Sensor Performance Specifications

By Dennis S. Bernstein

Measurement

How important is measurement? Consider this: Everything you buy in a grocery store is measured. A pound of cherries, a gallon of milk, and a dozen eggs. Just as eggs are measured by counting them, the cashier will measure the money you hand him by counting it. Everything is measured in this transaction.

Besides commerce, which is impossible without measurement, we need measurement for numerous everyday tasks. We measure our speed to avoid tickets (and accidents), and we measure the gasoline in the fuel tank so we don't run out. Cooking without measurement usually doesn't work out very well. Carpentry is another example: *Measure twice and cut once.* Clocks for measuring time are everywhere.

Engineering and science rely on measurement to an even greater extent. Without measurement, engineering would be impossible, and science would be philosophy. But, as alchemy shows, measurement alone doesn't make good science or engineering.

To make measurements, we need to recognize *dimensions*. There are really only four *physical dimensions*, namely, length, mass, time, and charge. We could argue that apples are different from bananas and introduce dimensions such as "pound of apples" or "pound of bananas." Although this distinction is useful if you're eating dinner, it isn't relevant to dynamics where inertia is more important than chemical composition.

In addition to dimensions, we need *units*. For length we can choose inches, centimeters, miles, or some other unit. Without units, measurement is meaningless, and no one, including ourselves, will know what we're talking about. Units are usually set by agreement so that two people who have never met can exchange measurement data. Although there are many shoe sizes, there is only one foot.

A *measurement* is information about a physical quantity, and to make measurements you need a *sensor*. A ruler is a good example. By merely placing the ruler next to an object, you can measure its length. Using your eye, you "read" the ruler, and the measurement is transmitted into your brain, where it is stored for as long as you can remember it or as long as it takes to record it on a piece of paper or in a computer.

The role of a sensor is to facilitate the transfer of information about a physical quantity into a display, computer register, or onto a piece of paper where it can be accessed. *Data acquisition* is the process of collecting and storing measurements.

But measurement is never perfect. The ruler may be slightly bent, the object may have a rough edge, the ruler might slip, or the object's edge might fall between two ruler markings. These and other problems can contribute to *measurement errors*.

A more subtle problem with measurement is that the information the sensor conveys about a physical quantity requires a transfer of energy. This is not a problem in measuring length by a ruler. In that case, you merely shine light on the object, which heats it slightly and changes its length by an inconsequential amount. When measuring the flow rate of a gas or liquid, your sensor may disrupt the flow and thereby obtain an erroneous reading with respect to the undisturbed flow.

In control system engineering, sensors play a critical role by providing measurements for feedback. The control system engineer must determine which quantities must be measured and how well they must be measured to achieve desired control system performance. In the following sections I explain *sensor specifications*, which quantify the ability of a sensor to provide measurements of physical variables. Sensor manufacturers typically provide information about sensor specifications in the form of a "spec sheet." The concepts described here will help you interpret the information on a spec sheet.

The concepts discussed below apply to all branches of science and engineering, not just control systems. However, the systems approach of control engineers provides a unique perspective on these topics and associated issues.

Measured Variables

Before delving into sensor specifications, it is useful to keep in mind the types of quantities that may be measured as well as the physics that sensors exploit to obtain measurements. Perhaps the most common measurement objective involves length, as in the ruler example. Depending on the application or context, length can be viewed as displacement, position, strain, or gap. Area and volume are extensions of length. All of these are time-*independent* translational notions in which rate plays no role. When the rate of change of displacement is of interest, we are concerned with time-*dependent* translation, which includes velocity and acceleration. Translational motion may be either *relative* (between stationary or moving points) or *inertial* (relative to the fixed stars).

The author (dsbaero@umich.edu) is with the Aerospace Engineering Department, Univeristy of Michigan, Ann Arbor, MI 48109-2140, U.S.A.

0272-1708/01/$10.00©2001IEEE

Many applications require measurement of angle, which is a nondimensional quantity involving the ratio of arc length to radius. An angle measurement can be viewed as a displacement, position, attitude, orientation, or gap, which concern time-independent rotation. Time-dependent rotation includes angular velocity and angular acceleration, which may be relative to specified bodies or an inertial frame.

Force and pressure (force per unit area) are commonly measured in many applications. Thermal properties (such as temperature) are important, as are mass and volume flow properties. Electrical properties such as charge, potential, current, resistance, capacitance, and inductance are obviously important, as are magnetic properties. Electromagnetic waves at optical or other frequencies are also relevant. Finally, chemical and molecular properties such as concentration require measurement.

Sensor Physics

Engineers and scientists have used amazing ingenuity to develop a vast array of sensors. These sensors exploit the laws of physics and exotic materials to measure physical variables. Since no sensor is ideal for all applications, many different sensors have been developed to measure the same physical variable. The selection of an appropriate sensor depends on cost, weight, power requirements, the need for tethering (for power and signal transmission), contacting or mounting requirements, environmental conditions (medium, temperature, electromagnetic fields, shock, and vibration), as well as performance issues.

The following are some of the physics exploited by sensors:

- Resistive: a voltage potential produces current (Ohm's law);
- Inductive: an inductance changes due to the position or velocity of a conducting target (Faraday's and Lenz's laws);
- Capacitive: a capacitance changes due to a conducting or insulating target;
- Piezoelectric: a strain produces a charge (PZT: zirconium titanate, PVDF: polyvinylidene fluoride);
- Hall effect: a magnetic field induces a potential;
- Thermoelectric: a thermal gradient induces a potential (Seebeck effect);
- Optical: a variety of techniques exploit intensity, interferometry, Doppler, triangulation, holography, modulation, and reflection;
- Ultrasonic: a reflected acoustic wave provides an image.

These physical phenomena are used in various ways to measure physical quantities. For example, translational position can be measured by various devices:

- Linear or rotary potentiometer
- Ultrasonic time of flight;
- Linear encoder tape
- Strain gauge (resistive, piezoelectric);
- Inductive (LVDT: linear variable displacement transducer);
- Capacitive (gap);
- Hall effect (gap);
- Optical (optical fiber reflection intensity, interferometry, holography, imaging);
- Radio frequency (time of flight, GPS).

