocular activity interferes with the desired brain signal, when electrodes are poorly attached to the body surface, or when an external source such as the sinusoidal 50/60 Hz powerline interferes with the signal. Hence, *noise reduction* represents a crucial objective of biomedical signal processing so as to mitigate the technical deficiencies of a recording, as well as to separate the desired physiological process from interfering processes. In fact, the desired signal is in certain situations so dramatically masked by noise that its very presence can only be revealed once appropriate signal processing has been applied. This is particularly evident for certain types of transient, very low-amplitude activity such as evoked potentials, which are part of brain signals, and late potentials, which are part of heart signals.

Certain diagnostic procedures require that a signal be recorded on a long timescale, sometimes lasting for several days. Such recordings are, for example, routinely done for the purpose of analyzing abnormal sleep patterns or to identify intermittently occurring disturbances in the heart rhythm. The resulting recording, which often involves many channels, amounts to huge data sizes, which quickly fill up hard disk storage space once a number of patients have been examined. Transmission of biomedical signals across public telephone networks is another, increasingly important application in which large amounts of data are involved. For both these situations, *data compression* of the digitized signal is essential and, consequently, another objective of biomedical signal processing. General-purpose methods of data compression, such as those used for sending documents over the internet, do not perform particularly well since the inherent characteristics of the biomedical signal are not at all exploited. Better performance can be obtained by applying tailored algorithms for data compression of biomedical signals. Data compression can also be understood in a wider sense as the process in which clinical information from a long-term recording is condensed into a smaller data set that is more manageable for the person analyzing the data. In this latter sense, it is highly desirable to develop signal processing algorithms which are able to determine and delimit clinically significant episodes.

Mathematical *signal modeling* and *simulation* constitute other important objectives in biomedical signal processing which can help to attain a bet-

ter understanding of physiological processes. With suitably defined model equations it is possible to simulate signals which resemble those recorded on the cellular level or on the body surface, thereby offering insight into the relationship between the model parameters and the characteristics of the observed signal. Examples of bioelectrical models include models of the head and brain for localizing sources of neural activity and models of the thorax and the heart for simulating different cardiac rhythms. Signal modeling is also central to the branch of signal processing called "model-based signal processing," where algorithm development is based on the optimization of an appropriately selected performance criterion. In employing the model-based approach, the suggested signal model is fitted to the observed signal by selecting those values of the model parameters which optimize the performance criterion. While model-based biomedical signal processing represents a systematic approach to the design of algorithms—to be frequently adopted in the present textbook—it does not always lead to superior performance; heuristic approaches may actually perform just as well and sometimes even better. It is a well-known fact that many commercial, medical devices rely on the implementation of ad hoc techniques in order to achieve satisfactory performance.

The complexity of a signal model depends on the problem to be solved. In most signal processing contexts, it is fortunately not necessary to develop a multilevel model which accounts for cellular mechanisms, current propagation in tissue, and other biological properties. Rather, it is often sufficient to develop a "phenomenological" model which only accounts for phenomena which are relevant to the specific problem at hand.

1.1.2 Contexts

The other purpose of this section is to point out the three major clinical contexts in which algorithms for biomedical signal processing are designed, namely, the contexts of

- *diagnosis*,
- *therapy*, and
- *monitoring*.

In the *diagnostic context*, medical conditions are identified from the examination of signal information, reflecting the function of an organ such as the brain or the heart, in combination with other symptoms and clinical signs. A signal is often acquired by a noninvasive procedure which makes the examination less taxing on the patient. Most of these procedures are also associated with inexpensive technology for acquisition and analysis, thus

increasing the likelihood that the technology can be disseminated to countries with less developed economies. A diagnostic decision rarely requires immediate availability of the results from signal analysis, but it is usually acceptable to wait a few minutes for the analysis to be completed. Hence, signal analysis can be done off-line on a personal computer, thus relying on standardized hardware and operating system, possibly supplemented with a digital signal processor (DSP) board for accelerating certain bottleneck computations. Algorithms for biomedical signal processing do not define the entire diagnostic computer system, but their scope ranges from performing a simple filtering operation to forming a more substantial part of the clinical decision-making.

Therapy generally signifies the treatment of disease and often involves drug therapy or surgery. With regard to biomedical signal processing, therapy may imply a narrower outlook in the sense that an algorithm is used to directly modify the behavior of a certain physiological process, for example, as the algorithms of a pacemaker do with respect to cardiac activity. In a therapeutic context, an algorithm is commonly designed for implementation in an implantable device like a heart defibrillator, and, therefore, it must, unlike an algorithm operating in a diagnostic context, strictly comply with the demands of on-line, real-time analysis. Such demands pose some serious constraints on algorithmic complexity as well as on the maximal acceptable time delay before a suitable action needs to be taken. Low power consumption is another critical factor to be considered in connection with devices that are implanted through a surgical procedure; for example, the battery of an implantable device is expected to last up to ten years. Hence, algorithms which involve computationally demanding signal processing techniques are less suitable for use in a therapeutic context.

Biomedical signal processing algorithms form an important part of real-time systems for *monitoring* of patients who suffer from a life-threatening condition. Such systems are usually designed to detect changes in cardiac or neurological function and to predict the outcome of a patient admitted to the intensive care unit (ICU). Since such changes may be reversible with early intervention, irreversible damage can sometimes be prevented. Similar to therapeutic contexts, the signal is processed during monitoring in an essentially sequential fashion such that past samples constitute the main basis for a decision, while just a few seconds of the future samples may also be considered—a property which usually stands in sharp contrast to signal processing for diagnostic purposes, where the signal is acquired in its entirety prior to analysis. Thus, a noncausal approach to signal analysis can only be adopted in the diagnostic context which mimics that of a human reader who interprets a signal by making use of both past and future properties. Constraints need to be imposed on the algorithmic design in terms of max-

imal delay time because the occurrence of a life-threatening event must be notified to the ICU staff within a few seconds. Another important issue to be considered is the implications of a clinical event that is missed by the algorithm or the implications of a nonevent that is falsely detected causing the staff to be notified.

1.2 Basics of Bioelectrical Signals

Although the scope of the present textbook is to present signal processing techniques useful for the analysis of electrical signals recorded on the *body surface*, it may still be well-motivated to consider the genesis of bioelectrical signals from a cellular perspective. Bioelectrical signals are related to ionic processes which arise as a result of electrochemical activity of a special group of cells having the property of *excitability*. The mechanisms which govern the activity of such cells are similar, regardless of whether the cells are part of the brain, the heart, or the muscles. In particular, the electrical force of attraction has central importance for the processing and transmission of information in the nervous system, as well as for sustaining the mechanical work done by the heart and the muscles. Since the origin of these voltages is only briefly described below, the interested reader is referred to textbooks on human physiology which offer a much more detailed description of the cellular aspects [1, 2]. The basic concepts introduced for mathematical modeling of bioelectrical phenomena are described in [3], while more comprehensive reading is found in [4–6].

