BIOMEDICAL TRANSDUCERS AND MEASUREMENTS

Sub Code: 18BM45
MODULE 1

Module -1 FUNDAMENTAL CONCEPTS & BASIC TRANSDUCERS: Introduction, Classification of Transducers, Measurement, Signals and Noise in the measurement-Measurement, signals and noise, signal to noise ratio, different types of noise. Characteristics of Measurement system -Transducer and measurement system, static characteristics, dynamic characteristics, standard and calibration, accuracy and error.

Module -1 FUNDAMENTAL CONCEPTS & BASIC TRANSDUCERS

1.1 Introduction,

1.2 Classification of Transducers,

1.3 Measurement,

1.4 Signals and Noise in the Measurement

1.5 Characteristics of the Measurement System

- **1.5.1 Transducer and Measurement System**
- **1.5.2 Static Characteristics**
- **1.5.2.1 Sensitivity, Resolution, and Reproducibility**
- **1.5.2.2 Measurement Range**
- **1.5.2.3 Linearity or Nonlinearity**
- **1.5.2.4 Hysteresis**
- **1.5.3 Dynamic Characteristics**
  - **1.5.3.1 Linear and Nonlinear Systems**
  - **1.5.3.2 Frequency Response**
  - **1.5.3.3 Time Constant, Response Time, Rise Time, and Settling Time**

1.6 Standard and Calibration

1.7 Accuracy and Error

1.1 INTRODUCTION

Transducers are devices which convert one form of energy into another. Because of the familiar advantages of electric and electronic methods of measurement, it is the usual practice to convert all non-electric phenomenon associated with the physiological events into electric quantities. Numerous methods have since been developed for this purpose and basic principles of physics have extensively been employed. Variation in electric circuit parameters like resistance, capacitance and inductance in accordance with the events to be measured, is the simplest of such methods. Peizo-electric and photoelectric transducers are also very common. Chemical events are detected by measurement of current flow through the electrolyte or by the potential changes developed across the membrane electrodes. A number of factors decide the choice of a particular transducer to be used for the study of a specific phenomenon. These factors include:

- The magnitude of quantity to be measured.
- The order of accuracy required.
- The static or dynamic character of the process to be studied.
- The site of application on the patient’s body, both for short-term and long-term monitoring.
- Economic considerations

1.2 CLASSIFICATION OF TRANSDUCERS

Many physical, chemical and optical properties and principles can be applied to construct transducers for applications in the medical field. The transducers can be classified in many ways, such as: (i) By the process used to convert the signal energy into an electrical signal. For this, transducers can be categorized as: Active Transducers—a transducer that converts one form of energy directly into another. For example: photovoltaic cell in which light energy is converted into electrical energy.

1.3 MEASUREMENT

A *measurement* is a procedure by which an observer determines the quantity that characterizes the property or state of an object. The quantity to be determined is the object quantity of the measurement.

In this book, physical or chemical quantities that contain physiological information are Considered as the object quantities. Sometimes such quantities can be estimated by human sense, for instance through visible observations. But, to obtain objective, reproducible, and quantitative results, instruments should be used where the results are given as the output of the measurement system.

The physical characteristics of the output depend on the type of instrument used. When electronic instruments are used, the output will be in the form of an electric potential. This output can also be converted into digital values if desired. In any case, the original physical or chemical quantity is converted into convenient forms such as an electric potential or a numerical value. To describe the object quantity correctly from the output of an instrument, the relation between the output of the instrument and the object quantity must be defined. To obtain this, adequate instrument calibration procedure is required as discussed in Section 1.3.1. The term measurement always implies the whole procedure by which the object quantity is correctly determined.

1.4 SIGNALS AND NOISE IN THE MEASUREMENT

1.4.1 Signals and Noise

In a measurement, the signal is defined as the component of a variable that contains information about the object quantity, whereas noise is defined as a component unrelated to the object quantity. Thus, the signal is the component that an observer wants to obtain, and noise is the unwanted component. Therefore, signals and noise are not uniquely defined by their physical nature, but depend on the intention of the observer. For example, the electromyogram (EMG), which is the potential generated by muscles, can be regarded as a signal by an observer who is interested in muscle activity. But the EMG is an unwanted component for another observer who is interested in obtaining nerve action potentials. In this situation, the EMG component is considered to be a kind of noise. In this sense, the definition of signals and noise is arbitrary in general measurement situations. The situation is different in communication technology, where a sender sends signals to a receiver, so that the signal is fully determined by the intention of the sender.

