1.Sensor Systems

Machines interact with their surroundings. Basically they manipulate it by imposing power. The imposed power changes the energy levels in system components. Eventually, the changes result in variation in observable state variables. This process is fundamental in mechatronics, where machines act aiming at some level of autonomy and intelligence.

Mechatronics systems use above manipulation routine purposefully. The purpose mentioned is actually a generic one but it can be abstracted any desired state in system. Some kind of information processing unit (computer in large) takes this desired state and perform logical and mathematical actions to activate some control electronics unit with a command signal. A powerful transducer at the output of the control circuitry injects power to the plant or process to be manipulated. Of course the one way action process is deeply questionable because continuous energy transfer may result in various scenarios so somehow it is necessary to now the reaction of the plant or process. The acquired information is then used for decision making and information processing. Gathering raw data, signal processing and obtaining information are basics steps in the information processing, which are carried on measurement or sensor systems.

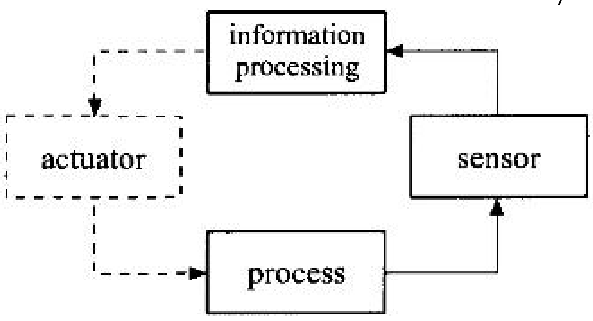


Figure 1. Mechatronics system information flow diagram

More technically, sensors and the associated measuring systems provide the required measurable information about the process in mechatronic systems. They represent an essential link between the process and the information-processing part of mechatronics systems.

Sensors that measure mechanical or thermal quantities and transform them into an electrical signal are of special importance for mechatronic systems.

2.Measuring System

The purpose of a measuring system is to observe and quantify a variable physical quantity (called a measurand) and to process the obtained information.

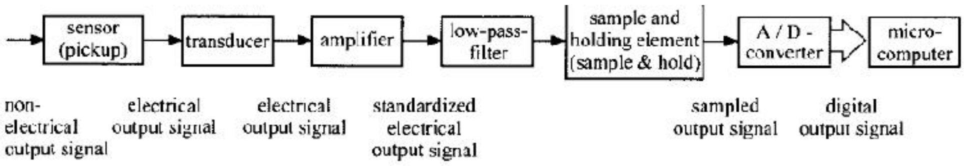


Figure 2. Measuring system components

- The first element of this system is the sensor or sensing element (increasingly used instead of "pickup"). Its primary function is to detect the measurand and transform it into a suitable signal, Mechatronic systems generally rely on sensors with an electrical output signal. The characteristics of the output signal depends on the measurement principle of the sensor.
- Transducers and amplifiers transform the electric sensor output signal into a standardized electrical signal, e.g., 0...20 mA or 4...20 mA or 0...10 V, which is more suitable for further processing.
- If high-frequency disturbances contaminate the usable signal, a low-pass filter is applied in order to decrease the influence.
- A sample and hold device and an analog-to-digital converter are necessary if the sensor signal is to be processed by a microcomputer.

Consumer goods and low-cost appliances do not require high precision measurement and a modular arrangement of the measuring system.

3.Classification Of Sensors

Because of the broad spectrum of metrology, it is difficult to classify sensors and the corresponding signal processing devices. An overview of the classification of some important measurands might be:

- Mechanical quantities;
- Thermal/caloric quantities;
- Electrical quantities;
- Chemical and physical quantities.

4. Sensor Properties

The transformation of non-electrical quantities into electrical ones depends on physical or chemical effects. These may be divided into main and side effects.

- The main effect is responsible for generating the desired measuring signal,e.g., the electrical voltage of a piezoelectric pressure sensor.
- The disturbing side effects are frequently superimposed, e.g., the influence of temperature changes. The design process for sensors needs to take these side effects (sometimes called "cross sensitivity") into account. Their influence should have only little effector should be compensated by appropriate measures.