1.2.1 On the Cellular Level

A cell is bounded by a plasma membrane which basically consists of lipid layers with poor ability to conduct an electrical current. The membrane possesses permeability properties which allow certain substances to pass from the inside of the cell to the outside through different channels, defined by body fluids, while other substances remain blocked. Intracellular and extracellular fluids mainly consist of water, which is electrically neutral; however, the fluids become electrically conductive since they contain several types of ions. The dominant ions in a nerve cell (neuron) are sodium (Na^+), potassium (K^+), and chloride (Cl^-). Other ions such as calcium (Ca^{2+}) are also present but play roles of varying importance depending on where the excitable cell is located; the calcium ion is much more important in the cells of the heart than in the nerves, for example.

Under resting conditions, the inside of a cell is negatively charged with respect to the outside, and, therefore, a negative transmembrane potential results since the outside is assumed to have zero voltage. The difference in

charge is due to the fact that the concentration of negatively charged ions inside the cell is higher than on the outside, whereas the opposite relation applies to the concentration of positive ions. In addition to the difference in ion concentration, the actual magnitude of the resting transmembrane potential is also determined by the permeability of the membrane to the different ions.

A potential arises when membrane channels open so that a certain ion may diffuse across the membrane. This process can be illustrated by the simplified situation in which potassium ions are assumed to be inside the cell and sodium ions outside and when the initial transmembrane potential is equal to zero. When the potassium channels are opened, an increase in positive electrical charge outside the cell is created as a result of the diffusion process; at the same time, the inside of the cell becomes increasingly negative and a potential arises across the membrane. This electrical potential constitutes the other force which causes ions to move across the membrane. As the outside of the cell becomes increasingly positive, the resulting potential will increasingly influence the outbound movement of potassium ions. The ion movement ceases when the concentration force balances the electrical force; an equilibrium potential is then said to have been reached. It should be noted that some other active transport mechanisms, not considered here, also come into play when a potential is created.

The resting transmembrane potential of a cell is determined by the equilibrium potentials of the different ions involved and is thus not equal to any of the equilibrium potentials of an individual type of ion. For the situation considered above with open potassium channels, the equilibrium potential for potassium in a nerve cell is found to be about -90 mV, while the equilibrium potential for sodium—assuming instead open sodium channels—is about $+60$ mV. The resting transmembrane potential is within the range of -60 to -100 mV, depending on the type of cell.

When a cell is stimulated by a current, rapid alterations in membrane ion permeability take place which give rise to a change in the membrane potential and generate a signal referred to as an *action potential*. The propagation of action potentials is the very mechanism which makes the heart contract and the nervous system communicate over short and long distances. The stimulus current must exceed a certain threshold level in order to elicit an action potential, otherwise the cell will remain at its resting potential. An excited cell exhibits nonlinear behavior: once a stimulus intensity exceeds the threshold level the resulting action potential is identical and independent of intensity—the *all-or-nothing principle*. An action potential consists mainly of two phases: *depolarization* during which the membrane potential changes toward zero so that the inside of the cell becomes less negative, and ultimately reverses to become positive, and *repolarization* during which the

potential returns to its resting level so that the inside again becomes more negative.

The membrane potential remains at its resting level until it is perturbed by some external stimulus, such as a current propagating from neighboring cells. Depolarization is then initiated, and the membrane permeability changes so that sodium channels are opened and the sodium ions can rush into the cell. At the same time, potassium ions try to exit since these are concentrated on the inside, but cannot, thereby causing the charge inside the cell to become increasingly positive, and eventually the membrane potential reverses polarity. Once the rush of sodium ions into the cell has stopped and the membrane potential approaches the sodium equilibrium potential, the peak amplitude of an action potential is reached. During repolarization, sodium channels close and potassium channels open so that the membrane potential can return to its resting, negative potential. The activity of a potassium channel is illustrated in Figure 1.1.

The duration of an action potential varies much more than its amplitude: the repolarization phase of a cardiac cell is much longer than the depolarization phase and lasts from 200 to 300 milliseconds, while for a neuron the two phases combined only last for about one millisecond with both phases having about the same duration. Figure 1.2 shows the action potentials for cells of the brain (motor neuron), the skeletal muscle, and the heart. From these waveforms, it can be observed that the cardiac action potential differs considerably from the others in its lack of an immediate repolarization phase. Instead, there is a plateau in the action potential because the membrane channels of the different ions open and close at different speeds.

Once an action potential has been elicited, the membrane cannot immediately respond to a new stimulus but remains in a "refractory" state for a certain period of time. The refractory period is related to changes that take place in sodium and potassium permeability of the membrane. Obviously, the refractory period imposes an upper limit on the frequency at which action potentials can be communicated through the nervous system or the heart can beat.

The propagation of an action potential exhibits special behavior since it travels a distance through the triggering of new action potentials rather than by traveling itself along the membrane. The current created by the initial membrane depolarization triggers an adjacent membrane so that a new action potential results, and so on. This process repeats itself until the membrane ends and delivers an action potential which is identical to the initial action potential. Due to the refractory period, the action potential travels away from membranes which recently have been excited and continues to do so until it reaches a point on the membrane where the voltage is insufficient for further stimulation.

