Reading 1:
Sensors and Information Systems

1.2 GENERALIZED MEDICAL INSTRUMENTATION SYSTEM

Every instrumentation system has at least some of the functional components shown in Figure 1.1. The primary flow of information is from left to right. Elements and relationships depicted by dashed lines are not essential. The major difference between this system of medical instrumentation and conventional instrumentation systems is that the source of the signals is living tissue or energy applied to living tissue.

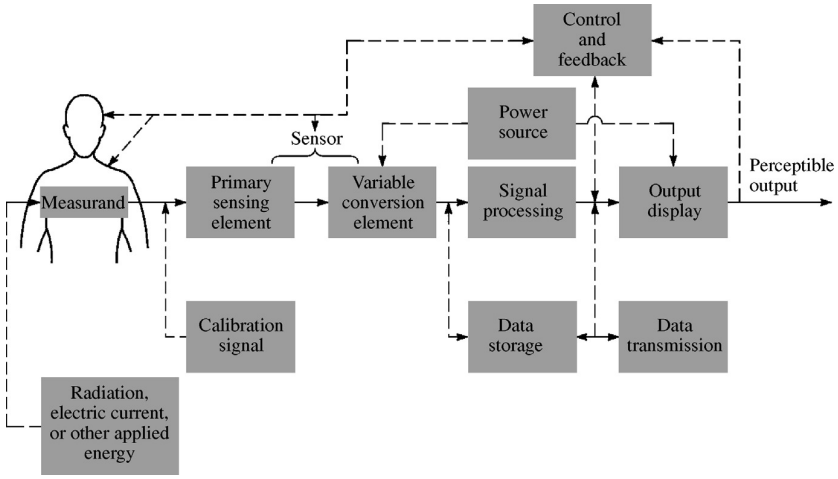

Figure 1.1 Generalized instrumentation system The sensor converts energy or information from the measurand to another form (usually electric). This signal is then processed and displayed so that humans can perceive the information. Elements and connections shown by dashed lines are optional for some applications.

MEASURAND

The physical quantity, property, or condition that the system measures is called the *measurand*. The accessibility of the measurand is important because it may be internal (blood pressure), it may be on the body surface (electrocardiogram potential), it may emanate from the body (infrared radiation), or it may be derived from a tissue sample (such as blood or a biopsy) that is removed from the body. Most medically important measurands can be grouped in the following categories: biopotential, pressure, flow, dimensions (imaging), displacement (velocity, acceleration, and force), impedance, temperature, and chemical concentrations. The measurand may be localized to a specific organ or anatomical structure.

SENSOR

Generally, the term *transducer* is defined as a device that converts one form of energy to another. A sensor converts a physical measurand to an electric output. The sensor should respond only to the form of energy present in the measurand, to the exclusion of all others. The sensor should interface with the living system in a way that minimizes the energy extracted, while being minimally invasive. Many sensors have a primary sensing element such as a diaphragm, which converts pressure to displacement. A variable-conversion element, such as a strain gage, then converts displacement to an electric voltage. Sometimes the sensitivity of the sensor can be adjusted over a wide range by altering the primary sensing element. Many variable-conversion elements need external electric power to obtain a sensor output.

SIGNAL CONDITIONING

Usually the sensor output cannot be directly coupled to the display device. Simple signal conditioners may only amplify and filter the signal or merely match the impedance of the sensor to the display. Often sensor outputs are converted to digital form and then processed by specialized digital circuits or a microcomputer (Tompkins and Webster, 1981). For example, signal filtering may reduce undesirable sensor signals. It may also average repetitive signals to reduce noise, or it may convert information from the time domain to the frequency domain.

OUTPUT DISPLAY

The results of the measurement process must be displayed in a form that the human operator can perceive. The best form for the display may be numerical or graphical, discrete or continuous, permanent or temporary—depending on

the particular measurand and how the operator will use the information. Although most displays rely on our visual sense, some information (Doppler ultrasonic signals, for example) is best perceived by other senses (here, the auditory sense). User controls and the output display should conform to the *Human Factors Engineering Guidelines and Preferred Practices for the Design of Medical Devices* (AAMI, 1993).

AUXILIARY ELEMENTS

A calibration signal with the properties of the measurand should be applied to the sensor input or as early in the signal-processing chain as possible. Many forms of control and feedback may be required to elicit the measurand, to adjust the sensor and signal conditioner, and to direct the flow of output for display, storage, or transmission. Control and feedback may be automatic or manual. Data may be stored briefly to meet the requirements of signal conditioning or to enable the operator to examine data that precede alarm conditions. Alternatively, data may be stored before signal conditioning, so that different processing schemes can be utilized. Conventional principles of communications can often be used to transmit data to remote displays at nurses' stations, medical centers, or medical data-processing facilities.

1.4 MEDICAL MEASUREMENT CONSTRAINTS

The medical instrumentation described throughout this book is designed to measure various medical and physiological parameters. The principal measurement and frequency ranges for each parameter are major factors that affect the design of all the instrument components shown in Figure 1.1. To get a brief overview of typical medical parameter magnitude and frequency ranges, refer to Table 1.1. Shown here are approximate ranges that are intended to include normal and abnormal values. Most of the parameter measurement ranges are quite low compared with nonmedical parameters. Note, for example, that most voltages are in the microvolt range and that pressures are low (about 100 mm Hg = 1.93 psi = 13.3 kPa). Also note that all the signals listed are in the audio-frequency range or below and that many signals contain direct current (dc) and very low frequencies. These general properties of medical parameters limit the practical choices available to designers for all aspects of instrument design.

Many crucial variables in living systems are inaccessible because the proper measurand–sensor interface cannot be obtained without damaging the system. Unlike many complex physical systems, a biological system is of such a nature that it is not possible to turn it off and remove parts of it during the measurement procedure. Even if interference from other physiological systems can be avoided, the physical size of many sensors prohibits the formation of a proper interface. Either such inaccessible variables must be measured indirectly, or corrections must be applied to data that are affected by the measurement process. The cardiac output is an important measurement that is obviously quite inaccessible.

Variables measured from the human body or from animals are seldom deterministic. Most measured quantities vary with time, even when all controllable factors are fixed. Many medical measurements vary widely among normal patients, even when conditions are similar. This inherent *variability* has been documented at the molecular and organ levels, and even for the whole body. Many internal anatomical variations accompany the obvious external differences among patients. Large tolerances on physiological measurements are partly the result of interactions among many physiological systems. Many feedback loops exist among physiological systems, and many of the interrelationships are poorly understood. It is seldom feasible to control or neutralize the effects of these other systems on the measured variable. The most common method of coping with this variability is to assume empirical statistical and probabilistic distribution functions. Single measurements are then compared with these *norms* (see Section 1.8).

