978-0-521-85455-9 - Continuous and Discrete Time Signals and Systems
Mrinal Mandal and Amir Asif

PART I

Introduction to signals and systems

Cambridge University Press
978-0-521-85455-9 - Continuous and Discrete Time Signals and Systems
Mrinal Mandal and Amir Asif
Excerpt
More information

CHAPTER

1 Introduction to signals

Signals are detectable quantities used to convey information about time-varying physical phenomena. Common examples of signals are human speech, temperature, pressure, and stock prices. Electrical signals, normally expressed in the form of voltage or current waveforms, are some of the easiest signals to generate and process.

Mathematically, signals are modeled as functions of one or more independent variables. Examples of independent variables used to represent signals are time, frequency, or spatial coordinates. Before introducing the mathematical notation used to represent signals, let us consider a few physical systems associated with the generation of signals. Figure 1.1 illustrates some common signals and systems encountered in different fields of engineering, with the physical systems represented in the left-hand column and the associated signals included in the right-hand column. Figure 1.1(a) is a simple electrical circuit consisting of three passive components: a capacitor C, an inductor L, and a resistor R. A voltage $v(t)$ is applied at the input of the RLC circuit, which produces an output voltage $y(t)$ across the capacitor. A possible waveform for $y(t)$ is the sinusoidal signal shown in Fig. 1.1(b). The notations $v(t)$ and $y(t)$ includes both the dependent variable, v and y, respectively, in the two expressions, and the independent variable t. The notation $v(t)$ implies that the voltage v is a function of time t. Figure 1.1(c) shows an audio recording system where the input signal is an audio or a speech waveform. The function of the audio recording system is to convert the audio signal into an electrical waveform, which is recorded on a magnetic tape or a compact disc. A possible resulting waveform for the recorded electrical signal is shown in Fig 1.1(d). Figure 1.1(e) shows a charge coupled device (CCD) based digital camera where the input signal is the light emitted from a scene. The incident light charges a CCD panel located inside the camera, thereby storing the external scene in terms of the spatial variations of the charges on the CCD panel. Figure 1.1(g) illustrates a thermometer that measures the ambient temperature of its environment. Electronic thermometers typically use a *therm*al res*istor*, known as a *thermistor*, whose resistance varies with temperature. The fluctuations in the resistance are used to measure the temperature. Figure 1.1(h)

