The Perception Layer

Week

I

1 What is a sensor?

Section 1. What is a sensor?
Section 2. Types of sensors
Section 3. Characteristics of a sensor

Understanding what is a sensor is the trailhead of this curricular unit, and sets the tone for the discussion we want to engage in during it.

What would you call a sensor?

In a broad perspective, one could argue that a sensor would be any abstract device that reacts to something that happens in the environment. Yet, with such a broad definition you would consider almost anything as a sensor. Going to the very fundamentals of physics, and focusing on the fundamental interactions (gravity, weak force, electromagnetic force, and strong nuclear force) a simple atom would be a sensor for any process in physics. This extreme example shows that the previously mentioned definition is overly broad, and as a result, it becomes meaningless. To better define what is a sensor, a better question we can ask ourselves is to start with why, i.e.

Why do you need sensors?

We need sensors because we want to know more about the environment so we can make decisions, build models, or construct machines that are aware of their surroundings. Therefore, being more strict, a sensor is a device that not only interacts with the environment but allows you to get information from it in order to make a **measurement**. Therefore:

> **Definition. Sensor:** A sensor is a device that detects and measures physical properties or changes in a physical system, producing a usable output signal (typically an electric signal) in response to a *stimulus*. For example, a microphone is an acoustic sensor that measures pressure variations that reach the diaphragm associated with sound. Another example is a temperature sensor that measures the ambient temperature and outputs a signal related to that temperature.

In science and engineering, sensors are also closely related to the transduction principle, i.e., the conversion of one form of energy into another. Transducers and actuators are therefore important concepts to complete the backbone of any measurement and control system.

Sensor is a device that provides usable output in response to a specified quantity that is measured.

Instrument Society of America

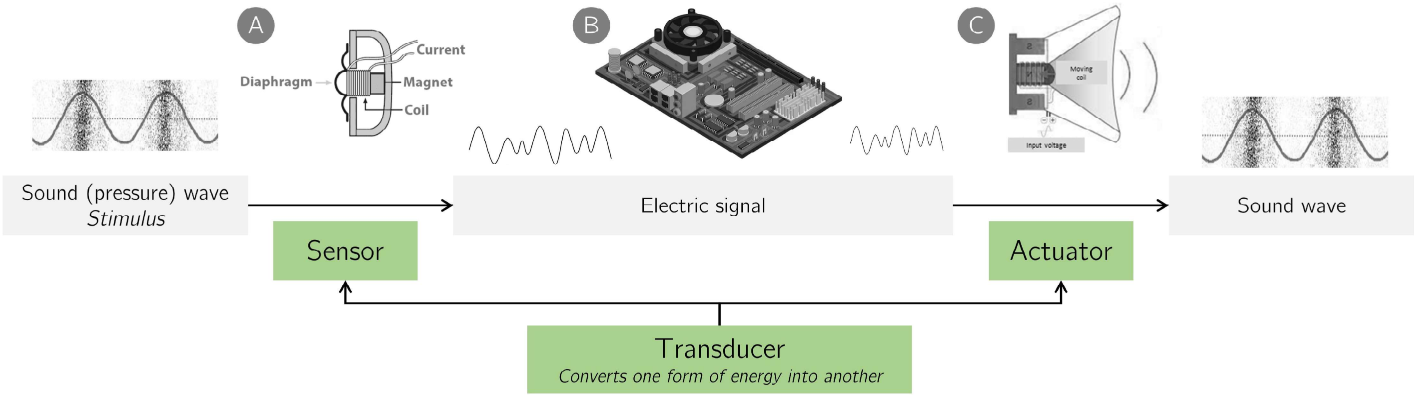

Figure 1. Sensors and actuators are both transducers. An illustrative and straightforward example is the sound components of our laptop. The sensor is the microphone, that converts sound (a signal transmitted as pressure waves) into electric signal. With this electric signal, some processing can be made to encode, decode, filter, amplify, storing, etc. The actuator in this case is the loudspeaker, which receives electrical signals and converts them into sound signals, with possibly altered properties.

Definition. Transducer: a device that converts one form of energy into another. For example, our microphone is a transducer that converts acoustic energy (sound waves) into an electrical signal. Yet, similarly, a loudspeaker can also be considered a transducer, but it is not a sensor.

Definition. Actuator: a device that produces a physical action or movement in response to a control signal. For example, an actuator may convert an electrical signal or other forms of energy into mechanical motion, as it happens in a loudspeaker. Another example is an electric motor in a robotic arm, that moves in response to signals from a control system.