Table 1.1 Medical and Physiological Parameters

Parameter or Measuring Technique	Principal Measurement Range of Parameter	Signal Frequency Range, Hz	Standard Sensor or Method
Ballistocardiography (BCG)	0–7 mg	dc–40	Accelerometer, strain gage
	0–100 μm	dc–40	Displacement linear variable differential transformer (LVDT)
Bladder pressure	1–100 cm H_2O	dc–10	Strain-gage manometer
Blood flow	1–300 ml/s	dc–20	Flowmeter (electromagnetic or ultrasonic)
Blood pressure, arterial			
Direct	10–400 mm Hg	dc–50	Strain-gage manometer
Indirect	25–400 mm Hg	dc–60	Cuff, auscultation
Blood pressure, venous	0–50 mm Hg	dc–50	Strain gage
Blood gases			
P_{O_2}	30–100 mm Hg	dc–2	Specific electrode, volumetric or manometric
P_{CO_2}	40–100 mm Hg	dc–2	Specific electrode, volumetric or manometric
P_{N_2}	1–3 mm Hg	dc–2	Specific electrode, volumetric or manometric
P_{CO}	0.1–0.4 mm Hg	dc–2	Specific electrode, volumetric or manometric
Blood pH	6.8–7.8 pH units	dc–2	Specific electrode
Cardiac output	4–25 liter/min	dc–20	Dye dilution, Fick
Electrocardiography (ECG)	0.5–4 mV	0.01–250	Skin electrodes
Electroencephalography (EEG)	5–300 μV	dc–150	Scalp electrodes
(Electrocorticography and brain depth)	10–5000 μV	dc–150	Brain-surface or depth electrodes
Electrogastrography (EGG)	10–1000 μV	dc–1	Skin-surface electrodes
	0.5–80 mV	dc–1	Stomach-surface electrodes
Electromyography (EMG)	0.1–5 mV	dc–10,000	Needle electrodes
Eye potentials			
Electro-oculogram (EOG)	50–3500 μV	dc–50	Contact electrodes
Electroretinogram (ERG)	0–900 μV	dc–50	Contact electrodes
Galvanic skin response (GSR)	1–500 kΩ	0.01–1	Skin electrodes
Gastric pH	3–13 pH units	dc–1	pH electrode; antimony electrode
Gastrointestinal pressure	0–100 cm H_2O	dc–10	Strain-gage manometer
Gastrointestinal forces	1–50 g	dc–1	Displacement system, LVDT
Nerve potentials	0.01–3 mV	dc–10,000	Surface or needle electrodes
Phonocardiography	Dynamic range 80 dB, threshold about 100 μPa	5–2000	Microphone
Plethysmography (volume change)	Varies with organ measured	dc–30	Displacement chamber or impedance change
Circulatory	0–30 ml	dc–30	Displacement chamber or impedance change
Respiratory functions	0–600 liter/min	dc–40	Pneumotachograph head and differential pressure
Pneumotachography (flow rate)			
Respiratory rate	2–50 breaths/min	0.1–10	Strain gage on chest, impedance, nasal thermistor
Tidal volume	50–1000 ml/breath	0.1–10	Above methods
Temperature of body	32–40°C 90–104 °F	dc–0.1	Themistor, thermocouple

SOURCE: Revised from *Medical Engineering*. C. D. Ray (ed.). Copyright © 1974 by Year Book Medical Publishers, Inc., Chicago. Used by permission.

Nearly all biomedical measurements depend either on some form of energy being applied to the living tissue or on some energy being applied as an incidental consequence of sensor operation. X-ray and ultrasonic imaging techniques and electromagnetic or Doppler ultrasonic blood flowmeters depend on externally applied energy interacting with living tissue. Safe levels of these various types of energy are difficult to establish, because many mechanisms of tissue damage are not well understood. A fetus is particularly vulnerable during the early stages of development. The heating of tissue is one effect that must be limited, because even reversible physiological changes can affect measurements. Damage to tissue at the molecular level has been demonstrated in some instances at surprisingly low energy levels.

Operation of instruments in the medical environment imposes important additional constraints. Equipment must be reliable, easy to operate, and capable of withstanding physical abuse and exposure to corrosive chemicals. Electronic equipment must be designed to minimize electric-shock hazards (Chapter 14). The safety of patients and medical personnel must be considered in all phases of the design and testing of instruments. The Medical Device Amendments of 1976 and the Safe Medical Devices Act of 1990 amend the

Federal Food, Drug, and Cosmetics Act to provide for the safety and effectiveness of medical devices intended for human use (Section 1.13).

1.5 CLASSIFICATIONS OF BIOMEDICAL INSTRUMENTS

The study of biomedical instruments can be approached from at least four viewpoints. Techniques of biomedical measurement can be grouped according to the *quantity that is sensed*, such as pressure, flow, or temperature. One advantage of this classification is that it makes different methods for measuring any quantity easy to compare.

A second classification scheme uses the *principle of transduction*, such as resistive, inductive, capacitive, ultrasonic, or electrochemical. Different applications of each principle can be used to strengthen understanding of each concept; also, new applications may be readily apparent.

Measurement techniques can be studied separately for each *organ system*, such as the cardiovascular, pulmonary, nervous, and endocrine systems. This approach isolates all important measurements for specialists who need to know only about a specific area, but it results in considerable overlap of quantities sensed and principles of transduction.

Finally, biomedical instruments can be classified according to the *clinical medicine specialties*, such as pediatrics, obstetrics, cardiology, or radiology. This approach is valuable for medical personnel who are interested in specialized instruments. Of course, certain measurements—such as blood pressure—are important to many different medical specialties.

1.6 INTERFERING AND MODIFYING INPUTS

Desired inputs are the measurands that the instrument is designed to isolate. *Interfering inputs* are quantities that inadvertently affect the instrument as a consequence of the principles used to acquire and process the desired inputs. If spatial or temporal isolation of the measurand is incomplete, the interfering input can even be the same quantity as the desired input. *Modifying inputs* are undesired quantities that indirectly affect the output by altering the performance of the instrument itself. Modifying inputs can affect processing of either desired or interfering inputs. Some undesirable quantities can act as both a modifying input and an interfering input.

A typical electrocardiographic recording system, shown in Figure 1.2, will serve to illustrate these concepts. The desired input is the electrocardiographic voltage v_{ecg} that appears between the two electrodes on the body surface. One interfering input is 60 Hz noise voltage induced in the shaded loop by environmental alternating current (ac) magnetic fields. The desired and the interfering voltages are in series, so both components appear at the input to the

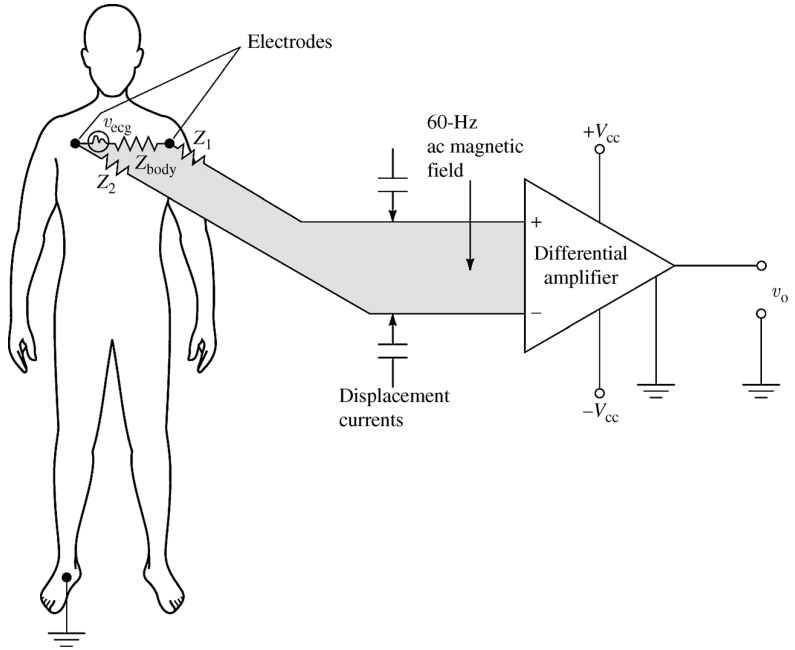

Figure 1.2 Simplified electrocardiographic recording system Two possible interfering inputs are stray magnetic fields and capacitively coupled noise. Orientation of patient cables and changes in electrode–skin impedance are two possible modifying inputs. Z_1 and Z_2 represent the electrode–skin interface impedances.

differential amplifier. Also, the difference between the capacitively coupled displacement currents flowing through each electrode and the body to ground causes an interfering voltage to appear across Z_{body} between the two electrodes and two interfering voltages across Z_1 and Z_2, the electrode impedances.