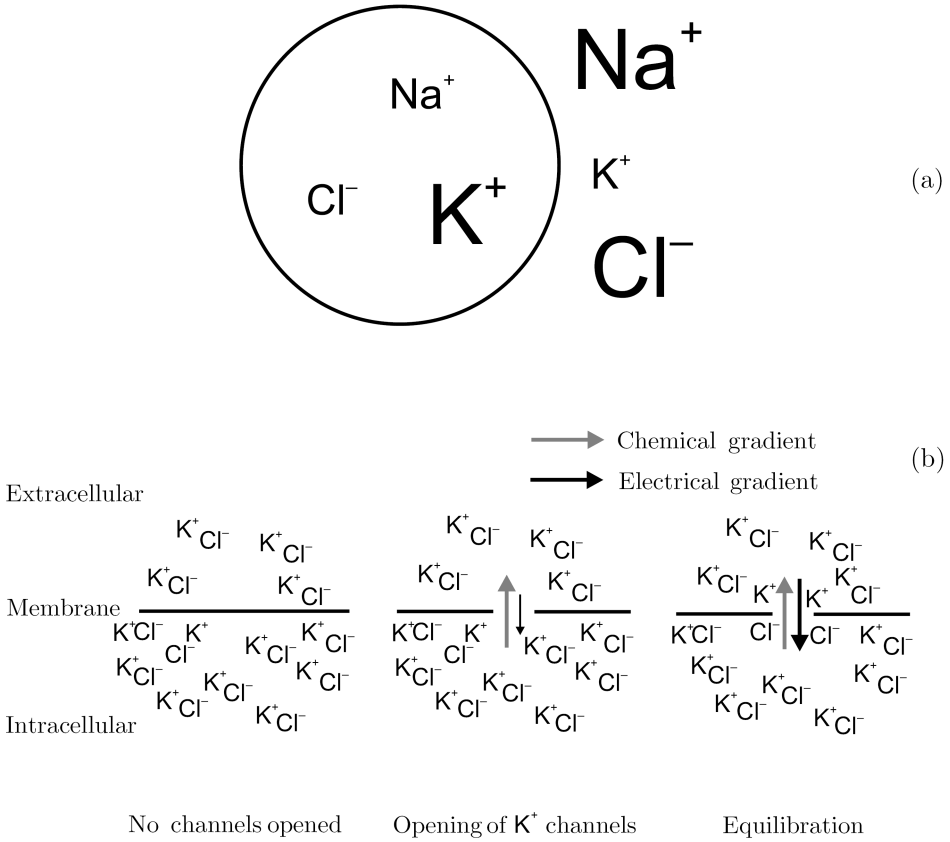

Figure 1.1: Cellular activity of potassium channels (which is similar for sodium but the reverse). (a) Concentration distribution of potassium (K^+), sodium (Na^+), and chloride (Cl^-) ions inside and outside a cell. (b) The relationship between chemical gradient and electrical gradient for K^+ ions and K^+ channels.

1.2.2 On the Body Surface

The ability of excitable cell membranes to generate action potentials causes a current to flow in the tissue that surrounds the cells. With the tissue being a conducting medium, commonly referred to as a volume conductor, the collective electrical activity of many cells can be measured noninvasively on the body surface [4–6]. The recording of a bioelectrical signal in clinical practice is done by attaching at least two electrodes to the body surface. In its simplest form, a signal is recorded by making use of two electrodes: the "exploring" electrode, placed close to the electrical source, and the "indifferent" electrode, placed elsewhere on the body surface [7]. Multiple electrode configurations are commonly used in clinical practice to obtain a spatial de-

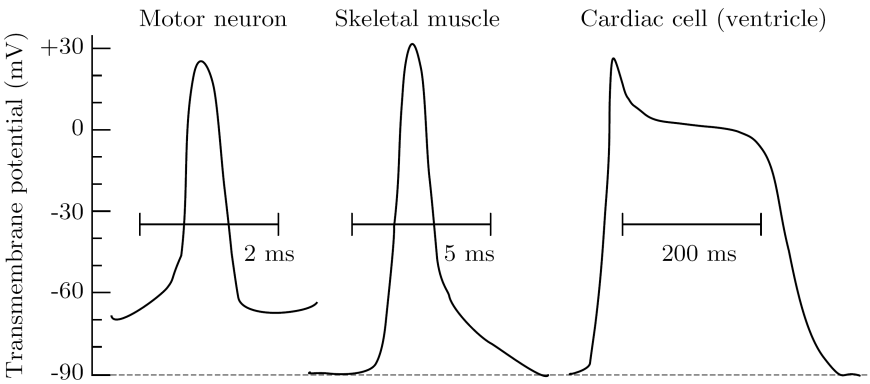

Figure 1.2: Examples of action potentials with shapes that range from the spike-like waveform of a nerve cell (left) to the much more extended waveform of a cardiac cell (right). The transmembrane potential difference was measured by placing one microelectrode inside the cell and another outside. It should be noted that the timescale differs from waveform to waveform.

scription of the bioelectrical phenomenon. Since the activity of excitable cells is viewed from a distance by the electrodes, with different tissues in between, such as blood, skeletal muscles, fat, and bone, it is impossible to noninvasively determine detailed information about cellular properties and propagation patterns. Nonetheless, significant empirical knowledge has over the years been acquired from analyzing the patterns of signals recorded on the body surface, which have been found crucial for clinical decision-making; this observation constitutes an important motivation for the writing of the present textbook.

The problem of characterizing the electrical source by noninvasive measurements has, in spite of the above-mentioned limitations, been the subject of considerable research due to the far-reaching clinical implications of its potential solution. In order to arrive at a meaningful solution, it is necessary to introduce a mathematical model in which the collective electrical cellular activity is treated as a volume source, i.e., it is defined by a fixed dipole, a multiple dipole, or some other source model. Furthermore, by introducing a model for the volume conductor which accounts for essential properties of the human body, such as geometry and resistivity, the electrical field measured on the body surface can be modeled. The important *inverse problem* consists of determining the electrical source from measurements on the body surface under the assumption that the geometry and electrical properties of the volume conductor are known [5].

1.2.3 Bioelectrical Signals

The present textbook deals with the processing of electrical signals that describe the activity of the brain, the heart, and the muscles. Some of these signals reflect spontaneous, ongoing activity, while others only occur as the result of external stimulation. The properties of these signals call for widely different processing techniques; an individual waveform can in some signals be directly linked to a specific clinical diagnosis, while in other signals the composite of many waveforms must be analyzed before a meaningful interpretation can be made.

The **electroencephalogram** (EEG) reflects the electrical activity of the brain as recorded by placing several electrodes on the scalp, see Figure 1.3(a). The EEG is widely used for diagnostic evaluation of various brain disorders such as determining the type and location of the activity observed during an epileptic seizure or for studying sleep disorders. The brain activity may also be recorded during surgery by attaching the electrodes directly to the uncovered brain surface; the resulting invasive recording is named an *electrocorticogram* (ECoG). The background to EEG signals is presented in Chapter 2 and is then followed by Chapter 3 where different EEG signal processing techniques are described.

Evoked potentials (EPs) constitute a form of brain activity which usually is evoked by a sensory stimulus such as one of visual or acoustic origin. Their clinical use includes the diagnosis of disorders related to the visual pathways and the brainstem. An EP, also referred to as an event-related potential, is a transient signal which consists of waves of such tiny amplitudes that its presence in the "background EEG" is typically invisible to the human eye, see Figure 1.4(a). Evoked potentials are recorded using an electrode configuration similar to that of an EEG. Chapter 4 contains an overview of methods developed for "revealing" EPs and for analyzing the resulting signal waveform.

The **electrocardiogram** (ECG) reflects the electrical activity of the heart and is obtained by placing electrodes on the chest, arms, and legs, see Figure 1.3(b). With every heartbeat, an impulse travels through the heart which determines its rhythm and rate and which causes the heart muscle to contract and pump blood. The ECG represents a standard clinical procedure for the investigation of heart diseases such as myocardial infarction. The *electrogram* (EG) is an intracardiac recording where the electrodes have been placed directly within the heart; the EG signal is used in implantable devices such as pacemakers and defibrillators. The background to ECG signals is presented in Chapter 6, while Chapters 7 and 8 present different ECG signal processing techniques.