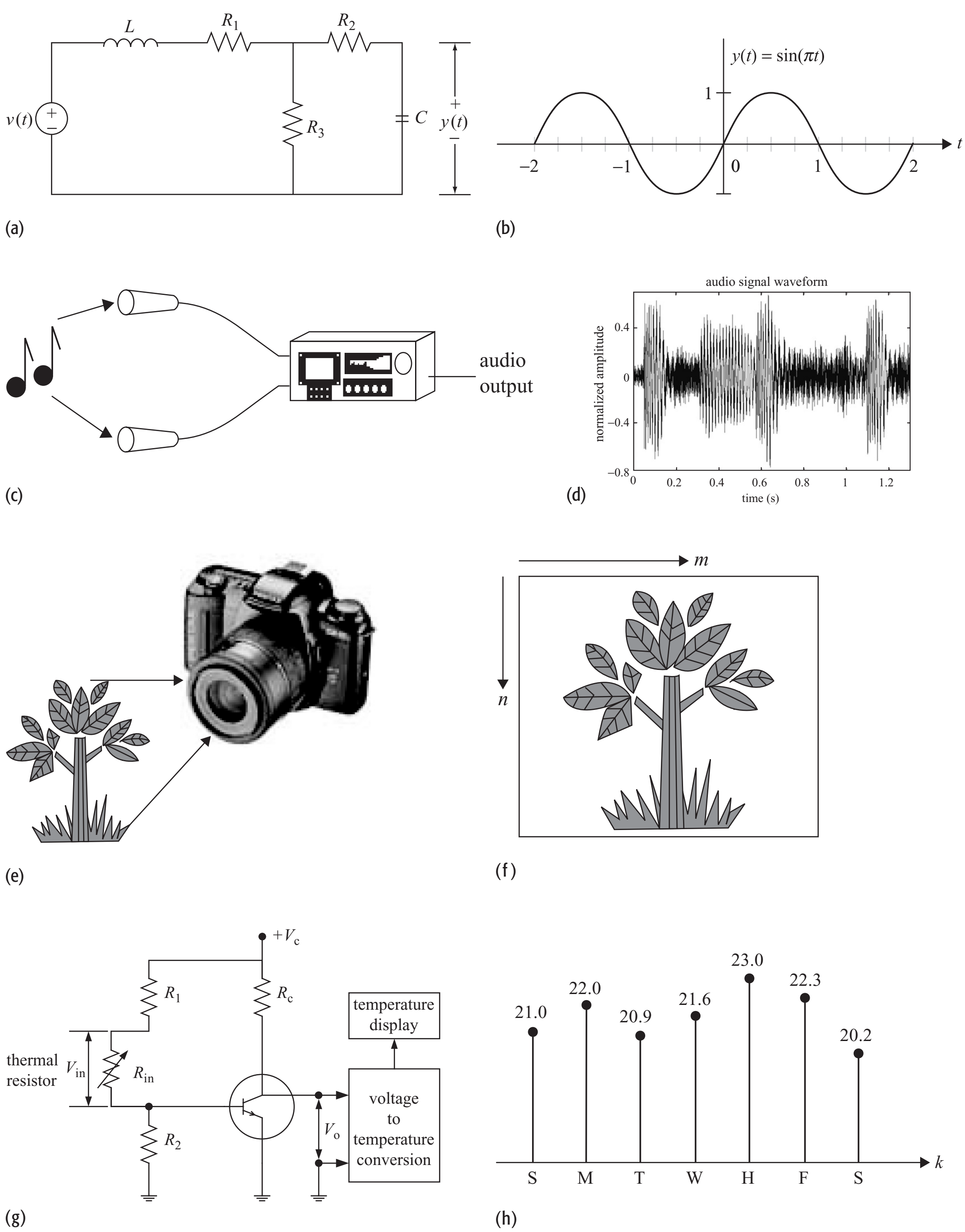

Fig. 1.1. Examples of signals and systems. (a) An electrical circuit; (c) an audio recording system; (e) a digital camera; and (g) a digital thermometer. Plots (b), (d), (f), and (h) are output signals generated, respectively, by the systems shown in (a), (c), (e), and (g).

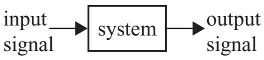

Fig. 1.2. Processing of a signal by a system.

plots the readings of the thermometer as a function of discrete time. In the aforementioned examples of Fig. 1.1, the RLC circuit, audio recorder, CCD camera, and thermometer represent different systems, while the information-bearing waveforms, such as the voltage, audio, charges, and fluctuations in resistance, represent signals. The output waveforms, for example the voltage in the case of the electrical circuit, current for the microphone, and the fluctuations in the resistance for the thermometer, vary with respect to only one variable (time) and are classified as one-dimensional (1D) signals. On the other hand, the charge distribution in the CCD panel of the camera varies spatially in two dimensions. The independent variables are the two spatial coordinates (m, n). The charge distribution signal is therefore classified as a two-dimensional (2D) signal.

The examples shown in Fig. 1.1 illustrate that typically every system has one or more signals associated with it. A *system* is therefore defined as an entity that processes a set of signals (called the *input* signals) and produces another set of signals (called the *output* signals). The voltage source in Fig. 1.1(a), the audio sound in Fig. 1.1(c), the light entering the camera in Fig. 1.1(e), and the ambient heat in Fig. 1.1(g) provide examples of the input signals. The voltage across capacitor C in Fig. 1.1(b), the voltage generated by the microphone in Fig. 1.1(d), the charge stored on the CCD panel of the digital camera, displayed as an image in Fig. 1.1(f), and the voltage generated by the thermistor, used to measure the room temperature, in Fig. 1.1(h) are examples of output signals.

Figure 1.2 shows a simplified schematic representation of a signal processing system. The system shown processes an input signal $x(t)$ producing an output $y(t)$. This model may be used to represent a range of physical processes including electrical circuits, mechanical devices, hydraulic systems, and computer algorithms with a single input and a single output. More complex systems have multiple inputs and multiple outputs (MIMO).

