1

Introduction to Biomedical Signals

1.1 THE NATURE OF BIOMEDICAL SIGNALS

Living organisms are made up of many component *systems* — the human body, for example, includes the nervous system, the cardiovascular system, and the musculoskeletal system, among others. Each system is made up of several subsystems that carry on many *physiological processes.* For example, the cardiac system performs the important task of rhythmic pumping of blood throughout the body to facilitate the delivery of nutrients, as well as pumping blood through the pulmonary system for oxygenation of the blood itself.

Physiological processes are complex phenomena, including nervous or hormonal stimulation and control; inputs and outputs that could be in the form of physical material, neurotransmitters, or information; and action that could be mechanical, electrical, or biochemical. Most physiological processes are accompanied by or manifest themselves as *signals* that reflect their nature and activities. Such signals could be of many types, including biochemical in the form of hormones and neurotransmitters, electrical in the form of potential or current, and physical in the form of pressure or temperature.

Diseases or defects in a biological system cause alterations in its normal physiological processes, leading to *pathological processes* that affect the performance, health, and general well-being of the system. A pathological process is typically associated with signals that are different in some respects from the corresponding normal signals. If we possess a good understanding of a system of interest, it becomes possible to observe the corresponding signals and assess the state of the system. The task is not very difficult when the signal is simple and appears at the outer surface of

the body. For example, most infections cause a rise in the temperature of the body, which may be sensed very easily, albeit in a relative and *qualitative* manner, via the palm of one's hand. Objective or *quantitative* measurement of temperature requires an instrument, such as a simple thermometer.

A single measurement x of temperature is a *scalar*, and represents the thermal state of the body at a *particular or single instant of time* t (and a particular position). If we record the temperature continuously in some form, say a strip-chart record, we obtain a *signal as a function of time*; such a signal may be expressed in *continuous-time* or *analog* form as $x(t)$. When the temperature is measured at *discrete* points of time, it may be expressed in *discrete-time* form as $x(nT)$ or $x(n)$, where n is the index or measurement sample number of the array of values, and T represents the uniform interval between the time instants of measurement. A discrete-time signal that can take amplitude values only from a limited list of *quantized* levels is called a *digital* signal; the distinction between discrete-time and digital signals is often ignored.

In intensive-care monitoring, the tympanic (ear drum) temperature may sometimes be measured using an infra-red sensor. Occasionally, when catheters are being used for other purposes, a temperature sensor may also be introduced into an artery or the heart to measure the *core* temperature of the body. It then becomes possible to obtain a continuous measurement of temperature, although only a few samples taken at intervals of a few minutes may be stored for subsequent analysis. Figure 1.1 illustrates representations of temperature measurements as a scalar, an array, and a signal that is a function of time. It is obvious that the graphical representation facilitates easier and faster comprehension of trends in the temperature than the numerical format. Long-term recordings of temperature can facilitate the analysis of temperature-regulation mechanisms [15, 16].

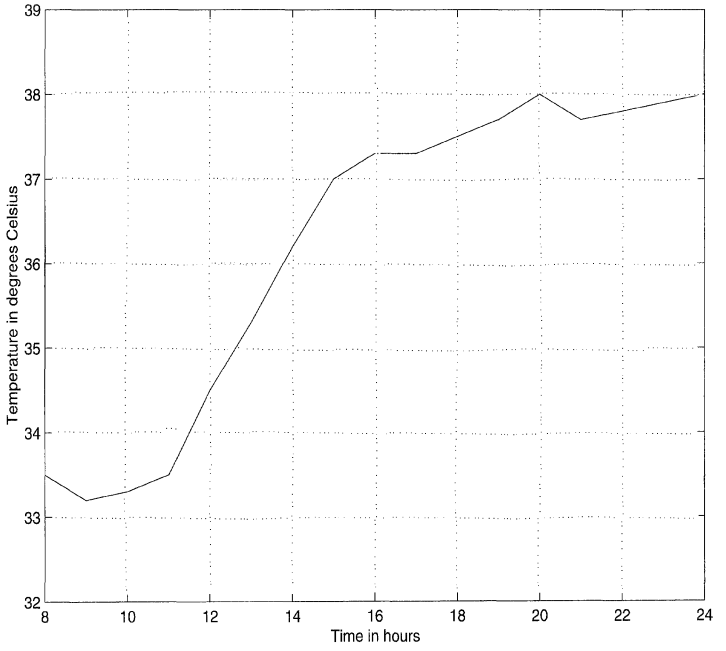
Let us now consider another basic measurement in health care and monitoring: that of blood pressure (BP). Each measurement consists of two values — the systolic pressure and the diastolic pressure. BP is measured in millimeters of mercury ($mm\ of\ Hg$) in clinical practice, although the international standard unit for pressure is the *Pascal*. A single BP measurement could thus be viewed as a *vector* $\mathbf{x} = [x_1, x_2]^T$ with two components: x_1 indicating the systolic pressure and x_2 indicating the diastolic pressure. When BP is measured at a few instants of time, we obtain an array of vectorial values $\mathbf{x}(n)$. In intensive-care monitoring and surgical procedures, a pressure transducer may sometimes be inserted into an artery (along with other intra-arterial or intra-venous devices). It then becomes possible to obtain the arterial systolic and diastolic BP on a continuous-time recording, although the values may be transferred to a computer and stored only at sampled instants of time that are several minutes apart. The signal may then be expressed as a function of time $\mathbf{x}(t)$. Figure 1.2 shows BP measurements as a single two-component vector, as an array, and as a function of time. It is clear that the plot as a function of time facilitates rapid observation of trends in the pressure.

33.5 °C

(a)

Time	08:00	10:00	12:00	14:00	16:00	18:00	20:00	22:00	24:00
°C	33.5	33.3	34.5	36.2	37.3	37.5	38.0	37.8	38.0

(b)



(c)

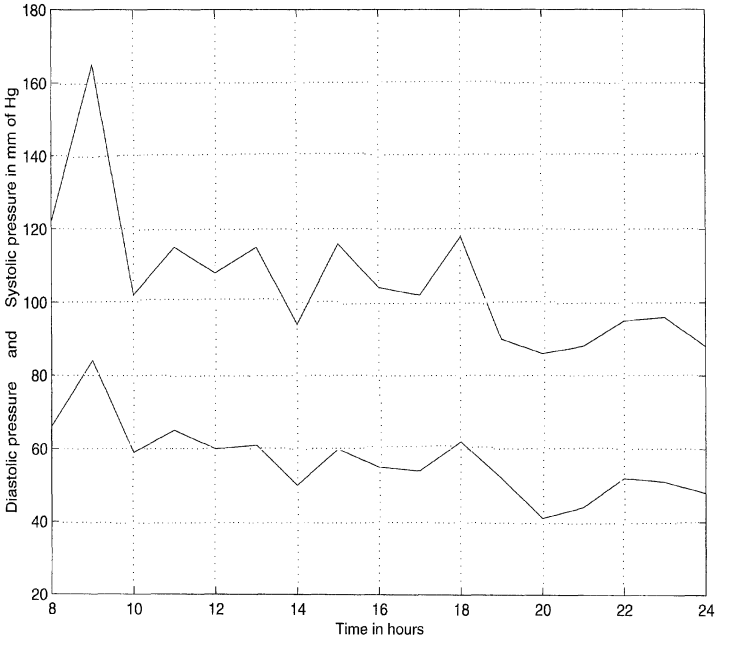
Figure 1.1 Measurements of the temperature of a patient presented as (a) a scalar with one temperature measurement x at a time instant t; (b) an array $x(n)$ made up of several measurements at different instants of time; and (c) a signal $x(t)$ or $x(n)$. The horizontal axis of the plot represents time in *hours*; the vertical axis gives temperature in *degrees Celsius*. Data courtesy of Foothills Hospital, Calgary.

$$\begin{bmatrix} 122 \\ 66 \end{bmatrix}$$

(a)

Time	08:00	10:00	12:00	14:00	16:00	18:00	20:00	22:00	24:00
Systolic	122	102	108	94	104	118	86	95	88
Diastolic	66	59	60	50	55	62	41	52	48

(b)



(c)

Figure 1.2 Measurements of the blood pressure of a patient presented as (a) a single pair or vector of systolic and diastolic measurements $\mathbf{x}$ in *mm of Hg* at a time instant t; (b) an array $\mathbf{x}(n)$ made up of several measurements at different instants of time; and (c) a signal $\mathbf{x}(t)$ or $\mathbf{x}(n)$. Note the use of boldface $\mathbf{x}$ to indicate that each measurement is a vector with two components. The horizontal axis of the plot represents time in *hours*; the vertical axis gives the systolic pressure (upper trace) and the diastolic pressure (lower trace) in *mm of Hg*. Data courtesy of Foothills Hospital, Calgary.

1.2 EXAMPLES OF BIOMEDICAL SIGNALS

The preceding example of body temperature as a signal is a rather simple example of a *biomedical signal*. Regardless of its simplicity, we can appreciate its importance and value in the assessment of the well-being of a child with a fever or that of a critically ill patient in a hospital. The origins and nature of a few other biomedical signals of various types are described in the following subsections, with brief indications of their usefulness in diagnosis. Further detailed discussions on some of the signals will be provided in the context of their analysis for various purposes in the chapters that follow.

1.2.1 The action potential

The action potential (AP) is the electrical signal that accompanies the mechanical contraction of a single cell when stimulated by an electrical current (neural or external) [10, 17, 18, 19, 20, 21]. It is caused by the flow of sodium (Na^+), potassium (K^+), chloride (Cl^-), and other ions across the cell membrane. The action potential is the basic component of all bioelectrical signals. It provides information on the nature of physiological activity at the single-cell level. Recording an action potential requires the isolation of a single cell, and microelectrodes with tips of the order of a few micrometers to stimulate the cell and record the response [10].

Resting potential: Nerve and muscle cells are encased in a semi-permeable membrane that permits selected substances to pass through while others are kept out. Body fluids surrounding cells are conductive solutions containing charged atoms known as ions. In their resting state, membranes of excitable cells readily permit the entry of K^+ and Cl^- ions, but effectively block the entry of Na^+ ions (the permeability for K^+ is 50–100 times that for Na^+). Various ions seek to establish a balance between the inside and the outside of a cell according to charge and concentration. The inability of Na^+ to penetrate a cell membrane results in the following [17]:

- Na^+ concentration inside the cell is far less than that outside.
- The outside of the cell is more positive than the inside of the cell.
- To balance the charge, additional K^+ ions enter the cell, causing higher K^+ concentration inside the cell than outside.
- Charge balance cannot be reached due to differences in membrane permeability for the various ions.
- A state of equilibrium is established with a potential difference, with the inside of the cell being negative with respect to the outside.

A cell in its resting state is said to be *polarized*. Most cells maintain a *resting potential* of the order of -60 to $-100\ mV$ until some disturbance or stimulus upsets the equilibrium.

Depolarization: When a cell is excited by ionic currents or an external stimulus, the membrane changes its characteristics and begins to allow Na^+ ions to enter the cell. This movement of Na^+ ions constitutes an ionic current, which further reduces the membrane barrier to Na^+ ions. This leads to an avalanche effect: Na^+ ions rush into the cell. K^+ ions try to leave the cell as they were in higher concentration inside the cell in the preceding resting state, but cannot move as fast as the Na^+ ions. The net result is that the inside of the cell becomes positive with respect to the outside due to an imbalance of K^+ ions. A new state of equilibrium is reached after the rush of Na^+ ions stops. This change represents the beginning of the *action potential*, with a peak value of about $+20\ mV$ for most cells. An excited cell displaying an action potential is said to be *depolarized*; the process is called *depolarization.*

Repolarization: After a certain period of being in the depolarized state the cell becomes polarized again and returns to its resting potential via a process known as *repolarization.* Repolarization occurs by processes that are analogous to those of depolarization, except that instead of Na^+ ions, the principal ions involved in repolarization are K^+ ions [19]. Membrane depolarization, while increasing the permeability for Na^+ ions, also increases the permeability of the membrane for K^+ ions via a specific class of ion channels known as voltage-dependent K^+ channels. Although this may appear to be paradoxical at first glance, the key to the mechanism for repolarization lies in the time-dependence and voltage-dependence of the membrane permeability changes for K^+ ions compared with that for Na^+ ions. The permeability changes for K^+ during depolarization occur considerably more slowly than those for Na^+ ions, hence the initial depolarization is caused by an inrush of Na^+ ions. However, the membrane permeability changes for Na^+ spontaneously decrease near the peak of the depolarization, whereas those for K^+ ions are beginning to increase. Hence, during repolarization, the predominant membrane permeability is for K^+ ions. Because K^+ concentration is much higher inside the cell than outside, there is a net efflux of K^+ from the cell, which makes the inside more negative, thereby effecting repolarization back to the resting potential.

It should be noted that the voltage-dependent K^+ permeability change is due to a distinctly different class of ion channels than those that are responsible for setting the resting potential. A mechanism known as the $Na^+ - K^+$ pump extrudes Na^+ ions in exchange for transporting K^+ ions back into the cell. However, this transport mechanism carries very little current in comparison with ion channels, and therefore makes a minor contribution to the repolarization process. The $Na^+ - K^+$ pump is essential for resetting the $Na^+ - K^+$ balance of the cell, but the process occurs on a longer time scale than the duration of an action potential.

Nerve and muscle cells repolarize rapidly, with an action potential duration of about $1\ ms$. Heart muscle cells repolarize slowly, with an action potential duration of $150 - 300\ ms$.

The action potential is always the same for a given cell, regardless of the method of excitation or the intensity of the stimulus beyond a threshold: this is known as the *all-or-none* or all-or-nothing phenomenon. After an action potential, there is a period during which a cell cannot respond to any new stimulus, known as the *absolute refractory period* (about $1\ ms$ in nerve cells). This is followed by a *relative*

refractory period (several ms in nerve cells), when another action potential may be triggered by a much stronger stimulus than in the normal situation.

Figure 1.3 shows action potentials recorded from individual rabbit ventricular and atrial myocytes (muscle cells) [19]. Figure 1.4 shows a ventricular myocyte in its relaxed and fully contracted states. The tissues were first incubated in digestive enzymes, principally collagenase, and then dispersed into single cells using gentle mechanical agitation. The recording electrodes were glass patch pipettes; a whole-cell, current-clamp recording configuration was used to obtain the action potentials. The cells were stimulated at low rates (once per 8 s); this is far less than physiological rates. Moreover, the cells were maintained at 20^o C, rather than body temperature. Nevertheless, the major features of the action potentials shown are similar to those recorded under physiological conditions.

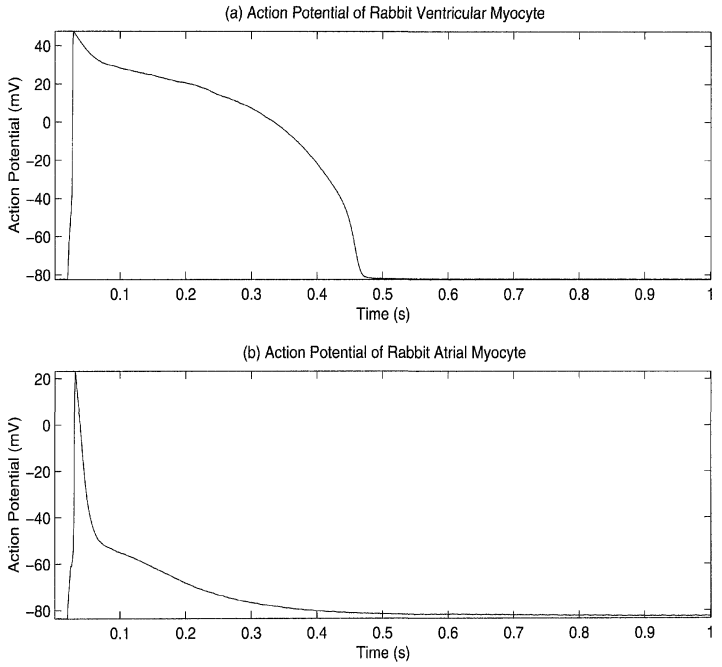
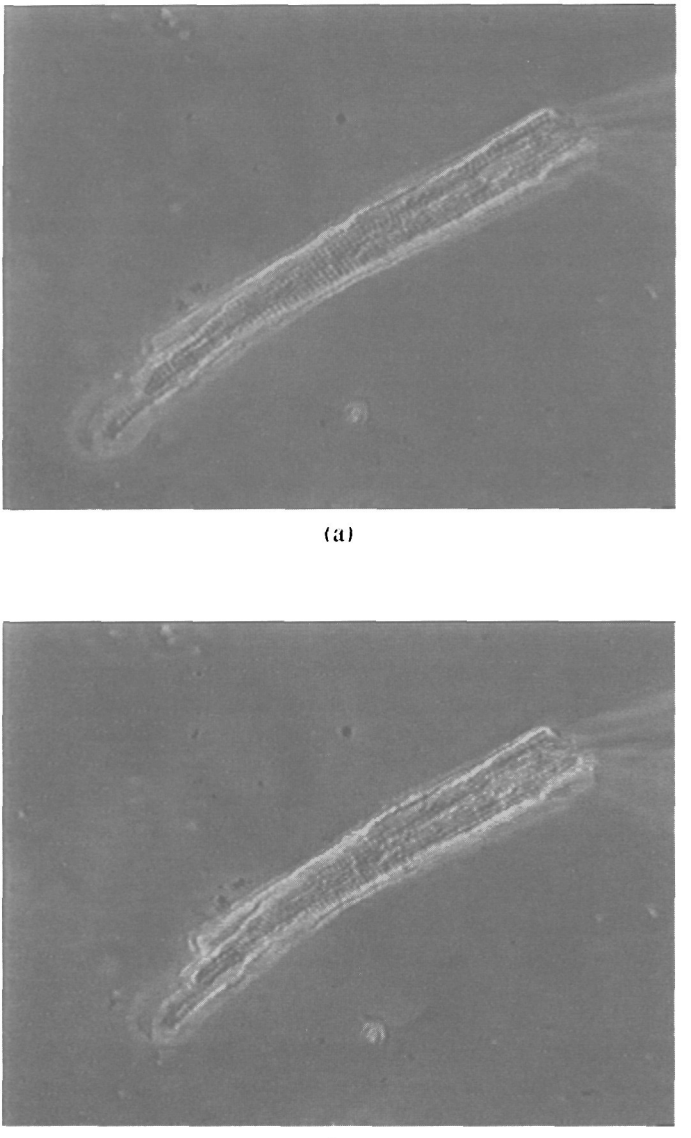


Figure 1.3 Action potentials of rabbit ventricular and atrial myocytes. Data courtesy of R. Clark, Department of Physiology and Biophysics, University of Calgary.