Translational velocity is more difficult to measure. Some devices are:

- Linear velocity transducer (inductive);
- Laser scanning vibrometer (Doppler).

Translational acceleration is measured using an accelerometer, which measures inertial acceleration, including gravitational acceleration. Mechanical accelerometers are based on the equation

$$m(\ddot{q}_{\mathrm{base}}+\ddot{q}_{\mathrm{rel}})+kq_{\mathrm{rel}}=0,$$

where q_{rel} is the relative displacement between a mass m mounted on a stiffness k relative to a base moving with displacement q_{base}. This sensor provides an approximate measure of base acceleration given by

$$\ddot{q}_{\mathrm{base}}\approx-(k/m)q_{\mathrm{rel}},$$

which is valid for low-frequency acceleration. A more sophisticated approach is based on the force rebalanced or servo accelerometer, which applies a force u to the system

$$m\ddot{q}+kq=f+u$$

to make $q=0$. Quartz and piezo accelerometers measure the charge induced by a strain but otherwise involve no moving parts. Since charge is induced by changes in acceleration only, this is an example of a non-dc response sensor, a concept discussed below. Piezoresistive and variable-capacitance accelerometers are accelerometer realizations having dc response.

Rotational displacement is measured by a variety of devices. Terrestrial attitude can be determined by means of a magnetometer, which senses the direction of the earth's magnetic field. This device uses three sensors to determine two of the three angles that characterize attitude. The third angle can be determined by using three dc accelerometers to determine the direction of the earth's gravity vector. In this case, each accelerometer serves as an inclinometer or tilt sensor.

Relative attitude and angles can be measured by a variety of techniques, including:

- Variable potentiometer (restricted rotation);
- Encoder (incremental or absolute, unrestricted rotation);

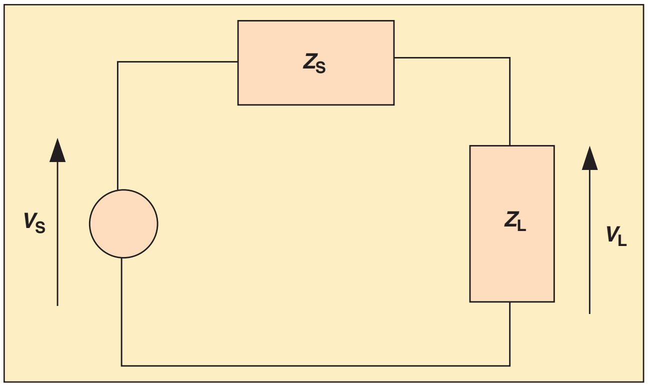

Figure 1. *The source impedance represents the system with output voltage V_S, whereas the load impedance represents the sensor with measurement V_L.*

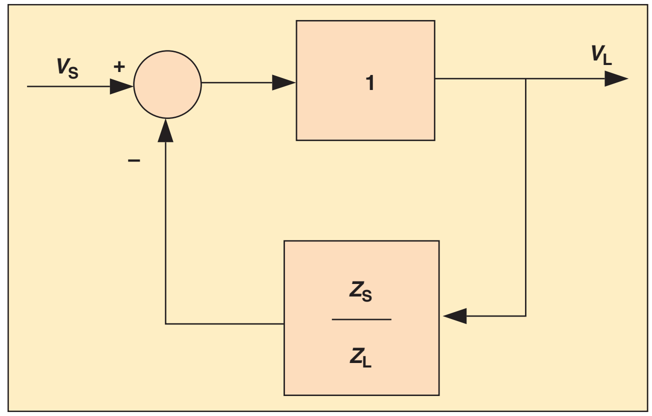

Figure 2. *A nonideal input or output impedance gives rise to a feedback loop involving the system and sensor.*

- Resolver (sine and cosine signals, unrestricted rotation);
- RVDT (rotational variable displacement transducer, inductive, restricted rotation);
- RCDT (rotational capacitive displacement transducer, capacitive, restricted rotation);
- Hall-effect angle sensor (restricted rotation).

A limitation of some of these sensors is the restriction to less than full rotation.

Angular velocity is measured by means of a gyro. A mechanical rate gyro measures inertial angular velocity and is based on a spinning rotor or vibrating crystal, which exploits the Coriolis effect as well as the fundamental law "torque equals change in angular momentum." An optical ring laser rate gyro uses a fiber-optic cable and measures the differential time of travel, which is the Sagnac effect. Angular acceleration can be measured by using a pair of translational accelerometers or a device based on the fact that torque is proportional to angular acceleration.

Force and torque can be measured indirectly by a load cell, which measures the strain due to force and torque. Pressure and stress are measured by a manometer, a microphone, and various other devices. Flow rate can be measured by a hot-wire anemometer, which is based on thermoresistive effects.

Connecting Systems and Sensors

Ideally, the connection between a sensor and a system is a perfect cascade, wherein the sensor does not affect the measured system. As discussed above, however, a sensor can disrupt or alter the physical variable it is being used to measure. To demonstrate how this occurs, let us consider a circuit configuration in which information is transferred in the form of a voltage. Analogous principles apply to noncircuit interconnections.

Consider two impedances as in Fig. 1. The load impedance Z_L represents the sensor and the source impedance Z_S represents the system to which the sensor is connected. The source voltage V_S is the sensor input, and the load voltage V_L is the output measurement. Consequently, the voltage drop across the load is

$$V_L = \frac{Z_L}{Z_S + Z_L} V_S.$$

Clearly, the output measurement V_L is not generally equal to the sensor input V_S. However, it can be seen that $V_L = V_S$ if and only if $Z_S = 0$ or $Z_L = \infty$, which correspond to the ideal cases of zero output impedance and infinite input impedance, respectively. In practice, however, nonideal output or input impedances affect the measured voltage input V_L so that $V_L \neq V_S$.

Redrawing Fig. 1 as Fig. 2, it can be seen that nonideal impedances give rise to a feedback interconnection rather than a cascade interconnection. Since true cascade does not occur in practice, sensor/system interconnections must be modeled as feedback when this effect is significant.

Static and Dynamic Sensor Response

A sensor is truly a dynamic system, and thus a good understanding of sensor performance requires that we understand its dynamic response. The dynamics of a sensor are relevant when the sensor input is a time-dependent signal. On the other hand, many applications involve sensor inputs that are constant over long time periods, in which case the dynamics of the sensor are less critical and static sensor performance is of interest.