An example of a modifying input is the orientation of the patient cables. If the plane of the cables is parallel to the ac magnetic field, magnetically introduced interference is zero. If the plane of the cables is perpendicular to the ac magnetic field, magnetically introduced interference is maximal.

1.8 BIOSTATISTICS

The application of statistics to medical data is used to design experiments and clinical studies; to summarize, explore, analyze, and present data; to draw inferences from data by estimation or hypothesis testing; to evaluate diagnostic procedures; and to assist clinical decision making (Dawson-Saunders and Trapp, 2004).

Medical research studies can be *observational studies*, wherein characteristics of one or more groups of patients are observed and recorded, or *experimental intervention studies*, wherein the effect of a medical procedure or treatment is investigated. The simplest observational studies are *case-series* studies that describe some characteristics of a group. These studies are without control subjects, in order only to identify questions for further research. *Case-control* observational studies use individuals selected because they have (or do not have) some outcome or disease and then look backward to find possible causes or risk factors. *Cross-sectional* observational studies analyze characteristics of patients at one particular time to determine the status of a disease or condition. *Cohort* observational studies prospectively ask whether a particular characteristic is a precursor or risk factor for an outcome or disease. Experimental clinical trials are *controlled* if the outcome for patients administered a drug, device, or procedure is compared to the outcome for patients given a placebo or another accepted treatment. The trials are *uncontrolled* if there is no such comparison. *Concurrent controls* are best, because patients are selected in the same way and for the same time period. *Double-blind* study with *randomized* selection of patients to treatment options is preferred, because this design minimizes investigator and patient bias. Medical outcome studies show cost-effective improvements in patient health are increasingly required prior to adoption and reimbursement for new medical technologies (Anonymous, 2001).

Quantitative data are measured on a continuous or discrete *numerical scale* with some precision. Qualitative data values that fit into categories are measured on *nominal scales* that show the names of the categories. An *ordinal scale* is used when the categories exhibit an inherent order. Descriptive statistics are useful to summarize data and their attributes. *Distributions* of empirical or theoretical data reflect the values of a variable or characteristic and the frequency of occurrence of those values.

Measures of the middle, or central tendency, include the well-known *mean*, which is the sum of observed values divided by the number of observations. The mean, found as follows,

$$\overline{X} = \frac{\sum X_i}{n} \tag{1.4}$$

works best as the measure of central tendency for symmetric distributions. The *median* is the value for which half the observations are smaller and half are larger; it is used for skewed numerical data or ordinal data. The *mode* is the observation that occurs most frequently; it is used for bimodal distributions. The *geometric mean* (GM) is the *n*th root of the product of the observations:

$$\text{GM} = \sqrt[n]{X_1 X_2 X_3 \cdots X_n} \tag{1.5}$$

It is used with data on a logarithmic scale.

Measures of spread or dispersion of data describe the variation in the observations. The *range*, which is the difference between the largest and smallest observations, is used to emphasize extreme values. The *standard deviation* is a measure of the spread of data about the mean. It is computed as follows:

$$s = \sqrt{\frac{\sum (X_i - \overline{X})^2}{n-1}} \tag{1.6}$$

It is used with the mean for symmetric distributions of numerical data. Regardless of the type of symmetric distribution, at least 75% of the values always lie between $\overline{X} - 2s$ and $\overline{X} + 2s$. The *coefficient of variation* (CV) is calculated as follows:

$$\text{CV} = \left(\frac{s}{\overline{X}}\right)(100\%) \tag{1.7}$$

It standardizes the variation, making it possible to compare two numerical distributions that are measured on different scales. A *percentile* gives the percentage of a distribution that is less than or equal to the percentile number; it may be used with the median for ordinal data or skewed numerical data. The *interquartile range* is the difference between the 25th and 75th percentiles, so it describes the central 50% of a distribution with any shape. The *standard error of the mean* (SEM) (i.e., standard deviation of the mean), $s_{\overline{X}} = s/\sqrt{n-1}$, expresses the variability to be expected among the *means* in future samples,

whereas the *standard deviation* describes the variability to be expected among *individuals* in future samples.

EXAMPLE 1.1 Your samples from a population are 1, 1, 3, 5, 5. Estimate the mean $\overline{X}$, the standard deviation s, and the standard deviation of the mean $s_{\overline{X}}$.

ANSWER Mean $\overline{X}$ = (sum of values)/(number of values) = $(1 + 1 + 3 + 5 + 5)/5 = 3$ standard deviation $s = \{[(1-3)^2 + (1-3)^2 + (3-3)^2 + (5-3)^2 + (5-3)^2]/(5-1)\}^{1/2} = (16/4)^{1/2} = 2$ standard deviation of mean $s_{\overline{X}} = s/\sqrt{n-1} = 2/\sqrt{5-1} = 1$.

Often we need to study relationships between two numerical characteristics. The *correlation coefficient* r is a measure of the relationship between numerical variables X and Y for paired observations.

$$r = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum(X_i - \overline{X})^2}\sqrt{\sum(Y_i - \overline{Y})^2}} \tag{1.8}$$

The correlation coefficient ranges from -1 for a negative linear relationship to $+1$ for a positive linear relationship; 0 indicates that there is no linear relationship between X and Y. The correlation coefficient is independent of the units employed to measure the variables, which can be different. Like the standard deviation, the correlation coefficient is strongly influenced by outlying values. Because the correlation coefficient measures only a straight-line relationship, it may be small for a strong curvilinear relationship. Of course, a high correlation does *not* imply a cause-and-effect relationship between the variables.

Estimation and hypothesis testing are two ways to make an inference about a value in a population of subjects from a set of observations drawn from a sample of such subjects. In estimation, *confidence intervals* are calculated for a statistic such as the mean. The confidence intervals indicate that a percentage—say 95%—of such confidence intervals contain the true value of the population mean. The confidence intervals indicate the degree of confidence we can have that they contain the true mean. Hypothesis testing reveals whether the sample gives enough evidence for us to reject the *null hypothesis*, which is usually cast as a statement that expresses the opposite of what we think is true. A *P-value* is the probability of obtaining, if the null hypothesis is true, a result that is at least as extreme as the one observed. The *P*-value indicates how often the observed difference would occur by chance alone if, indeed, nothing but chance were affecting the outcome. Recent trends favor using estimation and confidence intervals rather than hypothesis testing.

Methods for measuring the accuracy of a diagnostic procedure use three pieces of information. The *sensitivity* [TP/(TP + FN)] of a test is the probability of its yielding true positive (TP) results in patients who actually have the disease. A test with high sensitivity has a low *false-negative* (FN) rate. The *specificity* [TN/(TN + FP)]of a test is the probability of its yielding negative results in patients who do not have the disease. A test with high specificity has a low *false-positive* (FP) rate; it does not give a false positive (FP) result in many patients who do not have the disease. The third piece of information is the *prior probability*, or prevalence [(TP + FN)/(TN + TP + FN + FP)] of the condition prior to the test (all diseased persons divided by all persons). There are several methods for revising the probability that a patient has a condition on the basis of the results of a diagnostic test. Taking into consideration the results of a diagnostic procedure is only one part of the complex clinical decision-making process. Decision tree analysis and other forms of decision analysis that include economic implications are also used in an effort to make optimal decisions (Webster, 2004).

1.9 GENERALIZED STATIC CHARACTERISTICS

To enable purchasers to compare commercially available instruments and evaluate new instrument designs, quantitative criteria for the performance of instruments are needed. These criteria must clearly specify how well an instrument measures the desired input and how much the output depends on interfering and modifying inputs. Characteristics of instrument performance are usually subdivided into two classes on the basis of the frequency of the input signals.

Static characteristics describe the performance of instruments for dc or very low frequency inputs. The properties of the output for a wide range of constant inputs demonstrate the quality of the measurement, including nonlinear and statistical effects. Some sensors and instruments, such as piezoelectric devices, respond only to time-varying inputs and have no static characteristics.