The resting membrane potential of the cells (from 0 to 20 ms in the plots in Figure 1.3) is about -83 mV. A square pulse of current, 3 ms in duration and 1 nA in amplitude, was passed through the recording electrode and across the cell membrane, causing the cell to depolarize rapidly. The ventricular myocyte exhibits a depolarized potential of about $+40$ mV; it then slowly declines back to the resting potential level over an interval of about 500 ms. The initial, rapid depolarization of



(a)

(b)

Figure 1.4 A single ventricular myocyte (of a rabbit) in its (a) relaxed and (b) fully contracted states. The length of the myocyte is approximately 25 μm. The tip of the glass pipette, faintly visible at the upper-right end of the myocyte, is approximately 2 μm wide. Images courtesy of R. Clark, Department of Physiology and Biophysics, University of Calgary.

the atrial cell is similar to that of the ventricular cell, but does not overshoot zero membrane potential as much as the ventricular action potential; repolarization occurs much more quickly than is the case for the ventricular cell.

Propagation of an action potential: An action potential propagates along a muscle fiber or an unmyelinated nerve fiber as follows [22]: Once initiated by a stimulus, the action potential propagates along the whole length of a fiber without decrease in amplitude by progressive depolarization of the membrane. Current flows from a depolarized region through the intra-cellular fluid to adjacent inactive regions, thereby depolarizing them. Current also flows through the extra-cellular fluids, through the depolarized membrane, and back into the intra-cellular space, completing the local circuit. The energy to maintain conduction is supplied by the fiber itself.

Myelinated nerve fibers are covered by an insulating sheath of *myelin*. The sheath is interrupted every few millimeters by spaces known as the *nodes of Ranvier*, where the fiber is exposed to the interstitial fluid. Sites of excitation and changes of membrane permeability exist only at the nodes, and current flows by jumping from one node to the next in a process known as *saltatory conduction*.

1.2.2 The electroneurogram (ENG)

The ENG is an electrical signal observed as a stimulus and the associated nerve action potential propagate over the length of a nerve. It may be used to measure the velocity of propagation (or conduction velocity) of a stimulus or action potential in a nerve [10]. ENGs may be recorded using concentric needle electrodes or silver – silver-chloride electrodes ($Ag - AgCl$) at the surface of the body.

Conduction velocity in a peripheral nerve may be measured by stimulating a motor nerve and measuring the related activity at two points that are a known distance apart along its course. In order to minimize muscle contraction and other undesired effects, the limb is held in a relaxed posture and a strong but short stimulus is applied in the form of a pulse of about 100 V amplitude and $100 - 300$ μs duration [10]. The difference in the latencies of the ENGs recorded over the associated muscle gives the conduction time. Knowing the separation distance between the stimulus sites, it is possible to determine the conduction velocity in the nerve [10]. ENGs have amplitudes of the order of 10 μV and are susceptible to power-line interference and instrumentation noise.

Figure 1.5 illustrates the ENGs recorded in a nerve conduction velocity study. The stimulus was applied to the ulnar nerve. The ENGs were recorded at the wrist (marked "Wrist" in the figure), just below the elbow (BElbow), and just above the elbow (AElbow) using surface electrodes, amplified with a gain of 2,000, and filtered to the bandwidth $10 - 10,000$ Hz. The three traces in the figure indicate increasing latencies with respect to the stimulus time point, which is the left margin of the plots. The responses shown in the figure are normal, indicate a BElbow – Wrist latency of 3.23 ms, and lead to a nerve conduction velocity of 64.9 m/s.

Typical values of propagation rate or nerve conduction velocity are [22, 10, 23]:

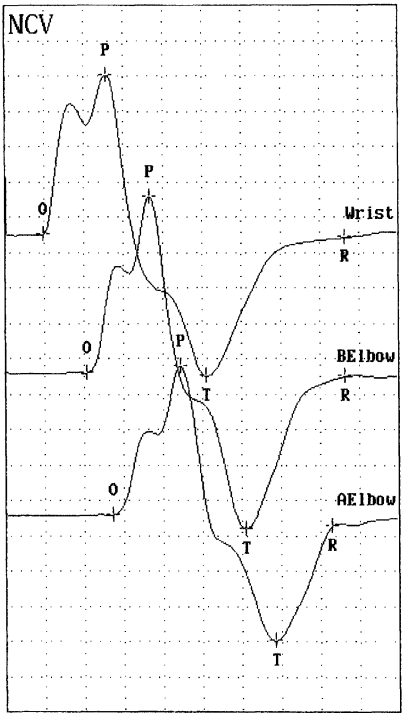


Figure 1.5 Nerve conduction velocity measurement via electrical stimulation of the ulnar nerve. The grid boxes represent 3 ms in width and 2 μV in height. AElbow: above the elbow. BElbow: below the elbow. O: onset. P: Peak. T: trough. R: recovery of base-line. Courtesy of M. Wilson and C. Adams, Alberta Children's Hospital, Calgary.

- $45 - 70\ m/s$ in nerve fibers;
- $0.2 - 0.4\ m/s$ in heart muscle;
- $0.03 - 0.05\ m/s$ in time-delay fibers between the atria and ventricles.

Neural diseases may cause a decrease in conduction velocity.

1.2.3 The electromyogram (EMG)

Skeletal muscle fibers are considered to be twitch fibers because they produce a mechanical twitch response for a single stimulus and generate a propagated action potential. Skeletal muscles are made up of collections of *motor units* (MUs), each of which consists of an anterior horn cell (or motoneuron or motor neuron), its axon, and all muscle fibers innervated by that axon. A motor unit is the smallest muscle unit that can be activated by volitional effort. The constituent fibers of a motor unit are activated synchronously. Component fibers of a motor unit extend lengthwise in loose bundles along the muscle. In cross-section, the fibers of a given motor unit are interspersed with the fibers of other motor units [22, 10, 24]. Figure 1.6 (top panel) illustrates a motor unit in schematic form [24].

Large muscles for gross movement have hundreds of fibers per motor unit; muscles for precise movement have fewer fibers per motor unit. The number of muscle fibers per motor nerve fiber is known as the *innervation ratio.* For example, it has been estimated that the platysma muscle (of the neck) has $1,826$ large nerve fibers controlling $27,100$ muscle fibers with $1,096$ motor units and an innervation ratio of 25, whereas the first dorsal interosseus (finger) muscle has 199 large nerve fibers and $40,500$ muscle fibers with 119 motor units and an innervation ratio of 340 [22]. The mechanical output (contraction) of a muscle is the net result of stimulation and contraction of several of its motor units.

When stimulated by a neural signal, each motor unit contracts and causes an electrical signal that is the summation of the action potentials of all of its constituent cells. This is known as the *single-motor-unit action potential* (SMUAP, or simply MUAP), and may be recorded using needle electrodes inserted into the muscle region of interest. Normal SMUAPs are usually biphasic or triphasic, $3 - 15\ ms$ in duration, $100 - 300\ \mu V$ in amplitude, and appear with frequency in the range of $6 - 30/s$ [10, 22]. The shape of a recorded SMUAP depends upon the type of the needle electrode used, its positioning with respect to the active motor unit, and the projection of the electrical field of the activity onto the electrodes. Figure 1.7 illustrates simultaneous recordings of the activities of a few motor units from three channels of needle electrodes [25]. Although the SMUAPs are biphasic or triphasic, the same SMUAP displays variable shape from one channel to another. (*Note:* The action potentials in Figure 1.3 are monophasic; the first two SMUAPs in Channel 1 in Figure 1.7 are biphasic, and the third SMUAP in the same signal is triphasic.)

The shape of SMUAPs is affected by disease. Figure 1.8 illustrates SMUAP trains of a normal subject and those of patients with neuropathy and myopathy. Neuropathy causes slow conduction and/or desynchronized activation of fibers, and a polyphasic

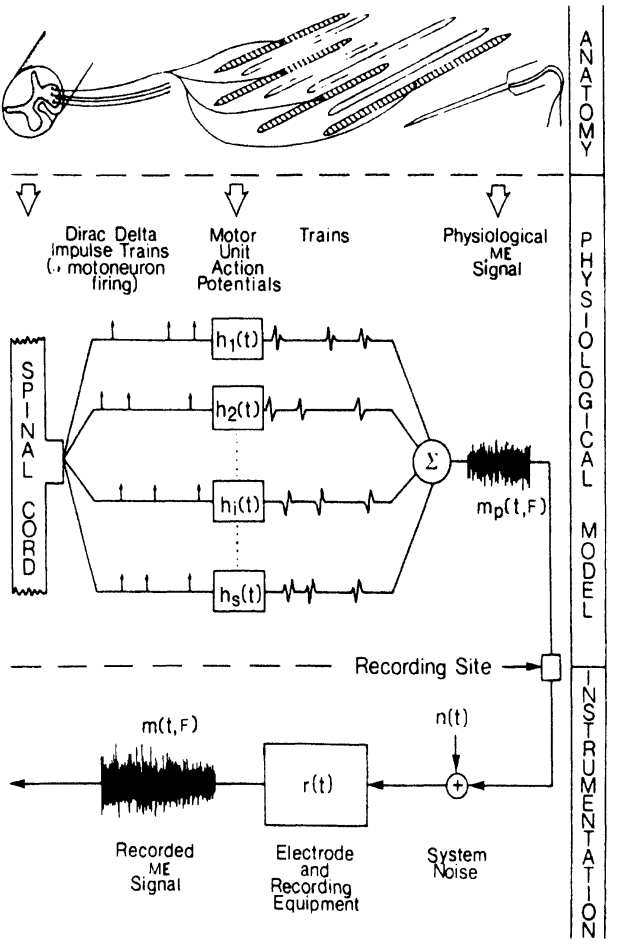


Figure 1.6 Schematic representation of a motor unit and model for the generation of EMG signals. Top panel: A motor unit includes an anterior horn cell or motor neuron (illustrated in a cross-section of the spinal cord), an axon, and several connected muscle fibers. The hatched fibers belong to one motor unit; the non-hatched fibers belong to other motor units. A needle electrode is also illustrated. Middle panel: The firing pattern of each motor neuron is represented by an impulse train. Each system $h_i(t)$ shown represents a motor unit that is activated and generates a train of SMUAPs. The net EMG is the sum of several SMUAP trains. Bottom panel: Effects of instrumentation on the EMG signal acquired. The observed EMG is a function of time t and muscular force produced F. Reproduced with permission from C.J. de Luca, Physiology and mathematics of myoelectric signals, *IEEE Transactions on Biomedical Engineering*, 26:313–325, 1979. ©IEEE.

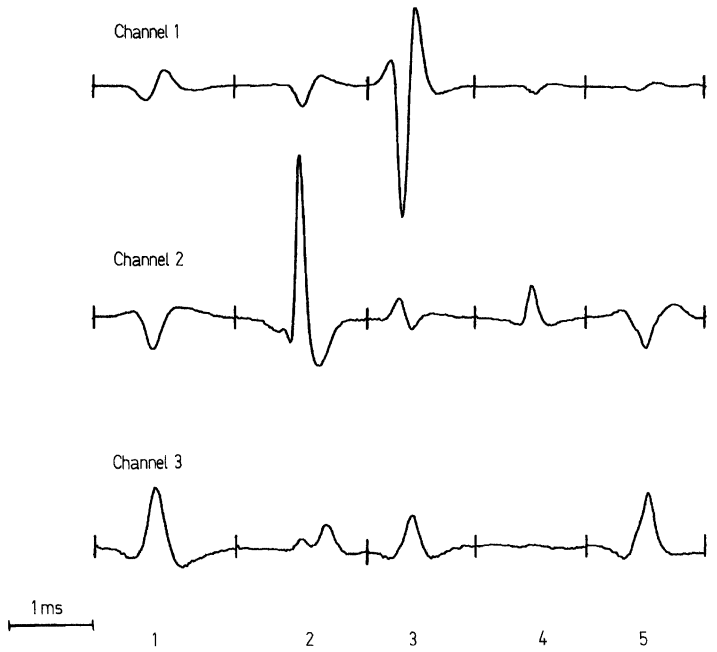


Figure 1.7 SMUAP trains recorded simultaneously from three channels of needle electrodes. Observe the different shapes of the same SMUAPs projected onto the axes of the three channels. Three different motor units are active over the duration of the signals illustrated. Reproduced with permission from B. Mambrito and C.J. de Luca, Acquisition and decomposition of the EMG signal, in *Progress in Clinical Neurophysiology*, Volume 10: Computer-aided Electromyography, Editor: J.E. Desmedt, pp 52–72, 1983. ©S. Karger AG, Basel, Switzerland.

SMUAP with an amplitude larger than normal. The same motor unit may be observed to fire at higher rates than normal before more motor units are recruited. Myopathy involves loss of muscle fibers in motor units, with the neurons presumably intact. *Splintering* of SMUAPs occurs due to asynchrony in activation as a result of patchy destruction of fibers (e.g., in muscular dystrophy), leading to polyphasic SMUAPs. More motor units may be observed to be recruited at low levels of effort.

Gradation of muscular contraction: Muscular contraction levels are controlled in two ways:

- *Spatial recruitment,* by activating new motor units with increasing effort; and
- *Temporal recruitment,* by increasing the frequency of discharge (firing rate) of each motor unit with increasing effort.

Motor units are activated at different times and at different frequencies causing asynchronous contraction. The twitches of individual motor units sum and fuse to form tetanic contraction and increased force. Weak volitional effort causes motor units to fire at about $5 - 15$ *pps* (pulses per second). As greater tension is developed, an *interference pattern* EMG is obtained, with the constituent and active motor units firing in the range of $25 - 50$ *pps*. Grouping of MUAPs has been observed as fatigue develops, leading to decreased high-frequency content and increased amplitude in the EMG [24].

Spatio-temporal summation of the MUAPs of all of the active motor units gives rise to the EMG of the muscle. EMG signals recorded using surface electrodes are complex signals including interference patterns of several MUAP trains and are difficult to analyze. An EMG signal indicates the level of activity of a muscle, and may be used to diagnose neuromuscular diseases such as neuropathy and myopathy.

Figure 1.9 illustrates an EMG signal recorded from the crural diaphragm of a dog using fine-wire electrodes sewn in-line with the muscle fibers and placed 10 *mm* apart [26]. The signal represents one period of breathing (inhalation being the active part as far as the muscle and EMG are concerned). It is seen that the overall level of activity in the signal increases during the initial phase of inhalation. Figure 1.10 shows the early parts of the same signal on an expanded time scale. SMUAPs are seen at the beginning stages of contraction, followed by increasingly complex interference patterns of several MUAPs.

Signal-processing techniques for the analysis of EMG signals will be discussed in Sections 5.2.4, 5.6, 5.9, 5.10, 7.2.1, and 7.3.