The dynamic structure of a sensor can have almost any form, the most general being

$$\dot{q} = f(q, x)$$
$$V = g(q, x),$$

where q is the internal sensor state (possibly a vector), x is the physical input to the sensor, and V is the sensor output,

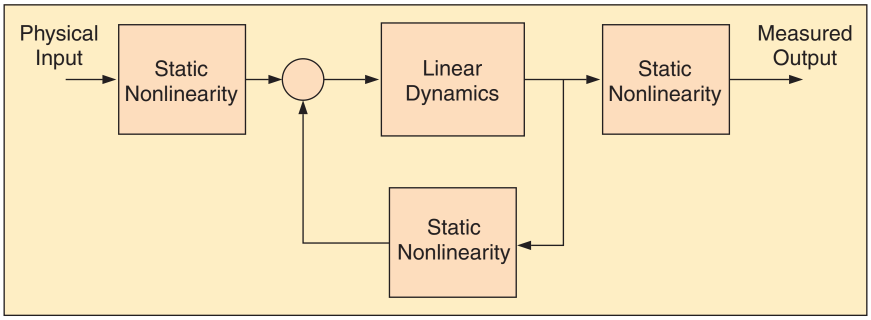

Figure 3. *This block-structured Hammerstein-Wiener nonlinear feedback dynamic sensor model involves a linear dynamic block surrounded by three static nonlinear blocks.*

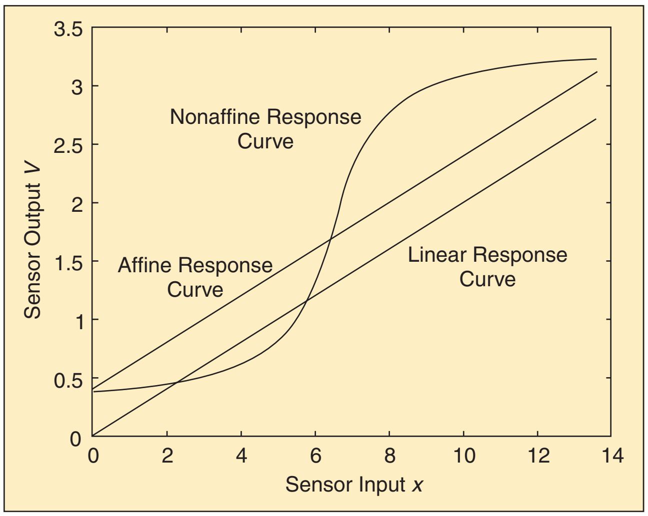

Figure 4. *For the linear static response curve, the sensor output is 0 V when the sensor input is zero. Although the affine static response curve is a line segment, the sensor output need not be 0 V when the sensor input is zero. A nonaffine static response curve need not be either linear or affine, but can have an arbitrary shape.*

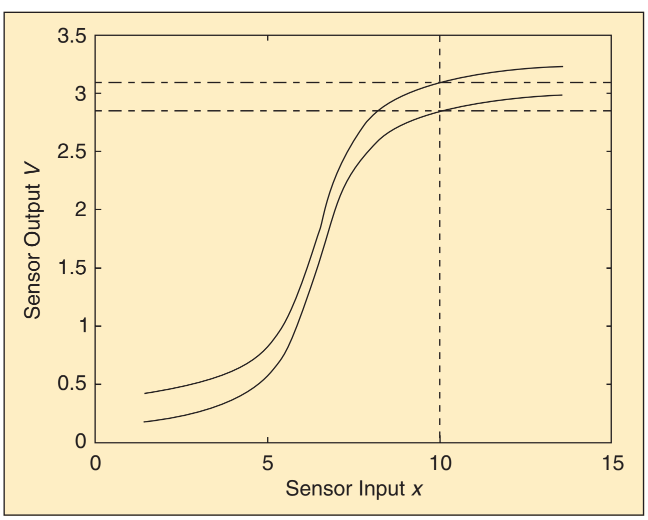

Figure 5. *The lower curve is the original static response curve; the upper curve results from an output bias drift; that is, a uniform shift in the output voltage.*

usually in volts. The feedback effects due to nonideal impedances (or their analogue for the particular application) are assumed to be included in this model.

A more structured, but less general, sensor model can be constructed by interconnecting static nonlinear blocks and dynamic linear blocks. The Hammerstein-Wiener nonlinear feedback model shown in Fig. 3 involves a linear dynamic block surrounded by three static nonlinear blocks.

The dc (static) response of a sensor can be determined by assuming that the sensor input is constant and by ignoring the sensor dynamics. The resulting *static response map* characterizes the output of the sensor after all transients have settled and the output has converged to a steady-state response. A dc sensor is one that responds to constant inputs, whereas a *non-dc sensor* is unable to distinguish constant inputs. A non-dc sensor with linear dynamics has a zero at the origin so that its frequency response rolls off at dc, whereas the frequency response of a dc sensor does not. At high frequencies, the sensor frequency response will roll off, which limits the ability of the sensor to respond to high-frequency sensor inputs.

As an example, a piezo accelerometer can sense *changing* acceleration but is not able to sense *constant* acceleration. Similarly, some gyros are unable to measure *constant* angular velocity. On the other hand, there are dc accelerometers that can measure constant acceleration and dc gyros that can measure constant angular rates. In choosing a sensor, it is important to determine whether it is dc or non-dc in accordance with your application requirements. The same distinction applies to the choice of signal and power amplifiers.

It is important to keep in mind that the ability of a sensor to respond to changing inputs does not change the *type* of sensor that it is. For example, the speedometer in your car can respond to *changing* speeds, but that does not make it an accelerometer, although it may be possible to infer acceleration information from the velocity measurements.

Henceforth we consider only dc sensors, and we are concerned with their static performance; that is, their steady-state response to constant inputs.

Static Response Curve

The steady-state response of a dc sensor to constant inputs is characterized by its *static response curve*. For convenience, we assume that the sensor output V is given in volts, whereas the sensor input x can have arbitrary physical dimensions.

