Chapter 2
Sensors Characteristics

Abstract In this chapter, static and dynamic characteristics of sensing systems will be presented. Their importance will be highlighted and their influence on the operation of sensing systems will be described.

2.1 Introduction

After receiving signals from a sensor, these signals need to be processed. The acceptable and accurate process of these signals requires: (a) full knowledge regarding the operation of the sensors and nature of signals, (b) *posteriori knowledge* regarding the received signals, and (c) information about the *dynamic* and *static characteristics* of the sensing systems.

(a) In order to be able to use signals' information correctly, the operation of a sensor, and the nature of signals they produce, should be well understood. By having this knowledge, we are able to choose the right tools for the acquisition of data from the sensor. For instance, if the sensor output is voltage, we utilize analogue-to-digital and sample and hold circuits, as well as a circuit that transfers the digits into the computer. If a sensor produces a time varying signal where the information is embedded in its frequency signatures, then a frequency counter and possibly a frequency analyzer are needed. If output of the sensor is a change in color then a visible spectrometer is necessary.
(b) A posteriori knowledge (a posteriori knowledge or justification is dependent on experience or empirical evidence) about the received signals is important in order to assure that the data will be interpreted correctly and that the right device is used in the measurement process. We need to have a good understanding for what is expected from the sensor and system. For instance, even during a simple DC voltage reading, if the DC input has been mixed with AC signals (may happen often due to the influence of unwanted electromagnetic waves), the measured value can be significantly different from the real measurand.

K. Kalantar-zadeh, *Sensors: An Introductory Course*,
DOI 10.1007/978-1-4614-5052-8_2, © Springer Science+Business Media New York 2013

In this system, the presence of unwanted AC signals can produce unrealistic and meaningless measurements. If knowledge regarding the presence of AC voltages were available (a posteriori knowledge), a filtering process could be efficiently used even before feeding the stimuli into the sensing system (e.g., electromagnetic shielding or filter to remove 50 Hz AC signals). Having knowledge about the characteristics of sensing systems also allows us to extract meaningful conclusions with minimal error. For example, we can avoid wrong readings at short time brackets; if we know that a gas sensor needs 5 min to respond to a target gas rather than 5 s.

(c) The characteristics of a sensor can be classified into two *static* and *dynamic* groups. Understanding the dynamic and static characteristics behaviors are imperative in correctly mapping the output versus input of a system (measurand). In the following sections, the static and dynamic characteristics will be defined and their importance in sensing systems will be illustrated.

2.2 Static Characteristics

Static characteristics are those that can be measured after all transient effects have been stabilized to their final or steady state values. Static characteristics relate to issues such as how a sensor's output change in response to an input change, how selective the sensor is, how external or internal interferences can affect its response, and how stable the operation of a sensing system can be.

Several of the most important static characteristics are as follows:

2.2.1 Accuracy

Accuracy of a sensing system represents the correctness of its output in comparison to the actual value of a measurand. To assess the accuracy, either the system is benchmarked against a standard measurand or the output is compared with a measurement system with a superior accuracy.

For instance considering a temperature sensing system, when the real temperature is 20.0 °C, the system is more accurate, if it shows 20.1 °C rather than 21.0 °C.

2.2.2 Precision

Precision represents capacity of a sensing system to give the same reading when repetitively measuring the same measurand under the same conditions. The precision is a statistical parameter and can be assessed by the standard deviation (or variance) of a set of readings of the system for similar inputs.

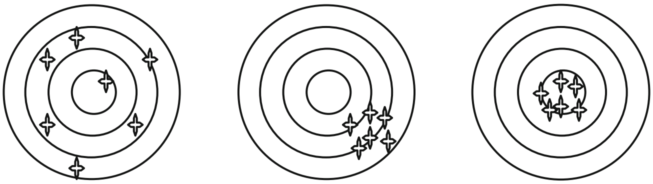

Fig. 2.1 The difference between accuracy and precision

For instance, a temperature sensing system is precise, if when the ambient temperature is 21.0 °C and it shows 22.0, 22.1, or 21.9 °C in three different consecutive measurements. It is not considered precise, if it shows 21.5, 21.0, and 20.5 °C although the measured values are closer to the actual temperature. The game of darts can be used as another good example of the difference between the accuracy and precision definitions (as can be seen in Fig. 2.1).

2.2.3 Repeatability

When all operating and environmental conditions remain constant, *repeatability* is the sensing system's ability to produce the same response for successive measurements. Repeatability is closely related to precision. Both long-term and short-term repeatability estimates can be important for a sensing system.

For a temperature sensing system, when ambient temperature remains constant at 21.0 °C, if the system shows 21.0, 21.1, and 21.0 °C in 1 min intervals, and shows 22.0, 22.1, and 22.2 °C after 1 h, in similar 1 min intervals, the system has a good short-term and poor long-term repeatability.