In actual measurement situations, there is no general rule by which signals and noise can be distinguished. Only detailed knowledge about the nature of the measurement object and possible disturbances in the system can help to distinguish signals from noise.

1.4.2 Signal-to-Noise Ratio

The *signal-to-noise ratio* is generally defined as the ratio of the value of the signal to that of the noise, and often denoted simply by *S/N* (or SNR). Commonly, the ratio of the power of the signal to that of the noise is used, but the ratio of peak-to-peak values or also root-mean-square amplitudes of the signal to that of the noise may be used as the signal-to-noise ratio, if it is previously defined.

In practical situations, the signal-to-noise ratio is considered in limited frequency ranges, and its value is always different for different frequency ranges because the signal and the noise always have different frequency spectra.

The ratio is often expressed in decibels (dB), which is defined as

$$10 \log_{10} S/N \quad (1.8)$$

when S and N denote the powers of the signal and the noise, respectively, while it is

$$20 \log_{10} S/N \quad (1.9)$$

When S and N denote the root-mean-square amplitudes of the signal and the noise, respectively. Due to these different definitions of signal-to-noise ratio for power and amplitude, the value expressed in decibels can be the same regardless of whether S and N are expressed in power or in amplitude.

1.4.3 Different Types of Noise

Different types of noise arising from different sources may appear in actual measurement situations and can be characterized by their power spectra, and can be treated theoretically.
In addition to these types of noise, all unwanted signals superimposed on the signal are classified as noise, irrespective of whether the sources are identified or not.

1.4.4 Thermal Noise

Thermal noise is a kind of noise that is caused by random thermal agitation. The power density of the thermal noise is uniformly distributed through the whole frequency range, and its power is proportional to the temperature. For example, across the terminals of a resistor of resistance R, the noise potential $v(t)$ appears in a frequency range Δf, so that it obeys the relation

$$v(t)^2 = 4kTR\Delta f \quad (1.10)$$

where
k is the Boltzmann constant (1.38×10^{-23} J/K)
T is absolute temperature

This relation is known as Nyquist's theorem.

1.4.5 1/*f* Noise

1/*f* noise is a kind of noise characterized by its power spectrum so that the power density is inversely proportional to the frequency in the lower frequency range. The 1/*f* noise may have different origins. When a current is passing through a semiconductor device, the 1/*f* noise is caused by the fluctuation of carriers in the semiconductor. This type of noise is also called flicker noise. Flicker noise is also generated in a resistor when a current is flowing through it.

In many unstable quantities, which may change over a long time interval up to days, months, or even years, it is often observed that the power density of the fluctuation in the quantity is almost inversely proportional to the frequency, and thus it is considered to be the 1/*f* noise. The drift is also considered as a very low frequency component of the 1/*f* noise.

Fluctuations in which the power density is inversely proportional to the frequency are found in physiological quantities such as fluctuation in the heart rate, which is considered to reflect physiological activities (Musha et al. 1983). If this measurement is used to study physiological activities, it cannot be considered as a noise but rather as a signal.

1.4.6 Interference

Interference is a kind of noise caused by physical or chemical events outside the object and the measurement system. The interference is sometimes caused by natural phenomenon such as lighting, but more commonly by artificial sources. The power line often causes interference by electromagnetic coupling to the object of measurement and for the measurement system. Sources other than electromagnetic equipments can cause interference. For example, fluorescent lamps may cause noise in optical measurement systems.

The power spectrum of noise due to interference depends on the source. The power line frequency (50/60 Hz) and its harmonics often appear in the power spectrum of the noise, when electric equipments supplied from the power line are used near the object and for the measurement system. Electric equipments which utilize pulse or switching operations generate noise in a wider frequency range. Machines having mechanical moving parts may generate vibration which interferes with mechanical measurement systems. The power spectrum of the mechanical interference may have peaks corresponding to mechanical resonance frequencies of the machine itself or materials excited mechanically.

Details of actual situations of interference and practical techniques for reducing noise can be found in appropriate textbooks (e.g., Morrison 1991).

1.4.7 Artifact

The term artifact usually refers to a component of noise superimposed on the object quantity and caused by external influences such as movement. The motion artifact often appears in biopotential measurement when using skin surface electrodes. It is considered that this is due in part to the potential generated by the epidermal layers of the skin and in part to the change in electrode potentials generated at the interfaces of the electrolyte and the metal. The waveform of the artifact depends on the nature of the external influence. Sometimes, the motion artifact resembles biopotential signals such as EEG and ECG, and thus it is difficult to remove the noise from the signal by simple methods such as the use of a band-pass filter.