The static response curve may be *linear* or *affine* (pronounced *uh-finé) or nonaffine* (not necessarily affine). A linear static response curve is a line segment with the property that if the input is zero, the sensor output is also zero. In practice, this situation rarely occurs since zero values are usually arbitrary, and it is more common to encounter an affine static response curve, which is a line segment with a zero-input voltage offset. More generally, a nonaffine static response curve need not be a line segment. These static response curves are illustrated in Fig. 4.

In our discussion of the static response curve, we are mainly interested in dc sensors. Since a non-dc sensor is unable to distinguish constant inputs, its static response curve is merely a horizontal line.

A static response curve is valid for a range of sensor inputs. This is the *sensor input range* given by $[x_{\min}, x_{\max}]$. Over this range of inputs, the *sensor output range* is denoted by $[V_{\min}, V_{\max}]$.

A static response curve is *one-to-one* if each sensor output is produced by exactly one sensor input. A quadratic or V-shaped static response curve is not one-to-one. To eliminate the ambiguity in determining the sensor input for a given sensor output, it is necessary to restrict the sensor input range so that the static sensor curve is one-to-one over the restricted range. This technique is used for some optical position sensors. In the remainder of this article we assume that the static response curve is one-to-one.

Quality of the Static Response Curve

The static response curve can be used to determine the sensor input once the sensor output has been measured. In reality, the static response curve is a theoretical idealization, and the true situation is more complex. We can identify three factors that undermine our ability to determine and work with the static response curve. These factors are drift, noise, and hysteresis.

Drift is the change in the static response curve over time. This change may be due to short-term environmental conditions, such as temperature change or humidity, or it may be due to long-term effects, such as aging, wear, fatigue, or oxidation. The most common type of drift is an *output bias drift* wherein the static response curve shifts by a constant voltage across the input range (see Fig. 5).

The usefulness of the static response curve is also undermined by the presence of noise that is superimposed on the measurement. Noise can arise from external disturbances acting on the system, or it may be inherent in the electrical circuits that are used to measure the sensor output voltage. In any event, the presence of noise effectively replaces the static response curve by an envelope of static response curves, so that the "true" static response curve cannot be distinguished.

The description of the noise may be either deterministic, that is, described by bounds on its size, or it may be stochastic, that is, described by a probability distribution. The *precision* of the sensor is determined by the noise; that is, a sensor has good precision if the noise level is low relative to the magnitude of the measurement. For a given sensor input, the *noise resolution* is the width of the possible sensor outputs due to the noise.

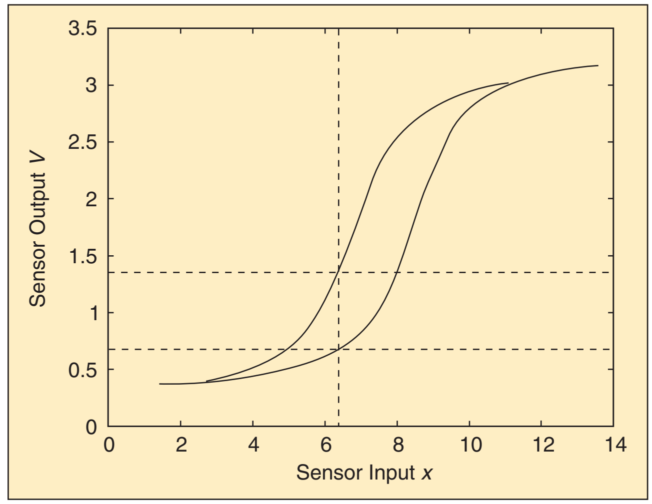

Figure 6. *A hysteretic static response curve arises when the sensor input moves away from a given value and then returns to it, with the second measurement different from the first.*

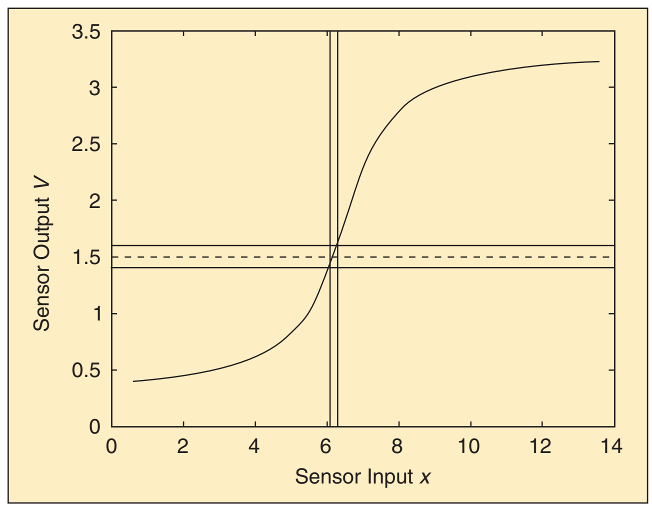

Figure 7. *Quantization error is the difference between the sensor output and the center of the bin as determined by the data acquisition system. The largest quantization error is half of the bin size. This figure shows a single bin centered at 1.5 V with bin size 0.2 V. For all sensor inputs for which the sensor output is between 1.4 V and 1.6 V, the measured value is 1.5 V. Therefore, the quantization error can be as large as 0.1 V.*

The static response curve can also be adversely affected if the sensor has *hysteretic* behavior. Hysteresis occurs when the input variable moves away from a given value and then returns to it, with the second measurement different from the first. In this case, as shown in Fig. 6, the static response curve is effectively multivalued.