Despite the wide scope of signals and systems, there is a set of fundamental principles that control the operation of these systems. Understanding these basic principles is important in order to analyze, design, and develop new systems. The main focus of the text is to present the theories and principles used in signals and systems. To keep the presentations simple, we focus primarily on signals with one independent variable (usually the time variable denoted by t or k), and systems with a single input and a single output. The theories that we develop for single-input, single-output systems are, however, generalizable to multidimensional signals and systems with multiple inputs and outputs.

1.1 Classification of signals

A signal is classified into several categories depending upon the criteria used for its classification. In this section, we cover the following categories for signals:

(i) continuous-time and discrete-time signals;
(ii) analog and digital signals;
(iii) periodic and aperiodic (or nonperiodic) signals;
(iv) energy and power signals;
(v) deterministic and probabilistic signals;
(vi) even and odd signals.

1.1.1 Continuous-time and discrete-time signals

If a signal is defined for all values of the independent variable t, it is called a *continuous-time* (CT) signal. Consider the signals shown in Figs. 1.1(b) and (d). Since these signals vary continuously with time t and have known magnitudes for all time instants, they are classified as CT signals. On the other hand, if a signal is defined only at discrete values of time, it is called a *discrete-time* (DT) signal. Figure 1.1(h) shows the output temperature of a room measured at the same hour every day for one week. No information is available for the temperature in between the daily readings. Figure 1.1(h) is therefore an example of a DT signal. In our notation, a CT signal is denoted by $x(t)$ with regular parenthesis, and a DT signal is denoted with square parenthesis as follows:

$$x[kT], \quad k = 0, \pm 1, \pm 2, \pm 3, \ldots,$$

where T denotes the time interval between two consecutive samples. In the example of Fig. 1.1(h), the value of T is one day. To keep the notation simple, we denote a one-dimensional (1D) DT signal x by $x[k]$. Though the sampling interval is not explicitly included in $x[k]$, it will be incorporated if and when required.

Note that all DT signals are not functions of time. Figure 1.1(f), for example, shows the output of a CCD camera, where the discrete output varies spatially in two dimensions. Here, the independent variables are denoted by (m, n), where m and n are the discretized horizontal and vertical coordinates of the picture element. In this case, the two-dimensional (2D) DT signal representing the spatial charge is denoted by $x[m, n]$.

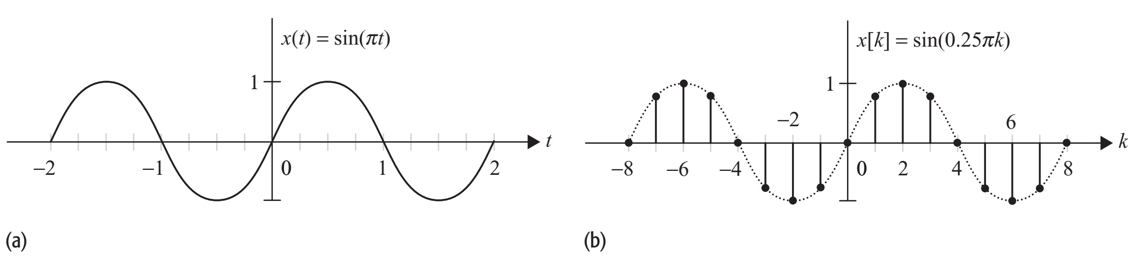

Fig. 1.3. (a) CT sinusoidal signal $x(t)$ specified in Example 1.1; (b) DT sinusoidal signal $x[k]$ obtained by discretizing $x(t)$ with a sampling interval $T = 0.25$ s.

Example 1.1

Consider the CT signal $x(t) = \sin(\pi t)$ plotted in Fig. 1.3(a) as a function of time t. Discretize the signal using a sampling interval of $T = 0.25$ s, and sketch the waveform of the resulting DT sequence for the range $-8 \leq k \leq 8$.