1.2.4 The electrocardiogram (ECG)

The ECG is the electrical manifestation of the contractile activity of the heart, and can be recorded fairly easily with surface electrodes on the limbs or chest. The ECG is perhaps the most commonly known, recognized, and used biomedical signal. The rhythm of the heart in terms of beats per minute (*bpm*) may be easily estimated by counting the readily identifiable waves. More important is the fact that the ECG

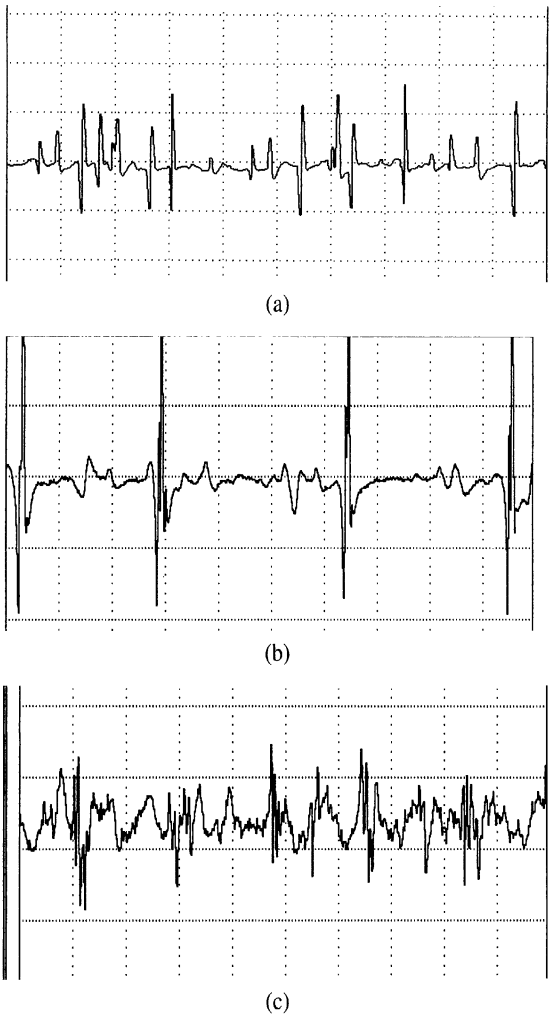


Figure 1.8 Examples of SMUAP trains. (a) From the right deltoid of a normal subject, male, 11 years; the SMUAPs are mostly biphasic, with duration in the range $3 - 5$ ms. (b) From the deltoid of a six-month-old male patient with brachial plexus injury (neuropathy); the SMUAPs are polyphasic and large in amplitude (800 μV), and the same motor unit is firing at a relatively high rate at low-to-medium levels of effort. (c) From the right biceps of a 17-year-old male patient with myopathy; the SMUAPs are polyphasic and indicate early recruitment of more motor units at a low level of effort. The signals were recorded with gauge 20 needle electrodes. The width of each grid box represents a duration of 20 ms; its height represents an amplitude of 200 μV. Courtesy of M. Wilson and C. Adams, Alberta Children's Hospital, Calgary.

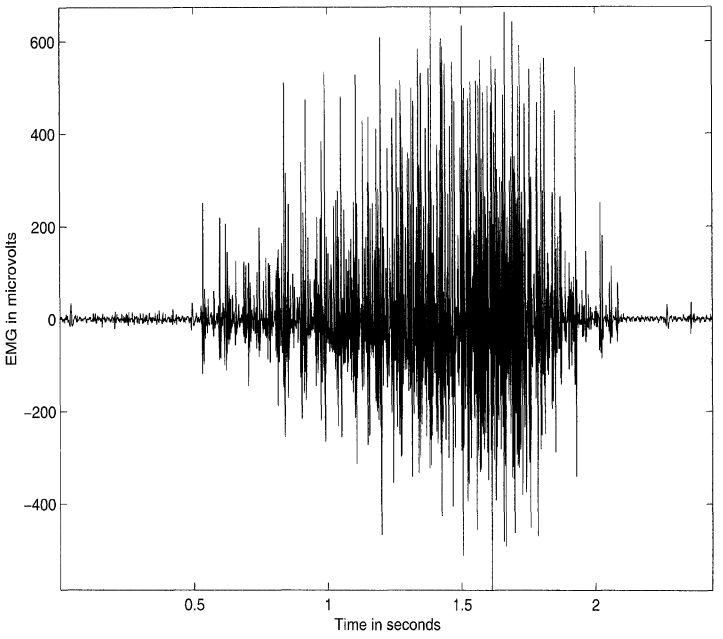


Figure 1.9 EMG signal recorded from the crural diaphragm muscle of a dog using implanted fine-wire electrodes. Data courtesy of R.S. Platt and P.A. Easton, Department of Clinical Neurosciences, University of Calgary.

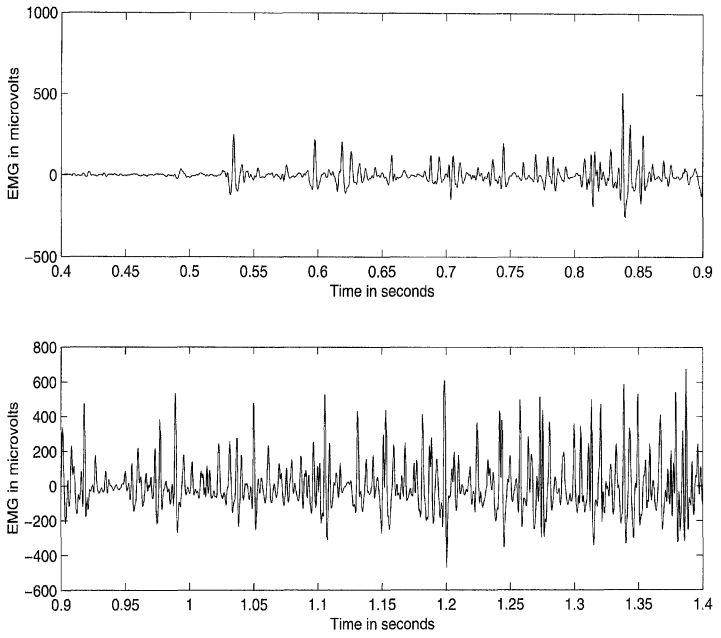


Figure 1.10 The initial part of the EMG signal in Figure 1.9 shown on an expanded time scale. Observe the SMUAPs at the initial stages of contraction, followed by increasingly complex interference patterns of several MUAPs. Data courtesy of R.S. Platt and P.A. Easton, Department of Clinical Neurosciences, University of Calgary.

waveshape is altered by cardiovascular diseases and abnormalities such as myocardial ischemia and infarction, ventricular hypertrophy, and conduction problems.

The heart: The heart is a four-chambered pump with two atria for collection of blood and two ventricles for pumping out of blood. Figure 1.11 shows a schematic representation of the four chambers and the major vessels connecting to the heart. The resting or filling phase of a cardiac chamber is called *diastole*; the contracting or pumping phase is called *systole*.

The right atrium (or auricle, RA) collects impure blood from the superior and inferior vena cavae. During atrial contraction, blood is passed from the right atrium to the right ventricle (RV) through the tricuspid valve. During ventricular systole, the impure blood in the right ventricle is pumped out through the pulmonary valve to the lungs for purification (oxygenation).

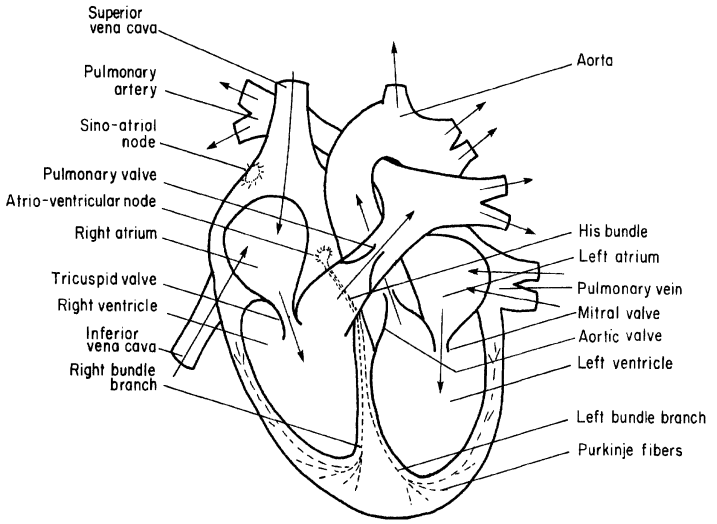


Figure 1.11 Schematic representation of the chambers, valves, vessels, and conduction system of the heart.

The left atrium (LA) receives purified blood from the lungs, which is passed on during atrial contraction to the left ventricle (LV) via the mitral valve. The left ventricle is the largest and most important cardiac chamber. The left ventricle contracts the strongest among the cardiac chambers, as it has to pump out the oxygenated blood through the aortic valve and the aorta against the pressure of the rest of the vascular system of the body. Due to the higher level of importance of contraction of the ventricles, the terms systole and diastole are applied to the ventricles by default.

The heart rate (HR) or cardiac rhythm is controlled by specialized pacemaker cells that form the sino-atrial (SA) node located at the junction of the superior vena cava and the right atrium [23]. The firing rate of the SA node is controlled by impulses

from the autonomous and central nervous systems leading to the delivery of the neurotransmitters acetylcholine (for vagal stimulation, causing a reduction in heart rate) or epinephrine (for sympathetic stimulation, causing an increase in the heart rate). The normal (resting) heart rate is about 70 *bpm*. The heart rate is lower during sleep, but abnormally low heart rates below 60 *bpm* during activity could indicate a disorder called *bradycardia*. The instantaneous heart rate could reach values as high as 200 *bpm* during vigorous exercise or athletic activity; a high resting heart rate could be due to illness, disease, or cardiac abnormalities, and is termed *tachycardia*.

The electrical system of the heart: Co-ordinated electrical events and a specialized conduction system intrinsic and unique to the heart play major roles in the rhythmic contractile activity of the heart. The SA node is the basic, natural cardiac pacemaker that triggers its own train of action potentials. The action potential of the SA node propagates through the rest of the heart, causing a particular pattern of excitation and contraction (see Figure 1.12). The sequence of events and waves in a cardiac cycle is as follows [23]:

1. The SA node fires.

2. Electrical activity is propagated through the atrial musculature at comparatively low rates, causing slow-moving depolarization (contraction) of the atria. This results in the P wave in the ECG (see Figure 1.13). Due to the slow contraction of the atria and their small size, the P wave is a slow, low-amplitude wave, with an amplitude of about $0.1 - 0.2\ mV$ and a duration of about $60 - 80\ ms$.

3. The excitation wave faces a propagation delay at the atrio-ventricular (AV) node, which results in a normally iso-electric segment of about $60 - 80\ ms$ after the P wave in the ECG, known as the PQ segment. The pause assists in the completion of the transfer of blood from the atria to the ventricles.

4. The His bundle, the bundle branches, and the Purkinje system of specialized conduction fibers propagate the stimulus to the ventricles at a high rate.

5. The wave of stimulus spreads rapidly from the apex of the heart upwards, causing rapid depolarization (contraction) of the ventricles. This results in the QRS wave of the ECG — a sharp biphasic or triphasic wave of about $1\ mV$ amplitude and $80\ ms$ duration (see Figure 1.13).

6. Ventricular muscle cells possess a relatively long action potential duration of $300 - 350\ ms$ (see Figure 1.3). The plateau portion of the action potential causes a normally iso-electric segment of about $100 - 120\ ms$ after the QRS, known as the ST segment.

7. Repolarization (relaxation) of the ventricles causes the slow T wave, with an amplitude of $0.1 - 0.3\ mV$ and duration of $120 - 160\ ms$ (see Figure 1.13).

Any disturbance in the regular rhythmic activity of the heart is termed *arrhythmia*. Cardiac arrhythmia may be caused by irregular firing patterns from the SA node, or

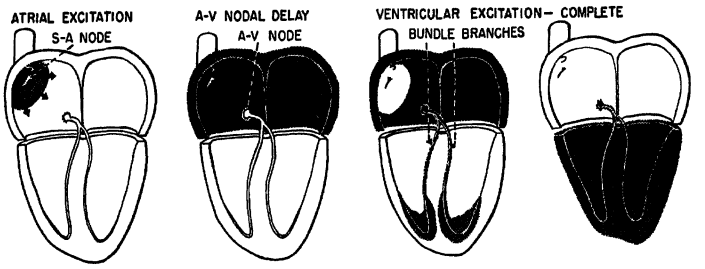


Figure 1.12 Propagation of the excitation pulse through the heart. Reproduced with permission from R.F. Rushmer, *Cardiovascular Dynamics*, 4th edition, ©W.B. Saunders, Philadelphia, PA, 1976.

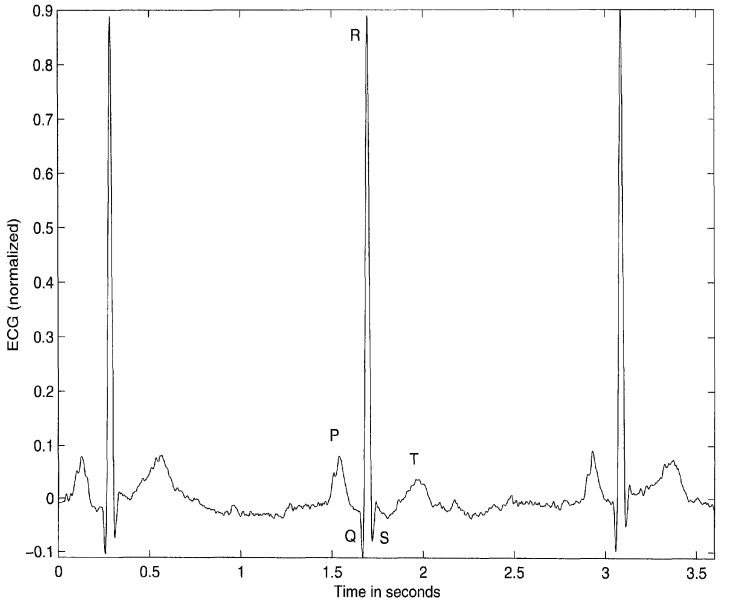


Figure 1.13 A typical ECG signal (male subject of age 24 years). (*Note:* Signal values are not calibrated, that is, specified in physical units, in many applications. As is the case in this plot, signal values in plots in this book are in arbitrary or normalized units unless specified.)

by abnormal and additional pacing activity from other parts of the heart. Many parts of the heart possess inherent rhythmicity and pacemaker properties; for example, the SA node, the AV node, the Purkinje fibers, atrial tissue, and ventricular tissue. If the SA node is depressed or inactive, any one of the above tissues may take over the role of the pacemaker or introduce *ectopic* beats. Different types of abnormal rhythm (arrhythmia) result from variations in the site and frequency of impulse formation. Premature ventricular contractions (PVCs) caused by ectopic foci on the ventricles upset the regular rhythm and may lead to ventricular dissociation and fibrillation — a state of disorganized contraction of the ventricles independent of the atria — resulting in no effective pumping of blood and possibly death. The waveshapes of PVCs are usually very different from that of the normal beats of the same subject due to the different conduction paths of the ectopic impulses and the associated abnormal contraction events. Figure 1.14 shows an ECG signal with a few normal beats and two PVCs. (See Figure 9.5 for an illustration of ventricular bigeminy, where every second pulse from the SA node is replaced by a PVC with a full compensatory pause.)

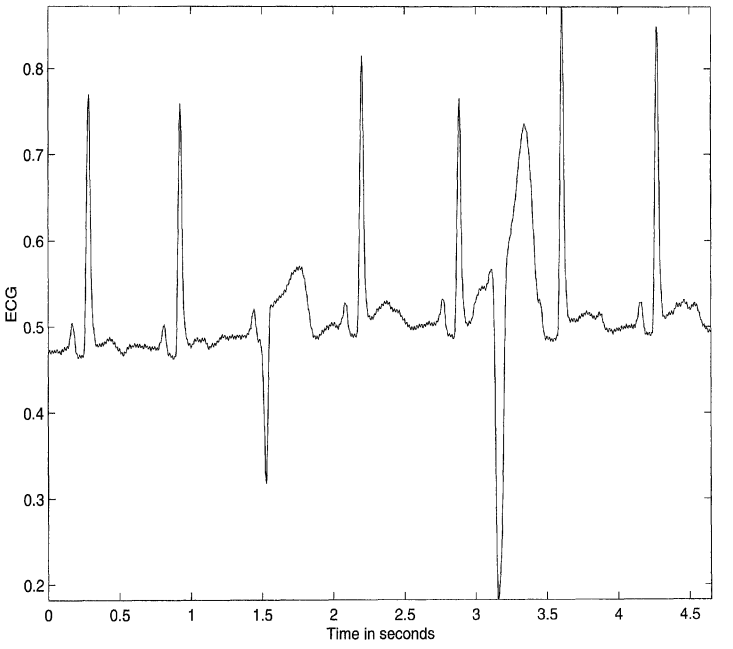


Figure 1.14 ECG signal with PVCs. The third and sixth beats are PVCs. The first PVC has blocked the normal beat that would have appeared at about the same time instant, but the second PVC has not blocked any normal beat triggered by the SA node. Data courtesy of G. Groves and J. Tyberg, Department of Physiology and Biophysics, University of Calgary.

The QRS waveshape is affected by conduction disorders; for example, bundle-branch block causes a widened and possibly jagged QRS. Figure 1.15 shows the ECG

signal of a patient with right bundle-branch block. Observe the wider-than-normal QRS complex, which also displays a waveshape that is significantly different from the normal QRS waves. Ventricular hypertrophy (enlargement) could also cause a wider-than-normal QRS.

The ST segment, which is normally iso-electric (flat and in line with the PQ segment) may be elevated or depressed due to myocardial ischemia (reduced blood supply to a part of the heart muscles due to a block in the coronary arteries) or due to myocardial infarction (dead myocardial tissue incapable of contraction due to total lack of blood supply). Many other diseases cause specific changes in the ECG waveshape: the ECG is a very important signal that is useful in heart-rate (rhythm) monitoring and the diagnosis of cardiovascular diseases.