All sensors are transducers, but not every transducer is a sensor.

At this point, our sensing machine has its backbone set but still incomplete. As we saw before, sensing is the process of getting information from the environment. Our definition of sensor utilizes two interesting concepts that we use every day in physics and engineering: **signal** and **measurement**. The signal definition is broad, but we can say that it refers to any quantity or process that conveys information, in the case of sensors, an electric signal. Measurement is the process of assigning a numerical value to a property or characteristic of an object or phenomenon (e.g. temperature, pressure, displacement). Ultimately, the sensor is only usable when we can translate a signal into a measurement.

Thus, measuring with a sensor has three steps:

1. the sensor is affected by the environment, outputting a response in the form of an electric signal;
2. the signal is transmitted to a processing unit;
3. the signal is processed and analyzed to convert raw information into the final measurement.

If we think about we will find this conceptual structure replicated everywhere, from man-made technology to nature's innerworkings. For example, human beings are equipped with 5 different types of sensors: eyes, which detect energy of the electromagnetic field[1]; ears that respond to acoustic pressure; a tongue and nose that are produce responses to the presence of given chemicals; and skin which can detect pressure and temperature. All the signals that come out of these sensory devices then travel through our body as electric signals using nervous pathways, reaching our nervous system where they are processed[2].

[1] mostly in the visible range, which is usually called light

[2] which may lead to a response by one of the many actuators in our body - such as a muscle.

2 Types of sensors

Grouping sensors into families is a challenge due to the vast diversity and specificity of sensors. Yet, such exercise offers significant learning opportunities while it also allows to establish common language across various fields of technology and science.

2.1 Active *vs* Passive

Active sensors require external power from an external source to operate. This signal is often referred to as an excitation signal. For example, a strain gauge is a sensor that explores the fact that a wire resistance changes when it is stretched or compressed. Thus, it is an active sensor as it requires passing a current in the sensing head to detect a resistance variation.

In contrast, passive sensors can output an electrical signal in response to the input stimulus without any additional power. The energy utilized is obtained directly from the measurand, as it happens for example in a thermocouple device, that makes use of the Seebeck effect.

2.2 Analogue *vs* Digital

An **analog sensor** generates an output signal that varies continuously within a specific range, typically producing a voltage that is directly proportional to the measured parameter. Parameters such as speed, temperature, pressure, and strain - inherently analog due to their continuous nature - align well with the capabilities of analog sensors. The output from these sensors changes in a smooth, continuous manner over time. However, the use of analog sensors may feature slower response times and often requires an external power supply for amplification (e.g. Op amp circuit) of the output signals.

On its turn, a **digital sensor** emits output signals in discrete digital form, essentially toggling between two distinct states: ON/1 and OFF/0. An illustrative example of a digital sensor is a push-button switch. Depending upon the bit depth employed to represent the measured parameter, digital sensors may feature higher accuracy compared to their analog counterparts. Besides, they are also preferable for direct integration into microcontroller-based systems. Note yet that it is still possible to integrate analog sensor outputs in similar manners by utilizing analog-to-digital converters.

2.3 Direct *vs* Hybrid

A sensor that incorporates more than one transducer to output an electric signal[3] is called **hybrid**, in opposition to a **direct** sensor.

[3] Can you give an example of a complex sensor? *Hint: Chemical sensors.*

2.4 External *vs* Internal

A sensor may be incorporated into a larger system with other sensors, transducers, processing units, and actuators. Considering the position in this system, the sensor is classified as **external** if it responds to the environment outside the system, or **internal** if it aims to monitor and measure the properties of its internal parts.

2.5 By operating physical principles

If we want to design a sensor from scratch, our better shot would be to look at physical laws and processes to act as the design unit of our sensing system. This approach entails identifying the conversion phenomena[4] that can be derived from various physical phenomena involving interrelated variables to react to a stimulus. This diversity can also be utilized to group sensors as shown below in the table.

[4] not exclusive

Exercise 1. Consider the sensors in your smartphone: name some of them, their function, and their type (active/passive, analog/digital, direct/hybrid, extrinsic/intrinsic).

3 Characteristics of a sensor

When we are selecting a sensor, the above qualitative classification is often secondary. Indeed, the characteristics/specifications of the sensor and how they compare with those required for the application are the most important factors. These characteristics of the sensors may be divided into **static** and **dynamic** characteristics.