The criteria for evaluating sensors are:

- static behavior;
- dynamic behavior;
- quality class,
- measuring range;
- overload capacity;
- compatibility with associated components;
- environmental influences;
- reliability.

A. Dynamic Behaivour

The dynamic behavior is described by a sensor's frequency response or simple characteristic values, e.g., cut-off frequencies or time constants. The sensor dynamics have to be adjusted to the process and the measuring task.
Mathematical model of the sensor system corresponds to a transfer operation, described via differential equations, termed as transfer function and produces an effect resulting in controlled output or response.

$$L\left\{\sum_{i=0}^{n} a_i x^{(i)} = \sum_{j=0}^{m} b_j u^{(j)}\right\} \Rightarrow G(s)\frac{x(s)}{u(s)} = \frac{b_m s^m + b_{m-1}s^{m-1} + \ldots + b_1 s + b_0}{a_n s^n + a_{n-1}s^{n-1} + \ldots + a_1 s + a_0}$$

Thus the cause and effect relationship between the output and input is related to each other through a transfer function. The transfer function of a system is defined as the ratio of the Laplace transform of the output variable to Laplace transform of the input variable assuming all initial conditions to be zero. It corresponds to differential equation in Laplace Domain.

$$\underbrace{a_n x^{(n)} + \ldots + a_1 \dot{x} + a_0 = b_n u^{(n)} + \ldots + b_1 u + b_0}_{\text{differential equation}} \Rightarrow \sum_{i=0}^{n} a_i x^{(i)} = \sum_{j=0}^{m} b_j u^{(j)}$$

$$L\left\{\sum_{i=0}^{n} a_i x^{(i)} = \sum_{j=0}^{m} b_j u^{(j)}\right\} \Rightarrow \frac{x(s)}{u(s)} = \overbrace{\frac{b_m s^m + b_{m-1}s^{m-1} + \ldots + b_1 s + b_0}{a_n s^n + a_{n-1}s^{n-1} + \ldots + a_1 s + a_0}}^{\text{transfer function}}$$

Figure 3. Transfer function and differential equation duality

<u>Time Response:</u> Time response is defined as the response of the system to the singular inputs like impulse and step. The characteristics of the inputs are that they represent a sudden change at initial time instant. Hence, the system response exhibits transient and steady state behavior showing the ability to system responding to sudden change. Time response are relatively easy to calculate using Laplace transformation and Heaviside formulas resulting following time signal as output.
Although the order of the differential equation is critical for time evaluation of the output signal, only two specific time response is assumed to have dominating importance. The first order and second order underdamped responses
The first order system is modelled with following differential equation:

$$a_1 \frac{dx}{dt} + a_0 x = b_0 u \quad (1)$$

with given solution to step input with o initial condition:

$$x(t) = K(1 - e^{\frac{t}{\tau}})$$

Where two characteristics of the system that effects the response are appeared as K: gain and τ:time constant.
Gain of the system is defined as the ratio of input quantity to output quantity and it is also named as sensitivity.
The time constant of the system, which has units of time, is the system parameter that establishes the time scale of system responses. It is defined as the duration from initial time to the time instant where the output reaches 63% of its final value. It also defines the steady state time of the system with a multiplication factor of 4τ(for 98%) and 5τ(for 99.3%).

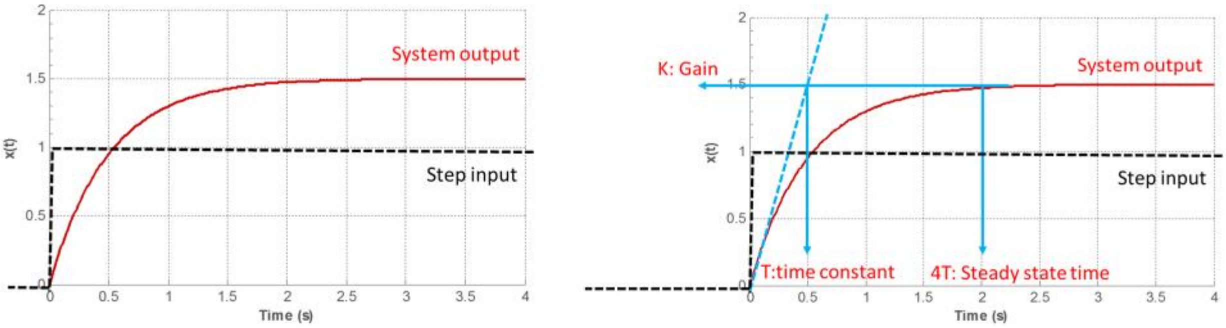


Figure 4. Time response of the first order system

The second order system response can also be modelled in a similar way.