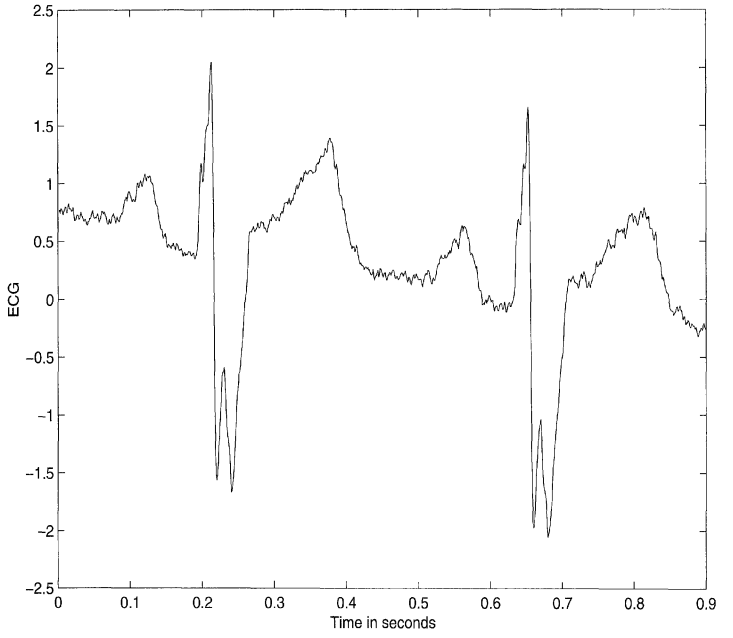


Figure 1.15 ECG signal of a patient with right bundle-branch block and hypertrophy (male patient of age 3 months). The QRS complex is wider than normal, and displays an abnormal, jagged waveform due to desynchronized contraction of the ventricles. (The signal also has a base-line drift, which has not been corrected for.)

ECG signal acquisition: In clinical practice, the standard 12-channel ECG is obtained using four limb leads and chest leads in six positions [23, 27]. The right leg is used to place the reference electrode. The left arm, right arm, and left leg are used to get leads I, II, and III. A combined reference known as *Wilson's central terminal* is formed by combining the left arm, right arm, and left leg leads, and is used as the reference for chest leads. The *augmented* limb leads known as aVR, aVL, and aVF

(aV for the augmented lead, R for the right arm, L for the left arm, and F for the left foot) are obtained by using the exploring electrode on the limb indicated by the lead name, with the reference being Wilson's central terminal without the exploring limb lead.

Figure 1.16 shows the directions of the axes formed by the six limb leads. The hypothetical equilateral triangle formed by leads I, II, and III is known as *Einthoven's triangle.* The center of the triangle represents Wilson's central terminal. Schematically, the heart is assumed to be placed at the center of the triangle. The six leads measure projections of the three-dimensional (3D) cardiac electrical vector onto the axes illustrated in Figure 1.16. The six axes sample the $0^o - 180^o$ range in steps of approximately 30^o. The projections facilitate viewing and analysis of the electrical activity of the heart and from different perspectives in the frontal plane.

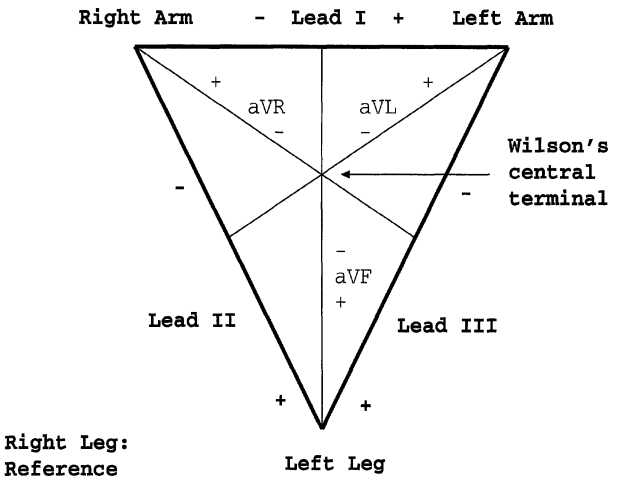


Figure 1.16 Einthoven's triangle and the axes of the six ECG leads formed by using four limb leads.

The six chest leads (written as V1 – V6) are obtained from six standardized positions on the chest [23] with Wilson's central terminal as the reference. The positions for placement of the precordial (chest) leads are indicated in Figure 1.17. The V1 and V2 leads are placed at the fourth intercostal space just to the right and left of the sternum, respectively. V4 is recorded at the fifth intercostal space at the left midclavicular line. The V3 lead is placed half-way between the V2 and V4 leads. The V5 and V6 leads are located at the same level as the V4 lead, but at the anterior axillary line and the midaxillary line, respectively. The six chest leads permit viewing the cardiac electrical vector from different orientations in a cross-sectional plane: V5 and V6 are most sensitive to left ventricular activity; V3 and V4 depict septal activity best; V1 and V2 reflect well activity in the right-half of the heart.

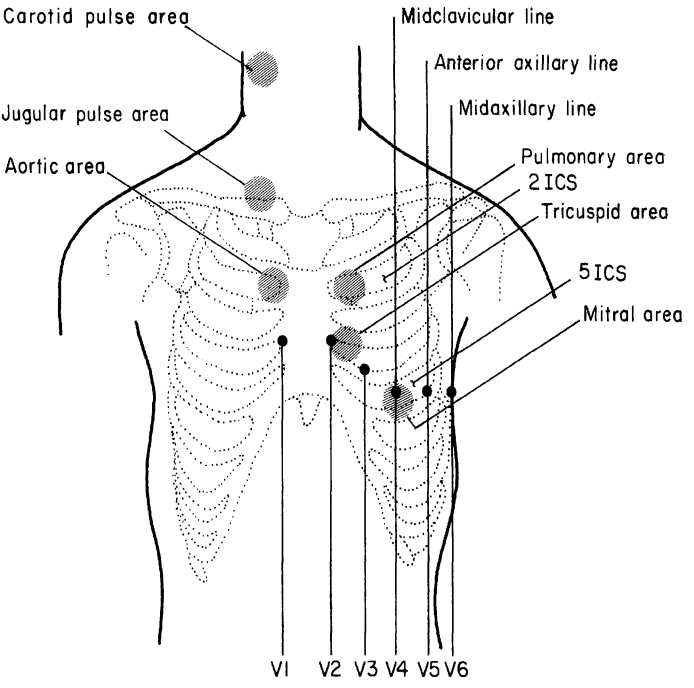


Figure 1.17 Positions for placement of the precordial (chest) leads V1 – V6 for ECG, auscultation areas for heart sounds, and pulse transducer positions for the carotid and jugular pulse signals. ICS: intercostal space.

In spite of being redundant, the 12-lead system serves as the basis of the standard clinical ECG. Clinical ECG interpretation is mainly empirical, based on experimental knowledge. A compact and efficient system has been proposed for *vectorcardiography* or VCG [28, 23], where loops inscribed by the 3D cardiac electrical vector in three mutually orthogonal planes, namely, the frontal, horizontal, and sagittal planes, are plotted and analyzed. Regardless, the 12-lead scalar ECG is the most commonly used procedure in clinical practice.

As the external ECG is a projection of the internal 3D cardiac electrical vector, the external recordings are not unique. Some of the lead inter-relationships are [23, 27]:

- II = I + III
- aVL = (I − III) / 2.

Some of the important features of the standard clinical ECG are:

- A rectangular calibration pulse of 1 mV amplitude and 200 ms duration is applied to produce a pulse of 1 cm height on the paper plot.
- The paper speed used is 25 mm/s, resulting in a graphical scale of 0.04 s/mm or 40 ms/mm. The calibration pulse width will then be 5 mm.
- The ECG signal peak value is normally about 1 mV.
- The amplifier gain used is 1,000.
- Clinical ECG is usually filtered to a bandwidth of about 0.05 − 100 Hz, with a recommended sampling rate of 500 Hz for diagnostic ECG. Distortions in the shape of the calibration pulse may indicate improper filter settings or a poor signal acquisition system.
- ECG for heart-rate monitoring could use a reduced bandwidth 0.5 − 50 Hz.
- High-resolution ECG requires a greater bandwidth of 0.05 − 500 Hz.

Figure 1.18 shows the 12-lead ECG of a normal male adult. The system used to obtain the illustration records three channels at a time: leads I, II, II; aVR, aVL, aVF; V1, V2, V3; and V4, V5, V6 are recorded in the three available channels simultaneously. Other systems may record one channel at a time. Observe the changing shape of the ECG waves from one lead to another. A well-trained cardiologist will be able to deduce the 3D orientation of the cardiac electrical vector by analyzing the waveshapes in the six limb leads. Cardiac defects, if any, may be localized by analyzing the waveshapes in the six chest leads.

Figure 1.19 shows the 12-lead ECG of a patient with right bundle-branch block with secondary repolarization changes. The increased QRS width and distortions in the QRS shape indicate the effects of asynchronous activation of the ventricles due to the bundle-branch block.

Signal-processing techniques to filter ECG signals will be presented in Sections 3.2, 3.3, 3.4, 3.5, and 3.8. Detection of ECG waveforms will be discussed

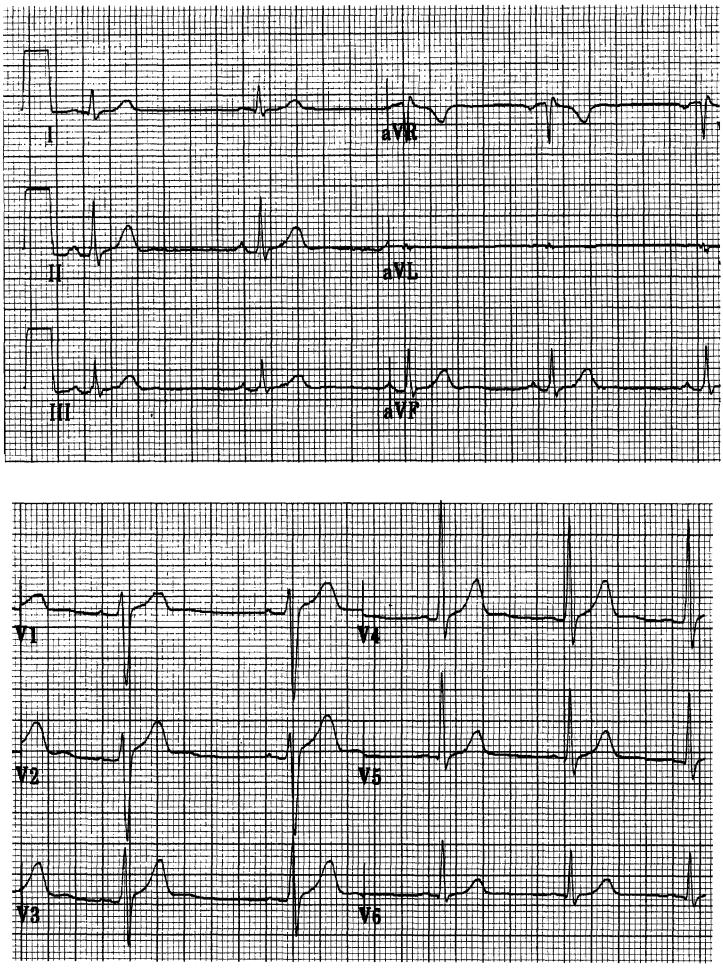


Figure 1.18 Standard 12-lead ECG of a normal male adult. Courtesy of E. Gedamu and L.B. Mitchell, Foothills Hospital, Calgary.

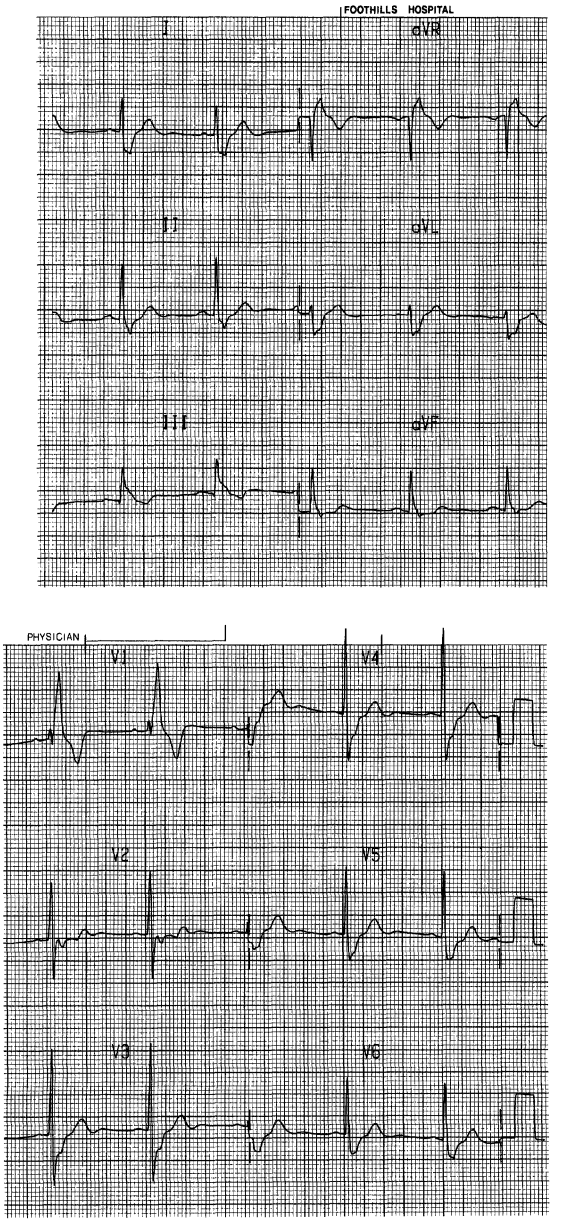


Figure 1.19 Standard 12-lead ECG of a patient with right bundle-branch block. Courtesy of L.B. Mitchell, Foothills Hospital, Calgary.

in Sections 4.2.1, 4.3.2, 4.7, and 4.9. Analysis of ECG waveform shape and classification of beats will be dealt with in Sections 5.2.1, 5.2.2, 5.2.3, 5.4, 5.7, 5.8, 9.2.1, and 9.12. Analysis of heart-rate variability will be described in Sections 7.2.2, 7.8, and 8.9. Reviews of computer applications in ECG analysis have been published by Jenkins [29, 30] and Cox et al. [31].

1.2.5 The electroencephalogram (EEG)

The EEG (popularly known as *brain waves*) represents the electrical activity of the brain [32, 33, 34]. A few important aspects of the organization of the brain are as follows: The main parts of the brain are the cerebrum, the cerebellum, the brain stem (including the midbrain, pons medulla, and the reticular formation), and the thalamus (between the midbrain and the hemispheres). The cerebrum is divided into two hemispheres, separated by a longitudinal fissure across which there is a large connective band of fibers known as the corpus callosum. The outer surface of the cerebral hemispheres, known as the cerebral cortex, is composed of neurons (grey matter) in convoluted patterns, and separated into regions by fissures (sulci). Beneath the cortex lie nerve fibers that lead to other parts of the brain and the body (white matter).

Cortical potentials are generated due to excitatory and inhibitory post-synaptic potentials developed by cell bodies and dendrites of pyramidal neurons. Physiological control processes, thought processes, and external stimuli generate signals in the corresponding parts of the brain that may be recorded at the scalp using surface electrodes. The scalp EEG is an average of the multifarious activities of many small zones of the cortical surface beneath the electrode.

In clinical practice, several channels of the EEG are recorded simultaneously from various locations on the scalp for comparative analysis of activities in different regions of the brain. The International Federation of Societies for Electroencephalography and Clinical Neurophysiology has recommended the $10-20$ system of electrode placement for clinical EEG recording [32], which is schematically illustrated in Figure 1.20. The name $10-20$ indicates the fact that the electrodes along the midline are placed at 10, 20, 20, 20, 20, and 10% of the total nasion – inion distance; the other series of electrodes are also placed at similar fractional distances of the corresponding reference distances [32]. The inter-electrode distances are equal along any antero-posterior or transverse line, and electrode positioning is symmetrical. EEG signals may be used to study the nervous system, monitoring of sleep stages, biofeedback and control, and diagnosis of diseases such as epilepsy.

Typical EEG instrumentation settings used are lowpass filtering at 75 Hz, and paper recording at 100 $\mu V/cm$ and 30 mm/s for $10-20$ minutes over $8-16$ simultaneous channels. Monitoring of sleep EEG and detection of transients related to epileptic seizures may require multichannel EEG acquisition over several hours. Special EEG techniques include the use of needle electrodes, naso-pharyngeal electrodes, recording the electrocorticogram (ECoG) from an exposed part of the cortex, and the use of intracerebral electrodes. Evocative techniques for recording the EEG include initial recording at rest (eyes open, eyes closed), hyperventilation (after breathing at

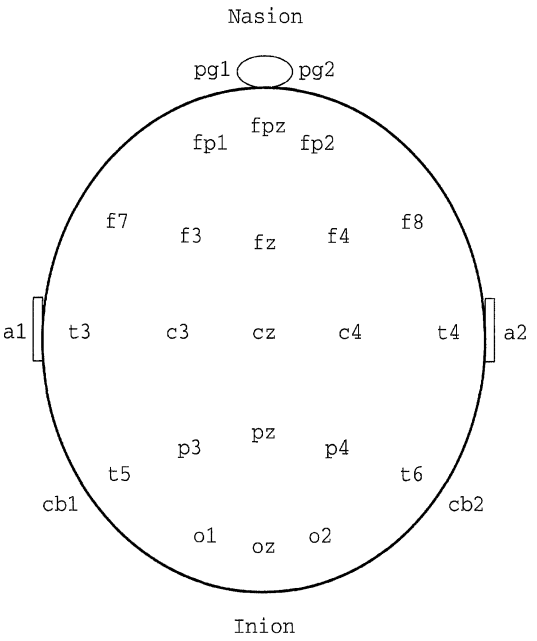


Figure 1.20 The 10 – 20 system of electrode placement for EEG recording [32]. Notes regarding channel labels: pg– naso-pharyngeal, a– auricular (ear lobes), fp– pre-frontal, f– frontal, p– pareital, c– central, o– occipital, t– temporal, cb– cerebellar, z– midline, odd numbers on the left, even numbers on the right of the subject.