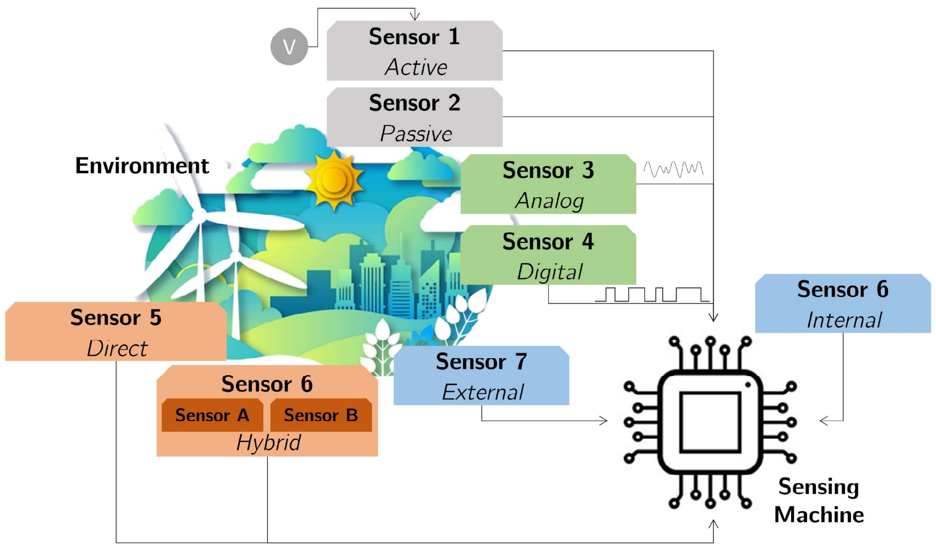

Figure 2. Wrapping up the qualitative classification of sensors into a single scheme.

Conversion Phenomena		Stimuli	
Physical	Photoelectric Photomagnetic Magnetoeletric Eletromagnetic Thermo-optic etc	Electric	Charge Current Voltage Conductivity Permittivity etc
Chemical	Chemical reaction Spectroscopy Electrochemical process Physical transformation etc.	Mechanic	Position Acceleration Force Mass etc.

Table 1. Examples of conversion phenomena and stimuli.

The **static** characteristics of a sensor entail the algebraic relation between the input - e.g. a stimulus S and its output - a signal E. A characteristic curve can then be written and graphically drawn as

$$E = F(S) \tag{3.1}$$

and the function F is denominated the **transfer function** of the sensor. Measuring is, therefore, nothing more than inverting the relation, i.e. $S_m = F^{-1}(E)$ being S_m the measurement. A sensor is normally provided by the

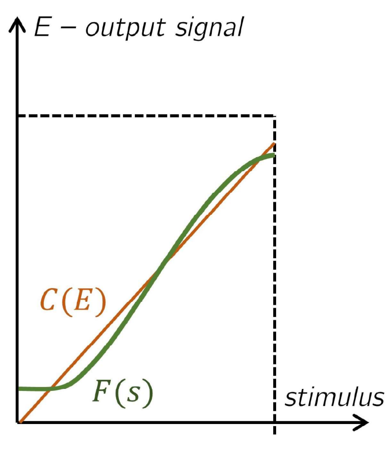

Figure 3. Transfer function and calibration.

manufacturer with a transfer function which is obtained using a calibration process. The provided mathematical model to approximate $F^{-1}(E)$ is called the **calibration curve**, which is normally a linear model $C(E) = mE + b$. Note however that the transfer function itself is often nonlinear[5] and may even be **multidimensional**.

[5] e.g. $F(S) = Ae^{bS}$

Dynamic characteristics on the other hand refer to the sensor response when the input changes. Usually, these are often expressed in terms of response to certain standard input signals (e.g. response to a step signal, frequency response bandwidth, possible damping, or overshooting - see figure 4).

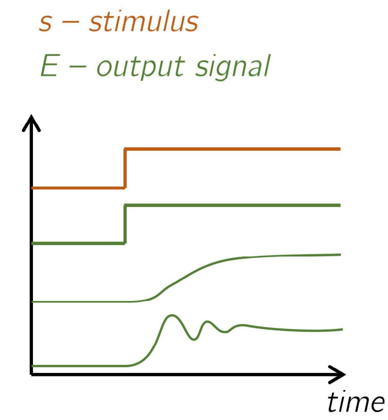

Figure 4. Various types of dynamic responses to a step stimulus.