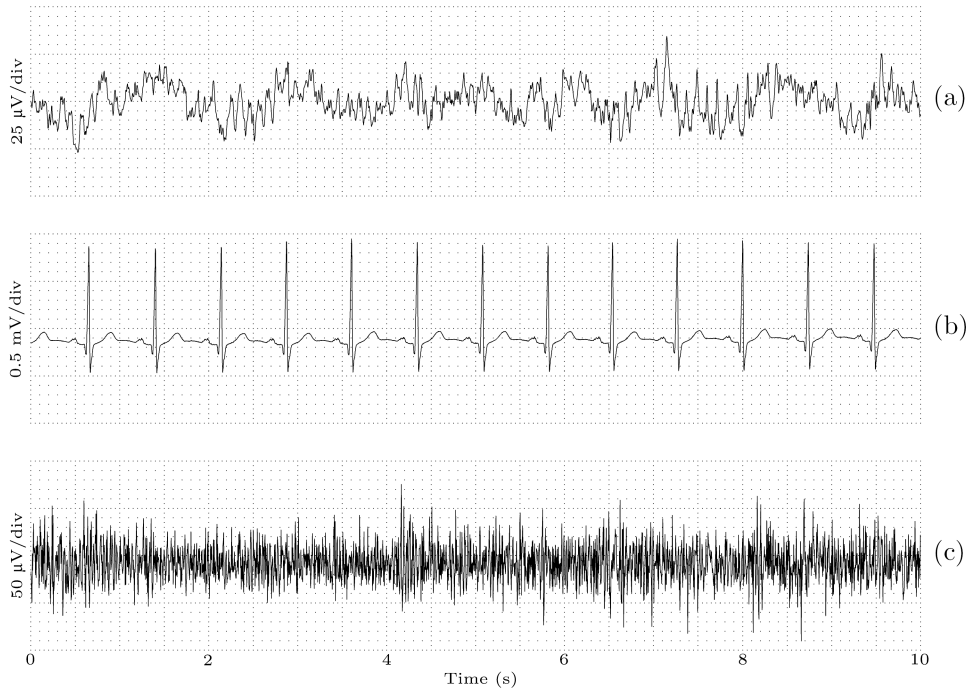

Figure 1.3: Examples of the three major bioelectrical signals recorded from the body surface: (a) an electroencephalogram (EEG) containing alpha activity, (b) an electrocardiogram (ECG) during sinus rhythm, and (c) an electromyogram (EMG) obtained from the chin in the waking state. All three signals were obtained from different normal subjects.

The **electromyogram** (EMG) records the electrical activity of skeletal muscles which produce an electrical current, usually proportional to the level of activity, see Figure 1.3(c). The EMG is used to detect abnormal muscular activity which occurs in many diseases such as muscular dystrophy, inflammation of muscles, and injury to nerves in arms and legs. Recording the surface EMG involves placing the electrodes on the skin overlying the muscle, whereas the intramuscular EMG involves inserting needle electrodes through the skin into the muscle to be examined. Chapter 5 presents an overview of EMG signal processing techniques.

Some other types of bioelectrical signals also deserve mentioning although their related signal analysis will not be further considered in the present textbook.

The *electroneurogram* (ENG) results from the stimulation of a peripheral nerve with an electric shock such that the response along the nerve can be measured. The ENG, usually acquired with needle electrodes, is used

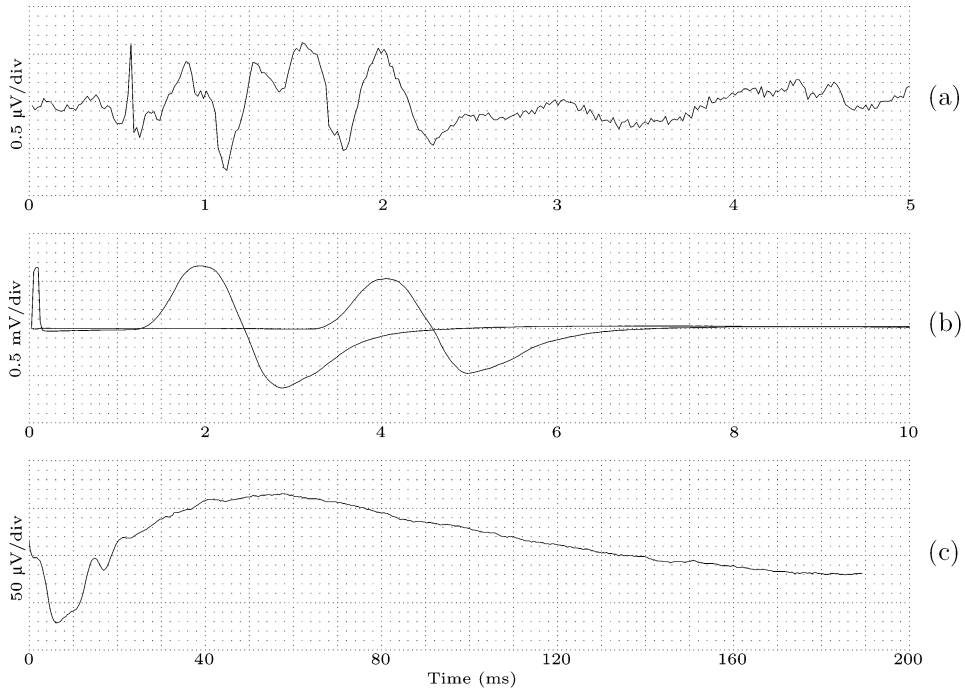

Figure 1.4: Examples of bioelectrical signals resulting from stimulation. (a) An evoked potential (EP) resulting from auditory stimulation (the brainstem response). The displayed signal is actually the result of averaging several responses in order to reduce the high noise level of the original signal; see Section 4.3 for details on noise reduction. (b) An electroneurogram (ENG) recorded at two electrode locations, where the delay between the two signals is used to estimate nerve conduction velocity. (c) An electroretinogram (ERG) obtained during stimulation with a flash of light.

to determine the conduction velocity of the nerve, thereby assisting in the diagnosis of nerve injury. By stimulating a nerve at two different sites separated by a well-defined distance, it is possible to estimate the conduction velocity from the distance by which the resulting two signal waveforms are separated, see the example in Figure 1.4(b). The ENG can be measured both invasively and noninvasively.

An *electroretinogram* (ERG) is used for studying the electrical potentials generated by the retina of the eye during light stimulation [8, 9], see Figure 1.4(c). The ERG is recorded by placing the exploring electrode, encapsulated in a contact lens, on the cornea. The ERG has been found useful for assessing the electrical response of the rods and cones, i.e., the visual cells at the back of the retina. A normal ERG shows appropriate responses

with increased light intensity, while an abnormal ERG is obtained in conditions such as arteriosclerosis of the retina or detachment of the retina. The algorithms described in Chapter 4 for signal processing of EPs are, by and large, also applicable to the analysis of ERGs.