2.2.4 Reproducibility

Reproducibility is the sensing system's ability to produce the same responses after measurement conditions have been altered.

For example, if a temperature sensing system shows similar responses; over a long time period, or when readings are performed by different operators, or at different laboratories, the system is reproducible.

2.2.5 Stability

Stability is a sensing system's ability to produce the same output value when measuring the same measurand over a period of time.

2.2.6 Error

Error is the difference between the actual value of the measurand and the value produced by the sensing system. Error can be caused by a variety of internal and external sources and is closely related to accuracy. Accuracy can be related to *absolute* or *relative error* as:

$$\begin{aligned} \text{Absolute error} &= \text{Output} - \text{True value}, \\ \text{Relative error} &= \frac{\text{Output} - \text{True value}}{\text{True value}}. \end{aligned} \tag{2.1}$$

For instance, in a temperature sensing system, if temperature is 21 °C and the system shows 21.1 °C, then the absolute and relative errors are equal to 0.1 °C and 0.0047 °C, respectively. While the absolute error has the same unit as the measurand, the relative error is unitless.

Errors are produced by fluctuations in the output signal and can be *systematic* (e.g., drift or interferences from other systems) or *random* (e.g., random noise).

2.2.7 Noise

The unwanted fluctuations in the output signal of the sensing system, when the measurand is not changing, are referred to as *noise*. The standard deviation value of the noise strength is an important factor in measurements. The mean value of the signal divided by this value gives a good benchmark, as how readily the information can be extracted. As a result, signal-to-noise ratio (S/N) is a commonly used figure in sensing applications. It is defined as:

$$\frac{S}{N} = \frac{\text{Mean value of signal}}{\text{Standard deviation of noise}}. \tag{2.2}$$

Noise can be caused by either *internal* or *external* sources. Electromagnetic signals such as those produced by transmission/reception circuits and power supplies, mechanical vibrations, and ambient temperature changes are all examples of external noise, which can cause systematic error. However, the nature of internal noises is quite different and can be categorized as follows:

1. *Electronic noise*: Thermal energy causes charge carriers to move in random motions, which results in random variations of current and/or voltage. It is unavoidable and is present in all sensing systems operating at temperatures higher than 0 K.

   One of the most commonly seen electronic noises in electronic instruments is caused by the thermal agitation of careers, which is called *thermal noise*.

It produces charge inhomogeneties, which in turn create voltage fluctuations that appear in the output signal. Thermal noise exists even in the absence of current. The magnitude of a thermal noise in a resistance of magnitude R (Ω) is extracted from thermodynamic calculations and is equal to:

$$\bar{v}_{\mathrm{rms}} = \sqrt{4kTR\Delta f}, \tag{2.3}$$

in which $\bar{v}_{\mathrm{rms}}$ is the root-mean-square of noise voltage, which is generated by the frequency component with the bandwidth of Δf, k is the Boltzmann's constant, which is equal to 1.38×10^{-23} JK^{-1}, and T is the temperature in Kelvin.

Example 1. The rise and fall time of a sensor signal are generally inversely proportional to its bandwidth. Assume that the rise time of a thermistor response is 0.05 s and the relation between the rise time and the bandwidth is $\tau_{\mathrm{rise}} = 1/2\Delta f$. (A) Calculate the magnitude of the thermal noise. The ambient temperature is 27 °C and the thermistor resistance is 5 kΩ at this temperature. (B) What is the signal-to-noise ratio, if the average of current passing through the resistor is 0.2 mA?

Answer:

(a) The bandwidth is equal to $\Delta f = 1/2\tau_{\mathrm{rise}} = 1/2 \times 0.05$ (s) $= 10$ Hz and according to (2.3), the rms value of the thermal noise voltage is equal to:

$$\begin{aligned}\bar{v}_{\mathrm{rms}} &= \sqrt{4 \times 1.38 \times 10^{-23} \times 300\ (\mathrm{K}) \times 5{,}000\ (\Omega) \times 10\ (\mathrm{Hz})} \\ &= 2.88 \times 10^{-8}\ (\mathrm{V}) = 0.0288\ \mathrm{mV} \text{ or } 20\log(\bar{v}_{\mathrm{rms}}) = -150.8\ \mathrm{dB}.\end{aligned}$$

(b) Current of 0.2 mA generates a voltage of 5,000 (kΩ) × 0.0002 (A) = 1 V in the thermistor. As a result, the signal-to-noise ratio is:

$$\frac{S}{N} = 1(\mathrm{V})/2.88 \times 10^{-8}(\mathrm{V}) = 3.47 \times 10^{9}.$$

2. *Shot noise*: The random fluctuations, which are caused by the carriers' random arrival time, produce shot noise. These signal carriers can be electrons, holes, photons, and phonons.