Solution

By substituting $t = kT$, the DT representation of the CT signal $x(t)$ is given by

$$x[kT] = \sin(\pi k \times T) = \sin(0.25\pi k).$$

For $k = 0, \pm 1, \pm 2, \ldots,$ the DT signal $x[k]$ has the following values:

$$x[-8] = x(-8T) = \sin(-2\pi) = 0, \qquad x[1] = x(T) = \sin(0.25\pi) = \frac{1}{\sqrt{2}},$$

$$x[-7] = x(-7T) = \sin(-1.75\pi) = \frac{1}{\sqrt{2}}, \qquad x[2] = x(2T) = \sin(0.5\pi) = 1,$$

$$x[-6] = x(-6T) = \sin(-1.5\pi) = 1, \qquad x[3] = x(3T) = \sin(0.75\pi) = \frac{1}{\sqrt{2}},$$

$$x[-5] = x(-5T) = \sin(-1.25\pi) = \frac{1}{\sqrt{2}}, \qquad x[4] = x(4T) = \sin(\pi) = 0,$$

$$x[-4] = x(-4T) = \sin(-\pi) = 0, \qquad x[5] = x(5T) = \sin(1.25\pi) = -\frac{1}{\sqrt{2}},$$

$$x[-3] = x(-3T) = \sin(-0.75\pi) = -\frac{1}{\sqrt{2}}, \qquad x[6] = x(6T) = \sin(1.5\pi) = -1,$$

$$x[-2] = x(-2T) = \sin(-0.5\pi) = -1, \qquad x[7] = x(7T) = \sin(1.75\pi) = -\frac{1}{\sqrt{2}},$$

$$x[-1] = x(-T) = \sin(-0.25\pi) = -\frac{1}{\sqrt{2}}, \qquad x[8] = x(8T) = \sin(2\pi) = 0,$$

$$x[0] = x(0) = \sin(0) = 0.$$

Plotted as a function of k, the waveform for the DT signal $x[k]$ is shown in Fig. 1.3(b), where for reference the original CT waveform is plotted with a dotted line. We will refer to a DT plot illustrated in Fig. 1.3(b) as a *bar* or a *stem* plot to distinguish it from the CT plot of $x(t)$, which will be referred to as a *line* plot.

Example 1.2

Consider the rectangular pulse plotted in Fig. 1.4. Mathematically, the rectangular pulse is denoted by

$$x(t) = \text{rect}\left(\frac{t}{\tau}\right) = \begin{cases} 1 & |t| \leq \tau/2 \\ 0 & |t| > \tau/2. \end{cases}$$

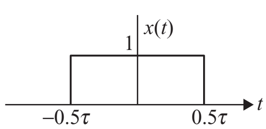

Fig. 1.4. Waveform for CT rectangular function. It may be noted that the rectangular function is discontinuous at $t = \pm\tau/2$.

From the waveform in Fig. 1.4, it is clear that $x(t)$ is continuous in time but has discontinuities in magnitude at time instants $t = \pm 0.5\tau$. At $t = 0.5\tau$, for example, the rectangular pulse has two values: 0 and 1. A possible way to avoid this ambiguity in specifying the magnitude is to state the values of the signal $x(t)$ at $t = 0.5\tau^-$ and $t = 0.5\tau^+$, i.e. immediately before and after the discontinuity. Mathematically, the time instant $t = 0.5\tau^-$ is defined as $t = 0.5\tau - \varepsilon$, where ε is an infinitely small positive number that is close to zero. Similarly, the time instant $t = 0.5\tau^+$ is defined as $t = 0.5\tau + \varepsilon$. The value of the rectangular pulse at the discontinuity $t = 0.5\tau$ is, therefore, specified by $x(0.5\tau^-) = 1$ and $x(0.5\tau^+) = 0$. Likewise, the value of the rectangular pulse at its other discontinuity $t = -0.5\tau$ is specified by $x(-0.5\tau^-) = 0$ and $x(-0.5\tau^+) = 1$.

A CT signal that is continuous for all t except for a finite number of instants is referred to as a piecewise CT signal. The value of a piecewise CT signal at the point of discontinuity t_1 can either be specified by our earlier notation, described in the previous paragraph, or, alternatively, using the following relationship:

$$x(t_1) = 0.5\left[x(t_1^+) + x(t_1^-)\right]. \tag{1.1}$$

Equation (1.1) shows that $x(\pm 0.5\tau) = 0.5$ at the points of discontinuity $t = \pm 0.5\tau$. The second approach is useful in certain applications. For instance, when a piecewise CT signal is reconstructed from an infinite series (such as the Fourier series defined later in the text), the reconstructed value at the point of discontinuity satisfies Eq. (1.1). Discussion of piecewise CT signals is continued in Chapter 4, where we define the CT Fourier series.