The *electrooculogram* (EOG) is the recording of the steady corneal–retinal potential which is proportional to vertical and horizontal movements of the eye, thus offering an objective way to quantify the direction of the gaze [5, 10], see Figure 1.5(a). The EOG is of particular interest in patients who suffer from sleep disorders, where the presence of rapid eye movement (REM) is important for determining certain sleep stages. The EOG is recorded when studying nystagmus, i.e., a rapid, involuntary oscillation of the eyeballs, for example, in patients suffering from vertigo and dizziness. The EOG is also useful in virtual reality environments where a device for eye-tracking may be needed. The EOG is briefly touched upon in Chapter 3 in connection with EEG signal processing since the electrical activity caused by eye movements often interferes with the EEG and, therefore, needs to be cancelled.

The *electrogastrogram* (EGG) is a recording of the impulses which propagate through the muscles of the stomach and which control their contractions [11], see Figure 1.5(b). The EGG is studied when the muscles of the stomach or the nerves controlling the muscles are not working normally, for example, when the stomach does not empty food normally. The EGG is recorded by attaching a number of electrodes over the stomach during fasting and subsequent to a meal. In normal individuals a regular "rhythmic" signal is generated by the muscles of the stomach, having an amplitude which increases after a meal; the normal frequency of the gastric rhythm is approximately 3 cycles/minute. However, in symptomatic patients the rhythm is often irregular and sometimes without the increase in amplitude that follows a meal. A small selection of papers describing technical means of analyzing the EGG signal can be found in [12–16].

1.3 Signal Acquisition and Analysis

The acquisition of bioelectrical signals is today accomplished by means of relatively low-cost equipment which appropriately amplifies and digitizes the signal. As a result, several clinical procedures based on bioelectrical signals are in widespread use in hospitals around the world. In many situations, PC-based systems can be utilized as an efficient and cost-effective solution for signal analysis, especially considering the availability of expansion cards for data acquisition. Such a system includes one or several sensors, external hardware for patient insulation and signal amplification, an acquisition

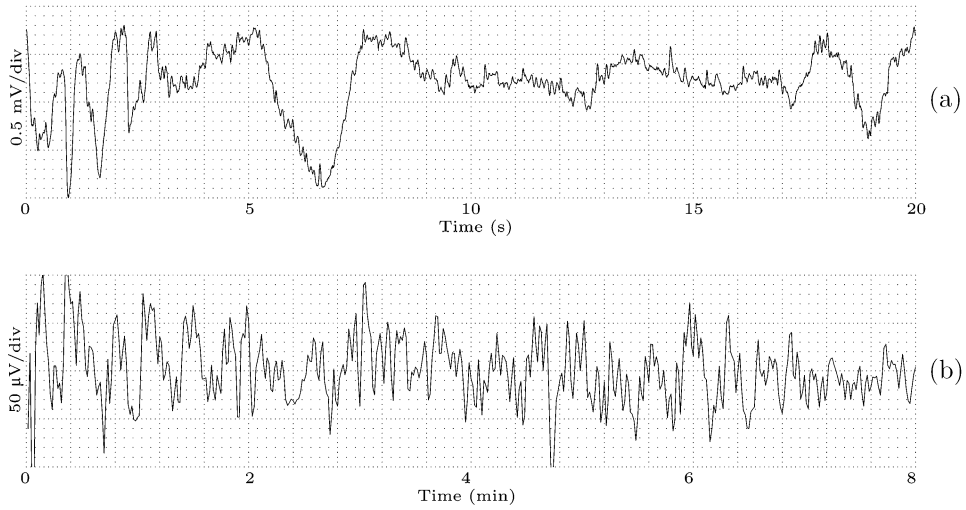

Figure 1.5: Recordings which exemplify (a) an electrooculogram (EOG) of the right eye and (b) an electrogastrogram (EGG). Note that the two timescales differ.

card with analog/digital (A/D) conversion, and software for signal analysis (Figure 1.6) [17]. In situations where the analysis is performed in an implantable device, the system design involves additional considerations, e.g., those related to the design of application-specific integrated circuitry and the selection of appropriate battery technology.

In the digitization process, it is usually sufficient to use 12–14 bits for amplitude quantization in order to cover the dynamic range of a signal; it is presumed that very slow, large-amplitude drift in the direct current (DC) level has been removed prior to quantization without modifying the physiological content of the signal. The amplitude of individual bioelectrical waveforms ranges from 0.1 μV, observed in certain types of EPs once subjected to noise reduction, to several millivolts, as observed in the ENG, ECG, and EOG.

Most bioelectrical signals recorded on the body surface have a spectral content confined to the interval well below 1 kHz, and thus the sampling rate—chosen to be at least the Nyquist rate—rarely exceeds a few kilohertz. However, since signals measured on the body surface are subjected to lowpass filtering caused by the intermediate tissue, invasively recorded signals, such as those on action potentials, generally exhibit a much higher frequency content.

In a PC-based system, signal analysis is often done locally by relying either on the internal CPU or an expansion digital signal processor (DSP) card.

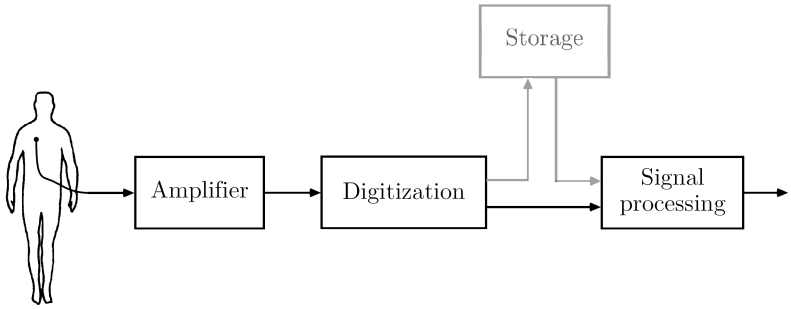

Figure 1.6: Block diagram describing the main steps in biomedical signal analysis. The signal is often processed at the time of its acquisition, but may also be stored on a local hard disk or a server on the web for later retrieval and processing.

However, with today's availability of web-based resources, it is no longer necessary to perform the entire signal analysis locally. It is equally possible to acquire the signal at one physical location, using the PC-based system, and then to process it at another location, i.e., relying on a client/server solution [18]. Since the acquired signal in most cases is stored in a database that resides on a server, it can be advantageous to also process the signal on the server since it may offer more computational power.

1.4 Performance Evaluation

Performance evaluation is an important and challenging part of biomedical signal processing required before any algorithm can be implemented in a clinical context. Unlike many other engineering applications where the information in the signal source is known a priori, the message "sent" by a bioelectrical source is unknown and has to be unmasked in some manual way in order to render performance evaluation possible. For example, the evaluation of an algorithm for detecting heartbeats is relatively straightforward since it is an easy task for a physician to determine the times of occurrence of the heartbeats; the performance figures would then be designed to reflect how well the output of the algorithm agrees with the manually obtained times of occurrence. The performance evaluation becomes much more complicated when the goal is to develop an algorithm that computes a parameter set which accurately discriminates signals obtained from healthy subjects and patients who suffer from a particular disease. In such cases, an assessment of the output of the algorithm cannot be carried out simply because the "truth" cannot be retrieved from the observed signal. Instead, the performance may be evaluated in terms of its ability to correctly discriminate between the two

Table 1.1: Definitions of the performance measures sensitivity, specificity, positive predictive value, and negative predictive value.