   Shot noise is a random and quantized event, which depends on the transfer of the individual electrons across the junction. Using the statistical calculations, the root-mean-square of the current fluctuation, generated by the shot noise, can be obtained as:

$$\bar{i}_{\mathrm{rms}} = \sqrt{2Ie\Delta f}, \tag{2.4}$$

where I is the average current passing through the junction, Δf is the bandwidth, and e is the charge of one electron, which is equal to 1.60×10^{-19} C.

Example 2. In a photodiode the bias current passing through the diode is 0.1 mA. (A) If the rise time of the photodiode is 0.2 ms and the relation between the rise time and the bandwidth is $\tau_{\text{rise}} = 1/4\Delta f$, calculate the rms value of the shot noise current fluctuation. (B) Calculate the magnitude of the shot noise voltage, when the junction resistance is equal to 250 Ω.

Answer: The bandwidth is equal to $\Delta f = 1/4\tau_{\text{rise}} = 1/[4 \times 0.0002\,(\text{s})] = 1{,}250$ Hz. According to (2.4) the rms value of the shot noise current fluctuation is equal to:

$$i_{\text{rms}} = \sqrt{2 \times 0.1 \times 10^{-3}(\text{A}) \times 1.6 \times 10^{-19}(C) \times 1,250\ (\text{Hz})} = 200 \times 10^{-12}\text{A}$$
$$= 200\ \text{pA}.$$

When the average resistance of the junction is equal to 250 Ω, this fluctuation current generates a rms voltage of 200×10^{-12} (A) $\times$ 250 (Ω) $= 50 \times 10^{-9}$ (V) $= 0.05$ μV or $20\log(\bar{v}_{\text{rms}}) = -146.02\,\text{dB}$.

3. *Generation-recombination* (or *g-r noise*): This type of noise is produced from the generation and recombination of electrons and holes in semiconductors. They are observed in junction electronic devices.
4. *Pink noise* (or *1/f noise*): In this type of noise the components of the frequency spectrum of the interfering signals are inversely proportional to the frequency. Pink noise is stronger at lower frequencies and each octave carries an equal amount of noise power. The origin of the pink signal is not completely understood.

A term, which is frequently seen in dealing with noise, is *white noise*. White noise has flat power spectral density, which means that the signal contains equal power for any frequency component. An infinite-bandwidth, white noise signal is purely theoretical, as by having power at all frequencies the total power is infinite.

2.2.8 Drift

Drift is observed when a gradual change in the sensing system's output is seen, while the measurand actually remains constant. Drift is the undesired change that is unrelated to the measurand. It is considered a systematic error, which can be attributed to interfering parameters such as mechanical instability and temperature instability, contamination, and the sensor's materials degradation. It is very common to assess the drift with respect to a sensor's *baseline*. Baseline is the output value, when the sensor is not exposed to a stimulus. Logically for a sensor with no drift, the baseline should remain constant.

For instance, in a semiconducting gas sensor, a gradual change of temperature may change the baseline. Additionally, gradual diffusion of the electrode's metal into substrate or sensitive layer may gradually change the conductivity of the sensitive element, which deteriorates the baseline value and causes a drift.

2.2.9 Resolution

Resolution (or *discrimination*) is the minimal change of the measurand that can produce a detectable increment in the output signal. Resolution is strongly limited by any noise in the signal.

A temperature sensing system with four digits has a higher resolution than three digits. When the ambient temperature is 21 °C, the higher resolution system (four digits) output is 21.00 °C while the lower resolution system (three digits) is 21.0 °C. Obviously, the lower resolution system cannot resolve any values between 21.01 °C and 21.03 °C.

2.2.10 Minimum Detectable Signal

In a sensing system, *minimum detectable signal* (*MDS*) is the minimum signal increment that can be observed, when all interfering factors are taken into account. When the increment is assessed from zero, the value is generally referred to as *threshold* or *detection limit*. If the interferences are large relative to the input, it will be difficult to extract a clear signal and a small MDS cannot be obtained.

2.2.11 Calibration Curve

A sensing system has to be calibrated against a known measurand to assure that the sensing results in correct outputs. The relationship between the measured variable (x) and the signal variable generated by the system (y) is called a calibration curve as shown in Fig. 2.2.