1.1.2 Analog and digital signals

A second classification of signals is based on their amplitudes. The amplitudes of many real-world signals, such as voltage, current, temperature, and pressure, change continuously, and these signals are called *analog* signals. For example, the ambient temperature of a house is an analog number that requires an infinite number of digits (e.g., 24.763 578. . .) to record the readings precisely. Digital signals, on the other hand, can only have a finite number of amplitude values. For example, if a digital thermometer, with a resolution of 1 °C and a range of [10 °C, 30 °C], is used to measure the room temperature at discrete time instants, $t = kT$, then the recordings constitute a digital signal. An example of a digital signal was shown in Fig. 1.1(h), which plots the temperature readings taken once a day for one week. This digital signal has an amplitude resolution of 0.1 °C, and a sampling interval of one day.

Figure 1.5 shows an analog signal with its digital approximation. The analog signal has a limited dynamic range between [−1, 1] but can assume any real value (rational or irrational) within this dynamic range. If the analog signal is sampled at time instants $t = kT$ and the magnitude of the resulting samples are quantized to a set of finite number of known values within the range [−1, 1], the resulting signal becomes a digital signal. Using the following set of eight uniformly distributed values,

$$[-0.875, -0.625, -0.375, -0.125, 0.125, 0.375, 0.625, 0.875],$$

within the range [−1, 1], the best approximation of the analog signal is the digital signal shown with the stem plot in Fig. 1.5.

Another example of a digital signal is the music recorded on an audio compact disc (CD). On a CD, the music signal is first sampled at a rate of 44 100

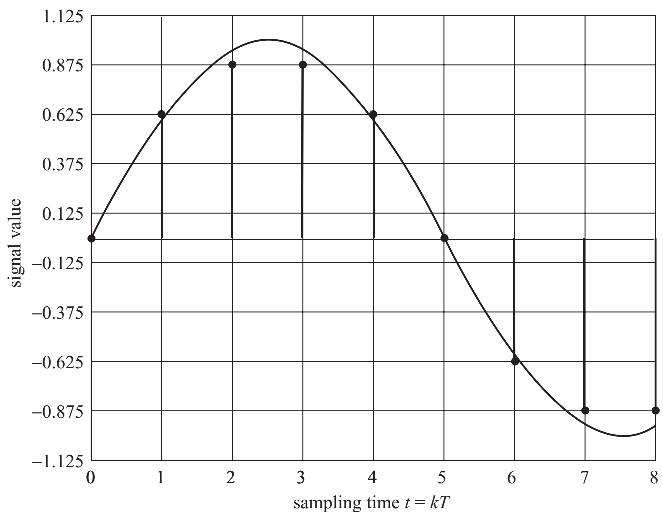

Fig. 1.5. Analog signal with its digital approximation. The waveform for the analog signal is shown with a line plot; the quantized digital approximation is shown with a stem plot.

samples per second. The sampling interval T is given by 1/44 100, or 22.68 microseconds (μs). Each sample is then quantized with a 16-bit uniform quantizer. In other words, a sample of the recorded music signal is approximated from a set of uniformly distributed values that can be represented by a 16-bit binary number. The total number of values in the discretized set is therefore limited to 2^{16} entries.

Digital signals may also occur naturally. For example, the price of a commodity is a multiple of the lowest denomination of a currency. The grades of students on a course are also discrete, e.g. 8 out of 10, or 3.6 out of 4 on a 4-point grade point average (GPA). The number of employees in an organization is a non-negative integer and is also digital by nature.

1.1.3 Periodic and aperiodic signals

A CT signal $x(t)$ is said to be *periodic* if it satisfies the following property:

$$x(t) = x(t + T_0), \tag{1.2}$$

at all time t and for some positive constant T_0. The smallest positive value of T_0 that satisfies the periodicity condition, Eq. (1.2), is referred to as the *fundamental period* of $x(t)$.