The *repeatability* of a sensor is determined by the hysteresis of the static response curve. A sensor has good *repeat-*

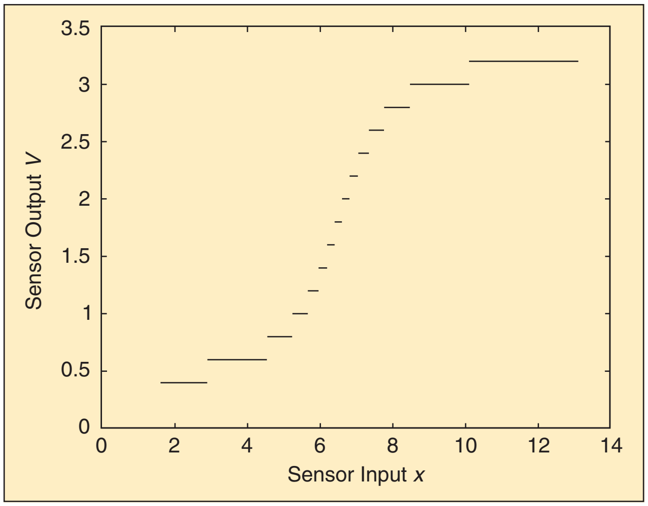

Figure 8. *The quantized sensor response curve is a step function, which gives the same sensor output for an interval of sensor inputs.*

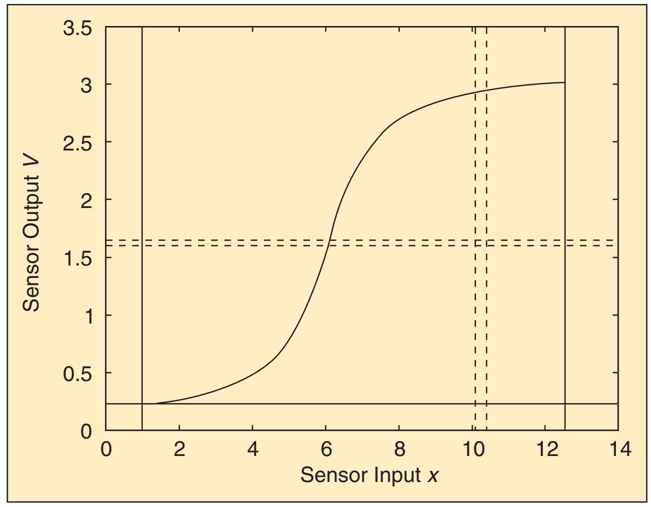

Figure 9. *Output dynamic range is the ratio between the output width and the smallest distinguishable output variation. Similarly, input dynamic range is the ratio between the input width and the smallest distinguishable input variation.*

ability if two measurements of the same input are close despite interim changes in the input. Noise also degrades repeatability. Hence, if a sensor has good repeatability, the noise and hysteresis are necessarily both small. We thus have the important relation

$$\text{Good Repeatability} \Rightarrow \text{Good Precision.}$$

Quantization and Resolution

An additional impediment to determining the static response curve is *quantization*, which is discretization of the output range. Most sensors provide an analog (continuously variable) output, which is quantized by the data acquisition system. Other sensors, such as encoders, provide digital output, which is inherently quantized.

Quantization involves partitioning the sensor output range into *bins*, which are usually of equal size. The bin size, or *quantization resolution*, is determined by

$$\text{quantization resolution} = \frac{\text{output width}}{\text{number of bins}},$$

where the *output width* is equal to $V_{\max} - V_{\min}$.

The *value* associated with a bin is determined by its center, whereas the *quantization error* is the error in the sensor output due to quantization. It can be seen from Fig. 7 that the largest quantization error is half of the bin size. By combining the effect of quantization with the sensor response curve, we obtain the *quantized sensor response curve* illustrated in Fig. 8.

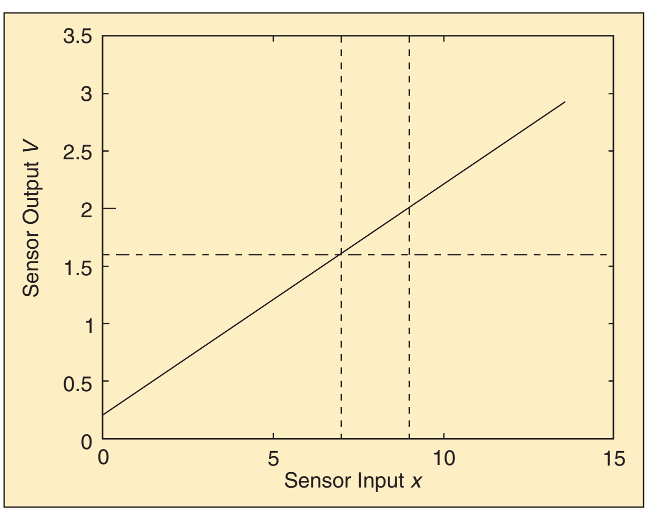

Figure 10. *Sensitivity is the slope of the sensor curve, which is constant if the sensor curve is affine.*

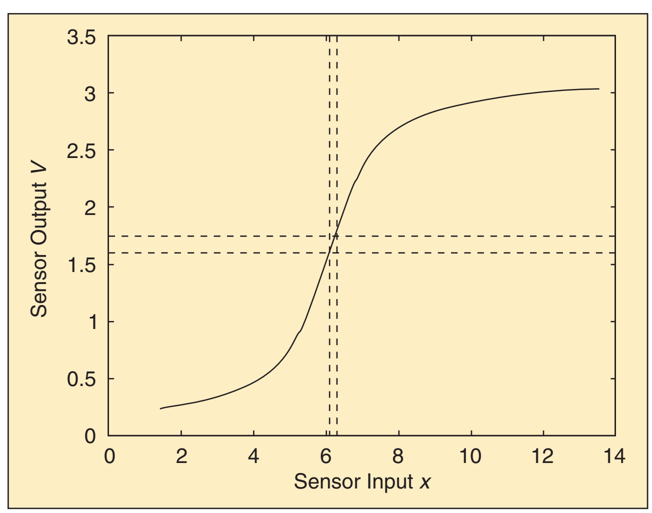

Figure 11. *Sensitivity determines the ratio between the input resolution and the output resolution.*

Once quantization has occurred, the output is effectively measured by counting bins. In principle, quantized quantities are less prone to error. As an example, money is quantized into pennies, which is why money is "counted" when it is measured.

As another example, a ruler is quantized by its markings. Since the human eye can usually interpolate a point between two markings with an accuracy of ±10%, the quantization resolution of a ruler marked in intervals of size 1/16 in is approximately 1/5 × 1/16 in = 1/80 in.

The effect of quantization on sensor specifications is related to the *output resolution*, which is defined by

$$\text{output resolution} = \max\{\text{quantization resolution, noise resolution}\},$$

which is the smallest distinguishable output variation. Output resolution determines the ability to distinguish close outputs.