Artifacts can be reduced by suppressing the process which couples the interference to the object or the measurement system. For example, motion artifacts at the recording electrodes can be reduced by the use of nonpolarizing electrodes. While the electrode potential varies due to the variation of the ionic concentration near the electrode surface, that variation is suppressed in nonpolarizing electrodes by employing adequate dissociation equilibrium. The abrasion or puncture of the epidermal part of the skin also reduces motion artifacts; as a result the potential developed in the skin is suppressed by short-circuiting the stratum corneum (Tam and Webster 1977, Ödman 1981).

1.5 Characteristics of the Measurement System

1.5.1 Transducer and Measurement System

In a measurement procedure, the observer obtains information about the object by using an appropriate measurement system. Usually, a measurement system consists of the sensor and the electronic instrument, as shown in Figure 1.1. The physical or chemical quantity which characterizes the object is detected by a sensor, and is converted into an electric quantity, which is displayed by using an appropriate electronic instrument which will transfer obtained results to the observer.

Sometimes, measurements require active procedures to be applied to the object such as excitation, illumination, irradiation, stimulation, application, or injection. Such procedures are considered a part of the measurement process and are provided by the sensor or another part of the measurement system.

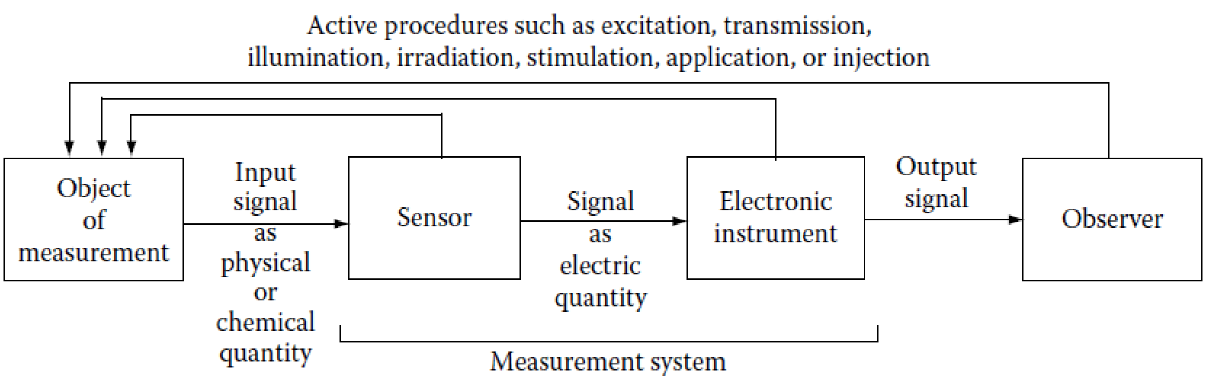

Figure 1.1 The general structure of a measurement system.

Although an active procedure is unavoidable in some measurements, its influence on the object should be minimized for two reasons, i.e., to minimize hazard and to minimize the change in the object quantity due to the active procedure. On the other hand, a situation often occurs in which measurement becomes much easier and more accurate if the applied energy or strength of the active procedure is increased. The level of the active procedure should then be determined as a compromise between minimized influence on the object and maximized performance of the measurement system.

The sensor is an essential part of the measurement system, because the quality of the measurement system is determined mostly by the sensor used. Signal-to-noise ratio, for example, is always determined mainly by the sensor, as long as adequately interfaced electronic circuits are employed.

Different types of object quantities require different kinds of sensors. Also different kinds of sensors are required according to the requirements of different measurement situations, such as different signal amplitudes and frequency ranges, accuracy requirements, limitations in size, shape or materials, or invasiveness of the measurement procedure. In biomedical measurements, sensors designed for other purposes are often unsuitable even if fundamental characteristics such as type of object quantities, measurement ranges, or frequency responses are acceptable. Actually, most sensors used in biomedical measurements are designed so that they can be applied to the body with minimum subsidiary effects, obtaining desired biological information correctly. Most parts of this book concern such sensors that are applicable to biomedical measurements.