Performance measure	Definition	Interpretation
Sensitivity	$\frac{N_{TP}}{N_{TP}+N_{FN}}$	The probability of a positive result for the diseased subjects
Specificity	$\frac{N_{TN}}{N_{FP}+N_{TN}}$	The probability of a negative result for the healthy subjects
Positive predictive value	$\frac{N_{TP}}{N_{TP}+N_{FP}}$	The probability of disease when the result is positive
Negative predictive value	$\frac{N_{TN}}{N_{FN}+N_{TN}}$	The probability of health when the result is negative

N_{TP} = the number of diseased subjects with a positive result (True Positive)
N_{TN} = the number of healthy subjects with a negative result (True Negative)
N_{FN} = the number of diseased subjects with a negative result (False Negative)
N_{FP} = the number of healthy subjects with a positive result (False Positive)

groups of healthy and diseased subjects. The most commonly used performance measures for describing such discrimination are those of sensitivity, specificity, positive predictive value, and negative predictive value, whose definitions are given in Table 1.1.

It has been pointed out that "while new analytic technologies seem very promising when they are first applied, the initial glitter often fades when the method is systematically evaluated" [19]. This statement not only underlines the importance of performance evaluation, but also that a great deal of effort must be devoted to algorithm development before satisfactory performance can be achieved.

1.4.1 Databases

The availability of signal databases is of vital importance for both development and evaluation of signal processing algorithms. The immense diversity of waveform patterns which exists among subjects necessitates evaluation of the algorithm on a database of considerable size before its performance can be judged as satisfactory for use in a clinical setting. Needless to say, one part of a database must be used for algorithm development, while the remaining part is kept for performance evaluation in order to assure that no learning of the evaluation data has taken place.

The word "database" is here interpreted as a collection of signals that has been obtained using the same recording protocol from suitably selected groups of healthy subjects and patients. A database often includes signals of one particular type, such as EEGs or ECGs, but may just as well include other types of concurrently recorded signals. Annotations are another important type of database information which define the time instants at which certain events occur in the signal, such as the presence of heartbeats or epileptic seizures. The annotations may also account for more complex signal properties as well as for nonphysiological information such as the presence of noise episodes and technical deficiencies due to poorly attached electrodes (Figure 1.7). The annotations are determined manually by one or several physicians who must carefully scrutinize the signal with respect to the properties to be annotated. The inclusion of several annotators generally implies that more reliable annotations are obtained. However, it is inevitable that discrepancies arise among the annotators which must be resolved by consensus, thus adding further labor to an already laborious process.

In addition to the signal and its annotation, the database may include additional information on subjects such as gender, race, age, weight, medication, and data from other clinical procedures which may be valuable when evaluating performance.

A substantial number of databases have been collected over the years for the purpose of addressing various clinical issues. The MIT–BIH arrhythmia database is the most widely used database for evaluation of methods designed for detecting abnormalities in cardiac rhythms and is almost certainly also the most popular database overall in biomedical signal processing [21, 22]. The MIT–BIH arrhythmia database contains ECG signals which have been recorded during ambulatory conditions such as working and eating. Another widely used ECG database is the AHA database, which was developed for evaluation of ventricular arrhythmia detectors [23]. More recent additions to the list of databases include the European ST–T and LTST databases, which were collected for the purpose of investigating the occurrence of insufficient blood supply to the cardiac muscle cells (myocardial ischemia) [20, 24]. An interesting adjunct to the MIT–BIH arrhythmia database is the MIT–BIH noise stress test database which contains several recordings of noise typically encountered in ambulatory conditions: by adding a calibrated amount of noise to a "clean" ECG signal, the noise immunity of an algorithm can be tested with this database [25].

Multimodal databases have also been collected and may include signals that reflect brain activity, heart activity, muscle activity, blood pressure, respiration, as well as other types of activity, see Figure 1.8. Some examples are the MIMIC database [26], the IMPROVE database [27], and the IBIS database [28, 29], which all contain continuously recorded data obtained

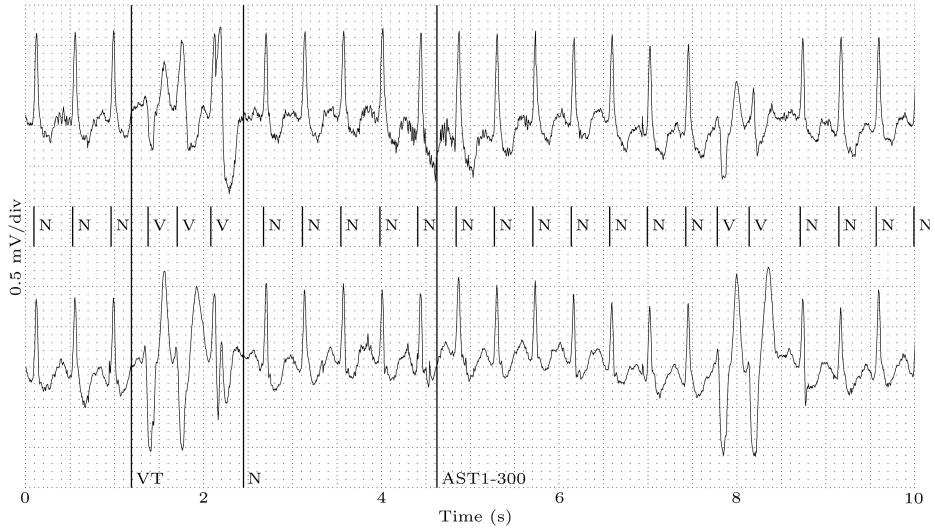

Figure 1.7: Example of a manually annotated, two-channel ECG from a patient with myocardial ischemia. The sequence of short, vertical bars shows the times of occurrence of the heartbeats, and the related labels "N" and "V" indicate whether the beat is of normal or ventricular origin. The three longer bars indicate the onset of a new rhythm (VT: ventricular tachycardia, N: sinus rhythm, and AST1-300: maximum ST depression of –300 μV). The signal was taken from the European ST–T database [20].

from intensive care monitoring, while other databases have been collected for investigating sleep disorders [30, 31]. Most databases described in the literature are publicly available, either at no cost or at a charge, while some remain the private property of those who collected the data. Databases of biomedical signals have proven to be equally valuable for researchers and instrument manufacturers.