Output and Input Dynamic Range

An important aspect of a sensor is its ability to respond to signals having both large and small amplitude variation. The ratio of the largest measurable output variation to the smallest distinguishable output variation (assumed to be constant over the output range) is the *output dynamic range*. This quantity is illustrated by Fig. 9, which contrasts the difference between the magnitudes of these quantities. It can be seen that the output dynamic range is given by

$$\text{output dynamic range} = \frac{\text{output width}}{\text{output resolution}}.$$

Dynamic range can be expressed in bits by

$$\text{output dynamic range in bits} = \log_2 \frac{\text{output width}}{\text{output resolution}}$$

or in decibels by

$$\text{output dynamic range in decibels} = 20\log_{10} \frac{\text{output width}}{\text{output resolution}}.$$

As an example, an output width of 8 V with a quantization resolution of 0.23 mV has an output dynamic range of 11.8 bits or 70.8 dB. Note that each bit corresponds to 6 dB.

The *input resolution* of a sensor is the smallest distinguishable input variation. In analogy with output dynamic range, the sensor's *input dynamic range* is the ratio of the largest measurable input variation to the smallest distinguishable input variation (assumed constant over the input range), as illustrated in Fig. 9. The input dynamic range is thus given by

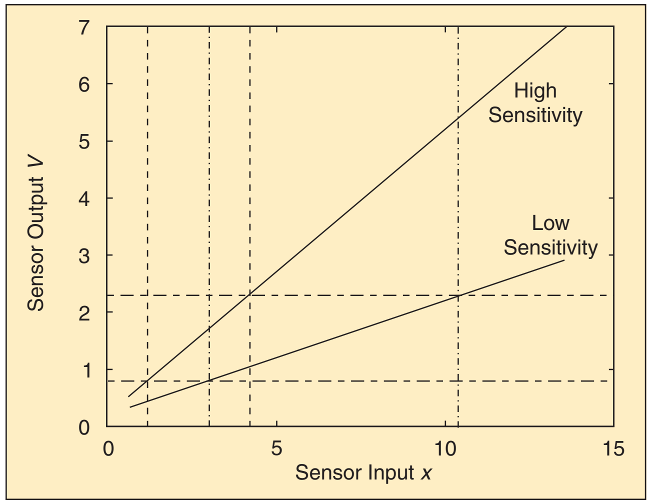

Figure 12. *Sensitivity affects the tradeoff between the input width and output width. For a fixed output width, a high-sensitivity static response curve has small input width, whereas a low-sensitivity static response curve has large input width.*

$$\text{input dynamic range} = \frac{\text{input width}}{\text{input resolution}},$$

where the *input width* is equal to $x_{max} - x_{min}$, which can be expressed in bits or decibels. The input dynamic range is degraded by noise and hysteresis.

Sensitivity and Sensitivity Tradeoffs

Another important aspect of sensor performance is the amount that the sensor output changes due to a change in

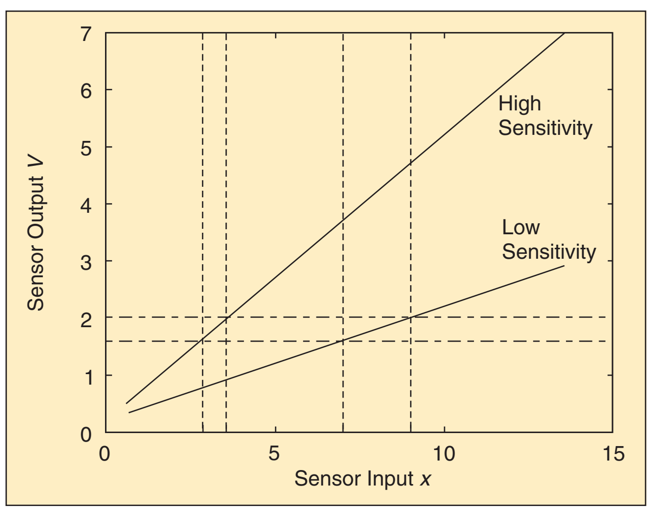

Figure 13. *Sensitivity affects the tradeoff between the input resolution and output resolution. For a fixed output resolution, a high-sensitivity static response curve has fine input resolution, whereas a low-sensitivity static response curve has coarse input resolution.*

the sensor input. The ratio of these quantities is the sensor *sensitivity*, which is the slope of the static response curve.

For an affine static response curve (see Fig. 10), the sensitivity is constant and can be determined by

$$\text{sensitivity} = \frac{\text{output width}}{\text{input width}}.$$

For a nonaffine static response curve, the sensitivity depends on the sensor input as determined by the slope of the static response curve for a given sensor input. A fundamental feature of the sensitivity is the fact that it relates the input resolution and the output resolution (see Fig. 11) according to

$$\text{sensitivity} = \frac{\text{output resolution}}{\text{input resolution}}.$$

Sensitivity is an important quantity since it affects fundamental sensor tradeoffs. To illustrate these tradeoffs, assume for the moment that the output *width* is fixed. In practice, the output width is constrained by the data acquisition system, which can operate over only a limited range of voltages. It can then be seen from Fig. 12 that a static response curve with high sensitivity will have a small input width, whereas a static response curve with low sensitivity will have a large input width; that is, for fixed output width

$$\text{High Sensitivity} \Rightarrow \text{Small Input Width}$$

and

$$\text{Low Sensitivity} \Rightarrow \text{Large Input Width.}$$

Alternatively, assume that the output *resolution* is fixed. In practice, the output width is constrained by the noise and quantization of the data acquisition system. It can then be seen from Fig. 13 that a static response curve with high sensitivity will have good input resolution, whereas a static response curve with low sensitivity will have poor input resolution; that is, for fixed output resolution

$$\text{High Sensitivity} \Rightarrow \text{Good Input Resolution}$$

and

$$\text{Low Sensitivity} \Rightarrow \text{Poor Input Resolution.}$$

Consequently, for a sensor and data acquisition system with fixed output width and fixed output resolution, there is a tradeoff between input width (the ability to measure inputs with large amplitude variation) and input resolution (the ability to distinguish between close inputs). This tradeoff is determined by the sensitivity of the static response curve.