$$a_2 \frac{d^2x}{dt^2} + a_1 \frac{dx}{dt} + a_0 x = b_0 u$$

The homogenous solution gives two poles as follows:

$$s_{1,2} = -\frac{a_1}{2a_2} \pm \frac{\sqrt{(a_1)^2 - 4a_2 a_0}}{2a_2} = -\xi\omega_n \pm \omega_n\sqrt{\xi^2 - 1}$$

The parameterized poles are expressed with two terms: ξ: damping ratio and ω_n:natural frequency.
The natural frequency defined as the frequency of the hypothetical oscillations that system represents when no input applied and there is no damping. It can be calculated with following formula (in rad/s):

$$\omega_n = \sqrt{\frac{a_0}{a_2}}$$

The damping ratio is defined as a ratio of the damping factor (generally represented by the coefficient of the first order term a_1) to the critical damping factor (a_1 value) that makes polynomial full square (double root).So the critical damping factor is calculated as follows:

$$C_{cr} = 2\sqrt{a_0 a_2}$$

Using the critical factor, the damping ratio formula is straight forward:

$$\xi = \frac{a_1}{2\sqrt{a_0 a_2}}$$

It is evident that the damping ratio may change the solution of differential equation in three ways: overdamped system (ξ>1), critically damped (ξ=1),under damped (ξ<1). Among three, the underdamped system response is of special response due to its oscilating nature. The time response function of underdamped second order system for step reponse is as follows:

$$s_{1,2} = -\xi\omega_n \pm i\omega_n\sqrt{1-\xi^2}$$

$$x(t) = K\left(1 \frac{1}{\sqrt{1-\xi^2}} e^{-\xi\omega_n t} \sin(\omega_n\sqrt{1-\xi^2}\, t + \phi)\right)$$

Where gain K is $K = \frac{b_0}{a_0}$ by definition and $\phi = \text{Arccos}(\xi) \Rightarrow \phi = \tan^{-1}\left(\frac{\sqrt{1-\xi^2}}{\xi}\right)$

These two charcteristics, natural frequency and damping ratio, effects observable features on sytem response as seen in figure 4b. These features are rise time, overshoot value and steady state time.

The definition of rise time is "that time taken for a system output to rise from 10% to 90% of its final value when stimulated by a step input". This measurement is useful because it is easy to measure on an oscilloscope and can be applied to any linear system. The rise time is inversely proportional to the system bandwidth, i.e. the wider bandwidth, the smaller the rise time. However, designing systems with wide bandwidth is costly, which indicates that systems with very fast response are expensive to design. <u>The rise time is inversely proportional to the system bandwidth, i.e. the wider bandwidth, the smaller the rise time. However, designing systems with wide bandwidth is costly, which indicates that systems with very fast response are expensive to design.</u>