3.1 Ranges and full-scales

A class of important static characteristics of a sensor aim to describe its range of operation. Depending on the perspective we are using or looking at the sensor, we may establish a few distinct characteristics:

Span/Full-scale Input
It indicates the difference between the minimum and maximum stimulus, s_{min} and s_{max} respectively, that produces a measurable response of the sensor, without causing unacceptably large error.

$$Span = s_{max} - s_{min}$$

Full-scale Output
For an analog output it indicates the difference between the minimum and maximum output signal generated. For a digital output, it is related to the maximum digital count that the convertor can resolve for the span.

$$FSO = E_{max} - E_{min}$$

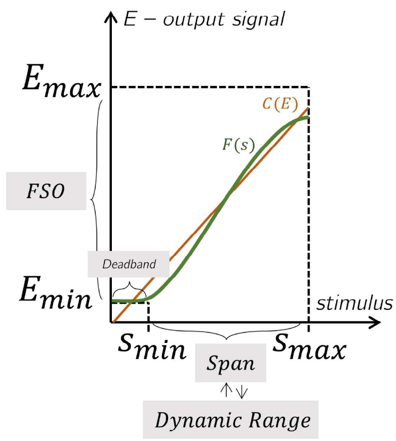

Figure 5. Range characteristics.

Dynamic Range
Usually, as the above scales span across multiple orders of magnitude, the ratio of the largest measurable stimulus to the smallest measurable stimulus may be a better metric for measuring the full-scale of a sensor. This is usually referred to as **dynamic range** and is given in decibels, which is a logarithmic measure of ratios of either powers or amplitudes. By definition, for powers, it is ten times the log of the ratio of powers,

$$\text{Dynamic Range} = 10 \log_{10} \frac{P_{max}}{P_{min}} dB \tag{3.2}$$

while for amplitudes(e.g.voltages) we have 20 times the log of the ratio of the minimum and maximum values, i.e.

$$\text{Dynamic Range} = 20 \log_{10} \frac{V_{max}}{V_{min}} dB \quad (3.3)$$

where P_{min} and V_{min} is the smallest measurable signal, typically defined as that equal to the noise level. In optics it is also common to use $P_{min} = 1mW$ with the unit being called **dBm** in that case.
Although the dynamic range or range normally applies to the span, it may also be applied to the full-scale output(it should be clearly stated in the spec sheet).

Deadband
A deadband (also known as a dead zone or a neutral zone) is a band of input values in the domain of a transfer function of a sensor for which the output value is 0 or equal to the noise (i.e. signal to noise ratio equal to 1).

3.2 Tranfer function characteristics

Another set of important static characteristics of a sensor aims to describe how the sensor responds to the stimulus and how it relates to the true value.

Sensitivity
Sensitivity is a relationship between the input physical signal and the output electrical signal. It is the ratio of the change in the output of the sensor to the input value and mathematically can be expressed as

$$Sensitivity = \frac{dE}{ds} \quad (3.4)$$

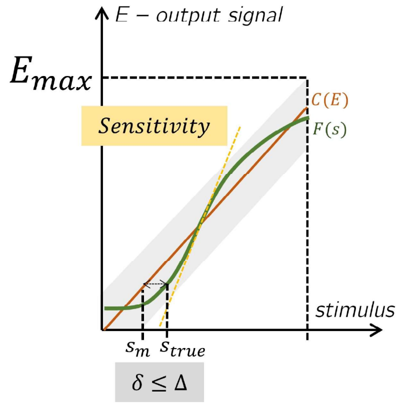

Figure 6. Sensitivity and accuracy.

Accuracy
It is the degree of exactness between actual measurement and true value. Note that the accuracy corresponds to a metric computed from the measured stimulus and not from the signal obtained itself, i.e. $\delta = s_m - s_{true}$. This means that accuracy may depend on the calibration function but also on the uncertainty of the measured signal. Besides, in typical real-world conditions, accuracy δ may vary with the stimulus as we will later discuss in this curricular unit.
Accuracy may be expressed in multiple manners, from absolute values to percentages of the full-scale input. The most common way to estimate a maximum value Δ is to calculate the calibration function and maximize

$|\delta|$ for the calibration dataset, which will be in already in the units of measurement.

Linearity
When we talk about linearity we are referring to how nonlinear is the transfer function of the model. This corresponds to the deviation of the curve of actual measurement from the curve of ideal linear measurement and can be expressed in percentage of the span. Normally it is tightly connected with the dynamic range.

Saturation
Saturation is intrinsically related to the operation limits of the sensor and is identified by the lack of variation of the output signal above a threshold of the input stimuli.