2.2.12 Sensitivity

Sensitivity is the ratio of the incremental change in the sensor's output (Δy) to the incremental change of the measurand in input (Δx). The slope of the calibration curve, $y = f(x)$, can be used for the calculation of sensitivity. As can be seen in Fig. 2.2, sensitivity can be altered depending on the calibration curve. In Fig. 2.2, the sensitivity for the lower values of the measurand ($\Delta y_1/\Delta x_1$) is larger than of

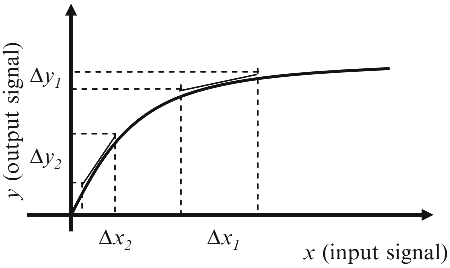

Fig. 2.2 Calibration curve: it can be used for the calculation of sensitivity

the other section of the curve ($\Delta y_2/\Delta x_2$). An ideal sensor has a large and preferably constant sensitivity in its operating range. An ideal sensor has a large and preferably constant sensitivity in its operating range. It is also seen that the sensor eventually reaches *saturation*, a state in which it can no longer respond to any changes.

For example, in an electronic temperature sensing system, if the output voltage increases by 1 V, when temperature changes by 0.1 °C, then the sensitivity will be 10 V/ °C.

2.2.13 Linearity

The closeness of the calibration curve to a specified straight line shows the *linearity* of a sensor. Its degree of resemblance to a straight line describes how linear a system is.

2.2.14 Selectivity

Selectivity is the sensing system's ability to measure a target measurand in the presence of others interferences.

For example, an oxygen gas sensor that does not show any response to other gas species, such as carbon dioxide or nitrogen oxide, is considered a very selective sensor.

2.2.15 Hysteresis

Hysteresis is the difference between output readings for the same measurand, depending on the trajectory followed by the sensor.

Hysteresis may cause false and inaccurate readings. Figure 2.3 represents the relation between output and input of a system with hysteresis. As can be seen, depending on whether path 1 or 2 is taken, two different outputs, for the same input, can be displayed by the sensing system.

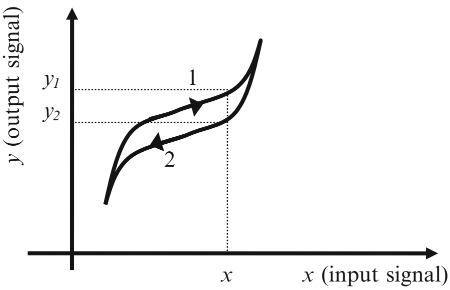

Fig. 2.3 An example of a hysteresis curve

2.2.16 Measurement Range

The maximum and minimum values of the measurand that can be measured with a sensing system are called the *measurement range*, which is also called *the dynamic range* or *span*. This range results in a meaningful and accurate output for the sensing system. All sensing systems are designed to perform over a specified range. Signals outside of this range may be unintelligible, cause unacceptably large inaccuracies, and may even result in irreversible damage to the sensor.

Generally the measurement range of a sensing system is specified on its technical sheet. For instance, if the measurement range of a temperature sensor is between -100 and 800 °C, exposing it to temperatures outside this range may cause damage or generate inaccurate readings.

2.2.17 Response and Recovery Time

When a sensing system is exposed to a measurand, the time it requires to reach a stable value is the response time. It is generally expressed as the time at which the output reaches a certain percentage (for instance, 95 %) of its final value, in response to a step change of the input. The "recovery time" is defined conversely.

2.3 Dynamic Characteristics

A sensing system response to a dynamically changing measurand can be quite different from when it is exposed to time invariable measurand. In the presence of a changing measurand, *dynamic characteristics* can be employed to describe the sensing system's transient properties. They can be used for defining how accurately the output signal is employed for the description of a time varying measurand. These characteristics deal with issues such as the rate at which the output changes in response to a measurand alteration and how these changes occur.

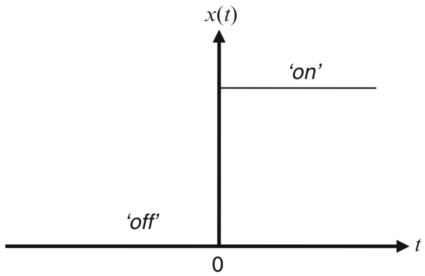

Fig. 2.4 Time variation of a step function

The reason for the presence of dynamic characteristics is the existence of energy-storing elements in a sensing system. They can be produced by electronic elements such as inductance and capacitance, mechanical elements such as vibration paths and mass, and/or thermal elements with heat capacity.

The most common method of assessing the dynamic characteristics is by defining a system's mathematical model and deriving the relationship between the input and output signal. Consequently, such a model can be utilized for analyzing the response to variable input waveforms such as impulse, step, ramp, sinusoidal, and white noise signals.