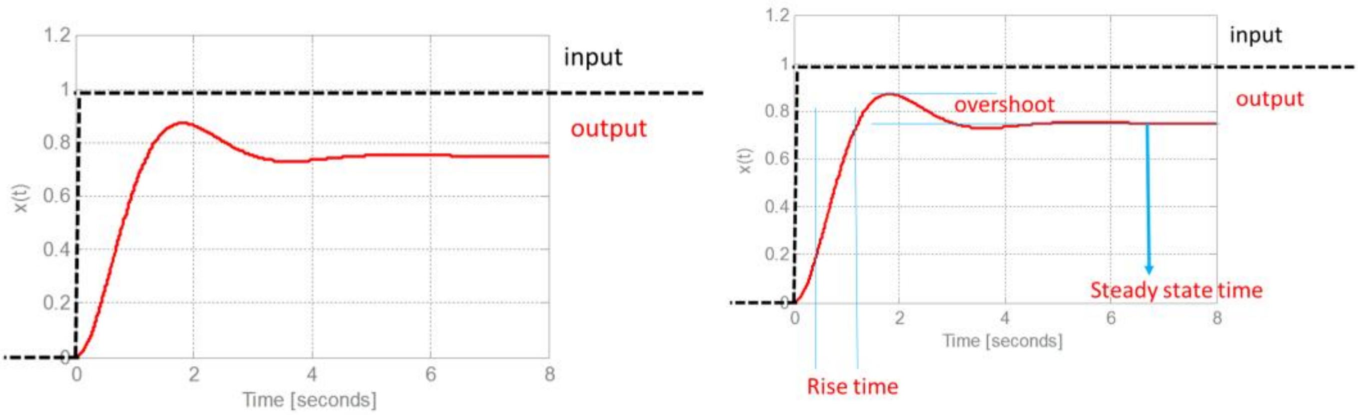


Figure 5. Time response of the first order system

The overshoot is the maximum peak value of the response curve measured from steady state value of the response. The maximum percent overshoot is an accepted measure for overshoot and iti is calculated as follows:

$$OV\% = \frac{Maximum\ Value - Steady\ State\ Value}{Steady\ State\ Value} \times 100$$

$$OV\% = e^{-\frac{\xi\pi}{\sqrt{1-\xi^2}}} \times 100$$

The settling time is the time required for the response curve to reach and stay within 2% of the final value.

$$T_{settling} = \frac{4}{\xi w_n}$$

<u>Frequency Response</u>: It is known from systems' theory that when a harmonic input is applied to a linear system, the output in steady state has also harmonic form with different amplitude and phase. The frequency response of a system ideally gives the amplitude and phase change of input for a specific angular frequency. Frequency response can be calculated by replacing s in transfer function with jw and rearranging the terms in form of complex function.

$$G(s) = \frac{b_m s^m + b_{m-1} s^{m-1} + \ldots + b_1 s + b_0}{a_n s^n + a_{n-1} s^{n-1} + \ldots + a_1 s + a_0} \Rightarrow F(jw) = G(s = jw)$$

$$F(jw) = Re(F(jw)) + jIm(F(jw))$$

$$F(jw) = F(w)\angle\phi(w)$$

In most cases, the frequency response analysis of a circuit or system is shown by plotting its gain, that is the size of its output signal to its input signal, Output/Input against a frequency scale over which the system is expected to operate. Graphical representations of frequency response curves are called Bode Plots and as such Bode plots are generally said to be a semi-logarithmic graphs because one scale (x-axis) is logarithmic and the other (y-axis) is linear (log-lin plot) as shown. Using given procedure gain plots for second and first order systems:

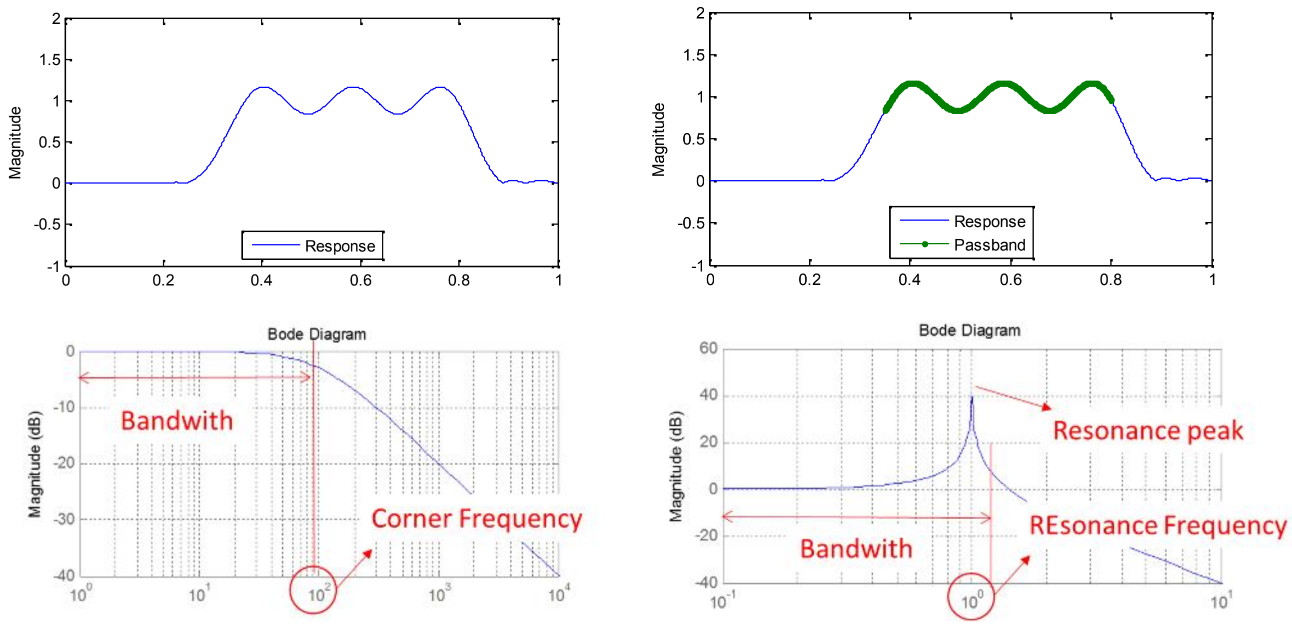


Figure 6. Bode gain plots of first and second order systems

Due to realization limitations all systems have a frequency band limited constant gain frequency response. The width of the flat region on frequency response in frequency space is called bandwidth. Frequency points f_L and f_H relate to the lower corner or cut-off frequency and the upper corner or cut-off frequency points respectively were the circuits gain falls off at high and low frequencies. These points on a frequency response curve are known commonly as the -3dB (decibel) points. These -3dB corner frequency points define the frequency at which the output gain is reduced to 70.71% of its maximum value. Then we can correctly say that the -3dB point is also the frequency at which the systems gain has reduced to 0.707 of its maximum value. The -3dB point is also know as the half-power points since the output power at this corner frequencies will be half that of its maximum 0dB value.

For second underdamped systems, at a specific frequency, the gain coming from frequency response becomes maximum. This phenomenon is called resonance, in which a vibrating system or external force drives another system to oscillate with greater amplitude at a specific preferential frequency. Frequencies at which the response amplitude is a relative maximum are known as the system's resonant frequencies or resonance frequencies. At resonant frequencies, small periodic driving forces have the ability to produce large amplitude oscillations.

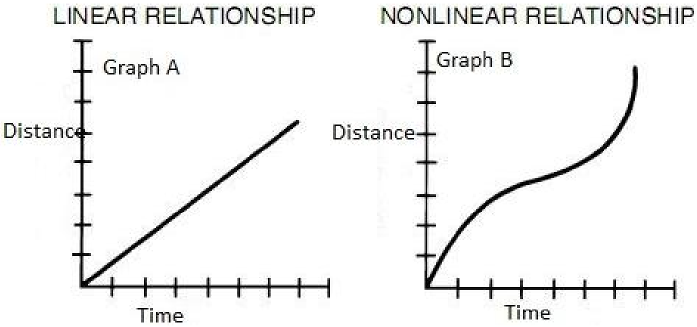
$$\omega_R = \omega_n \sqrt{1-2\xi^2}$$

$$M_R = \frac{1}{2\xi\sqrt{1-\xi^2}}$$

B. Static Behaivour

A sensor's static behavior ignores transient state of the sensor response and gives directly its input output relationship. In otherwords, the input excitation is related to steady state value of the system response. The static beahivor of the system can be expressed with different models such as exponential or power equation model.

For linear systems, the differential equations turn to line equation. The first thing which must be investigated in this case is to validity of the line equation. The static behavior characteristics "linearity" represents this property.



Linearity: The term "linearity" actually means "nonlinearity and Nonlinearity is specified for sensors whose transfer function may be approximated by a straight line and mathematically defined as maximum deviation (L) of a real transfer function from the approximation straight line.