20 respirations per minute for 2 – 4 minutes), photic stimulation (with 1 – 50 flashes of light per second), auditory stimulation with loud clicks, sleep (different stages), and pharmaceuticals or drugs.

EEG signals exhibit several patterns of rhythmic or periodic activity. (*Note:* The term *rhythm* stands for different phenomena or events in the ECG and the EEG.) The commonly used terms for EEG frequency (f) bands are:

- Delta (δ): $0.5 \leq f < 4\ Hz$;
- Theta (θ): $4 \leq f < 8\ Hz$;
- Alpha (α): $8 \leq f \leq 13\ Hz$; and
- Beta (β): $f > 13\ Hz$.

Figure 1.21 illustrates traces of EEG signals with the rhythms listed above.

EEG rhythms are associated with various physiological and mental processes [33, 34]. The alpha rhythm is the principal resting rhythm of the brain, and is common in wakeful, resting adults, especially in the occipital area with bilateral synchrony. Auditory and mental arithmetic tasks with the eyes closed lead to strong alpha waves, which are suppressed when the eyes are opened (that is, by a visual stimulus); see Figure 1.21(e) [32].

The alpha wave is replaced by slower rhythms at various stages of sleep. Theta waves appear at the beginning stages of sleep; delta waves appear at deep-sleep stages. High-frequency beta waves appear as background activity in tense and anxious subjects. The depression or absence of the normal (expected) rhythm in a certain state of the subject could indicate abnormality. The presence of delta or theta (slow) waves in a wakeful adult would be considered to be abnormal. Focal brain injury and tumors lead to abnormal slow waves in the corresponding regions. Unilateral depression (left – right asymmetry) of a rhythm could indicate disturbances in cortical pathways. Spikes and sharp waves could indicate the presence of epileptogenic regions in the corresponding parts of the brain.

Figure 1.22 shows an example of eight channels of the EEG recorded simultaneously from the scalp of a subject. All channels display high levels of alpha activity. Figure 1.23 shows 10 channels of the EEG of a subject with spike-and-wave complexes. Observe the distinctly different waveshape and sharpness of the spikes in Figure 1.23 as compared to the smooth waves in Figure 1.22. EEG signals also include spikes, transients, and other waves and patterns associated with various nervous disorders (see Figure 4.1 and Section 4.2.4). Detection of events and rhythms in EEG signals will be discussed in Sections 4.4, 4.5, and 4.6. Spectral analysis of EEG signals will be dealt with in Sections 6.4.3 and 7.5.2. Adaptive segmentation of EEG signals will be described in Section 8.2.2, 8.5, and 8.7.

1.2.6 Event-related potentials (ERPs)

The term *event-related potential* is more general than and preferred to the term *evoked potential*, and includes the ENG or the EEG in response to light, sound,

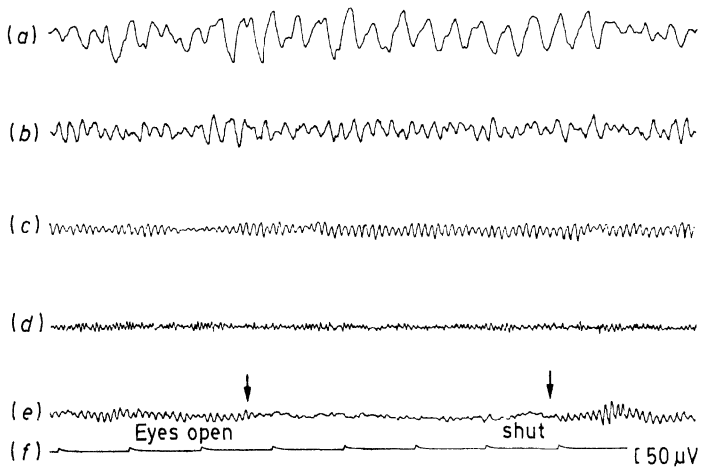


Figure 1.21 From top to bottom: (a) delta rhythm; (b) theta rhythm; (c) alpha rhythm; (d) beta rhythm; (e) blocking of the alpha rhythm by eye opening; (f) 1 s time markers and 50 μV marker. Reproduced with permission from R. Cooper, J.W. Osselton, and J.C. Shaw, *EEG Technology*, 3rd Edition, 1980. ©Butterworth Heinemann Publishers, a division of Reed Educational & Professional Publishing Ltd., Oxford, UK.

electrical, or other external stimuli. Short-latency ERPs are predominantly dependent upon the physical characteristics of the stimulus, whereas longer-latency ERPs are predominantly influenced by the conditions of presentation of the stimuli.

Somatosensory evoked potentials (SEPs) are useful for noninvasive evaluation of the nervous system from a peripheral receptor to the cerebral cortex. Median nerve short-latency SEPs are obtained by placing stimulating electrodes about $2 - 3$ cm apart over the median nerve at the wrist with electrical stimulation at $5 - 10$ pps, each stimulus pulse being of duration less than 0.5 ms with an amplitude of about 100 V (producing a visible thumb twitch). The SEPs are recorded from the surface of the scalp. The latency, duration, and amplitude of the response are measured.

ERPs and SEPs are weak signals typically buried in ongoing activity of associated systems. Examples of ERPs are provided in Figures 3.2 and 3.12. Signal-to-noise ratio (SNR) improvement is usually achieved by synchronized averaging and filtering, which will be described in Section 3.3.1.

1.2.7 The electrogastrogram (EGG)

The electrical activity of the stomach consists of rhythmic waves of depolarization and repolarization of its constituent smooth muscle cells [35, 36, 37]. The activity originates in the mid-corpus of the stomach, with intervals of about 20 s in humans. The waves of activity are always present and are not directly associated

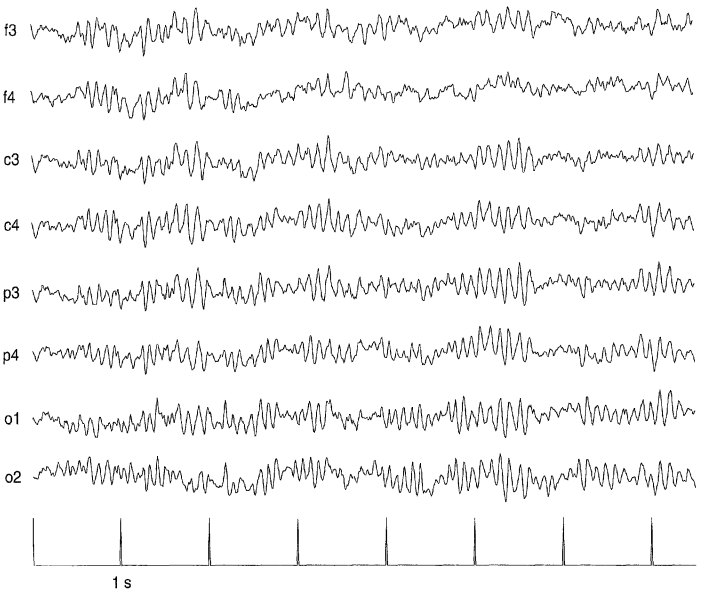


Figure 1.22 Eight channels of the EEG of a subject displaying alpha rhythm. See Figure 1.20 for details regarding channel labels. Data courtesy of Y. Mizuno-Matsumoto, Osaka University Medical School, Osaka, Japan.

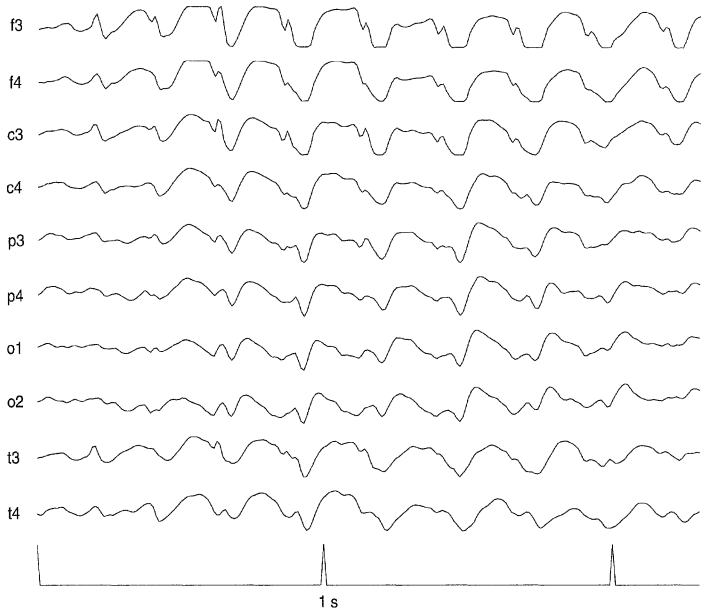


Figure 1.23 Ten channels of the EEG of a subject displaying spike-and-wave complexes. See Figure 1.20 for details regarding channel labels. Data courtesy of Y. Mizuno-Matsumoto, Osaka University Medical School, Osaka, Japan. Note that the time scale is expanded compared to that of Figure 1.22.

with contractions; they are related to the spatial and temporal organization of gastric contractions.

External (cutaneous) electrodes can record the signal known as the electrogastrogram (EGG). Chen et al. [38] used the following procedures to record cutaneous EGG signals. With the subject in the supine position and remaining motionless, the stomach was localized by using a 5 MHz ultrasound transducer array, and the orientation of the distal stomach was marked on the abdominal surface. Three active electrodes were placed on the abdomen along the antral axis of the stomach with an inter-electrode spacing of 3.5 cm. A common reference electrode was placed 6 cm away in the upper right quadrant. Three bipolar signals were obtained from the three active electrodes in relation to the common reference electrode. The signals were amplified and filtered to the bandwidth of $0.02 - 0.3$ Hz with 6 $dB/octave$ transition bands, and sampled at 2 Hz.

The surface EGG is believed to reflect the overall electrical activity of the stomach, including the electrical control activity and the electrical response activity. Chen et al. [38] indicated that gastric dysrhythmia or arrhythmia may be detected via analysis of the EGG. Other researchers suggest that the diagnostic potential of the signal has not yet been established [35, 36]. Accurate and reliable measurement of the electrical activity of the stomach requires implantation of electrodes within the stomach [39], which limits its practical applicability.

1.2.8 The phonocardiogram (PCG)

The heart sound signal is perhaps the most traditional biomedical signal, as indicated by the fact that the stethoscope is the primary instrument carried and used by physicians. The PCG is a vibration or sound signal related to the contractile activity of the cardiohemic system (the heart and blood together) [23, 40, 41, 42, 43, 44], and represents a recording of the heart sound signal. Recording of the PCG signal requires a transducer to convert the vibration or sound signal into an electronic signal: microphones, pressure transducers, or accelerometers may be placed on the chest surface for this purpose. The normal heart sounds provide an indication of the general state of the heart in terms of rhythm and contractility. Cardiovascular diseases and defects cause changes or additional sounds and murmurs that could be useful in their diagnosis.

The genesis of heart sounds: It is now commonly accepted that the externally recorded heart sounds are not caused by valve leaflet movements *per se*, as earlier believed, but by vibrations of the whole cardiovascular system triggered by pressure gradients [23]. The cardiohemic system may be compared to a fluid-filled balloon, which, when stimulated at any location, vibrates as a whole. Externally, however, heart sound components are best heard at certain locations on the chest individually, and this localization has led to the concept of *secondary sources* on the chest related to the well-known auscultatory areas: the mitral, aortic, pulmonary, and tricuspid areas [23]. The standard auscultatory areas are indicated in Figure 1.17. The mitral area is near the apex of the heart. The aortic area is to the right of the sternum, in the second right-intercostal space. The tricuspid area is in the fourth intercostal space

near the right sternal border. The pulmonary area lies at the left parasternal line in the second or third left-intercostal space [23].

A normal cardiac cycle contains two major sounds — the first heart sound (S1) and the second heart sound (S2). Figure 1.24 shows a normal PCG signal, along with the ECG and carotid pulse tracings. S1 occurs at the onset of ventricular contraction, and corresponds in timing to the QRS complex in the ECG signal.

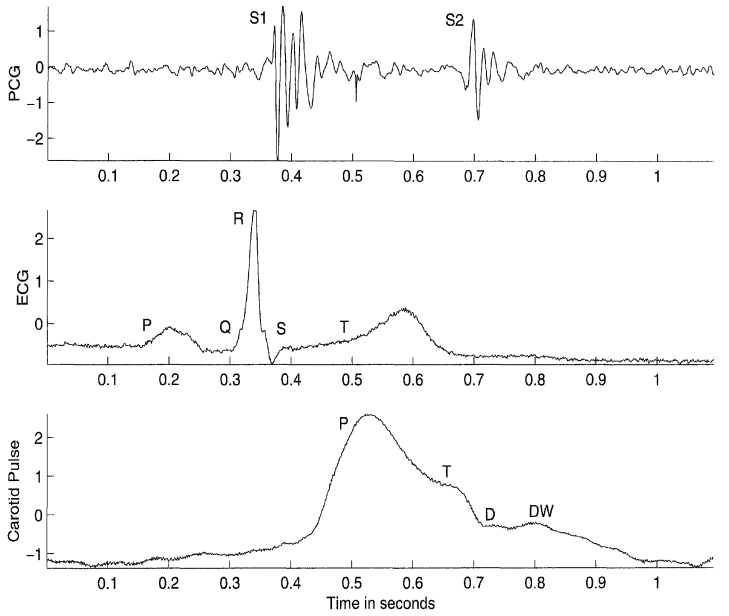


Figure 1.24 Three-channel simultaneous record of the PCG, ECG, and carotid pulse signals of a normal male adult.

The initial vibrations in S1 occur when the first myocardial contractions in the ventricles move blood toward the atria, sealing the atrio-ventricular (AV — mitral and tricuspid) valves (see Figure 1.25). The second component of S1 begins with abrupt tension of the closed AV valves, decelerating the blood. Next, the semilunar (aortic and pulmonary) valves open and the blood is ejected out of the ventricles. The third component of S1 may be caused by oscillation of blood between the root of the aorta and the ventricular walls. This is followed by the fourth component of S1, which may be due to vibrations caused by turbulence in the ejected blood flowing rapidly through the ascending aorta and the pulmonary artery.

Following the systolic pause in the PCG of a normal cardiac cycle, the second sound S2 is caused by the closure of the semilunar valves. While the primary vibrations occur in the arteries due to deceleration of blood, the ventricles and atria also vibrate, due to transmission of vibrations through the blood, valves, and the valve rings. S2 has two components, one due to closure of the aortic valve (A2)

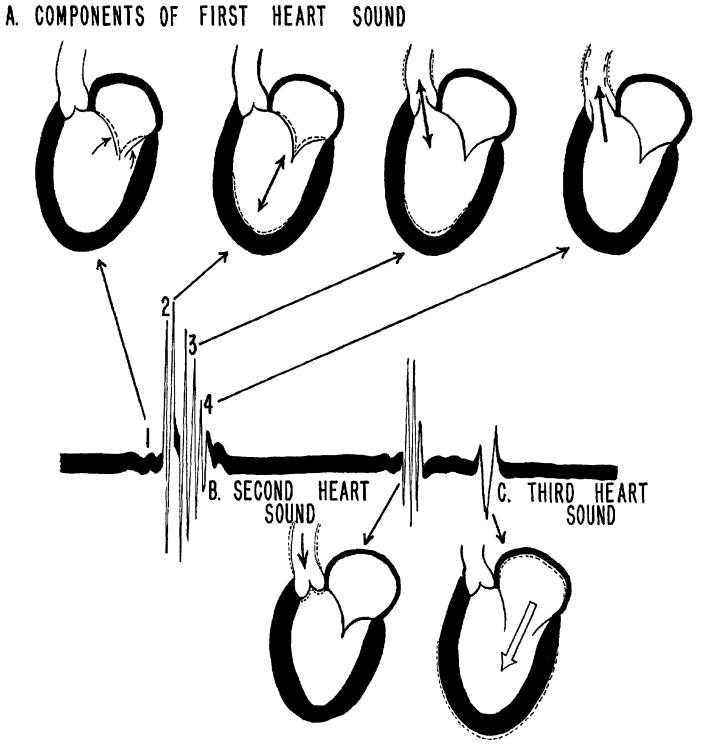


Figure 1.25 Schematic representation of the genesis of heart sounds. Only the left portion of the heart is illustrated as it is the major source of the heart sounds. The corresponding events in the right portion also contribute to the sounds. The atria do not contribute much to the heart sounds. Reproduced with permission from R.F. Rushmer, *Cardiovascular Dynamics*, 4th edition, ©W.B. Saunders, Philadelphia, PA, 1976.

and another due to closure of the pulmonary valve (P2). The aortic valve normally closes before the pulmonary valve, and hence A2 precedes P2 by a few milliseconds. Pathologic conditions could cause this gap to widen, or may also reverse the order of occurrence of A2 and P2. The A2 – P2 gap is also widened in normal subjects during inspiration. (*Note:* The PCG signal in Figure 1.24 does not show the A2 and P2 components separately.)

In some cases a third heart sound (S3) may be heard, corresponding to sudden termination of the ventricular rapid-filling phase. Because the ventricles are filled with blood and their walls are relaxed during this part of diastole, the vibrations of S3 are of very low frequency. In late diastole, a fourth heart sound (S4) may be heard sometimes, caused by atrial contractions displacing blood into the distended ventricles. In addition to these sounds, valvular clicks and snaps are occasionally heard.