The increasing availability of databases certainly makes it more convenient and less time-consuming to pursue projects on algorithm development. Because of the easy access to databases, now available on different sites on the World Wide Web, it is possible to develop and evaluate signal processing algorithms without having to deal with the cumbersome and often labor-intensive task of data collection. The *PhysioNet* (`www.physionet.org`) is a website which constitutes a tremendous leap forward, being a resource where various types of physiological signals are freely available for download [32]. The PhysioNet maintains different classes of databases, ranging from those which are carefully scrutinized and thoroughly annotated to those which are unannotated and sometimes not yet completely acquired.

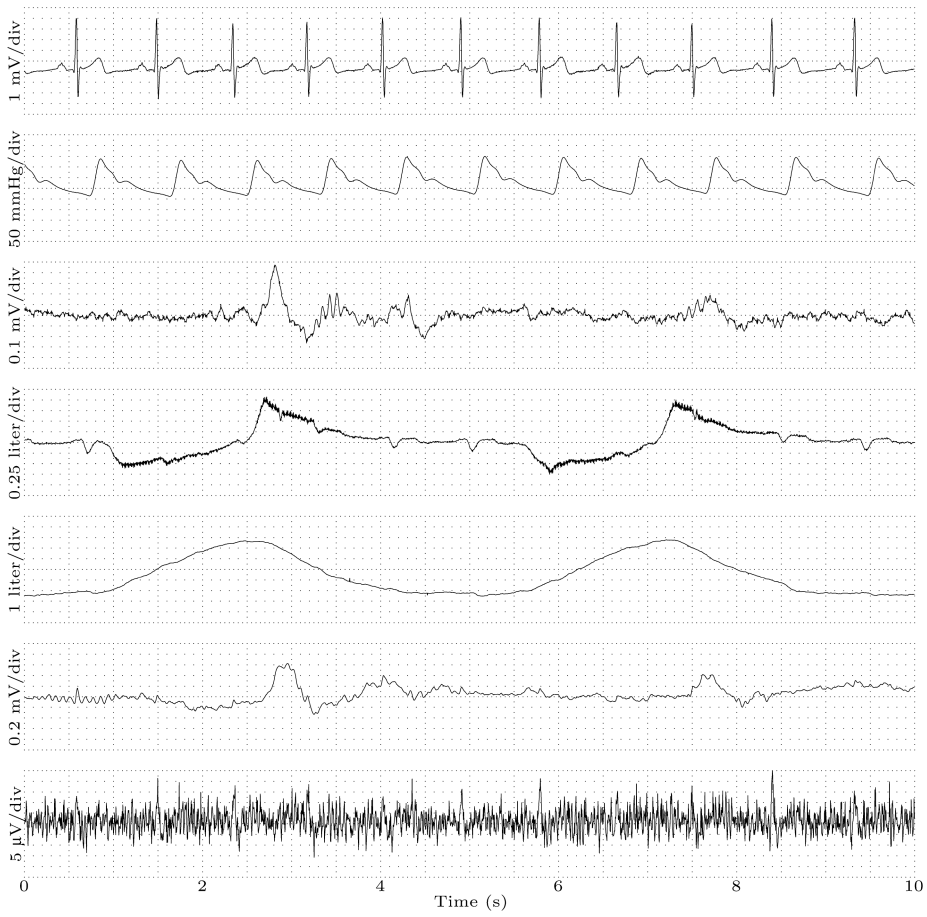

Figure 1.8: Concurrently recorded signals from a multimodal database, from top to bottom: ECG, blood pressure, EEG, nasal respiration, abdominal respiration, EOG, and EMG. This type of recording is used for studying sleep disorders, see Section 2.4.2. The signals were taken from the MIT–BIH polysomnographic database [30].

With the easy availability of databases comes also the potential risk of omitting medical expertise from projects since hospital-based activities are no longer needed. If the worst comes to the worst, a lack of clinically experienced eyes may lead to the introduction of clinically unacceptable distortion into the signal via the algorithm, rather than improving its interpretation. Hence, it is always important for the project's outcome to establish a viable liaison between engineers and physicians. Another potential risk when downloading a database is that its original clinical purpose is tweaked into

answering questions for which the database was never intended. In cases when no suitable database is available, it is necessary to develop an appropriate recording protocol for data collection of one's own and then, of course, to perform the actual signal acquisition. Anyone embarking on a project in biomedical signal processing is, in addition to considering the use of available databases, strongly encouraged to also deal with the details of collecting signals.

1.4.2 Simulation

A simulation describes quantitatively some physiological behavior by mathematical equations and is used to replicate signals which are generated in the body. An advantage of simulations is the possibility to investigate conditions which are difficult to deal with experimentally. Another advantage, of particular relevance for algorithmic performance evaluation, is that the properties of a simulated signal can be exactly controlled by a set of parameters. As a result, the agreement between the "true" parameter values of the simulated signal and those determined by an estimation method can be quantitatively assessed and expressed in terms of a suitable performance measure. The exact definition of such a measure depends on the case at hand and may involve rates of missed events and false events in detection problems and bias and variance in parameter estimation problems.

Signal modeling and simulation are intimately linked together because a simulation is based on an existing model. Models producing highly realistic-looking signals are often associated with high complexity and do not easily lend themselves to parameter estimation. Simpler models, which can only account for a partial phenomenon of the signal, are still very useful for algorithm development and have, in fact, often been considered.

In biomedical signal processing, a model of the physiological, "clean" signal is often accompanied by a model of the noise sources, and the combination of the two models makes it possible to simulate signals observed on the body surface. The term "noise" is here used in a wide sense which includes physiological activities other than the one under study which may interfere with the desired signal. Simulated signals with different signal-to-noise ratios (SNRs) can be easily produced using this approach. While performance evaluation is mostly concerned with accuracy, i.e., the difference between the true value and the estimated value, it is also important to study the reproducibility of an algorithm. Reproducibility is the ability of an algorithm to produce repeated measurements which cohere, obviously under the assumption that the same signal conditions apply to all measurements. Although reproducibility is best investigated by sequentially repeating an experiment on the same patient, simulated signals represent a powerful and

much more manageable means of evaluating the reproducibility of an algorithm. The performance is then evaluated by using the algorithm to process a series of simulated signals, each time by adding a different noise realization to the clean signal.

In addition to simulations based on mathematical models for both signal and noise, it may in certain cases be appropriate to evaluate the performance by employing simulated signals to which "real world" noise is instead added. The reverse situation with "real world" signals and simulated noise may sometimes also be of interest.

We conclude this section by noting that the simulation approach represents a useful step in algorithm development, provided of course that the signal model is adequate. However, databases consisting of collected signals must constitute the lion's share of the evaluation work so that the clinical utility of an algorithm can be thoroughly established.

Bibliography

[1] A. C. Guyton and J. E. Hall, *Textbook of Medical Physiology*. Philadelphia: W. B. Saunders, 10th ed., 2000.