1.5.2 Static Characteristics

In most measurement systems, the output of the measurement system at every moment can be fully determined by its input as the object quantity at that moment, if the change of the object quantity is slow enough. In such a situation, the input–output relationship of the measurement system can be uniquely determined without depending on the time course. The object quantity and the characteristics that represent the relation between the object quantity and the output of the measurement system are called the static characteristics. Fundamental features of a measurement system can be characterized by the static characteristics.

1.5.2.1 Sensitivity, Resolution, and Reproducibility

The term *sensitivity* is always used in such a way that the sensitivity of a sensor or a measurement system is high when a small change in object quantity causes a large change in its output. But quantitative definition of sensitivity is not unique. In some fields, the sensitivity is defined as the ratio of the output to the input. In this definition, the numerical value that represents the sensitivity is large when the sensitivity is high. In other fields, the sensitivity is defined as the ratio of the input to the output. This factor corresponds to the amount of change in the object quantity that produces a unit change in the output. By this definition, the numerical value is small when the sensitivity is high. The sensitivity has a dimension when the dimension of the object quantity and that of the output are different. Sensitivities for different object's quantities are represented in different units such as mV/kPa, μA/K, mV/pH, etc.

Sensitivity can be a constant value when the change in the output is linearly related to the change in the object quantity, whereas it cannot be constant when the response is nonlinear. In such a case, sensitivity depends on the absolute value of the object quantity.

The *resolution* is the least value of the object quantity that can be distinguished at the output of the measurement system. A change in the object quantity, which is smaller than the resolution of the measurement system, will not produce a detectable change in its output which can be distinguished from noise. The numerical value of the resolution is small when the resolution is high. The resolution has the same dimension as that of the object quantity.

The *reproducibility* describes how close to one another repeated outputs are when the same quantity is measured repeatedly. Quantitatively, the reproducibility of a measurement system is defined as a range in the object quantity so that the results of successive measurements for the same quantity fall into that range with a given probability. If the probability level is not specified, it is usually understood to be 95%. When the range is narrow, the reproducibility is high. The term repeatability is also used to express the similar concept of reproducibility, but repeatability is understood as the reproducibility in a short time interval, when these two terms are distinguished.

1.5.2.2 Measurement Range

The *measurement range* is the total range of the object quantity within which the measurement system works so as to meet the nominal performance of the measurement system. Thus the measurement range depends on performance requirements such as sensitivity, resolution, or reproducibility. If the requirements are severe, the measurement range will be narrow. Sometimes different measurement ranges are specified for different requirements. For example, in a thermometer, measurement range is from 30°C to 40°C for reproducibility of 0.1°C, and that from 0°C to 50°C for reproducibility of 0.5°C.

The measurement range states the maximum allowable change of the object quantity as long as the nominal performance of the measurement system is expected. On the other hand, the minimum detectable change of the object quantity is stated by the resolution. The ratio of measurement range to resolution is called the *dynamic range*. The dynamic range is a nondimensional value and is sometimes expressed in decibels (dB).

The dynamic range has to be considered when the signal is converted into a digital quantity. For example, the number of bits of an analog-to-digital converter and the data format or number of digits of the digital display are determined so that the maximum digital number usable by these devices is large enough compared with the dynamic range.

1.5.2.3 Linearity or Nonlinearity

The *linearity* describes how close the input–output relationship of the measurement system is to an appropriate straight line. Different definitions of linearity are used depending on which straight line is considered. The straight line defined by the least square fit to the input–output relation can be used, whereas other straight lines defined by the least square fit positioned to pass either the origin, or the terminal point, or both, can also be used. When the straight line passing through the origin is used, the linearity of this specific definition is sometimes called zero-based linearity or proportionality.

As a quantitative measure of the linearity, the ratio of the maximum deviation from the input–output relation curve from a straight line is used. However, the term nonlinearity is conventionally used to indicate this value, because the numerical value when using this definition is large when the deviation of the input–output relation curve from the straight line is significant.

When the linearity is high (or the nonlinearity is small), the input–output relationship can be regarded as a straight line, and thus the sensitivity can be regarded as constant. On the other hand, when the linearity is low (or nonlinearity is large), the sensitivity depends on the input level.

Although higher linearity is desired in most measurement systems, accurate measurement is possible even if the response is nonlinear as long as the input–output relation is fully determined. By using a computer, the input can be estimated at every sampling interval from the output and knowledge of the input–output relation.