Heart murmurs: The intervals between S1 and S2, and S2 and S1 of the next cycle (corresponding to ventricular systole and diastole, respectively) are normally silent. Murmurs, which are caused by certain cardiovascular defects and diseases, may occur in these intervals. Murmurs are high-frequency, noise-like sounds that arise when the velocity of blood becomes high as it flows through an irregularity (such as a constriction or a baffle). Typical conditions in the cardiovascular system that cause turbulence in blood flow are valvular stenosis and insufficiency. A valve is said to be stenosed when, due to the deposition of calcium or other reasons, the valve leaflets are stiffened and do not open completely, and thereby cause an obstruction or baffle in the path of the blood being ejected. A valve is said to be insufficient when it cannot close effectively and causes reverse leakage or regurgitation of blood through a narrow opening.

Systolic murmurs (SM) are caused by conditions such as ventricular septal defect (VSD — essentially a hole in the wall between the left ventricle and the right ventricle), aortic stenosis (AS), pulmonary stenosis (PS), mitral insufficiency (MI), and tricuspid insufficiency (TI). Semilunar valvular stenosis (aortic stenosis, pulmonary stenosis) causes an obstruction in the path of blood being ejected during systole. AV valvular insufficiency (mitral insufficiency, tricuspid insufficiency) causes regurgitation of blood to the atria during ventricular contraction.

Diastolic murmurs (DM) are caused by conditions such as aortic or pulmonary insufficiency (AI, PI), and mitral or tricuspid stenosis (MS, PS). Other conditions causing murmurs are atrial septal defect (ASD), patent ductus arteriosus (PDA), as well as certain physiological or functional conditions that cause increased cardiac output or blood velocity.

Features of heart sounds and murmurs, such as intensity, frequency content, and timing, are affected by many physical and physiological factors such as the recording site on the thorax, intervening thoracic structures, left ventricular contractility, position of the cardiac valves at the onset of systole, the degree of the defect present, the heart rate, and blood velocity. For example, S1 is loud and delayed in mitral stenosis; right bundle-branch block causes wide splitting of S2; left bundle-branch block results in reversed splitting of S2; acute myocardial infarction causes a pathologic S3; and severe mitral regurgitation (MR) leads to an increased S4 [40, 41, 42, 43, 44].

Although murmurs are noise-like events, their features aid in distinguishing between different causes. For example, aortic stenosis causes a diamond-shaped midsystolic murmur, whereas mitral stenosis causes a decrescendo – crescendo type diastolic – presystolic murmur. Figure 1.26 illustrates the PCG, ECG, and carotid pulse signals of a patient with aortic stenosis; the PCG displays the typical diamond-shaped murmur in systole.

Recording PCG signals: PCG signals are normally recorded using piezoelectric contact sensors that are sensitive to displacement or acceleration at the skin surface. The PCG signals illustrated in this section were obtained using a Hewlett Packard HP21050A transducer, which has a nominal bandwidth of $0.05 - 1,000\ Hz$. The carotid pulse signals shown in this section were recorded using the HP21281A pulse transducer, which has a nominal bandwidth of $0 - 100\ Hz$. PCG recording is normally performed in a quiet room, with the patient in the supine position with the head resting on a pillow. The PCG transducer is placed firmly on the desired position on the chest using a suction ring and/or a rubber strap.

Use of the ECG and carotid pulse signals in the analysis of PCG signals will be described in Sections 2.2.1, 2.2.2, and 2.3. Segmentation of the PCG based on events detected in the ECG and carotid pulse signals will be discussed in Section 4.10. A particular type of synchronized averaging to detect A2 in S2 will be the topic of Section 4.11. Spectral analysis of the PCG and its applications will be presented in Sections 6.2.1, 6.4.5, 6.6, and 7.10. Parametric modeling and detection of S1 and S2 will be described in Sections 7.5.2 and 7.9. Modeling of sound generation in stenosed coronary arteries will be discussed in Section 7.7.1. Adaptive segmentation of PCG signals with no other reference signal will be explored in Section 8.8.

1.2.9 The carotid pulse (CP)

The carotid pulse is a pressure signal recorded over the carotid artery as it passes near the surface of the body at the neck. It provides a pulse signal indicating the variations in arterial blood pressure and volume with each heart beat. Because of the proximity of the recording site to the heart, the carotid pulse signal closely resembles the morphology of the pressure signal at the root of the aorta; however, it cannot be used to measure absolute pressure [41]. The carotid pulse is a useful adjunct to the PCG and can assist in the identification of S2 and its components.

The carotid pulse rises abruptly with the ejection of blood from the left ventricle to the aorta, reaching a peak called the percussion wave (P, see Figure 1.24). This is followed by a plateau or a secondary wave known as the tidal wave (T), caused by a reflected pulse returning from the upper body. Next, closure of the aortic valve causes a notch known as the dicrotic notch (D). The dicrotic notch may be followed by the dicrotic wave (DW, see Figure 1.24) due to a reflected pulse from the lower body [41]. The carotid pulse trace is affected by valvular defects such as mitral insufficiency and aortic stenosis [41]; however, it is not commonly used in clinical diagnosis.

The carotid pulse signals shown in this section were recorded using the HP21281A pulse transducer, which has a nominal bandwidth of $0 - 100\ Hz$. The carotid pulse

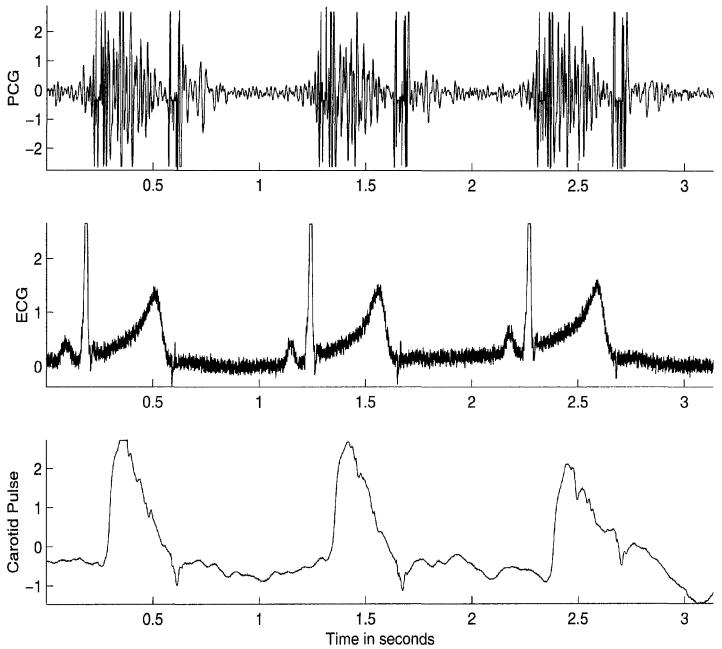


Figure 1.26 Three-channel simultaneous record of the PCG, ECG, and carotid pulse signals of a patient (female, 11 years) with aortic stenosis. Note the presence of the typical diamond-shaped systolic murmur and the split nature of S2 in the PCG.

signal is usually recorded with the PCG and ECG signals. Placement of the carotid pulse transducer requires careful selection of a location on the neck as close to the carotid artery as possible, where the pulse is felt the strongest, usually by a trained technician (see Figure 1.17).

Details on intervals that may be measured from the carotid pulse and their use in segmenting the PCG will be presented in Sections 2.2.2 and 2.3. Signal-processing techniques for the detection of the dicrotic notch will be described in Section 4.3.3. Use of the dicrotic notch for segmentation of PCG signals will be explored in Sections 4.10 and 4.11. Application of the carotid pulse to averaging of PCG spectra in systole and diastole will be proposed in Section 6.4.5.

1.2.10 Signals from catheter-tip sensors

For very specific and close monitoring of cardiac function, sensors placed on catheter tips may be inserted into the cardiac chambers. It then becomes possible to acquire several signals such as left ventricular pressure, right atrial pressure, aortic (AO) pressure, and intracardiac sounds [43, 44]. While these signals provide valuable and accurate information, the procedures are invasive and are associated with certain risks.

Figures 1.27 and 1.28 illustrate multi-channel aortic, left ventricular, and right ventricular pressure recordings from a dog using catheter-tip sensors. The ECG signal is also shown. Observe in Figure 1.27 that the right ventricular and left ventricular pressures increase exactly at the instant of each QRS complex. The aortic pressure peaks slightly after the increase in the left ventricular pressure. The notch (incisura) in the aortic pressure signal is due to closure of the aortic valve. (The same notch propagates through the vascular system and appears as the dicrotic notch in the carotid pulse signal.) The left ventricular pressure range ($10 - 110\ mm\ of\ Hg$) is much larger than the right ventricular pressure range ($5 - 25\ mm\ of\ Hg$). The aortic pressure range is limited to the vascular BP range of $80 - 120\ mm\ of\ Hg$.

The signals in Figure 1.28 display the effects of PVCs. Observe the depressed ST segment in the ECG signal in the figure, likely due to myocardial ischemia. (It should be noted that the PQ and ST segments of the ECG signal in Figure 1.27 are iso-electric, even though the displayed values indicate a non-zero level. On the other hand, in the ECG in Figure 1.28, the ST segment stays below the corresponding iso-electric PQ segment.) The ECG complexes appearing just after the $2\ s$ and $3\ s$ markers are PVCs arising from different ectopic foci, as evidenced by their markedly different waveforms. Although the PVCs cause a less-than-normal increase in the left ventricular pressure, they do not cause a rise in the aortic pressure, as no blood is effectively pumped out of the left ventricle during the ectopic beats.

1.2.11 The speech signal

Human beings are social creatures by nature, and have an innate need to communicate. We are endowed with the most sophisticated vocal system in nature. The speech signal

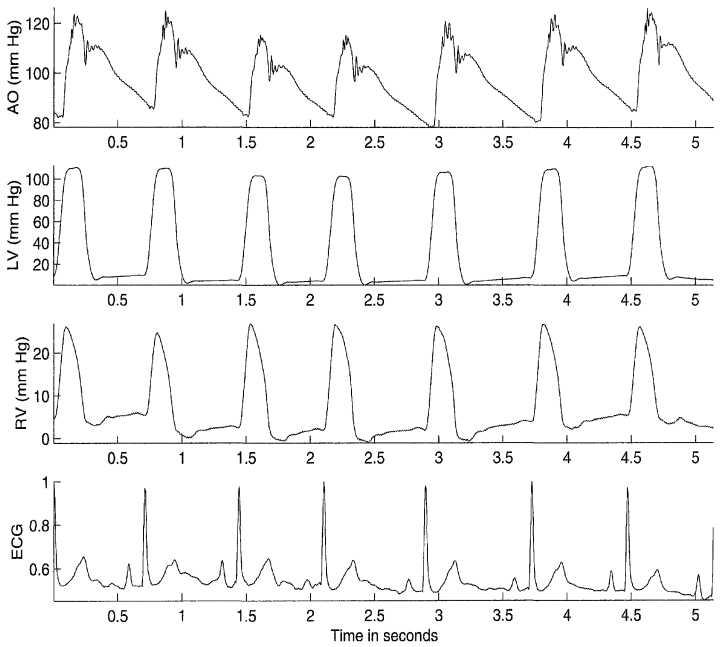


Figure 1.27 Normal ECG and intracardiac pressure signals from a dog. AO represents aortic pressure near the aortic valve. Data courtesy of R. Sas and J. Tyberg, Department of Physiology and Biophysics, University of Calgary.

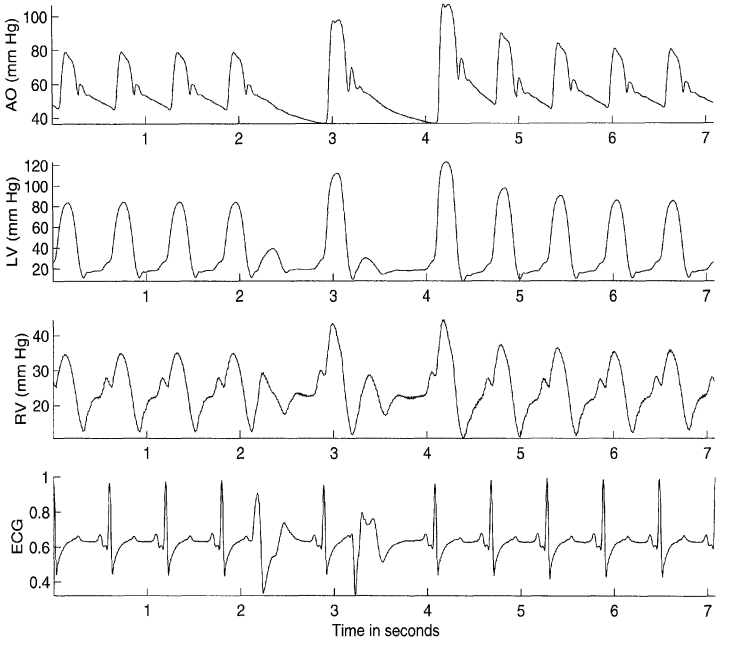


Figure 1.28 ECG and intracardiac pressure signals from a dog with PVCs. Data courtesy of R. Sas and J. Tyberg, Department of Physiology and Biophysics, University of Calgary.

is an important signal, although it is more commonly considered as a communication signal than a biomedical signal. However, the speech signal can serve as a diagnostic signal when speech and vocal-tract disorders need to be investigated [45].

Speech sounds are produced by transmitting puffs of air from the lungs through the vocal tract (as well as the nasal tract for certain sounds) [46]. The vocal tract starts at the vocal cords or glottis in the throat and ends at the lips and the nostrils. The shape of the vocal tract is varied to produce different types of sound units or *phonemes* which, when concatenated, form speech. In essence, the vocal tract acts as a filter that modulates the spectral characteristics of the input puffs of air. It is evident that the system is dynamic, and that the filter, and therefore the speech signal produced, have time-varying characteristics, that is, they are nonstationary (see Section 3.1.2).

Speech sounds may be classified mainly as voiced, unvoiced, and plosive sounds [46]. Voiced sounds involve the participation of the glottis: air is forced through the vocal cords held at a certain tension. The result is a series of quasi-periodic pulses of air which is passed through the vocal tract. The input to the vocal tract may be treated as an impulse train that is almost periodic. Upon convolution with the impulse response of the vocal tract, which is held steady at a certain configuration for the duration of the voiced sound desired, a quasi-periodic signal is produced with a characteristic waveshape that is repeated. All vowels are voiced sounds. Figure 1.29 shows the speech signal of the word "safety" spoken by a male. Figure 1.30 shows, in the upper trace, a portion of the signal corresponding to the /E/ sound (the letter "a" in the word). The quasi-periodic nature of the signal is evident. Features of interest in voiced signals are the pitch (average interval between the repetitions of the vocal-tract impulse response or basic wavelet) and the resonance or formant frequencies of the vocal-tract system.

An unvoiced sound (or fricative) is produced by forcing a steady stream of air through a narrow opening or constriction formed at a specific position along the vocal tract. The result is a turbulent signal that appears almost like random noise. In fact, the input to the vocal tract is a broadband random signal, which is filtered by the vocal tract to yield the desired sound. Fricatives are unvoiced sounds, as they do not involve any activity (vibration) of the vocal cords. The phonemes /S/, /SH/, /Z/, and /F/ are examples of fricatives. The lower trace in Figure 1.30 shows a portion of the signal corresponding to the /S/ sound in the word "safety". The signal has no identifiable structure, and appears to be random (see also Figures 3.1, 3.3, and 3.4, as well as Section 3.1.2). The transfer function of the vocal tract, as evidenced by the spectrum of the signal itself, would be of interest in analyzing a fricative.

Plosives, also known as stops, involve complete closure of the vocal tract, followed by an abrupt release of built-up pressure. The phonemes /P/, /T/, /K/, and /D/ are examples of plosives. The sudden burst of activity at about 1.1 s in Figure 1.29 illustrates the plosive nature of /T/. Plosives are hard to characterize as they are transients; their properties are affected by the preceding phoneme as well. For more details on the speech signal, see Rabiner and Schafer [46].

Signal-processing techniques for extraction of the vocal-tract response from voiced speech signals will be described in Section 4.8.3. Frequency-domain characteristics of speech signals will be illustrated in Section 7.6.3 and 8.4.1.

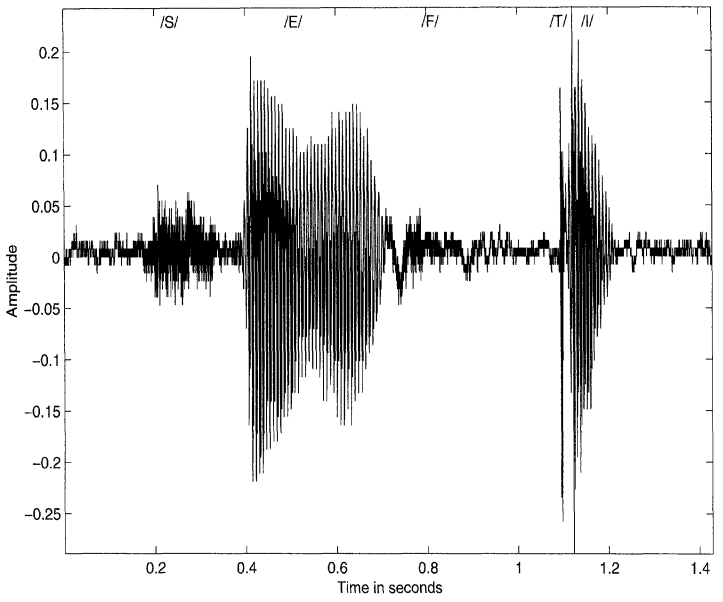


Figure 1.29 Speech signal of the word "safety" uttered by a male speaker. Approximate time intervals of the various phonemes in the word are /S/: $0.2 - 0.35$ *s*; /E/: $0.4 - 0.7$ *s*; /F/: $0.75 - 0.95$ *s*; /T/: transient at 1.1 *s*; /I/: $1.1 - 1.2$ *s*. Background noise is also seen in the signal before the beginning and after the termination of the speech, as well as during the stop interval before the plosive /T/.