[2] A. J. Vander, J. H. Sherman, and D. S. Luciano, *Human Physiology. The Mechanisms of Body Function*. New York: McGraw–Hill, 5th ed., 1990.

[3] J. Enderle, "Bioelectric phenomena," in *Introduction to Biomedical Engineering* (J. Enderle, S. Blanchard, and J. Bronzino, eds.), ch. 3, pp. 79–138, San Diego: Academic Press, 2000.

[4] R. Plonsey and R. C. Barr, *Bioelectric Phenomena: A Quantitative Approach*. New York: Plenum, 1988.

[5] J. Malmivuo and R. Plonsey, *Bioelectromagnetism*. Oxford: Oxford University Press, 1995.

[6] R. M. Gulrajani, *Bioelectricity and Biomagnetism*. Montreal: John Wiley & Sons, 1998.

[7] M. R. Neuman, "Biopotential electrodes," in *Medical Instrumentation. Application and Design* (J. G. Webster, ed.), ch. 5, pp. 183–232, New York: John Wiley & Sons, 1998.

[8] J. C. Armington, *The Electroretinogram*. New York: Academic Press, 1974.

[9] J. R. Heckenlively and G. B. Arden (eds.), *Principles and Practices of Clinical Electrophysiology of Vision*. St. Louis, MO: Mosby Year Book, 1991.

[10] R. H. S. Carpenter, *Movements of the Eyes*. London: Pion, 2nd ed., 1988.

[11] W. C. Alvarez, "The electrogastrogram and what it shows," *JAMA*, vol. 28, pp. 1116–1118, 1922.

[12] D. A. Linkens and S. P. Datardina, "Estimation of frequencies of gastrointestinal electrical rhythms using autoregressive modelling," *Med. Biol. Eng. & Comput.*, vol. 16, pp. 262–268, 1978.

[13] R. H. Smallwood, D. A. Linkens, H. L. Kwok, and C. J. Stoddard, "Use of autoregressive-modelling techniques for the analysis of colonic myoelectrical activity in man," *Med. Biol. Eng. & Comput.*, vol. 18, pp. 591–600, 1980.

[14] N. Mirizzi and U. Scafoglieri, "Optimal direction of the electrogastrographic signal in man," *Med. Biol. Eng. & Comput.*, vol. 21, pp. 385–389, 1983.

[15] J. D. Z. Chen, W. R. Stewart Jr., and R. W. McCallum, "Spectral analysis of episodic rhythmic variations in the cutaneous electrogastrogram," *IEEE Trans. Biomed. Eng.*, vol. 40, pp. 128–135, 1993.

[16] Z. M. Zhou, Z. Hui, R. Shaw, and F. S. Barnes, "Real-time multichannel computerized electrogastrograph," *IEEE Trans. Biomed. Eng.*, vol. 44, pp. 1228–1236, 1997.

[17] J. G. Webster, *Medical Instrumentation. Application and Design*. Boston: John Wiley & Sons, 1998.

[18] N. H. Lovell, F. Magrabi, B. G. Celler, K. Huynh, and H. Garsden, "Web-based acquisition, storage, and retrieval of biomedical signals," *IEEE Eng. Med. Biol. Mag.*, vol. 20, pp. 38–44, 2001.

[19] A. S. Gevins and A. Rémond, *Handbook of Electroencephalography and Clinical Neurophysiology: Methods of Analysis of Brain Electrical and Magnetic Signals*, vol. 1. Amsterdam/New York: Elsevier, 1987.

[20] A. Taddei, G. Distante, M. Emdin, P. Pisani, G. B. Moody, C. Zeelenberg, and C. Marchesi, "The European ST-T database: Standards for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography," *Eur. Heart J.*, vol. 13, pp. 1164–1172, 1992.

[21] R. G. Mark, P. S. Schluter, G. B. Moody, P. H. Devlin, and D. Chernoff, "An annotated ECG database for evaluating arrhythmia detectors," in *Proc. IEEE Frontiers Eng. Health Care*, pp. 205–210, 1982.

[22] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database. History, lessons learned, and its influence on current and future databases," *IEEE Eng. Med. Biol. Mag.*, vol. 20, pp. 45–50, 2001.

[23] R. E. Hermes, D. B. Geselowitz, and G. C. Oliver, "Development, distribution, and use of the American Heart Association database for ventricular arrhythmia detector evaluation," in *Proc. Computers in Cardiology*, pp. 263–266, IEEE Computer Society Press, 1980.

[24] F. J. Jager, A. Taddei, G. B. Moody, M. Emdin, G. Antolič, R. Dorn, A. Smrdel, C. Marchesi, and R. G. Mark, "Long-term ST database: A reference for the development and evaluation of automated ischaemia detectors and for the study of the dynamics of myocardial ischaemia," *Med. Biol. Eng. & Comput.*, vol. 41, pp. 172–183, 2003.

[25] G. B. Moody, W. K. Muldrow, and R. G. Mark, "A noise stress test for arrhythmia detectors," in *Proc. Computers in Cardiology*, pp. 381–384, IEEE Computer Society Press, 1984.

[26] G. B. Moody and R. G. Mark, "A database to support development and evaluation of intelligent intensive care monitoring," in *Proc. Computers in Cardiology*, pp. 657–660, IEEE Press, 1996.

[27] I. Korhonen, J. Ojaniemi, K. Nieminen, M. van Gils, A. Heikelä, and A. Kari, "Building the IMPROVE data library," *IEEE Eng. Med. Biol. Mag.*, vol. 16, pp. 25–32, 1997.

[28] J. Gade, I. Korhonen, M. van Gils, P. Weller, and L. Pesu, "Technical description of the IBIS data library," *Comp. Meth. Prog. Biomed.*, vol. 3, pp. 175–186, 2000.

[29] S. M. Jakob, K. Nieminen, J. Karhu, and J. Takala, "IBIS data library: Clinical description of the Finnish database," *Comp. Meth. Prog. Biomed.*, vol. 3, pp. 161–166, 2000.

[30] Y. Ichimaru and G. B. Moody, "Development of the polysomnographic database on CD-ROM," *Psychiatry Clin. Neurosci.*, vol. 53, pp. 175–177, 1999.

[31] G. Klösch, B. Kemp, T. Penzel, A. Schlögl, P. Rappelsberger, E. Trenker, G. Gruber, J. Zeitlhofer, B. Saletu, W. M. Herrmann, S. L. Himanen, D. Kunz, M. L. Barbanoj, J. Röschke, A. Värri, and G. Dorffner, "The SIESTA project polygraphic and clinical database," *IEEE Eng. Med. Biol. Mag.*, vol. 20, pp. 51–57, 2001.

[32] G. B. Moody, R. G. Mark, and A. L. Goldberger, "PhysioNet: A web-based resource for study of physiologic signals," *IEEE Eng. Med. Biol. Mag.*, vol. 20, pp. 70–75, 2001.