In modeling a system the initial simplification is always an important step. The simplest and most studied sensing systems are *linear time invariant* (*LTI*) systems. The properties of such systems do not change in time, hence time invariant, and should satisfy the properties of superposition (addition of two different inputs produces the addition of their individual outputs) and scaling (when input is amplified, the output is also amplified by the same amount), hence linear.

The relationship between the input and output of any LTI sensing system can be described as:

$$\begin{aligned} & a_n \frac{\mathrm{d}^n y(t)}{\mathrm{d}t^n} + a_{n-1} \frac{\mathrm{d}^{n-1} y(t)}{\mathrm{d}t^{n-1}} + \cdots + a_1 \frac{\mathrm{d}y(t)}{\mathrm{d}t} + a_0 y(t) \\ & \quad = b_m \frac{\mathrm{d}^{m-1} x(t)}{\mathrm{d}t^{m-1}} + b_{m-1} \frac{\mathrm{d}^{m-2} x(t)}{\mathrm{d}t^{m-2}} + \cdots + b_2 \frac{\mathrm{d}x(t)}{\mathrm{d}t} + b_1 x(t) + b_0, \end{aligned} \tag{2.5}$$

where $x(t)$ is the measured (input signal) and $y(t)$ is the output signal and $a_0, \ldots, a_n$, $b_0, \ldots, b_m$ are constants, which are defined by the system's parameters.

$x(t)$ can have different forms such as impulse, step, sinusoidal, and exponential functions. As a simple example, $x(t)$ may be considered to be a step function similar as depicted in Fig. 2.4. This means that a measurand suddenly appears at the sensor. This is an over simplification, as there is generally a rise time when a stimulant appears.

When the input signal is a step change, all derivatives of $x(t)$ with respect to t are zero and (2.5) is reduced to:

$$a_n \frac{\mathrm{d}^n y(t)}{\mathrm{d}t^n} + a_{n-1} \frac{\mathrm{d}^{n-1} y(t)}{\mathrm{d}t^{n-1}} + \cdots + a_1 \frac{\mathrm{d}y(t)}{\mathrm{d}t} a_{n-1} + a_0 y(t) = b_1, \tag{2.6}$$

for $t \geq 0$ (b_0 is also considered zero in this case. If not zero, a baseline is added to the system response).

Equation (2.6) is a differential equation that models a sensing system response to a step function. Classical solutions to this equation can be readily found in differential equations textbooks and references.

2.3.1 Zero-Order Systems

A perfect *zero-order system* can be considered, if output shows a without-delay response to the input signal. In this case, all a_i coefficients except a_0 are zero. Equation (2.5) can then be simplified to:

$$a_0 y(t) = b_1 \text{ or simply : } y(t) = K. \tag{2.7}$$

where $K = b_1/a_0$ is defined as the *static sensitivity* for a linear system.

2.3.2 First-Order Systems

An order of complexity can be introduced when the output approaches its final value gradually. Such a system is called a *first-order system*. A first-order system is mathematically described as:

$$a_1 \frac{\mathrm{d}y(t)}{\mathrm{d}t} + a_0 y(t) = b_1, \tag{2.8}$$

or after rearranging:

$$\frac{a_1}{a_0} \frac{\mathrm{d}y(t)}{\mathrm{d}t} + y(t) = \frac{b_1}{a_0}. \tag{2.9}$$

If $\tau = a_1/a_0$ is defined as the time constant, the equation will take the form of a *first-order ordinary differential equation*:

$$\tau \frac{\mathrm{d}y(t)}{\mathrm{d}t} + y(t) = K. \tag{2.10}$$

This equation can be solved by obtaining the *homogenous* and *particular solutions*.

Solving (2.10) reveals that in response to the step function $x(t)$, $y(t)$ is reaching K via an exponential rate. τ is the time that the output value requires to reach approximately 63% [$(1 - 1/\mathrm{e}^{-1}) = 0.6321$] of its final value K. A typical response for a first-order system is shown in Fig. 2.5.