Dynamic characteristics require the use of differential and/or integral equations to describe the quality of the measurements. Although dynamic characteristics usually depend on static characteristics, the nonlinearities and statistical variability are usually ignored for dynamic inputs, because the differential equations become difficult to solve. Complete characteristics are approximated by the sum of static and dynamic characteristics. This necessary oversimplification is frequently responsible for differences between real and ideal instrument performance.

ACCURACY

The *accuracy* of a single measured quantity is the difference between the true value and the measured value divided by the true value. This ratio is usually expressed as a percent. Because the true value is seldom available, the

accepted true value or reference value should be traceable to the National Institute of Standards and Technology.

The accuracy usually varies over the normal range of the quantity measured, usually decreases as the full-scale value of the quantity decreases on a multirange instrument, and also often varies with the frequency of desired, interfering, and modifying inputs. Accuracy is a measure of the total error without regard to the type or source of the error. The possibility that the measurement is low and that it is high are assumed to be equal. The accuracy can be expressed as percent of reading, percent of full scale, ± number of digits for digital readouts, or $\pm 1/2$ the smallest division on an analog scale. Often the accuracy is expressed as a sum of these, for example, on a digital device, $\pm 0.01\%$ of reading $\pm 0.015\%$ of full-scale ± 1 digit. If accuracy is expressed simply as a percentage, full scale is usually assumed. Some instrument manufacturers specify accuracy only for a limited period of time.

PRECISION

The *precision* of a measurement expresses the number of distinguishable alternatives from which a given result is selected. For example, a meter that displays a reading of 2.434 V is more precise than one that displays a reading of 2.43 V. High-precision measurements do not imply high accuracy, however, because precision makes no comparison to the true value.

RESOLUTION

The smallest incremental quantity that can be measured with certainty is the *resolution*. If the measured quantity starts from zero, the term *threshold* is synonymous with *resolution*. Resolution expresses the degree to which nearly equal values of a quantity can be discriminated.

REPRODUCIBILITY

The ability of an instrument to give the same output for equal inputs applied over some period of time is called *reproducibility* or *repeatability*. Reproducibility does not imply accuracy. For example, a broken digital clock with an AM or PM indicator gives very reproducible values that are accurate only once a day.

STATISTICAL CONTROL

The accuracy of an instrument is not meaningful unless all factors, such as the environment and the method of use, are considered. Statistical control ensures that random variations in measured quantities that result from all factors that influence the measurement process are tolerable. Any systematic errors or bias can be removed by calibration and correction factors, but random variations pose a more difficult problem. The measurand and/or the instrument may introduce statistical variations that make outputs unreproducible. If the cause of this variability cannot be eliminated, then statistical analysis must be used to determine the error variation. Making multiple measurements and averaging the results can improve the estimate of the true value.

STATIC SENSITIVITY

A static calibration is performed by holding all inputs (desired, interfering, and modifying) constant except one. This one input is varied incrementally over the normal operating range, resulting in a range of incremental outputs. The static sensitivity of an instrument or system is the ratio of the incremental output quantity to the incremental input quantity. This ratio is the static component of G_d for desired inputs within the range of the incremental inputs. The incremental slope can be obtained from either the secant between two adjacent points or the tangent to one point on the calibration curve. The static sensitivity may be constant for only part of the normal operating range of the instrument, as shown in Figure 1.3(a). For input–output data that indicate a straight-line calibration curve, the slope m and intercept b for the line with the minimal sum of the squared differences between data points and the line are given by the following equations:

$$m = \frac{n\sum x_d y - (\sum x_d)(\sum y)}{n\sum x_d^2 - (\sum x_d)^2} \tag{1.9}$$

$$b = \frac{(\sum y)(\sum x_d^2) - (\sum x_d y)(\sum x_d)}{n\sum x_d^2 - (\sum x_d)^2} \tag{1.10}$$

$$y = mx_d + b \tag{1.11}$$

where n is the total number of points and each sum is for all n points. The static sensitivity for modulating sensors is usually given per volt of excitation, because the output voltage is proportional to the excitation voltage. For example, the static sensitivity for a blood-pressure sensor containing a strain-gage bridge might be $50\,\mu V \cdot V^{-1}\,mm\,Hg^{-1}$.

ZERO DRIFT

Interfering and/or modifying inputs can affect the static calibration curve shown in Figure 1.3(a) in several ways. Zero drift has occurred when all output values increase or decrease by the same absolute amount. The slope of the sensitivity curve is unchanged, but the output-axis intercept increases or decreases as shown in Figure 1.3(b). The following factors can cause zero drift: manufacturing misalignment, variations in ambient temperature, hysteresis, vibration, shock, and sensitivity to forces from undesired directions. A change in the de-offset voltage at the electrodes in the electrocardiograph example in Figure 1.2 is an example of zero drift. Slow changes in the de-offset voltage do not cause a

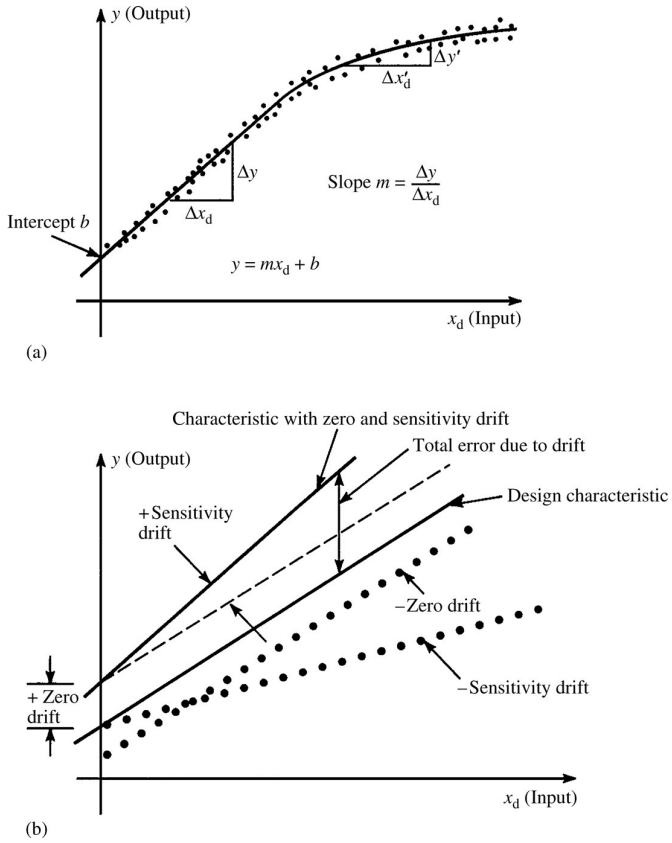

Figure 1.3 (a) Static-sensitivity curve that relates desired input x_d to output y. Static sensitivity may be constant for only a limited range of inputs. (b) Static sensitivity: zero drift and sensitivity drift. Dotted lines indicate that zero drift and sensitivity drift can be negative. [Part (b) modified from *Measurement Systems: Application and Design*, E. O. Doebelin. Copyright © 1990 by McGraw-Hill, Inc. Used with permission of McGraw-Hill Book Co.]

problem, because the ECG amplifier is ac-coupled. Fast changes due to motion of the subject do cause low-frequency artifact to appear at the output.

SENSITIVITY DRIFT

When the slope of the calibration curve changes as a result of an interfering and/or modifying input, a drift in sensitivity results. Sensitivity drift causes error that is proportional to the magnitude of the input. The slope of the calibration curve can either increase or decrease, as indicated in Figure 1.3(b). Sensitivity drift can result from manufacturing tolerances, variations in power supply, nonlinearities, and changes in ambient temperature and pressure. Variations in the electrocardiograph-amplifier voltage gain as a result of fluctuations in dc power-supply voltage or change in temperature are examples of sensitivity drift.