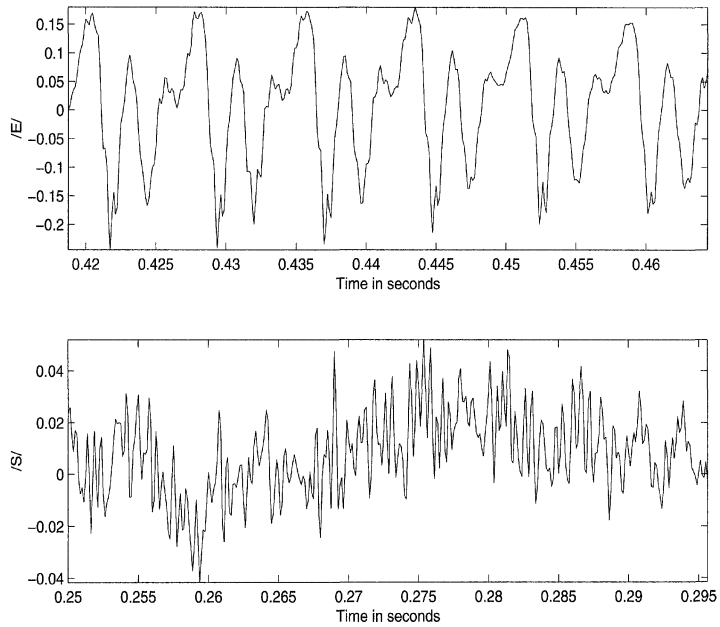


Figure 1.30 Segments of the signal in Figure 1.29 on an expanded scale to illustrate the quasi-periodic nature of the voiced sound /E/ in the upper trace, and the almost-random nature of the fricative /S/ in the lower trace.

1.2.12 The vibromyogram (VMG)

The VMG is the direct mechanical manifestation of contraction of a skeletal muscle, and is a vibration signal that accompanies the EMG. The signal has also been named as the sound-, acoustic-, or phono-myogram. Muscle sounds or vibrations are related to the change in dimensions (contraction) of the constituent muscle fibers (see Figure 1.4), and may be recorded using contact microphones or accelerometers (such as the Dytran 3115A accelerometer, Dytran, Chatsworth, CA) placed on the muscle surface [47, 48]. The frequency and intensity of the VMG have been shown to vary in direct proportion to the contraction level. The VMG, along with the EMG, may be useful in studies related to neuromuscular control, muscle contraction, athletic training, and biofeedback. VMG signal analysis, however, is not as well established or popular as EMG analysis.

Simultaneous analysis of the VMG and EMG signals will be discussed in Section 2.2.5. Adaptive cancellation of the VMG from knee-joint vibration signals will be the topic of Sections 3.6.2, 3.6.3, and 3.10. Analysis of muscle contraction using the VMG will be described in Section 5.10.

1.2.13 The vibroarthrogram (VAG)

The knee joint: As illustrated in Figure 1.31, the knee joint is formed between the femur, the patella, and the tibia. The knee joint is the largest articulation in the human body that can effectively move from 0^o extension to 135^o flexion, together with 20^o to 30^o rotation of the flexed leg on the femoral condyles. The joint has four important features: (1) a joint cavity, (2) articular cartilage, (3) a synovial membrane, and (4) a fibrous capsule [49, 50]. The knee joint is known as a synovial joint, as it contains a lubricating substance called the synovial fluid. The patella (knee cap), a sesamoid bone, protects the joint, and is precisely aligned to slide in the groove (trochlea) of the femur during leg movement. The knee joint is made up of three compartments: (1) the patello-femoral, (2) the lateral tibio-femoral, and (3) the medial tibio-femoral compartments. The patello-femoral compartment is classified as a synovial gliding joint and the tibio-femoral as a synovial hinge joint [51]. The anterior and posterior cruciate ligaments as well as the lateral and medial ligaments bind the femur and tibia together, give support to the knee joint, and limit movement of the joint. The various muscles around the joint help in the movement of the joint and contribute to its stability.

The knee derives its physiological movement and its typical rolling – gliding mechanism of flexion and extension from its six degrees of freedom: three in translation and three in rotation. The translations of the knee take place on the anterior – posterior, medial – lateral, and proximal – distal axes. The rotational motion consists of flexion – extension, internal – external rotation, and abduction – adduction.

Although the tibial plateaus are the main load-bearing structures in the knee, the cartilage, menisci, and ligaments also bear loads. The patella aids knee extension by lengthening the lever arm of the quadriceps muscle throughout the entire range of motion, and allows a better distribution of compressive stresses on the femur [52].

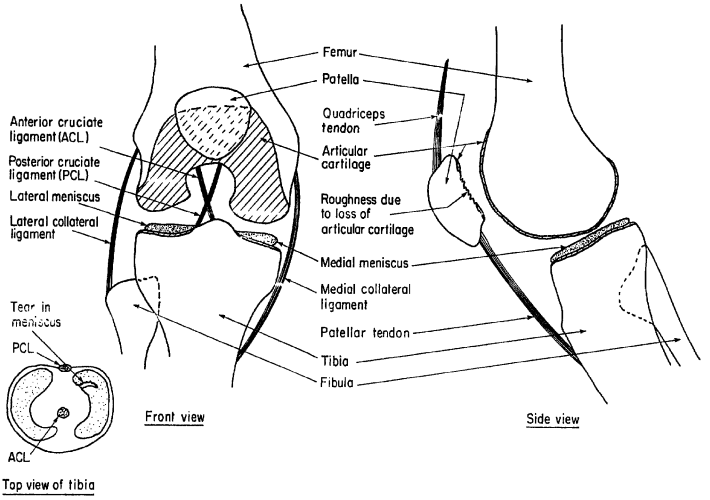


Figure 1.31 Front and side views of the knee joint (the two views are not mutually orthogonal). The inset shows the top view of the tibia with the menisci.

Articular cartilage: Two types of cartilage are present in the knee joint: the *articular cartilage*, which covers the ends of bones, and the wedge-shaped fibrocartilaginous structure called the *menisci*, located between the femur and the tibia [53]. The shock-absorbing menisci are composed of the medial meniscus and the lateral meniscus, which are two crescent-shaped plates of fibrocartilage that lie on the articular surface of the tibia.

The articular surfaces of the knee joint are the large curved condyles of the femur, the flattened condyles (medial and lateral plateaus) of the tibia, and the facets of the patella. There are three types of articulation: an intermediate articulation between the patella and the femur, and lateral and medial articulation between the femur and the tibia. The articular surfaces are covered by cartilage, like all the major joints of the body. Cartilage is vital to joint function because it protects the underlying bone during movement. Loss of cartilage function leads to pain, decreased mobility, and in some instances, deformity and instability.

Knee-joint disorders: The knee is the most commonly injured joint in the body. Arthritic degeneration of injured knees is a well-known phenomenon, and is known to result from a variety of traumatic causes. Damage to the stabilizing ligaments of the knee, or to the shock-absorbing fibrocartilage pads (the menisci) are two of the most common causes of deterioration of knee-joint surfaces. Impact trauma to the articular cartilage surfaces themselves could lead to surface deterioration and secondary osteoarthritis.

Non-traumatic conditions of the knee joint include the extremely common idiopathic condition known as chondromalacia patella (soft cartilage of the patella),

in which articular cartilage softens, fibrillates, and sheds off the undersurface of the patella. Similarly, the meniscal fibrocartilage of the knee can apparently soften, which could possibly lead to degenerative tears and secondary changes in the regional hyaline surfaces.

Knee-joint sounds: Considerable noise is often associated with degeneration of knee-joint surfaces. The VAG is the vibration signal recorded from a joint during movement (articulation) of the joint. Normal joint surfaces are smooth and produce little or no sound, whereas joints affected by osteoarthritis and other degenerative diseases may have suffered cartilage loss and produce grinding sounds. Detection of knee-joint problems via the analysis of VAG signals could help avoid unnecessary exploratory surgery, and also aid better selection of patients who would benefit from surgery [54, 55, 56, 57, 58, 59, 60]. The VAG signal, however, is not yet well understood, and is a difficult signal to analyze due to its complex nonstationary characteristics.

Further details on the VAG signal will be provided in Sections 2.2.6, 3.2.6, and 8.2.3. Modeling of a specific type of VAG signal known as patello-femoral crepitus will be presented in Sections 7.2.4, 7.3, and 7.7.2. Adaptive filtering of the VAG signal to remove muscle-contraction interference will be described in Sections 3.6.2, 3.6.3, and 3.10. Adaptive segmentation of VAG signals into quasi-stationary segments will be illustrated in Sections 8.6.1 and 8.6.2. The role of VAG signal analysis in the detection of articular cartilage diseases will be discussed in Section 9.13.

1.2.14 Oto-acoustic emission signals

The oto-acoustic emission (OAE) signal represents the acoustic energy emitted by the cochlea either spontaneously or in response to an acoustic stimulus. The discovery of the existence of this signal indicates that the cochlea not only receives sound but also produces acoustic energy [61]. The OAE signal could provide objective information on the micromechanical activity of the preneural or sensory components of the cochlea that are distal to the nerve-fiber endings. Analysis of the OAE signal could lead to improved noninvasive investigative techniques to study the auditory system. The signal may also assist in screening of hearing function and in the diagnosis of hearing impairment.

1.3 OBJECTIVES OF BIOMEDICAL SIGNAL ANALYSIS

The representation of biomedical signals in electronic form facilitates computer processing and analysis of the data. Figure 1.32 illustrates the typical steps and processes involved in computer-aided diagnosis and therapy based upon biomedical signal analysis.

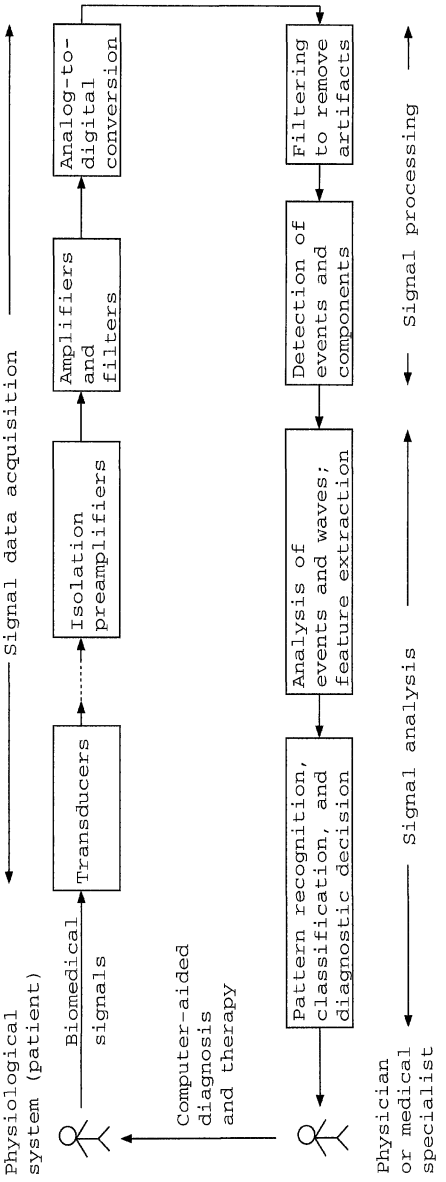


Figure 1.32 Computer-aided diagnosis and therapy based upon biomedical signal analysis.

The major objectives of biomedical instrumentation and signal analysis [17, 13, 10, 11, 12] are:

- *Information gathering* — measurement of phenomena to interpret a system.
- *Diagnosis* — detection of malfunction, pathology, or abnormality.
- *Monitoring* — obtaining continuous or periodic information about a system.
- *Therapy and control* — modification of the behavior of a system based upon the outcome of the activities listed above to ensure a specific result.
- *Evaluation* — objective analysis to determine the ability to meet functional requirements, obtain proof of performance, perform quality control, or quantify the effect of treatment.

Signal acquisition procedures may be categorized as being invasive or noninvasive, and active or passive.

Invasive versus noninvasive procedures: Invasive procedures involve the placement of transducers or other devices inside the body, such as needle electrodes to record MUAPs, or insertion of catheter-tip sensors into the heart via a major artery or vein to record intracardiac signals. Noninvasive procedures are desirable in order to minimize risk to the subject. Recording of the ECG using limb or chest electrodes, the EMG with surface electrodes, or the PCG with microphones or accelerometers placed on the chest are noninvasive procedures.

Note that making measurements or imaging with x-rays, ultrasound, and so on, may be classified as invasive procedures, as they involve penetration of the body with externally administered radiation, even though the radiation is invisible and there is no visible puncturing or invasion of the body.

Active versus passive procedures: Active data acquisition procedures require external stimuli to be applied to the subject, or require the subject to perform a certain activity to stimulate the system of interest in order to elicit the desired response or signal. For example, recording an EMG signal requires contraction of the muscle of interest, say the clenching of a fist; recording the VAG signal from the knee requires flexing of the leg over a certain joint angle range; recording visual ERP signals requires the delivery of flashes of light to the subject. While these stimuli may appear to be innocuous, they do carry risks in certain situations for some subjects: flexing the knee beyond a certain angle may cause pain for some subjects; strobe lights may trigger epileptic seizures in some subjects. The investigator should be aware of such risks, factor them in a *risk – benefit analysis*, and be prepared to manage adverse reactions.

Passive procedures do not require the subject to perform any activity. Recording of the ECG using limb or chest electrodes, the EEG during sleep using scalp-surface electrodes, or the PCG with microphones or accelerometers placed on the chest are passive procedures, but require contact between the subject and the instruments. Note that although the procedure is passive, the system of interest is active under its own natural control in these procedures. Acquiring an image of a subject with reflected

natural light (with no flash from the camera) or with the natural infra-red (thermal) emission could be categorized as a passive and non-contact procedure.

Most organizations require ethical approval by specialized committees for experimental procedures involving human or animal subjects, with the aim of minimizing the risk and discomfort to the subject and maximizing the benefits to both the subjects and the investigator.

The human – instrument system: The components of a *human – instrument system* [17, 13, 10, 11, 12] are:

- *The subject or patient:* It is important always to bear in mind that the main purpose of biomedical instrumentation and signal analysis is to provide a certain benefit to the subject or patient. All systems and procedures should be designed so as not to unduly inconvenience the subject, and not to cause any harm or danger. In applying invasive or risky procedures, it is extremely important to perform a risk – benefit analysis and determine if the anticipated benefits of the procedure are worth placing the subject at the risks involved.

- *Stimulus or procedure of activity:* Application of stimuli to the subject in active procedures requires instruments such as strobe light generators, sound generators, and electrical pulse generators. Passive procedures require a standardized protocol of the desired activity to ensure repeatability and consistency of the experiment.

- *Transducers:* electrodes, sensors.

- *Signal-conditioning equipment:* amplifiers, filters.

- *Display equipment:* oscilloscopes, strip-chart or paper recorders, computer monitors, printers.

- *Recording, data processing, and transmission equipment:* analog instrumentation tape recorders, analog-to-digital converters (ADCs), digital-to-analog converters (DACs), digital tapes, compact disks (CDs), diskettes, computers, telemetry systems.

- *Control devices:* power supply stabilizers and isolation equipment, patient intervention systems.

The science of measurement of physiological variables and parameters is known as *biometrics*. Some of the aspects to be considered in the design, specification, or use of biomedical instruments [17, 13, 10, 11, 12] are:

- *Isolation of the subject or patient* — of paramount importance so that the subject is not placed at the risk of electrocution.

- *Range of operation* — the minimum to maximum values of the signal or parameter being measured.

- *Sensitivity* — the smallest signal variation measurable. This determines the resolution of the system.

- *Linearity* — desired over at least a portion of the range of operation. Any nonlinearity present may need to be corrected for at later stages of signal processing.

- *Hysteresis* — a lag in measurement due to the direction of variation of the entity being measured. Hysteresis may add a bias to the measurement, and should be corrected for.

- *Frequency response* — represents the variation of sensitivity with frequency. Most systems encountered in practice exhibit a lowpass behavior, that is, the sensitivity of the system decreases as the frequency of the input signal increases. Signal restoration techniques may be required to compensate reduced high-frequency sensitivity.

- *Stability* — an unstable system could preclude repeatability and consistency of measurements.

- *Signal-to-noise ratio (SNR)* — power-line interference, grounding problems, thermal noise, and so on, could compromise the quality of the signal being acquired. A good understanding of the signal-degrading phenomena present in the system is necessary in order to design appropriate filtering and correction procedures.

- *Accuracy* — includes the effects of errors due to component tolerance, movement, or mechanical errors; drift due to changes in temperature, humidity, or pressure; reading errors due to, for example, parallax; and zeroing or calibration errors.