A straight line input output equation specifies two static behavior characteristics:

Offset of an instrument is a specified as its error when measuring zero magnitude input

Sensitivity of an instrument is the ratio of the readout form the instrument to the change in the measured variable causing this. Sensitivity is also defined as the output over the input and therefore it is named as the gain of the instrument.

In many cases, a nonlinear sensor may be considered linear over a limited range. Over the extended range, a nonlinear transfer function may be modeled by several straight lines. This is called a piecewise linear approximation via Taylor Series Linearization method with formula:

$$F(x = x_0) = f(x_0) + \frac{df(x)}{dx}(x - x_0)$$

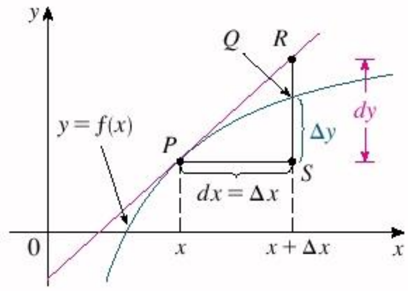


Figure 8. Taylor Series Approximation

In piecewise linear approximation, the sensitivity and offset values change with operating (linearization) points.

The other characteristics related to the sensor's static behavior are:

Span (Full-Scale Input): A dynamic range of stimuli which may be converted by a sensor is called a span or an input full scale (FS). It represents the highest possible input value that can be applied to the sensor without causing an unacceptably large inaccuracy

Full-scale output (FSO) is the algebraic difference between the electrical output signals measured with maximum input stimulus and the lowest input stimulus applied

A *hysteresis error* is a deviation of the sensor's output at a specified point of the input signal when it is approached from the opposite directions (Fig. 2.4). For example, a displacement sensor when the object moves from left to right at a certain point produces a voltage which differs by 20 mV from that when the object moves from right to left. If the sensitivity of the sensor is 10 mV/mm, the hysteresis error in terms of displacement units is 2 mm. Typical causes for hysteresis are friction and structural changes in the materials.

Resolution is the smallest change in input signal needed to produce a change in the output signal.

In the real world, any sensor performs with some kind of imperfection. The real transfer functions of the sensors differs from manufacturer specs.

Calibration means the determination of specific variables that describe the overall transfer function.In a calibration process, a known input value is applied to a measurement device/instrument for the purpose of observing the system output value. It establishes the relationship between the input and output values. The known value used for the calibration is what is defined in the standard.

C. Statistical Propeerties

The quality class gives a basic measure about a sensor's accuracy. It is the percentage maximum error of a measurement with reference to the full scale.

Accuracy: Accuracy is the closeness of agreement between a measured value and the true value. Accuracy error is formally defined as the measured value minus the true value. The accuracy error of a reading (which may also be called inaccuracy or uncertainty) represents a combination of bias and precision errors.The overall accuracy error (or the overall inaccuracy) of a set of readings is defined as the average of all readings minus the true value. Thus, overall accuracy error is identical to systematic or bias error. Root mean bias error can be calculated for a sensor as a measure of accuracy of a sensor system.

$$\mathrm{RMBE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{(x_i - x_{true})}{x_{true}}\right)^2}$$

Precision: Precision characterizes the random error of the instrument's output. Precision error (of one reading) is defined as the reading minus the average of readings. Thus, precision error is identical to random error. The most common measure for calculating precision is to use standart deviation:

$$S = \sqrt{\frac{\sum_{i=1}^{n} d_i^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}}$$

Applications for consumer goods don't need a high accuracy (2% to 5% is sufficient). Industrial applications, on the other hand, require a much higher accuracy (0.05% to 1%).

D. Other Properties

The measuring range describes the range in which the sensor's specifications are met. The overload capacity specifies the range in which a sensor may be operated with acceptable changes in the sensor's characteristics without damaging to itself. Typical overload capacities are between 200% and 500%.

A sensor's compatibility depends on the output signal type. Environmental influences, e.g., temperature, acceleration, corrosion, contamination, wear and tear, are especially important.

The reliability of a sensor is described by characteristic parameters, e.g., the "mean time between failures" (MTBF in [h] or its reciprocal value the mean failure rate ([h-~])).