LINEARITY

A system or element is linear if it has properties such that if y_1 is the response to x_1 and y_2 is the response to x_2, then $y_1 + y_2$ is the response to $x_1 + x_2$, and Ky_1 is the response to Kx_1. These two requirements for system linearity are restated in Figure 1.4(a).

They are clearly satisfied for an instrument with a calibration curve that is a straight line.

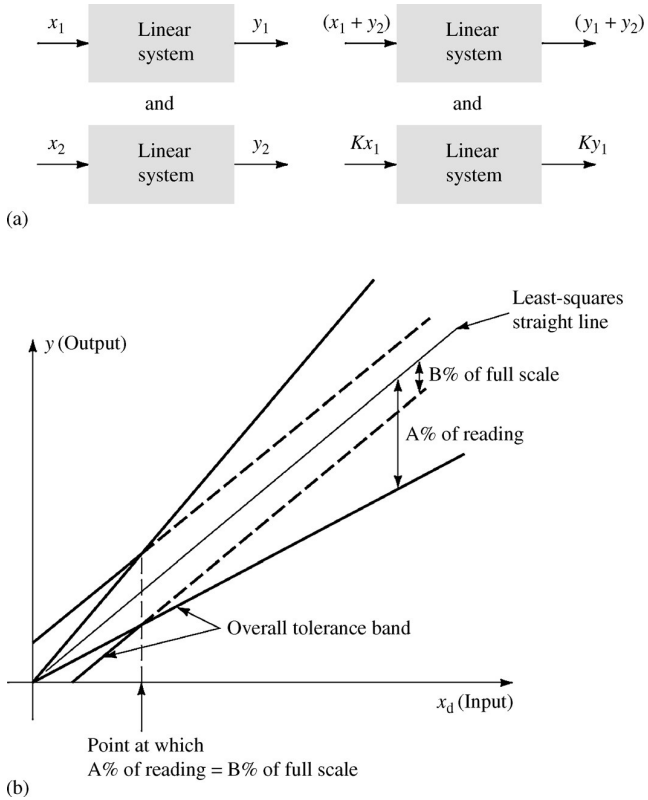

Figure 1.4 (a) Basic definition of linearity for a system or element. The same linear system or element is shown four times for different inputs. (b) A graphical illustration of independent nonlinearity equals ±A% of the reading, or ±B% of full scale, whichever is greater (that is, whichever permits the larger error). [Part (b) modified from *Measurement Systems: Application and Design*, E. O. Doebelin. Copyright © 1990 by McGraw-Hill, Inc. Used with permission of McGraw-Hill Book Co.]

Keep in mind, however, that high accuracy does not necessarily imply linearity. In practice, no instrument has a perfect linear response, so a measure of deviation from linearity is needed. *Independent nonlinearity* expresses the maximal deviation of points from the least-squares fitted line as either $\pm A\%$ of the reading or $\pm B\%$ of full scale, whichever is greater (that is, whichever permits the larger error). This linearity specification is shown in Figure 1.4(b). For up-scale readings, the percent-of-reading figure is desirable because most errors are proportional to the reading. For small readings near zero, however, percentage of full scale is more realistic because it is not feasible to test for small percent-of-reading deviations near zero. All data points must fall inside the ''funnel'' shown in Figure 1.4(b). For most instruments that are essentially linear, if other sources of error are minimal, the accuracy is equal to the nonlinearity.

INPUT RANGES

Several maximal ranges of allowed input quantities are applicable for various conditions. Minimal resolvable inputs impose a lower bound on the quantity to be measured. The normal linear operating range specifies the maximal or near-maximal inputs that give linear outputs.

The static linear range and the dynamic linear range may be different. The maximal operating range is the largest input that does not damage the instrument. Operation in the upper part of this range is more likely to be nonlinear. Finally, storage conditions specify environmental and interfering input limits that should not be exceeded when the instrument is not being used. These ranges are not always symmetric with respect to zero input, particularly for storage conditions. Typical operating ranges for blood-pressure sensors have a positive bias, such as +200 mm Hg to −60 mm Hg (+26.6 to −8.0 kPa).

INPUT IMPEDANCE

Because biomedical sensors and instruments usually convert nonelectric quantities into voltage or current, we introduce a generalized concept of input impedance. This is necessary so that we can properly evaluate the degree to which instruments disturb the quantity being measured. For every desired input X_{d1} that we seek to measure, there is another implicit input quantity X_{d2} such that the product $X_{d1} \cdot X_{d2}$ has the dimensions of power. This product represents the instantaneous rate at which energy is transferred across the tissue–sensor interface. The generalized input impedance Z_x is the ratio of the phasor equivalent of a steady-state sinusoidal *effort* input variable (voltage, force, pressure) to the phasor equivalent of a steady-state sinusoidal *flow* input variable (current, velocity, flow).

$$Z_x = \frac{X_{d1}}{X_{d2}} = \frac{\text{effort variable}}{\text{flow variable}} \tag{1.12}$$

The power P is the time rate of energy transfer from the measurement medium.

$$P = X_{d1}X_{d2} = \frac{X_{d1}^2}{Z_x} = Z_x X_{d2}^2 \tag{1.13}$$

To minimize P, when measuring effort variables X_{d1}, we should make the generalized input impedance as large as possible. This is usually achieved by minimizing the flow variable. However, most instruments function by measuring minute values of the flow variable, so the flow variable cannot be reduced to zero. On the other hand, when we are measuring flow variables X_{d2}, small input impedance is needed to minimize P. The loading caused by measuring devices depends on the magnitude of the input impedance $|Z_x|$ compared with the magnitude of the source impedance $|Z_s|$ for the desired input. Unfortunately, biological source impedances are usually unknown, variable, and difficult to measure and control. Thus the instrument designer must usually focus on maximizing the input impedance Z_x for effort-variable measurement. When the measurand is a flow variable instead of an effort variable, it is more convenient to use the admittance $Y_x = 1/Z_x$ than the impedance.

1.10 GENERALIZED DYNAMIC CHARACTERISTICS

Only a few medical measurements, such as body temperature, are constant or slowly varying quantities. Most medical instruments must process signals that are functions of time. It is this time-varying property of medical signals that requires us to consider dynamic instrument characteristics. Differential or integral equations are required to relate dynamic inputs to dynamic outputs for continuous systems. Fortunately, many engineering instruments can be described by ordinary linear differential equations with constant coefficients. The input $x(t)$ is related to the output $y(t)$ according to the following equation:

$$a_n \frac{d^n y}{dt^n} + \cdots + a_1 \frac{dy}{dt} + a_0 y(t) = b_m \frac{d^m x}{dt^m} + \cdots + b_1 \frac{dx}{dt} + b_0 x(t) \tag{1.14}$$

where the constants $a_i (i = 0, 1, \ldots, n)$ and $b_j (j = 0, 1, \ldots, m)$ depend on the physical and electric parameters of the system. By introducing the differential operator $D^k \equiv d^k()/dt^k$, we can write this equation as

$$(a_n D^n + \cdots + a_1 D + a_0) y(t) = (b_m D^m + \cdots + b_1 D + b_0) x(t) \tag{1.15}$$

Readers familiar with Laplace transforms may recognize that D can be replaced by the Laplace parameter s to obtain the equation relating the transforms $Y(s)$ and $X(s)$. This is a *linear* differential equation, because the linear properties stated in Figure 1.4(a) are assumed and the coefficients

a_i and b_j are not functions of time or the input $x(t)$. The equation is *ordinary*, because there is only one independent variable y. Essentially such properties mean that the instrument's methods of acquiring and analyzing the signals do not change as a function of time or the quantity of input. For example, an autoranging instrument may violate these conditions.

Most practical instruments are described by differential equations of zero, first, or second order; thus $n = 0, 1, 2$, and derivatives of the input are usually absent, so $m = 0$.