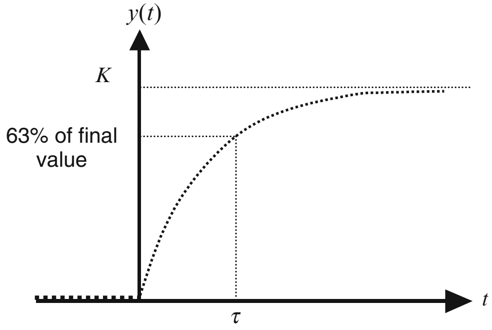

Fig. 2.5 Response of a first-order system to a step function

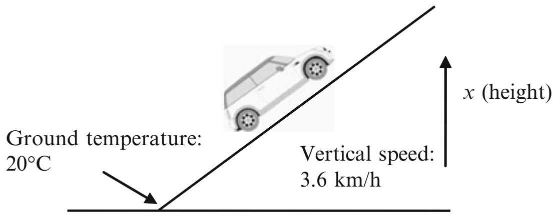

Fig. 2.6 Example 3: first-order response of a car temperature sensing system

Example 3. A car is equipped with altitude and temperature sensors and associated measurement systems. It is traveling up a hill at a constant speed. This road resembles a ramp (Fig. 2.6) with a constant slope. The temperature at the bottom of the hill is 20 °C. The true temperature at the altitude of x meters is given by: $T_x(x) = 20\ °C - 0.1x$.

This means that the temperature drops 1 °C for every 10 m of vertical height increase. The altitude measurement system has an ideal (zero-order) response. However, the temperature sensing system has a first-order response with a time constant of $\tau = 10$ s (the delay time).

(a) If the altitude of the car increases with a speed of 3.6 km h^{-1}, what will be the temperature and height measurements at 10, 20, 30, and 40 s?
(b) What will be the values demonstrated by the temperature sensor, if its time constant is reduced to $\tau = 1$ s?

Answer:

(a) The car's altitude is a zero-order function of time t and can be obtained using x (m) = [3,600 (m h^{-1})/3,600 (s h^{-1})] $\times t = t$ (s), which means that the altitude increases by 1 m every second. As a result, $T_x(t)$ can also represents the actual ambient temperature as a function of time: $T_x(t) = 20\ °C - 0.1\ t$.

The measured temperature (the output of a first-order system) is $T_m(t)$, which can be obtained from a first-order differential equation:

$$\tau \frac{\mathrm{d}T_m(t)}{\mathrm{d}t} + T_m(t) = T_x(t).$$

Table 2.1 Temperature sensor responses at different time intervals for the time constant of 10 s

Time (s)	Altitude (m)	Real temp (°C)	Measured temperature (°C)	Temperature error (°C)
0	0	20	20	0
10	10	19	19.6321	0.6321
20	20	18	18.8647	0.8647
30	30	17	17.9502	0.9502
40	40	16	16.9817	0.9817

By substituting the value of the time constant and the function describing the ambient temperature, the resulting equation is:

$$10\frac{\mathrm{d}T_m(t)}{\mathrm{d}t} + T_m(t) = 20 - 0.1t.$$

A first-order equation's answer is the addition of two general solutions: homogenous (natural response of the equation) and particular integral (generated by the step function in this example). For this example, the homogenous part of the solution is:

$$T_{m-h}(t) = A\mathrm{e}^{\frac{-t}{10}}.$$

The particular-integral part of the solution is given by:

$$T_{m-pi}(t) = -0.1t + 21.$$

Consequently, the total solution can be obtained as:

$$T_m(t) = A\mathrm{e}^{\frac{-t}{10}} + 21 - 0.1t.$$

Applying the initial condition of $T_{mr}(0) = 20$ °C, it can be found that $A = -1$. As a result, the $T_m(t)$ can be calculated by:

$$T_r(t) = -1 \times \mathrm{e}^{\frac{-t}{10}} + 21 - 0.1t.$$

Using the above formula, Table 2.1 can be established which presents the values of the ambient temperature (real temperature) and the measured temperature at different intervals.

As can be observed, the error increases with time, approaching 1 °C. This is due to the large time constant value of 10 s. The car has a constant speed, hence constant change of temperature, and there is always a lag in the sensor response. After a while, the system reaches a stable condition where a steady and constant error always exists.

(b) Using a smaller time constant, $\tau = 1$ s, the response of the temperature sensor is faster (a fast processing-responding system) and the differential equation is transformed into:

Table 2.2 Temperature sensor responses at different time intervals for the time constant of 1 s

Time (s)	Altitude (m)	Real temperature (°C)	Measured temperature (°C)	Temperature error (°C)
0	0	20	20	0
10	10	19	19.1	0.1
20	20	18	18.1	0.1
30	30	17	17.1	0.1
40	40	16	16.1	0.1

$$\frac{dT_m(t)}{dt} + T_m(t) = 20 - 0.1t,$$

which has the homogenous solution of:

$$T_{m-h}(t) = Ae^{-t},$$

and the particular-integral solution is given by:

$$T_{m-pi}(t) = -0.1t + 20.1.$$

In this case, the total solution is obtained as:

$$T_m(t) = Ae^{-t} + 20.1 - 0.1t.$$

Using the initial condition of $T_{mr}(0) = 20$ °C, we obtain $A = -0.1$. As a result, the output temperature equation can be rewritten as:

$$T_m(t) = -0.1 \times e^{-t} + 20.1 - 0.1t.$$

Similarly, Table 2.2 can now be established, which presents the values of the ambient temperature (real temperatures) and the measured temperature at different intervals.