1.5.2.4 Hysteresis

Hysteresis is a kind of phenomenon in which different output values appear corresponding to the same input no matter how slow the speed of change of the input is. If a measurement system has hysteresis, the input–output relation curve is not unique but depends on the direction and the range of successive input values. There are different causes of hysteresis, such as backlash in mechanical coupling parts, viscoelasticity or creep of mechanical elements, magnetization of ferromagnetic materials, or adsorption and desorption of materials on electrochemical devices. While hysteresis due to backlash is independent of the range of variations of the input, hysteresis because of other causes depends on the variations of the input, so that large variations cause large hysteresis. To minimize hysteresis, larger inputs appearing at a transient, or by the artifacts beyond the normal variation range of the signal, should be avoided. The sensor design in which the input is limited within the normal measurement range is advantageous not only to protect sensing elements in the sensor from destruction but also to reduce hysteresis. For example, the stopper for the diaphragm or beam in mechanical sensors is employed for this reason.

1.5.3 Dynamic Characteristics

The *dynamic characteristics* of a measurement system describe the input–output relation in a transient, whereas the static characteristics describe the relation when the input remains constant or changes slowly. The dynamic characteristics are required where the response to time-varying inputs is of concern. The time-varying pattern of the object quantity is observed as a waveform, but true waveforms will not be reproduced unless its dynamic characteristics are excellent.

The dynamic characteristics are particularly important when the sensor consists of a part of a control system. Instability or oscillation may occur due to a poor dynamic response of the sensor. The most common cause that affects the dynamic characteristics of the measurement system is the presence of elements which store and release energy when the object quantity varies. For example, inertial elements such as masses, capacitances and inductances, and compliant elements such as springs, electric and heat capacitances are such system components. If the displacements of mechanical parts and fluids cause significant time delays, they will also affect the dynamic characteristics of the system. Besides the measurement system, the object matter of the measurement and the interfacing media may also affect the dynamic characteristics of the whole measurement process. In such a situation, the dynamic characteristics should be discussed, including the object or the interfacing media, as in the case of the catheter-sensor systems in pressure measurements (see Section 2.2.2).

Although some important concepts and terms regarding the dynamic characteristics of the measurement systems are explained briefly in the following sections, we also recommend the study of other textbooks (e.g., Doebelin 1990, Pallas-Areny and Webster 1991).

1.5.3.1 Linear and Nonlinear Systems

The term *linear system*, or the expression that a system is linear, always refers to a system in which the response to simultaneous inputs is the sum of their independent inputs. A system which does not meet this condition is called a nonlinear system.

In linear systems, the dynamic characteristics are the same regardless of the input's amplitude. The amplitude of the response is simply proportional to the input amplitude, because a large input can be regarded as a sum of smaller inputs, and hence the response corresponding to the large input is the sum of the responses corresponding to the small inputs. This property is important because many convenient parameters, which characterize the system, can be defined regardless of the signal amplitude.

Real systems cannot always be linear when the input increases far beyond the measurement range. On the other hand, most measurement systems can be regarded as linear if the variations of the input are small. Even in a nonlinear system, an appropriate linear system can be assumed as a result of the approximation in a small measurement range.

In a linear system, the response to a sinusoidal input is also sinusoidal with the same frequency as that of the input, while other frequency components such as higher harmonics may appear in a nonlinear system.

1.5.3.2 Frequency Response

The *frequency response* refers to the distribution of the amplitude and the phase shift of the output to sinusoidal inputs of unit amplitude over the whole frequency range in which the dynamic characteristics are considered. Usually, the frequency response is defined only for linear systems.

The output of a linear system can be described as the sum of the responses corresponding to sinusoidal inputs having different frequencies, because the input is expressed as the sum of sinusoidal functions such as Equation 1.1 or 1.5. Therefore, the frequency response provides complete information about the output of the system for any input.

When the input–output relation of a system is described by a constant-coefficient first-order differential equation, the system is called a *first-order system*. The differential equation which describes a first-order system is written as

$$a_1 \frac{dy(t)}{dt} + a_0 y(t) = x(t) \qquad (1.11)$$

where

$x(t)$ and $y(t)$ are the input and the output of the system

$a0$ and $a1$ are constants

Then the frequency response of this system can be represented as shown in Figure 1.2a and b, where *fc* is given as *a*0/2π*a*1 and called the *cut-off frequency*.