The input $x(t)$ can be classified as transient, periodic, or random. No general restrictions are placed on $x(t)$, although, for particular applications, bounds on amplitude and frequency content are usually assumed. Solutions for the differential equation depend on the input classifications. The step function is the most common transient input for instrumentation. Sinusoids are the most common periodic function to use because, through the Fourier-series expansion, any periodic function can be approximated by a sum of sinusoids. Band-limited white noise (uniform-power spectral content) is a common random input because one can test instrument performance for all frequencies in a particular bandwidth.

TRANSFER FUNCTIONS

The transfer function for a linear instrument or system expresses the relationship between the input signal and the output signal mathematically. If the transfer function is known, the output can be predicted for any input. The *operational transfer function is the ratio* $y(D)/x(D)$ as a function of the differential operator D.

$$\frac{y(D)}{x(D)} = \frac{b_m D^m + \cdots + b_1 D + b_0}{a_n D^n + \cdots + a_1 D + a_0} \tag{1.16}$$

This form of the transfer function is particularly useful for transient inputs. For linear systems, the output for transient inputs, which occur only once and do not repeat, is usually expressed directly as a function of time, $y(t)$, which is the solution to the differential equation.

The *frequency transfer function* for a linear system is obtained by substituting $j\omega$ for D in (1.16).

$$\frac{Y(j\omega)}{X(j\omega)} = \frac{b_m(j\omega)^m + \cdots b_1(j\omega) + b_0}{a_n(j\omega)^n + \cdots a_1(j\omega) + a_0} \tag{1.17}$$

where $j = +\sqrt{-1}$ and ω is the angular frequency in radians per second. The input is usually given as $x(t) = A_x \sin(\omega t)$, and all transients are assumed to have died out. The output $y(t)$ is a sinusoid with the same frequency, but the amplitude and phase depend on ω; that is, $y(t) = B(\omega) \sin[\omega t + \phi(\omega)]$. The frequency transfer function is a complex quantity having a magnitude that is the ratio of the magnitude of the output to the magnitude of the input and a phase angle ϕ that is the phase of the output $y(t)$ minus the phase of the input $x(t)$. The phase angle for most instruments is negative. We do not usually express the output of the system as $y(t)$ for each frequency, because we know that it is just a sinusoid with a particular magnitude and phase. Instead, the amplitude ratio and the phase angle are given separately as functions of frequency.

The dynamic characteristics of instruments are illustrated below by examples of zero-, first-, and second-order linear instruments for step and sinusoidal inputs.

ZERO-ORDER INSTRUMENT

The simplest nontrivial form of the differential equation results when all the a's and b's are zero except a_0 and b_0.

$$a_0 y(t) = b_0 x(t) \tag{1.18}$$

This is an algebraic equation, so

$$\frac{y(D)}{x(D)} = \frac{Y(j\omega)}{X(j\omega)} = \frac{b_0}{a_0} = K = \text{static sensitivity} \tag{1.19}$$

where the single constant K replaces the two constants a_0 and b_0. This zero-order instrument has ideal dynamic performance, because the output is proportional to the input for all frequencies and there is no amplitude or phase distortion.

A linear potentiometer is a good example of a zero-order instrument. Figure 1.5 shows that if the potentiometer has pure uniform resistance, then the output voltage $y(t)$ is directly proportional to the input displacement $x(t)$, with no time delay for any frequency of input. In practice, at high frequencies, some parasitic capacitance and inductance might cause slight distortion. Also, low-resistance circuits connected to the output can load this simple zero-order instrument.

FIRST-ORDER INSTRUMENT

If the instrument contains a single energy-storage element, then a first-order derivative of $y(t)$ is required in the differential equation.

$$a_1 \frac{dy(t)}{dt} + a_0 y(t) = b_0 x(t) \tag{1.20}$$

This equation can be written in terms of the differential operator D as

$$(\tau D + 1) y(t) = K x(t) \tag{1.21}$$

where $K = b_0/a_0 =$ static sensitivity, and $\tau = a_1/a_0 =$ time constant.

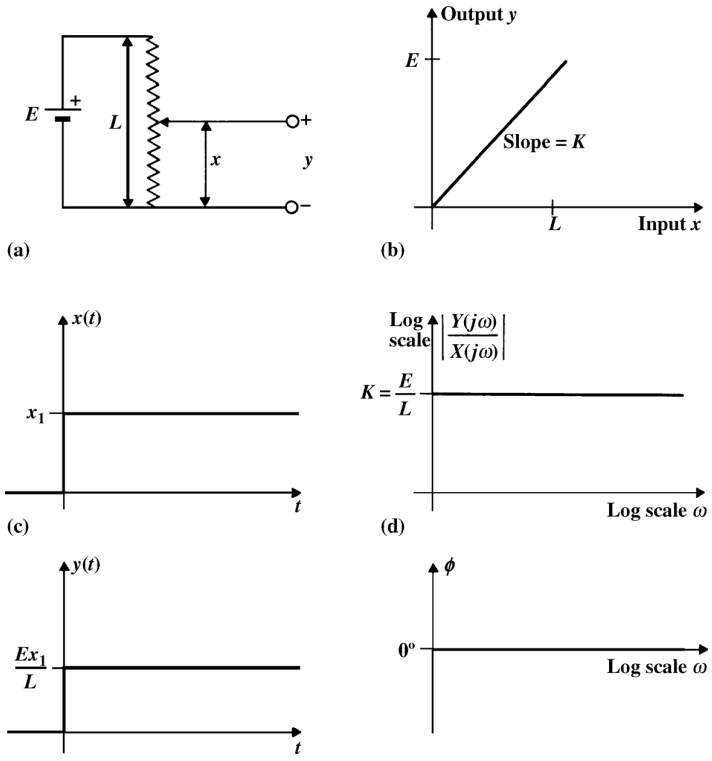

Figure 1.5 (a) A linear potentiometer, an example of a zero-order system. (b) Linear static characteristic for this system. (c) Step response is proportional to input. (d) Sinusoidal frequency response is constant with zero phase shift.

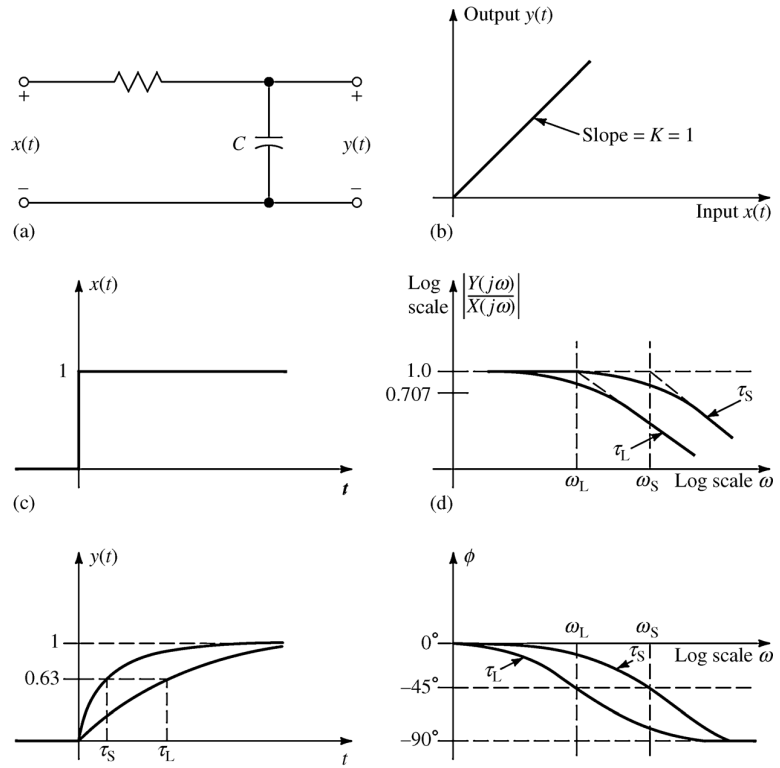

Figure 1.6 (a) A low-pass *RC* filter, an example of a first-order instrument. (b) Static sensitivity for constant inputs. (c) Step response for large time constants (τ_L) and small time constants (τ_S). (d) Sinusoidal frequency response for large and small time constants.