As can be seen for the time constant of 10 s, the error converges to 1 °C but for a time constant an order of magnitude smaller, 1 s, it converges to 0.1 °C.

Example 4. A photodiode sensor has a first-order response with $\tau = 1$ ms. The calibration curve of this sensor is linear and it generates a current of 1 mA at full sun (1 sun) and 0 mA at the fully dark condition. Graph the response of this sensor in time for the following condition: it has been keep in dark for a long time (the initial stable condition), exposed to 0.5 sun for 2 ms, and then turn off the light source to produce the dark condition after that (Fig. 2.7).

Answer: For a first-order system the differential equation is defined as follows:

$$\tau \frac{dy(t)}{dt} + y(t) = K. \tag{2.11}$$

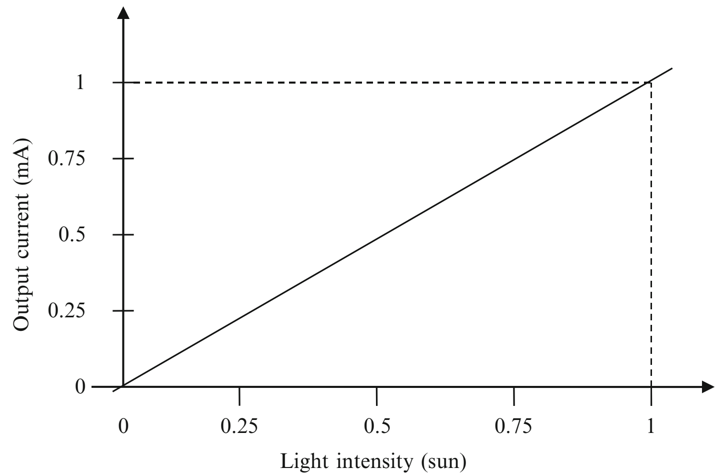

Fig. 2.7 Calibration curve for Example 4

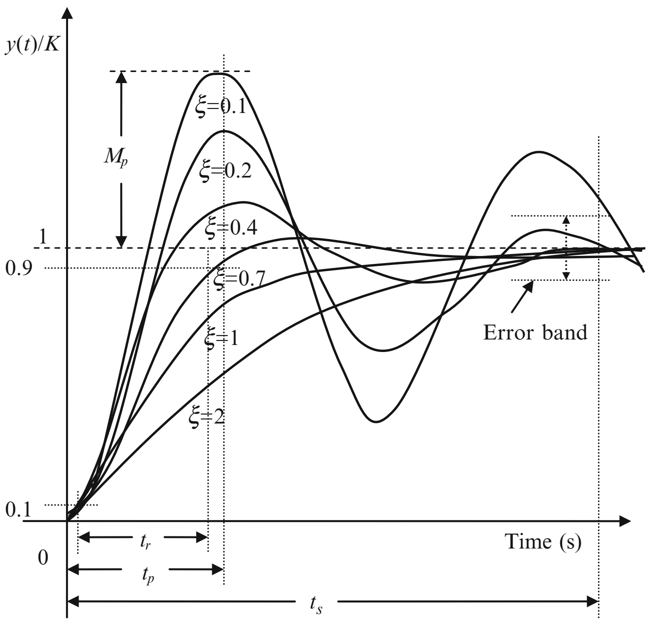

Fig. 2.8 Responses of a second-order sensing system to a step function at different damping ratios

Again, this equation can be solved by obtaining the homogenous and particular solutions.

The value of K can be obtained from the calibration curve as demonstrated in Fig. 2.8.

The sensor has been in zero light exposure ($t < 0$), to 0.5 sun (hence generating 0.5 mA for $0 \le t \le 2$ ms), and to zero sun again after $t > 2$ s. Considering the calibration curve, $x(t)$ will be defined as:

$$x(t) = \begin{cases} 0 & t<0 \\ 0.5\,\text{mA} & 0 \le t \le 2\,\text{ms} \\ 0 & t>2\,\text{ms} \end{cases}.$$

The solution can be found using the first-order differential equation for each of these divisors in time.

For $t < 0$ the sensor has long been kept in the dark condition so $y(t) = 0$, as it generates no current.