1.4 DIFFICULTIES ENCOUNTERED IN BIOMEDICAL SIGNAL ACQUISITION AND ANALYSIS

In spite of the long history of biomedical instrumentation and its extensive use in health care and research, many practical difficulties are encountered in biomedical signal acquisition, processing, and analysis [17, 13, 10, 11, 12]. The characteristics of the problems, and hence their potential solutions, are unique to each type of signal. Particular attention should be paid to the following issues.

Accessibility of the variables to measurement: Most of the systems and organs of interest, such as the cardiovascular system and the brain, are located well within the body (for good reasons!). While the ECG may be recorded using limb electrodes, the signal so acquired is but a projection of the true 3D cardiac electrical vector of the heart onto the axis of the electrodes. Such a signal may be sufficient for rhythm monitoring, but could be inadequate for more specific analysis of the cardiac system

such as atrial electrical activity. Accessing the atrial electrical activity at the source requires insertion of an electrode close to the atrial surface or within the atria.

Similarly, measurement of blood pressure using a pressure cuff over an arm gives an estimate of the brachial arterial pressure. Detailed study of pressure variations within the cardiac chambers or arteries over a cardiac cycle would require insertion of catheters with pressure sensors into the heart. Such invasive procedures provide access to the desired signals at their sources and often provide clear and useful signals, but carry high risks.

The surface EMG includes the interference pattern of the activities of several motor units even at very low levels of muscular contraction. Acquisition of SMUAPs requires access to the specific muscle layer or unit of interest by insertion of fine-wire or needle electrodes. The procedure carries risks of infection and damage to muscle fibers, and causes pain to the subject during muscular activity.

An investigator should assess the system and variables of interest carefully and determine the minimal level of intervention absolutely essential to the data acquisition procedure. The trade-off to be performed is that of integrity and quality of the information acquired versus the pain and risks to the subject.

Variability of the signal source: It is evident from the preceding sections that the various systems that comprise the human body are dynamic systems with several variables. Biomedical signals represent the dynamic activity of physiological systems and the states of their constituent variables. The nature of the processes or the variables could be deterministic or random (stochastic); a special case is that of periodicity or quasi-periodicity.

A normal ECG exhibits a regular rhythm with a readily identifiable waveshape (the QRS complex) in each period, and under such conditions the signal may be referred to as a deterministic and periodic signal. However, the cardiovascular system of a heart patient may not stay in a given state over significant periods and the waveshape and rhythm may vary over time.

The surface EMG is the summation of the MUAPs of the motor units that are active at the given instant of time. Depending upon the level of contraction desired (at the volition of the subject), the number of active motor units varies, increasing with increasing effort. Furthermore, the firing intervals or the firing rate of each motor unit also vary in response to the level of contraction desired, and exhibit stochastic properties. While the individual MUAPs possess readily identifiable and simple monophasic, biphasic, or triphasic waveshapes, the interference pattern of several motor units firing at different rates will appear as an almost random signal with no visually recognizable waves or waveshapes.

The dynamic nature of biological systems causes most signals to exhibit stochastic and nonstationary behavior. This means that signal statistics such as mean, variance, and spectral density change with time. For this reason, signals from a dynamic system should be analyzed over extended periods of time including various possible states of the system, and the results should be placed in the context of the corresponding states.

Inter-relationships and interactions among physiological systems: The various systems that compose the human body are not mutually independent; rather, they are

inter-related and interact in various ways. Some of the interactive phenomena are compensation, feedback, cause-and-effect, collateral effects, loading, and take-over of function of a disabled system or part by another system or part. For example, the second heart sound exhibits a split during active inspiration in normal subjects due to reduced intra-thoracic pressure and decreased venous return to the left side of the heart [41] (but not during expiration); this is due to normal physiological processes. However, the second heart sound is split in both inspiration and expiration due to delayed right ventricular contraction in right bundle-branch block, pulmonary valvular stenosis or insufficiency, and other conditions [41]. Ignoring this inter-relationship could lead to misinterpretation of the signal.

Effect of the instrumentation or procedure on the system: The placement of transducers on and connecting a system to instruments could affect the performance or alter the behavior of the system, and cause spurious variations in the parameters being investigated. The experimental procedure or activity required to elicit the signal may lead to certain effects that could alter signal characteristics. This aspect may not always be obvious unless careful attention is paid. For example, the placement of a relatively heavy accelerometer may affect the vibration characteristics of a muscle and compromise the integrity of the vibration or sound signal being measured. Fatigue may set in after a few repetitions of an experimental procedure, and subsequent measurements may not be indicative of the true behavior of the system; the system may need some rest between procedures or their repetitions.

Physiological artifacts and interference: One of the pre-requisites for obtaining a good ECG signal is for the subject to remain relaxed and still with no movement. Coughing, tensing of muscles, and movement of the limbs cause the corresponding EMG to appear as an undesired artifact. In the absence of any movement by the subject, the only muscular activity in the body would be that of the heart. When chest leads are used, even normal breathing could cause the associated EMG of the chest muscles to interfere with the desired ECG. It should also be noted that breathing causes beat-to-beat variations in the RR interval, which should not be mistaken to be sinus arrhythmia. An effective solution would be to record the signal with the subject holding breath for a few seconds. This simple solution does not apply in long-term monitoring of critically ill patients or in recording the ECG of infants; signal-processing procedures would then be required to remove the artifacts.

A unique situation is that of acquiring the ECG of a fetus through surface electrodes placed over the mother's abdomen: the maternal ECG appears as an interference in this situation. No volitional or external control is possible or desirable to prevent the artifact in this situation, which calls for more intelligent adaptive cancellation techniques using multiple channels of various signals [62].

Another example of physiological interference or cross-talk is that of muscle-contraction interference (MCI) in the recording of the knee-joint VAG signal [63]. The rectus femoris muscle is active (contracting) during the swinging movement of the leg required to elicit the joint vibration signal. The VMG of the muscle is propagated to the knee and appears as an interference. Swinging the leg mechanically using a mechanical actuator is a possible solution; however, this represents an unnatural situation, and may cause other sound or vibration artifacts from the machine. Adap-

tive filtering using multi-channel vibration signals from various points is a feasible solution [63].

Energy limitations: Most biomedical signals are generated at microvolt or millivolt levels at their sources. Recording such signals requires very sensitive transducers and instrumentation with low noise levels. The connectors and cables need to be shielded as well, in order to obviate pickup of ambient electromagnetic (EM) signals. Some applications may require transducers with integrated amplifiers and signal conditioners so that the signal leaving the subject at the transducer level is much stronger than ambient sources of potential interference.

When external stimuli are required to elicit a certain response from a system, the level of the stimulus is constrained due to safety factors and physiological limitations. Electrical stimuli to record the ENG need to be limited in voltage level so as to not cause local burns or interfere with the electrical control signals of the cardiac or nervous systems. Auditory and visual stimuli are constrained by the lower thresholds of detectability and upper thresholds related to frequency response, saturation, or pain.

Patient safety: Protection of the subject or patient from electrical shock or radiation hazards is an unquestionable requirement of paramount importance. The relative levels of any other risks involved should be assessed when a choice is available between various procedures, and analyzed against their relative benefits. Patient safety concerns may preclude the use of a procedure that may yield better signals or results than others, or require modifications to a procedure that may lead to inferior signals. Further signal-processing steps would then become essential in order to improve signal quality or otherwise compensate for the initial loss.

1.5 COMPUTER-AIDED DIAGNOSIS

Physicians, cardiologists, neuroscientists, and health-care technologists are highly trained and skilled practitioners. Why then would we want to use computers or electronic instrumentation for the analysis of biomedical signals? The following points provide some arguments in favor of the application of computers to process and analyze biomedical signals.

- Humans are highly skilled and fast in the analysis of visual patterns and waveforms, but are slow in arithmetic operations with large numbers of values. The ECG of a single cardiac cycle (heart beat) could have up to 200 numerical values; the corresponding PCG up to 2,000. If signals need to be processed to remove noise or extract a parameter, it would not be practical for a person to perform such computation. Computers can perform millions of arithmetic operations per second. It should be noted, however, that recognition of waveforms and images using mathematical procedures typically requires huge numbers of operations that could lead to slow responses in such tasks from low-level computers.

- Humans could be affected by fatigue, boredom, and environmental factors, and are susceptible to committing errors. Long-term monitoring of signals, for example, the heart rate and ECG of a critically ill patient, by a human observer watching an oscilloscope or computer tracing is neither economical nor feasible. A human observer could be distracted by other events in the surrounding areas and may miss short episodes or transients in the signal. Computers, being inanimate but mathematically accurate and consistent machines, can be designed to perform computationally specific and repetitive tasks.

- Analysis by humans is usually subjective and qualitative. When comparative analysis is required between the signal of a subject and another or a standard pattern, a human observer would typically provide a qualitative response. For example, if the QRS width of the ECG is of interest, a human observer may remark that the QRS of the subject is wider than the reference or normal. More specific or objective comparison to the accuracy of the order of a few milliseconds would require the use of electronic instrumentation or a computer. Derivation of quantitative or numerical features from signals with large numbers of samples would certainly demand the use of computers.

- Analysis by humans is subject to inter-observer as well as intra-observer variations (with time). Given that most analyses performed by humans are based upon qualitative judgment, they are liable to vary with time for a given observer, or from one observer to another. The former could also be due to lack of diligence or due to inconsistent application of knowledge, and the latter due to variations in training and level of understanding. Computers can apply a given procedure repeatedly and whenever recalled in a consistent manner. It is further possible to encode the knowledge (to be more specific, the logic) of many experts into a single computational procedure, and thereby enable a computer with the collective intelligence of several human experts in the area of interest.

- Most biomedical signals are fairly slow (lowpass) signals, with their bandwidth limited to a few tens to a few thousand Hertz. Typical sampling rates for digital processing of biomedical signals therefore range from 100 Hz to $10-20$ kHz. Sampling rates as above facilitate *on-line, real-time* analysis of biomedical signals with even low-end computers. Note that the term "real-time analysis" may be used to indicate the processing of each sample of the signal before the next sample arrives, or the processing of an epoch or episode such as an ECG beat before the next one is received in its entirety in a buffer. Heart-rate monitoring of critically ill patients would certainly demand real-time ECG analysis. However, some applications do not require on-line, real-time analysis: for example, processing a VAG signal to diagnose cartilage degeneration, and analysis of a long-term ECG record obtained over several hours using an ambulatory system do not demand immediate attention and results. In such cases, computers could be used for *off-line* analysis of pre-recorded signals with sophisticated signal-processing and time-consuming modeling

techniques. The speed required for real-time processing and the computational complexities of modeling techniques in the case of off-line applications both would rule out the possibility of performance of the tasks by humans.

One of the important points to note in the above discussion is that *quantitative analysis* becomes possible by the application of computers to biomedical signals. The logic of medical or clinical diagnosis via signal analysis could then be *objectively* encoded and *consistently* applied in routine or repetitive tasks. However, it should be emphasized at this stage that the end-goal of biomedical signal analysis should be seen as computer-*aided* diagnosis and not automated diagnosis. A physician or medical specialist typically uses a significant amount of information in addition to signals and measurements, including the general physical appearance and mental state of the patient, family history, and socio-economic factors affecting the patient, many of which are not amenable to quantification and logistic rule-based processes. Biomedical signals are, at best, indirect indicators of the state of the patient; most cases lack a direct or unique signal – pathology relationship [31]. The results of signal analysis need to be integrated with other clinical signs, symptoms, and information by a physician. Above all, the *intuition* of the specialist plays an important role in arriving at the final diagnosis. For these reasons, and keeping in mind the realms of practice of various licensed and regulated professions, liability, and legal factors, the final diagnostic decision is best left to the physician or medical specialist. It is expected that quantitative and objective analysis facilitated by the application of computers to biomedical signal analysis will lead to a more accurate diagnostic decision by the physician.

On the importance of quantitative analysis:

"When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of *science*."

— Lord Kelvin (William Thomson, 1824 – 1907) [64]

On assumptions made in quantitative analysis:

"Things do not in general run around with their measure stamped on them like the capacity of a freight car; it requires a certain amount of investigation to discover what their measures are ... What most experimenters take for granted before they begin their experiments is infinitely more interesting than any results to which their experiments lead."

— Norbert Wiener (1894 – 1964)

1.6 REMARKS

We have taken a general look at the nature of biomedical signals in this chapter, and seen a few signals illustrated for the purpose of gaining familiarity with their typical

appearance and features. Specific details of the characteristics of the signals and their processing or analysis will be dealt with in subsequent chapters.

We have also stated the objectives of biomedical instrumentation and signal analysis. Some practical difficulties that arise in biomedical signal investigation were discussed in order to draw attention to the relevant practical issues. The suitability and desirability of the application of computers for biomedical signal analysis were discussed, with emphasis on objective and quantitative analysis toward the end-goal of computer-aided diagnosis. The remaining chapters will deal with specific techniques and applications.

1.7 STUDY QUESTIONS AND PROBLEMS

(*Note:* Some of the questions may require background preparation with other sources on the ECG (for example, Rushmer [23]), the EMG (for example, Goodgold and Eberstein [22]), and biomedical instrumentation (for example, Webster [10].)

1. Give two reasons to justify the use of electronic instruments and computers in medicine.
2. State any two objectives of using biomedical instrumentation and signal analysis.
3. Distinguish between open-loop and closed-loop monitoring of a patient.
4. List three common types or sources of artifact in a biomedical instrument.
5. A nerve cell has an action potential of duration 10 ms including the refractory period. What is the maximum rate (in pulses per second) at which this cell can transmit electrical activity?
6. Consider a myocardial cell with an action potential of duration 300 ms including its refractory period. What is the maximum rate at which this cell can be activated (fired) into contraction?
7. Distinguish between spatial and temporal recruitment of motor units to obtain increasing levels of muscular activity.
8. Consider three motor units with action potentials (SMUAPs) that are of different biphasic and triphasic shapes. Consider the initial stages of contraction of the related muscle. Draw three plots of the net EMG of the three motor units for increasing levels of contraction with the spatial and temporal recruitment phenomena invoked individually and in combination. Assume low levels of contraction and that the SMUAPs do not overlap.
9. Draw a typical ECG waveform over one cardiac cycle indicating the important component waves, their typical durations, and the typical intervals between them. Label each wave or interval with the corresponding cardiac event or activity.
10. Draw the waveform corresponding to two cycles of a typical ECG signal and indicate the following waves and periods: (a) the P, QRS, and T waves; (b) the RR interval; (c) atrial contraction; (d) atrial relaxation; (e) ventricular contraction; and (f) ventricular relaxation.
11. Explain why the P and T waves are low-frequency signals whereas the QRS complex is a high-frequency signal. Include diagrams of action potentials and an ECG waveform in your reasoning.

12. Explain the reasons for widening of the QRS complex in the case of certain cardiac diseases.
13. Give two examples that call for the use of electronic instruments and/or computers in ECG analysis.
14. A heart patient has a regular SA node pulse (firing) pattern and an irregular ectopic focus. Over a period of 10 s, the SA node was observed to fire regularly at $t = 0, 1, 2, 3, 4, 5, 6, 7, 8$, and 9 s. The ectopic focus was observed to fire at $t = 1.3, 2.8, 6.08$, and 7.25 s.

    Draw two impulse sequences corresponding to the firing patterns of the SA node and the ectopic focus. Draw a schematic waveform of the resulting ECG of the patient. Explain the source of each beat (SA node or ectopic focus) and give reasons.
15. A patient has ventricular bigeminy, where every second pulse from the SA node is replaced by a premature ventricular ectopic beat with a full compensatory pause. (See Figure 9.5 for an illustration of bigeminy.) The SA-node firing rate is regular at 80 beats a minute, and each ectopic beat precedes the blocked SA node pulse by 100 ms.

    (a) Draw a schematic trace of the ECG for 10 beats, marking the time scale in detail.

    (b) Draw a histogram of the RR intervals for the ECG trace.

    (c) What is the average RR interval computed over the 10 beats?
16. Draw a typical PCG (heart sound signal) waveform over one cardiac cycle indicating the important component waves, their typical durations, and the typical intervals between them. Label each wave or interval with the corresponding cardiac event or activity.
17. Give two examples that require the application of electronic instruments and/or computers in EEG analysis.
18. Distinguish between ECG rhythms and EEG rhythms. Sketch one example of each.

1.8 LABORATORY EXERCISES AND PROJECTS

1. Visit an ECG, EMG, or EEG laboratory in your local hospital or health sciences center. View a demonstration of the acquisition of a few biomedical signals. Request a specialist in a related field to explain how he or she would interpret the signals. Volunteer to be the experimental subject and experience first-hand a biomedical signal acquisition procedure!
2. Set up an ECG acquisition system and study the effects of the following conditions or actions on the quality and nature of the signal: loose electrodes; lack of electrode gel; the subject holding his/her breath or breathing freely during the recording procedure; and the subject coughing, talking, or squirming during signal recording.
3. Using a stethoscope, listen to your own heart sounds and those of your friends. Examine the variability of the sounds with the site of auscultation. Study the effects of heavy breathing and speaking by the subject as you are listening to the heart sound signal.
4. Record speech signals of vowels (/A/, /I/, /U/, /E/, /O/), diphthongs (/EI/, /OU/), fricatives (/S/, /F/), and plosives (/T/, /P/), as well as words with all three types of sounds (for example, safety, explosive, hearty, heightened, house). You may be able to perform this experiment with the microphone on your computer workstation. Study the waveform and characteristics of each signal.