Finally, for an affine static response curve

$$\begin{aligned}\text{output dynamic range} &= \frac{\text{output width}}{\text{output resolution}}\\ &= \frac{\text{sensitivity} \times \text{input width}}{\text{output resolution}}\\ &= \frac{\text{input width}}{\text{input resolution}}\\ &= \text{input dynamic range.}\end{aligned}$$

Hence, for an affine static response curve, the output dynamic range and the input dynamic range are equal and are called the *dynamic range*. For a nonaffine static

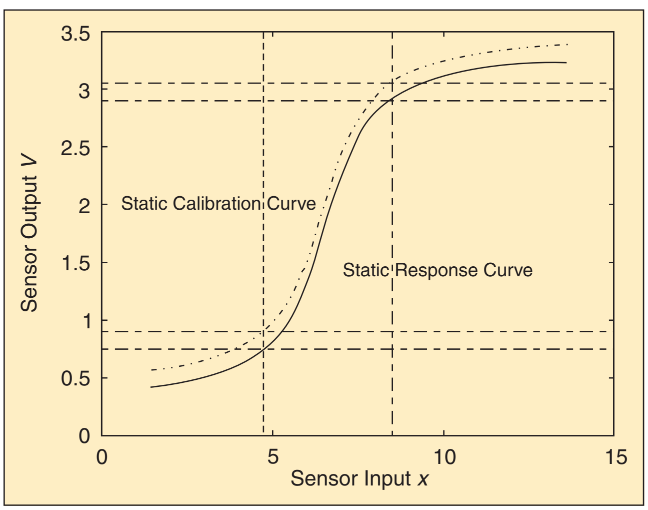

Figure 14. *For each input value, the static response curve output and the static calibration curve output differ by a constant voltage due to an output bias drift. Consequently, the output calibration error is constant over the input range.*

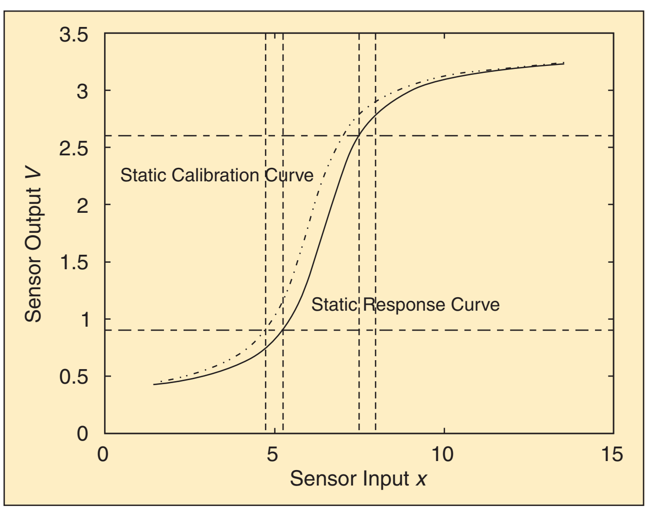

Figure 15. *For each output voltage, the static response curve input and the static calibration curve input differ by a constant value due to an input bias drift. Consequently, the input calibration error is constant over the output range.*

response curve, this equality is valid over a range of inputs within which the static response curve is approximately affine.

Static Calibration Curve

Since the static response curve of a sensor is generally unknown, one of the most important tasks that engineers face is determining an approximation of it. This process is known as *calibration*, and the approximation of the static response curve is the *static calibration curve*. Using the sensor output, the static calibration curve is used to obtain an estimate of the sensor input.

Any discrepancy between the static response curve and the static calibration curve gives rise to *calibration error*. As discussed above, drift, noise, hysteresis, and quantization corrupt the static response curve and make it difficult to obtain a static calibration curve with a satisfactory calibration error. Since the static response curve is unknown, it is also important to have *uncertainty specifications* that quantify the size of the calibration error.

When you purchase a sensor, the manufacturer usually provides specifications on a spec sheet. Sometimes these specifications are incomplete or questionable, and thus it may be necessary to verify them. To perform the calibration, you will need instrumentation that allows you to control the sensor input in a precise manner, or at least measure it accurately. This process may require extensive analysis, as well as a more expensive, precalibrated "truth" sensor, which you can use to determine the sensor input. The main point here is that determining a good static calibration curve and quantifying the calibration error in terms of uncertainty specifications may require substantial engineering effort.

A static calibration curve is *accurate* if it is a good approximation of the sensor response curve. The accuracy of a static calibration curve can be quantified in terms of the input and output calibration errors. For each value of the sensor input, the difference between the sensor output determined by the static response curve and the value given by the static calibration curve is the *output calibration error*; that is,

$$\text{output calibration error} = \text{sensor output} - \text{calibrated output},$$

where the *calibrated output* is the output determined by the static calibration curve for the given sensor input. Similarly, the *input calibration error* is defined by

$$\text{input calibration error} = \text{sensor input} - \text{calibrated input},$$

where the *calibrated input* is the estimate of the sensor input determined by the static calibration curve for the given sensor output. Since the sensor input is not known, neither of these quantities can be evaluated in practice. Therefore, it is important to determine uncertainty specifications in the form of bounds on the calibration errors.

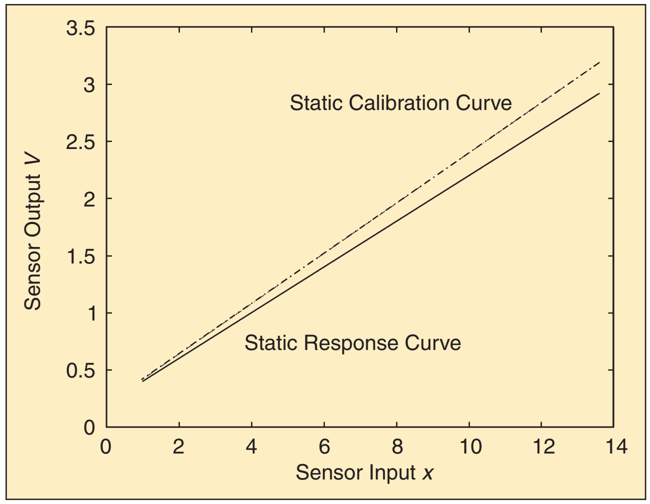

Figure 16. *For an affine static response curve and an affine static calibration curve, a mismatch between the sensitivity of the static response curve and the scale factor of the static calibration curve results in output and input calibration errors that increase or decrease over the output and input ranges.*