Hysteresis
The hysteresis of a sensor can be defined as the maximum difference of the measurable value when approaching the point in opposite directions, i.e. first with increasing and then with decreasing the input stimuli. Hysteresis is commonly related to design, friction, or structural changes in materials, such as plastic deformation for example.

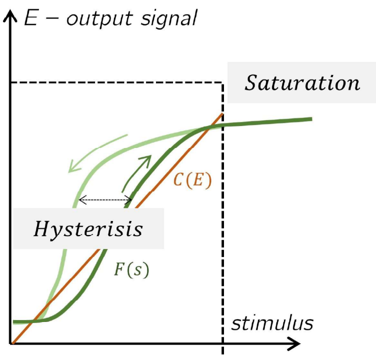

Figure 7. Hysterisis and Saturation.

3.3 Operation characteristics

Repeatability
Repeatability or reproducibility is the inability of the sensor to produce the same output for the same stimuli under similar conditions. This may be related with thermal noise or charge accumulation for example. Typically it is computed as the maximum difference of measured stimulus for two distinct runs and can be represented in the same units or in a percentage of the scale.

Stability and Drift
Over time, the characteristics of a sensor can change, which may affect the accuracy of the measurements. This may happen in short (typically minutes and hours) or long (days to years) timescales, for which the degradation in accuracy is commonly denominated stability (also short-term stability) or drift respectively.

Reliability

Reliability is closely related to the concept of stability and measures the sensor's ability to perform as expected over a period of time. Common metrics are presented in time instead of accuracy, for example, the mean-time-to-failure

$$MTTF = \frac{1}{n}\sum_i t_f^i \tag{3.5}$$

with t_f the time of failure of device i of n devices in total, starting the test at $t = 0$. It may also include parameters under extreme conditions(such as temperature, humidity, corrosion, etc.).

3.4 Dynamic Characteristics

Dynamic characteristics include a vast number of characteristics that allow one to characterize the response of a sensor to a time-varying signal. Ultimately they are related to the components constituents, principles of operation, and governing equations[6]. There are multiple parameters and strongly depend on the final application. For the sake of simplicity, we will discuss two of the most important ones: the rise time and bandwidth.

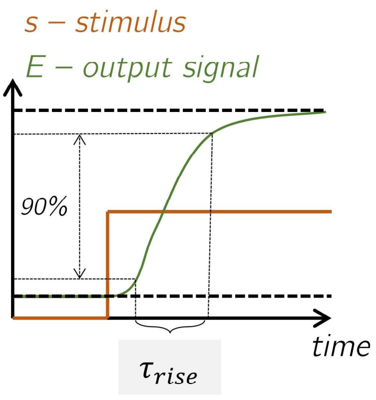

Figure 8. Rise time.

[6] e.g. first order or second order differential equations

Rise Time

The rise time τ_{rise} is defined as the time it takes for the output signal waveform transitions from 10% to 90% (sometimes 20% to 80%) of the final value when the input is a step signal. It can be connected with the concept of **slew rate** in electronics, which is the resulting rate of change of the output signal, i.e. $\Delta V/\tau_{rise}$.

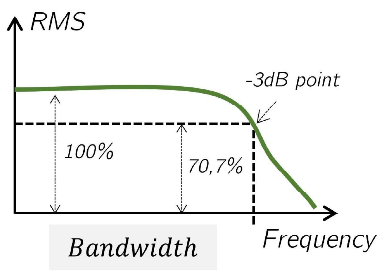

Figure 9. Bandwidth.

Bandwidth

The bandwidth of a sensor indicates how sensors respond to signals at different frequencies. It is usually measured at the -3dB point, i.e. the output signal drops $\approx 30\%$ in amplitude with the increase of frequency. Normally we only look at the upper frequency limit and assume that the sensor works well in continuous response.

Exercise 2. In the class we will distribute a sensor for each group, accompanied by a component spec sheet if possible. In groups, prepare a short presentation about your sensor, focusing on the *stimulus* to be measured, outlining the working physical principle, classifying the type of sensor qualitatively, and presenting some characteristics (from 3 to 5) of the sensor.

Concluding remarks

In this first week, we focused on understanding what is a sensor, and how to characterize one in qualitative and quantitative terms. As we have seen a sensor is not complete without a calibration function, that tries to approximate the transfer function of the sensor. This is only possible when one applies known stimuli, and analyzes and processes the output signal, extracting the necessary information from the signal that allows one to construct the mathematical model by numerical curve fitting. That is the motivation for next week.