The *second-order system* is a system which can be described by a second-order constant-coefficient differential equation written as

$$a_2 \frac{d^2 y(t)}{dt^2} + a_1 \frac{dy(t)}{dt} + a_0 y(t) = x(t) \qquad (1.12)$$

where *a*0, *a*1, and *a*2 are constants. The frequency response of this system can be represented as shown in Figure 1.2c and d, where

$$f_0 = \frac{1}{2\pi}\sqrt{\frac{a_0}{a_2}} \qquad (1.13)$$

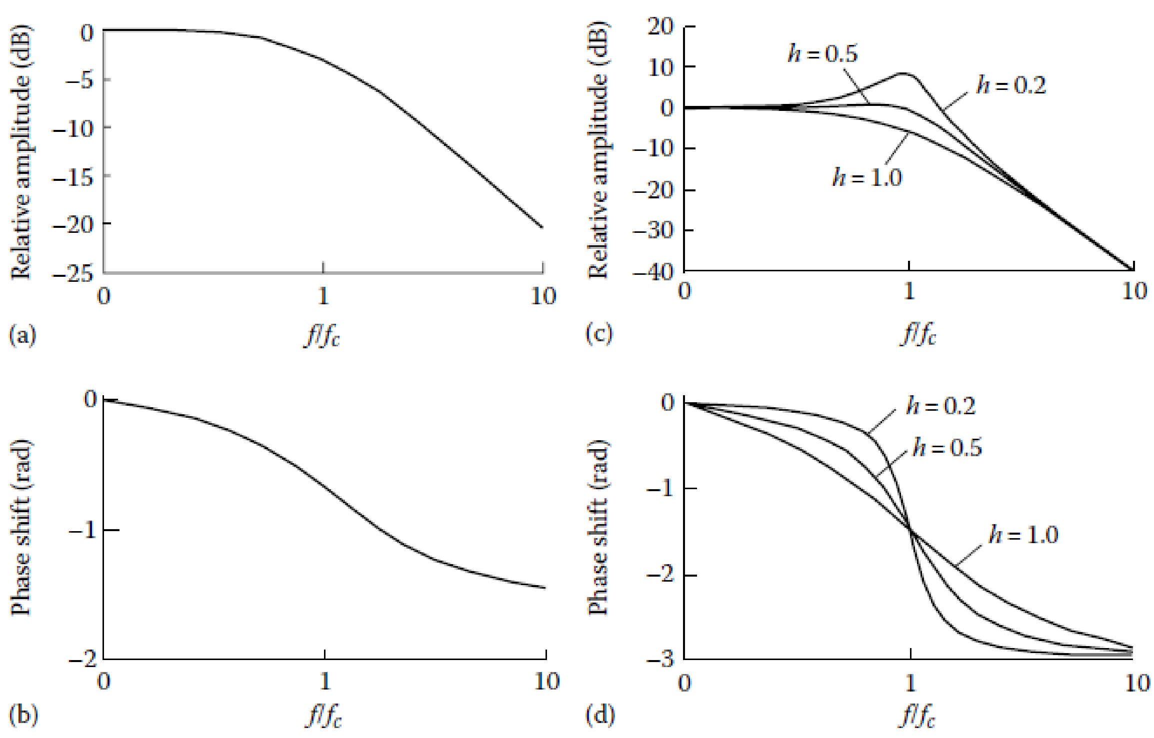

Figure 1.2 Frequency responses of first- and second-order linear systems. Amplitude and phase responses of the first-order system (a) and (b), and those of the second-order system (c) and (d).