For $0 \le t \le 2$ ms, the 0.5 mA current is fed into the right-hand side of the first-order differential equation. Therefore, the first-order equation takes the format: $(1\,\text{ms})\frac{\text{d}y(t)}{\text{d}t} + y(t) = 0.5\,\text{mA}$, solving this equation, the *homogenous* and *particular solutions* will be $y_h(t) = A(\text{e}^{-t/1\ \text{ms}})$ and $y_p(t) = 0.5$ mA, respectively, and:

$y(t) = y_h(t) + y_p(t) = A(\text{e}^{-t/1\ \text{ms}}) + 0.5$ mA. As at $t = 0$, $y(t) = 0$ then $A = -0.5$ mA or $y(t) = 0.5\ \text{mA} \times (1 - \text{e}^{-t/1\ \text{ms}})$.

Using this equation $y(2\ \text{ms}) = 0.5\ \text{mA}(1 - \text{e}^{-2}) = 0.432$ mA.

For $t > 0$ the sensor is placed in a dark ambient again, and the input on the right-hand side of the equation will eventually be equal to zero. In this case, only the *homogenous* solution exists as:

$y(t) = y_h(t) = B \times (\text{e}^{-(t\ -\ 2\ \text{ms})/1\ \text{ms}})$. As the current is continuous, $y(2\ \text{ms}) = B = 0.432$ mA, which results in:

$$y(t) = 0.432\ \text{mA} \times (\text{e}^{-(t-2\text{ms})/1\text{ms}}).$$

Subsequently, the description of the sensor response is as below:

$$y(t) = \begin{cases} 0 & t<0, \\ 0.5\left(1 - \text{e}^{\frac{-t\ \text{ms}}{1\ \text{ms}}}\right) & 0 \le t \le 2\text{s}, \\ 0.432\left(\text{e}^{\frac{-(t-2\ \text{ms})}{1\ \text{ms}}}\right) & t>2\text{s}. \end{cases}$$

2.3.3 Second-Order Systems

The response of a system can be more complicated. In response to a step function, it may oscillate before reaches its final value. The response can be *overdamped* or *underdamped*. Such responses can be better described by a second-order system approximation.

The response of a *second-order system* to a step change is shown as:

$$a_2 \frac{\mathrm{d}^2 y(t)}{\mathrm{d}t^2} + a_1 \frac{\mathrm{d}y(t)}{\mathrm{d}t} + a_0 y(t) = b_1. \tag{2.12}$$

By defining the undamped natural frequency as $\omega^2 = a_0/a_2$, and the dampening ratio as $\xi = a_1/2(a_0 a_2)^{1/2}$, (2.12) reduces to:

$$\frac{1}{\omega^2}\frac{\mathrm{d}^2 y(t)}{\mathrm{d}t^2} + \frac{2\xi}{\omega}\frac{\mathrm{d}y(t)}{\mathrm{d}t} + y(t) = K. \tag{2.13}$$

This is a standard second-order system in response to a step function for which $K = b_1/a_0$.

The damping ratio and natural frequency play pivotal roles in the shape of the response as seen in Fig. 2.8. If $\xi = 0$ there is no damping and the output shows a constant sinusoidal oscillation with a frequency equal to the natural frequency. If ε is relatively small then the damping is light, and the oscillation takes a long time to vanish, *underdamped*. When $\xi = 0.707$ the system is *critically damped*. A critically damped system converges to zero faster than any other conditions without any oscillation. When ξ is large the response is *overdamped*. Other response parameters include: rise time (t_r), peak overshoot (M_p), time to first peak (t_p), and settling time (t_s) (the time elapsed from when the step input is applied to the time at which the amplifier output remains within a specified error band).

Many sensing systems follow the second-order equations. For such systems responses that are not near critically damped condition ($0.6 < \xi < 0.8$) are highly undesirable as they are either slow or oscillatory.

The majority of sensing systems can be nicely described either with the first or second-order equations. However, more complexity can be added when describing a dynamic response of such systems with unusual behaviors. For instance, very often in semiconducting gas sensors, after the initial interactions of the gas with the surface, which is generally a first-order response, many other interactions might occur to change the order of the system. Gas molecules might further diffuse into the bulk of the materials, the morphology of the sensitive material might change, and several stages of interaction might occur. As a result in such systems, obtaining the mathematical description of the dynamic responses can be quite a challenging task.

2.4 Summary

A comprehensive overview of static and dynamic parameters that are used in sensing systems was presented in this chapter.

The most important static characteristics including accuracy, precision, repeatability, reproducibility, stability, error, noise, drift, resolution, MDS, calibration curve, sensitivity, linearity, selectivity, hysteresis, measurement range, as well as response and recovery time were described.

Dynamic characteristics of sensing systems were then presented. Differential equations that describe such systems were discussed with the emphasis of zero-, first-, and second-order systems.

http://www.springer.com/978-1-4614-5051-1

Sensors
An Introductory Course
Kalantar-zadeh, K.
2013, XII, 196 p., Hardcover
ISBN: 978-1-4614-5051-1