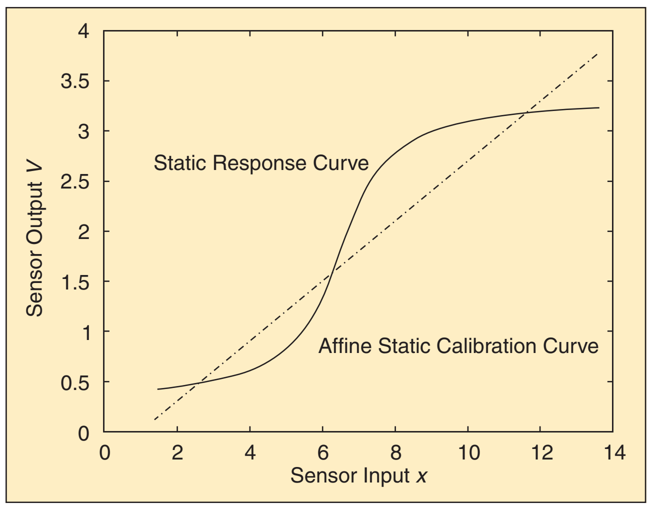

Figure 17. *Nonlinearity refers to the input or output calibration error that arises when an affine static calibration curve is used to approximate a nonaffine static response curve.*

If a sensor has an accurate static calibration curve, then it has good *accuracy*. It is easy to see that if a sensor has good accuracy, then it must have both good repeatability and thus good precision; that is,

$$\text{Good Accuracy} \Rightarrow \text{Good Repeatability} \Rightarrow \text{Good Precision}.$$

I leave it to the reader to construct counterexamples to show that the converse of these implications is not true.

The output and input errors may vary arbitrarily over the output and input ranges. However, it sometimes occurs that the static calibration curve is determined with good accuracy, but there is a subsequent output bias drift. In this case, as can be seen from Fig. 14, the output calibration error is constant over the output range. Similarly, environmental factors can affect the sensor in such a way that there is an *input bias drift*. In this case, Fig. 15 shows that the input calibration error is constant over the input range.

Another source of calibration error, which often occurs in practice, is a mismatch between the slope of an affine static response curve (i.e., its sensitivity) and the slope of the static calibration curve (i.e., its *scale factor*). In this case, Fig. 16 shows that a *scale factor error* results in output and input calibration errors that increase or decrease over the output and input ranges.

In practice, it is desirable to use an affine static calibration curve even if the static response curve is known to not be affine. In this case, the sensor accuracy is limited. However, an affine static calibration curve is easy to store in a computer, requiring only a single point value and slope. The *output nonlinearity* is the worst-case output calibration error associated with an affine static calibration curve; that is,

$$\text{output nonlinearity} = \max_{[\text{output range}]} \frac{\text{affine output calibration error}}{\text{sensor output}},$$

whereas the *input nonlinearity* is the worst-case input calibration error associated with an affine static calibration curve; that is,

$$\text{input nonlinearity} = \max_{[\text{input range}]} \frac{\text{affine input calibration error}}{\text{sensor input}}.$$

Nonlinearity is illustrated by Fig. 17, where an affine static calibration curve is used to approximate a static response curve that is not affine.

Conclusions

Measurement is one of the most fundamental activities of science and engineering since it is the process by which data about the real world are obtained. Sensors are needed for measurement, and the quality of the data depends on the sensor performance specifications. In this column, I have explained the concepts that quantify sensor performance specifications. It is my hope that this tutorial will be of value to students and engineers in working with technology and will increase appreciation for sensors as crucial components of virtually every engineering system.

Sensors and Measurement References

Historical perspectives on the development of measurement ideas and techniques are given in

- H.A. Klein, *The Science of Measurement: A Historical Survey*. New York: Dover, 1974.
- J.F. Keithley, *The Story of Electrical and Magnetic Measurements*. Piscataway, NJ: IEEE Press, 1999.

An extensive treatment of measurement concepts and techniques is given in

- E.O. Doebelin, *Measurement Systems: Application and Design*, 4th ed. New York: McGraw-Hill, 1990.

Additional references on sensors and measurement include

- P. Elgar, *Sensors for Measurement and Control*. Essex: Addison Wesley Longman, 1998.
- J. Fraden, *Handbook of Modern Sensors: Physics, Designs, and Applications*, 2nd ed. New York: American Institute of Physics, 1996.
- R.E. Fraser, *Process Measurement and Control: Introduction to Sensors, Communication, Adjustment, and Control*. New York: Prentice-Hall, 2001.
- E.E. Herceg, *Handbook of Measurement and Control*. New York: Schaevitz Engineering, 1976.
- S. Merhav, *Aerospace Sensor Systems and Applications*. New York: Springer, 1996.
- K.I.T. Richardson, *The Gyroscope Applied*. New York: Hutchinson's, 1954.
- P.H. Savet, Ed., *Gyroscopes: Theory and Design*. New York: McGraw-Hill, 1961.
- R.M. Setbacken, *Feedback Devices in Motion Control Systems*. Santa Barbara: Fithian, 1997.
- W. Stadler, *Analytical Robotics and Mechatronics*. New York: McGraw-Hill, 1995.
- C.P. Wright, *Applied Measurement Engineering*. New York: Prentice-Hall, 1995.

Although data analysis is the subject of probability and statistics, there are surprisingly few books aimed at the practical analysis of experimental data. Some useful references are

- E.L. Crow, F.A. Davis, and M.W. Maxfield, *Statistics Manual,* reprinted by New York: Dover, 1960.
- J. Mandel, *The Statistical Analysis of Experimental Data*. New York: Wiley, 1964, reprinted by New York: Dover, 1984.
- S. Rabinovich, *Measurement Errors: Theory and Practice*. New York: American Institute of Physics, 1995.
- B.P. Roe, *Probability and Statistics in Experimental Physics*. New York: Springer, 1992.

Acknowledgments

I would like to thank David Wilson and Tom Pilutti for helpful suggestions.

Dennis Bernstein teaches in the Aerospace Engineering Department at the University of Michigan in Ann Arbor. His current research interests include identification and adaptive control with application to aerospace systems. He has been a visiting faculty member at the University of Leeds and the University of Glasgow. His secondary interests include the history of technology, Native American history, and Celtic music.