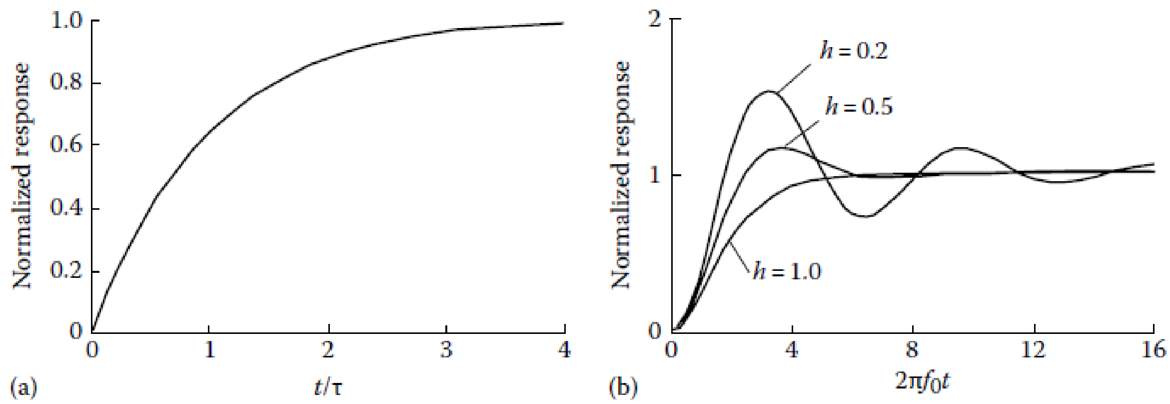

Figure 1.3 Responses to a step input in the first-order system (a) and that of the second-order system (b).

which is called the *natural frequency*, and

$$h = \frac{a_1}{2\sqrt{a_0 a_2}} \tag{1.14}$$

which is called the *damping coefficient*.

1.5.3.3 Time Constant, Response Time, Rise Time, and Settling Time

When the input of a system changes abruptly from one level to another, the behavior of the output can be characterized by some specific parameters according to the type of the system. Such parameters can be determined by the response to a unit step input in which the input is zero before a specific time, and unity after that time.

The *time constant* is defined in the first-order system. As shown in Figure 1.3a, the response of a first-order system to a unit step input is a process approaching exponentially to the final value, and the time constant τ is defined as the time required for the output to reach to 1 – 1/e ≈ 0.673 of the final value, and is given as $a1/a0$ for the system represented by the Equation 1.11.

In the second-order system, the response to a unit step input varies with the damping coefficient as shown in Figure 1.3b. Some parameters are used to express how quickly the system can follow the input. The *response time* is usually defined as the time needed to reach 90% of the final value, and the rise time is usually defined as the time that the output changes from 10% to 90% of the final value to a unit step input. The *settling time* is defined as the time required for the output to settle within a definite range near the final value, for example, the range is defined as ±5% of the final value to a unit step input.

1.6 Standard and Calibration

Any measurement system can be calibrated by comparing with either an intrinsic standard or a reliable standard instrument. There are some convenient standards for different quantities. For example, a mercury column can be a standard of pressure. Because the density of pure mercury is known, and the gravity of the earth is also known, the pressure developed at the bottom of a mercury column of a definite height can be a standard of pressure. Ice point of pure water, which is 0°C, or melting point of gallium, which is 29.771°C, are used as intrinsic temperature standards.

An instrument which is stable enough and correctly calibrated can also be a standard, so that other measurement systems can be calibrated comparing with the standard instrument. For example, commercial standard thermometers having crystal-resonator temperature sensors can be used to determine temperatures absolutely within a deviation of 0.01°C, which is precise enough for most biomedical measurements.

When the measurement system to be calibrated is nonlinear, calibration has to be performed at many points in the measurement range. On the contrary, when the measurement system is linear, calibration at two points is enough. Even if slight nonlinearity exists in the whole measurement range of the system, as long as the variation of the object quantity is limited to a narrow range and the system can be regarded as linear in that limited range, two-point calibration in that range is enough for all practical purposes. In a case where the measurement system is linear and its sensitivity is stable but drift remains to some extent, occasional one-point calibration after initial two-point calibration is recommended.

Even if a measurement system is nonlinear but is stable enough and its input–output relation can be approximated by a simple formula having few parameters, what is required in calibration is to determine all parameters in the formula. For example, if the input–output relation is well approximated by a quadratic formula, the number of parameters to be determined is three, and thus the fitting curve can be determined by three-point calibration. Such a curve-fitting computation can be performed easily even in real time when the output of the measurement system is connected to a computer.

1.7 Accuracy and Error

The term *accuracy* describes how close the measured value is to the true value. The difference between them is termed the *error*.

The error may depend on the level of the object quantity, especially when the measurement range includes small to large values of the object quantity. The error may be small when the object quantity is small, while it may be large when the object quantity is large. In such a situation, the relative error, which is defined as the ratio of the error to the true value, can be a convenient figure of the performance of the measurement system.

When the measurement system is calibrated adequately, the error will be reduced to a limit which is determined by the reproducibility. Sometimes, the accuracy of the measurement system becomes poor due to the drift in a long period of time even though the reproducibility in a short period of time has not changed very much. In such a situation, the accuracy can be recovered to the initial level by a recalibration. By repeated calibration procedures, the accuracy can be maintained within a definite range for a long time period.