Exponential functions offer solutions to this equation when appropriate constants are chosen. The operational transfer function is

$$\frac{y(D)}{x(D)} = \frac{K}{1 + \tau D} \tag{1.22}$$

and the frequency transfer function is

$$\frac{Y(j\omega)}{X(j\omega)} = \frac{K}{1 + j\omega\tau} = \frac{K}{\sqrt{1 + \omega^2\tau^2}} \underline{/\phi = \arctan\ (-\omega\tau/1)} \tag{1.23}$$

The *RC* low-pass filter circuit shown in Figure 1.6(a) is an example of a first-order instrument. The input is the voltage $x(t)$, and the output is the voltage $y(t)$ across the capacitor. The first-order differential equation for this circuit is $RC[dy(t)/dt] + y(t) = Kx(t)$. The static-sensitivity curve given in Figure 1.6(b) shows that static outputs are equal to static inputs. This is verified by the differential equation because, for static conditions, $dy/dt = 0$. The step response in Figure 1.6(c) is exponential with a time constant $\tau = RC$.

$$y(t) = K(1 - e^{-t/\tau}) \tag{1.24}$$

The smaller the time constant, the faster the output approaches the input. For sinusoids, (1.23) and Figure 1.6(d) show that the magnitude of the output decreases as frequency increases. For larger time constants, this decrease occurs at lower frequency.

When $\omega = 1/\tau$, the magnitude is $1/\sqrt{2} = 0.707$ times smaller, and the phase angle is $-45°$. This particular frequency ω is known as the *corner, cutoff*, or *break* frequency. Figure 1.6(d) verifies that this is a low-pass filter; low-frequency sinusoids are not severely attenuated, whereas high-frequency sinusoids produce very little output voltage. The ordinate of the frequency-response magnitude in Figure 1.6(d) is usually plotted on a log scale and may be given in units of decibels (dB), which are defined as $\text{dB} = 20 \log_{10}|Y(j\omega)/X(j\omega)|$. A mercury-in-glass thermometer is another example of a low-pass first-order instrument.

EXAMPLE 1.2 A first-order low-pass instrument has a time constant of 20 ms. Find the maximal sinusoidal input frequency that will keep output error due to frequency response less than 5%. Find the phase angle at this frequency.

ANSWER

$$\frac{Y(j\omega)}{X(j\omega)} = \frac{K}{1 + j\omega\tau}$$

$$\left|\frac{K}{1 + j\omega\tau}\right| = \frac{K}{\sqrt{1 + \omega^2\tau^2}} = 0.95K$$

$$(\omega^2\tau^2 + 1)(0.95)^2 = 1$$

$$\omega^2 = \frac{1 - (0.95)^2}{(0.95)^2(20 \times 10^{-3})^2}$$

$$\omega = 16.4 \text{ rad/s}$$

$$f = \frac{\omega}{2\pi} = 2.62 \text{ Hz}$$

$$\phi = \tan^{-1}\left(\frac{-\omega\tau}{1}\right) = -18.2°$$

If R and C in Figure 1.6(a) are interchanged, the circuit becomes another first-order instrument known as a *high-pass filter*. The static characteristic is zero for all values of input, and the step response jumps immediately to the step voltage but decays exponentially toward zero as time increases. Thus $y(t) = Ke^{-t/\tau}$. Low-frequency sinusoids are severely attenuated, whereas high-frequency sinusoids are little attenuated. The sinusoidal transfer function is $Y(j\omega)/X(j\omega) = j\omega\tau/(1 + j\omega\tau)$.

EXAMPLE 1.3 From a 2 kV source in series with a 20 kΩ resistor, calculate the time required to charge a 100 μF defibrillator capacitor to 1.9 kV

ANSWER Circuit is shown in Figure 1.6(a). Use (1.24) $v_C = V - Ve^{-\frac{t}{RC}}$

$$1900 \text{ V} = 2000 \text{ V} - 2000 \text{ V} \cdot e^{-\frac{t}{(20{,}000\Omega)(100\times10^{-6}\text{ F})}}$$

$$-100 \text{ V} = -2000 \text{ V} \cdot e^{-\frac{t}{(20{,}000\Omega)(100\times10^{-6}\text{ F})}}$$

$$0.05 = e^{-\frac{t}{(20{,}000\Omega)(100\times10^{-6}\text{ F})}}$$

$$\ln 0.05 = -\frac{t}{2\,\Omega \cdot \text{F}}$$

$$t = 5.99 \text{ s}$$

SECOND-ORDER INSTRUMENT

An instrument is second order if a second-order differential equation is required to describe its dynamic response.

$$a_2\frac{d^2y(t)}{dt^2} + a_1\frac{dy(t)}{dt} + a_0y(t) - b_0x(t) \tag{1.25}$$

Many medical instruments are second order or higher, and low pass. Furthermore, many higher-order instruments can be approximated by second-order characteristics if some simplifying assumptions can be made. The four constants in (1.25) can be reduced to three new ones that have physical significance:

$$\left[\frac{D^2}{\omega_n^2} + \frac{2\zeta D}{\omega_n} + 1\right]y(t) = Kx(t) \tag{1.26}$$

where

$$K = \frac{b_0}{a_0} = \text{static sensitivity, output units divided by input units}$$

$$\omega_n = \sqrt{\frac{a_0}{a_2}} = \text{undamped natural frequency, rad/s}$$

$$\zeta = \frac{a_1}{2\sqrt{a_0a_1}} = \text{damping ratio, dimensionless}$$

Again exponential functions offer solutions to this equation, although the exact form of the solution varies as the damping ratio becomes greater than, equal to, or less than unity. The operational transfer function is

$$\frac{y(D)}{x(D)} = \frac{K}{\frac{D^2}{\omega_n^2} + \frac{2\zeta D}{\omega_n} + 1} \tag{1.27}$$

and the frequency transfer function is

$$\frac{Y(j\omega)}{X(j\omega)} = \frac{K}{(j\omega/\omega_n)^2 + (2\zeta j\omega/\omega_n) + 1}$$
$$= \frac{K}{\sqrt{[1-(\omega/\omega_0)^2]^2 + 4\zeta^2\omega^2/\omega_n^2}} \Big/ \phi = \arctan\frac{2\zeta}{\omega/\omega_n - \omega_n/\omega} \tag{1.28}$$

A mechanical force-measuring instrument illustrates the properties of a second-order instrument (Doebelin, 1990). Mass, spring, and viscous-damping elements oppose the applied input force $x(t)$, and the output is the resulting displacement $y(t)$ of the movable mass attached to the spring [Figure 1.7(a)]. If the natural frequency of the spring is much greater than the frequency components in the input, the dynamic effect of the spring can be included by adding one-third of the spring's mass to the mass of the moving elements to obtain the equivalent total mass M.