Likewise, a DT signal $x[k]$ is said to be *periodic* if it satisfies

$$x[k] = x[k + K_0] \tag{1.3}$$

at all time k and for some positive constant K_0. The smallest positive value of K_0 that satisfies the periodicity condition, Eq. (1.3), is referred to as the fundamental period of $x[k]$. A signal that is not periodic is called an *aperiodic* or *non-periodic* signal. Figure 1.6 shows examples of both periodic and aperiodic

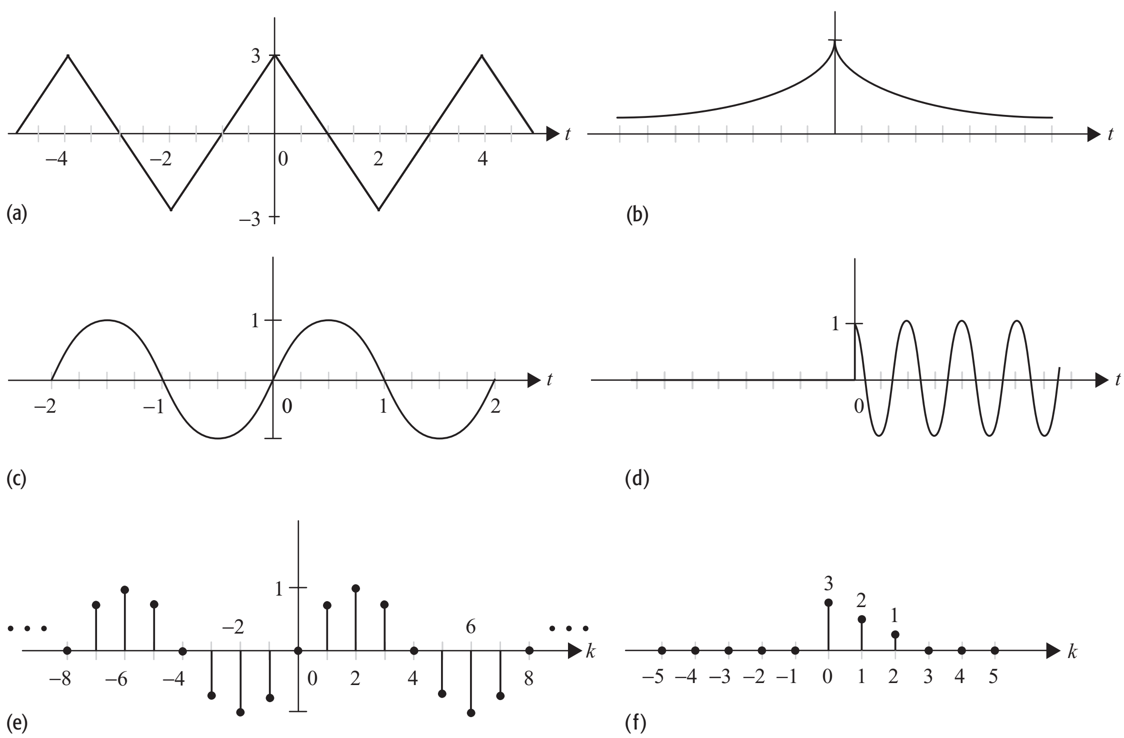

Fig. 1.6. Examples of periodic ((a), (c), and (e)) and aperiodic ((b), (d), and (f)) signals. The line plots (a) and (c) represent CT periodic signals with fundamental periods T_0 of 4 and 2, while the stem plot (e) represents a DT periodic signal with fundamental period $K_0 = 8$.

signals. The reciprocal of the fundamental period of a signal is called the *fundamental frequency*. Mathematically, the fundamental frequency is expressed as follows

$$f_0 = \frac{1}{T_0}, \text{ for CT signals,} \quad \text{or} \quad f_0 = \frac{1}{K_0}, \text{ for DT signals,} \tag{1.4}$$

where T_0 and K_0 are, respectively, the fundamental periods of the CT and DT signals. The frequency of a signal provides useful information regarding how fast the signal changes its amplitude. The unit of frequency is *cycles per second* (c/s) or *hertz* (Hz). Sometimes, we also use *radians per second* as a unit of frequency. Since there are 2π radians (or $360°$) in one cycle, a frequency of f_0 hertz is equivalent to $2\pi f_0$ radians per second. If radians per second is used as a unit of frequency, the frequency is referred to as the *angular frequency* and is given by

$$\omega_0 = \frac{2\pi}{T_0}, \text{ for CT signals,} \quad \text{or} \quad \Omega_0 = \frac{2\pi}{K_0}, \text{ for DT signals.} \tag{1.5}$$

A familiar example of a periodic signal is a sinusoidal function represented mathematically by the following expression:

$$x(t) = A \sin(\omega_0 t + \theta).$$