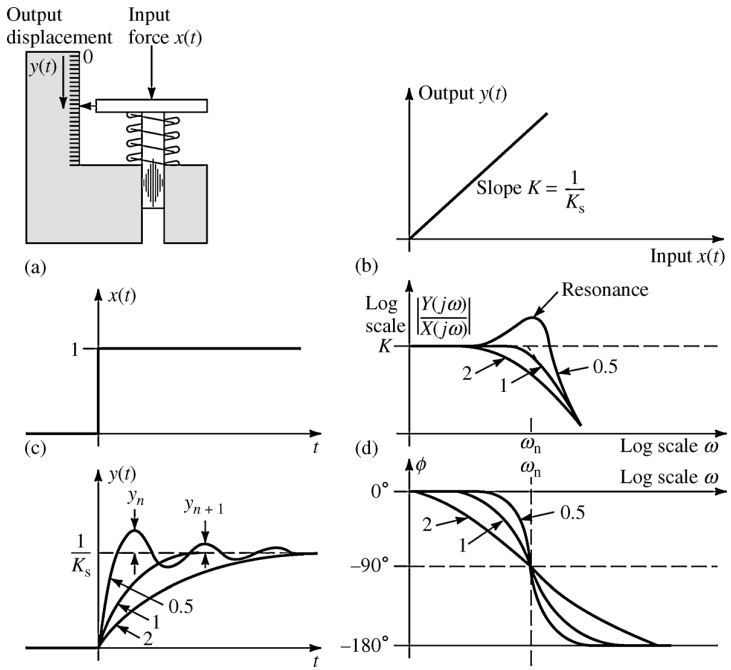

Figure 1.7 (a) Force-measuring spring scale, an example of a second-order instrument. (b) Static sensitivity. (c) Step response for overdamped case $\zeta = 2$, critically damped case $\zeta = 1$, underdamped case $\zeta = 0.5$. (d) Sinusoidal steady-state frequency response, $\zeta = 2$, $\zeta = 1$, $\zeta = 0.5$. [Part (a) modified from *Measurement Systems: Application and Design*, E. O. Doebelin. Copyright © 1990 by McGraw-Hill, Inc. Used with permission of McGraw-Hill Book Co.]

Hooke's law for linear springs is assumed, so the spring constant is K_s. Dry friction is neglected and perfect viscous friction is assumed, with constant B.

To eliminate gravitational force from the equation, we adjust the scale until $y = 0$ when $x = 0$. Then the sum of the forces equals the product of mass and acceleration.

$$x(t) - B\frac{dy(t)}{dt} - K_s y(t) = M\frac{d^2y(t)}{dt^2} \tag{1.29}$$

This equation has the same form as (1.26) when the static sensitivity, undamped natural frequency, and damping ratio are defined in terms of K_s, B, and M, as follows:

$$K = \frac{1}{K_s} \tag{1.30}$$

$$\omega_n = \sqrt{K_s/M} \tag{1.31}$$

$$\zeta = \frac{B}{2\sqrt{K_m M}} \tag{1.32}$$

The static response is $y(t) = Kx(t)$, as shown in Figure 1.7(b). The step response can have three forms, depending on the damping ratio. For a unit-step input, these three forms are

Overdamped, $\zeta > 1$:

$$y(t) = -\frac{\zeta + \sqrt{\zeta^2-1}}{2\sqrt{\zeta^2-1}} Ke^{\left(-\zeta+\sqrt{\zeta^2-1}\right)\omega_n t} + \frac{\zeta - \sqrt{\zeta^2-1}}{2\sqrt{\zeta^2-1}} Ke^{\left(-\zeta-\sqrt{\zeta^2-1}\right)\omega_n t} + K \tag{1.33}$$

Critically damped, $\zeta = 1$:

$$y(t) = -(1+\omega_n t)Ke^{-\omega_n t} + K \tag{1.34}$$

Underdamped, $\zeta < 1$:

$$y(t) = -\frac{e^{-\zeta\omega_n t}}{\sqrt{1-\zeta^2}} K\sin\left(\sqrt{1-\zeta^2}\omega_n t + \phi\right) + K$$
$$\phi = \arcsin\sqrt{1-\zeta^2} \tag{1.35}$$

Examples of these three step responses are represented in Figure 1.7(c). Only for damping ratios less than unity does the step response overshoot the final

value. Equation (1.35) shows that the frequency of the oscillations in the underdamped response in Figure 1.7(c) is the damped natural frequency $\omega_d = \omega_n\sqrt{1-\zeta^2}$. A practical compromise between rapid rise time and minimal overshoot is a damping ratio of about 0.7.

EXAMPLE 1.4 For underdamped second-order instruments, find the damping ratio ζ from the step response.

ANSWER To obtain the maximums for the underdamped response, we take the derivative of (1.35) and set it to zero. For $\zeta < 0.3$ we approximate positive maximums when the sine argument equals $3\pi/2$, $7\pi/2$, and so forth. This occurs at

$$t_n = \frac{3\pi/2 - \phi}{\omega_n\sqrt{1-\zeta^2}} \quad \text{and} \quad t_{n+1} = \frac{7\pi/2 - \phi}{\omega_n\sqrt{1-\zeta^2}} \tag{1.36}$$

The ratio of the first positive overshoot y_n to the second positive overshoot y_{n+1} [Figure 1.7(c)] is

$$\begin{aligned}\frac{y_n}{y_{n+1}} &= \frac{\dfrac{K}{\sqrt{1-\zeta^2}}\exp\left(-\zeta\omega_n\dfrac{(3\pi/2-\phi)}{\omega_n\sqrt{1-\zeta^2}}\right)}{\dfrac{K}{\sqrt{1-\zeta^2}}\exp\left(-\zeta\omega_n\dfrac{(7\pi/2-\phi)}{\omega_n\sqrt{1-\zeta^2}}\right)} \\ &= \exp\frac{2\pi\zeta}{\sqrt{1-\zeta^2}} \\ \ln\frac{y_n}{y_{n+1}} &= \Lambda = \frac{2\pi\zeta}{\sqrt{1-\zeta^2}}\end{aligned} \tag{1.37}$$

where Λ is defined as *logarithmic decrement*. Solving for ζ yields

$$\zeta = \frac{\Lambda}{\sqrt{4\pi^2 + \Lambda^2}} \tag{1.38}$$

For sinusoidal steady-state responses, the frequency transfer function (1.28) and Figure 1.7(d) show that low-pass frequency responses result. The rate of decline in the amplitude frequency response is twice the rate of that decline for first-order instruments. Note that resonance phenomena can occur if the damping ratio is too small. Also note that the output phase lag can be as much as 180°, whereas for single-order instruments, the maximal phase lag is 90°.

TIME DELAY

Instrument elements that give an output that is exactly the same as the input, except that it is delayed in time by τ_d, are defined as *time-delay elements*. The mathematical expression for these elements is

$$y(t) = K_x(t - \tau_d), \quad t > \tau_d \tag{1.39}$$

These elements may also be called analog delay lines, transport lags, or dead times. Although first-order and second-order instruments have negative phase angles that imply time delays, the phase angle varies with frequency, so the delay is not constant for all frequencies. For time delays, the static characteristic is the constant K, the step response is specified by (1.39), and the sinusoidal frequency response for magnitude and phase is

$$\frac{Y(j\omega)}{X(j\omega)} = Ke^{-j\omega\tau_d} \tag{1.40}$$

Time delays are present in transmission lines (electric, mechanical, hydraulic blood vessels, and pneumatic respiratory tubing), magnetic tape recorders, and some digital signal-processing schemes. Usually these time delays are to be avoided, especially in instruments or systems that involve feedback, because undesired oscillations may result.

If the instrument is used strictly for measurement and is not part of a feedback-control system, then some time delay is usually acceptable. The transfer function for undistorted signal reproduction with time delay becomes $Y(j\omega)/X(j\omega) = K\underline{/-\omega\tau_d}$. Our previous study of time-delay elements shows that the output magnitude is K times the input magnitude for all frequencies and that the phase lag increases linearly with frequency.

The transfer-function requirements concern the *overall* instrument transfer function. The overall transfer function of linear elements connected in series is the product of the transfer functions for the individual elements. Many combinations of nonlinear elements can produce the overall linear transfer function required. Various forms of modulation and demodulation are used, and unavoidable sensor nonlinearities can sometimes be compensated for by other instrument elements.