Lectures 1 and 2: Instrument Systems and Limits to Measurement

A system is completely defined by its input and output specifications. For example, an instrument system might measure, record, and display a physical variable. The exact way this is done is unimportant as long as the specifications are met. We often deal with components of an instrument system such as sensors, amplifiers, or recorders but in general these are not complete and their characteristics depend on what they are connected to. The concept of a system being defined by its specification is useful because it frees us from thinking about specific implementations. Instead we may treat the system as a black box and concern ourselves solely with the inputs and outputs.

The input to an instrument system is the physical variable to be sensed. Thus the input to a thermometer is temperature, to an acoustic receiver is sound pressure, and to a current meter is velocity field. That the thermometer has self-heating and also may be velocity dependent is a problem to be met inside the system but should not concern us at this stage. For simple measurements, there may be little benefit in thinking of the physical variable independent of the sensor but in more complicated measurements there is an advantage. For example, the physical quantity may be highly variable and require that a long observation be made to obtain the desired statistical significance. This must be recognized immediately and not be masked by the averaging characteristics of some sensor. Furthermore, a clear understanding of the physical variable will aid in sensor design.

The sensor in an instrument system converts the physical signal to a more easily manipulated form. The use of the term signal in this context implies a separation of the physical variable into a meaningful part, the physical signal, and a non-meaningful part, noise. This separation often starts in the sensor. Any introduction of false information at the sensor or loss of true information is generally impossible to correct with subsequent processing. So the behavior of the sensor is one of the most critical concerns in an instrument system. The physical signal is composed of direct variations in a physical property such as pressure, concentration, or temperature. The output of the sensor is a voltage, displacement, resistance change, or other easily amplified, averaged, or stored characteristic.

After the physical signal is transformed to some other form by the sensor, it can be conditioned by amplifying, filtering, correlating, or sampling. Amplifying can generally be done with little degradation of signal to noise ratio. The term signal to noise ratio defines the ratio of useful information to unwanted information. Extra information added by the amplifier is unwanted. So if it adds information, this is noise. Similarly if in addition to amplifying the signal, it loses a part of it, this reduces the signal to noise ratio. Broadband electrical noise is generated by thermal excitation of electrons in a resistor. Because electronic amplifiers have resistive components, this thermal noise is introduced by amplification. It becomes the limit of the system in those cases where the sensor signal has very low level (as in the case of radio or acoustic receivers) and care must be taken to deal with this noise source and not permit it to become worse than the theoretical limit.

The thermal noise of a resistor is called Johnson noise. Considered as a current source in series with the resistor, or as a voltage source in parallel with the resistor, two expressions for the noise can be written:

$$i_n^2 = 4KTB/R; \quad e_n^2 = 4KTBR$$

where i_n^2 = mean squared noise current (amps)
e_n^2 = mean squared noise voltage (volts)
K = Boltzmann's constant - $1.38 * 10^{-23}$ J/K
T = temperature (K)
B = bandwidth (Hz)
R = resistance (ohms)

From this it can be seen that cooling the amplifier may be required for achieving the best noise figure and this is done for certain space radio receivers and in some photo detectors such as CCD arrays. However for the rest of us, the only reasonable thing to do is lower the bandwidth.

Bandwidth is the range in frequency passed by the system. Analog filtering and sampling determine the bandwidth. In a communication system, the bandwidth determines how fast information can be transferred and there is a tradeoff between a fast system and a reliable one. The simplest way to limit bandwidth is with a low pass filter since the bandwidth below a certain frequency is limited to that frequency while the bandwidth above is infinite. In high frequency systems, a narrow band filter is used.

When the signal has been filtered, it is often sampled. In modern instruments, it is generally digitized because the signal to noise ratio for digitized data is very high and can be maintained through storage and subsequent processing with little degradation. At this point the problem of aliasing arises. The sampled data contains limited information about signal at frequencies higher than the sampling frequency. In fact, if there is signal at higher frequency than the sampling frequency, it will appear as energy at some lower frequency and thus degrades the data at the lower frequency. This is almost always undesirable. The solution is to filter before sampling and sample at twice the frequency of the high frequency limit of the signal passed by the filter. The high frequency limit may be imposed by the nature of the physical signal, by the response of the sensor, or by a filter in the signal conditioner.

A system should have an output that is either directly coupled to a human observer or is indirectly coupled to one through subsequent processing. This defines the output of the system. It might be a display or a data port. But before final presentation, some human engineering is required to match the display to the capabilities of the observer. Modern commercial instruments do this reasonably well but prototype instruments are sometimes incomplete in this area.

Recording in temporary form is required in a large class of instruments used in oceanography, those unattended and submerged. Other systems must deal with permanent data storage. Recorders are used to do this storage. They have data limits. However the earlier in the processing sequence that the data can be stored, the more chance there is for subsequent processing. This is in conflict with the need to extend the deployment time and the fixed data capacity. Important compromises must be made here. An extreme case is sometimes heard where it is desired to record everything, everywhere, all the time. This does not require an instrument system; it is the universe running in real time. Consideration of the requirements of an eventual human observer helps resolve some of these conflicts. If the observer can only hope to deal with a small fraction of the possible observables, some reduction in what is recorded may be reasonable. In fact the increase in data capacity of hard drives and compact flash memory brings home the realization that the bottleneck in data capacity may be the human ability to unpack and absorb stored data.

In the 1950's recording was limited to film, chart paper, and punched paper tape with a few poor tape recorders. The bathythermograph smoked glass slide was very much with us as a mechanical recorder. At the end of the 1960's the digital cassette tape recorder appeared and was used in the 1970's extensively. In that decade, some giant audio tape recorders were packaged for ambient noise recording. Solid state memory began to be used in the 1980's for recording small amounts of data and bubble memory was tried for larger amounts of data at WHOI in 1983. The digital cassette tape with 2 megabytes (MBytes) of data was the standard for a decade. Digital streaming recorders and optical disks with up to 100 MBytes were used with moderate difficulty in the late 1980s. In situ processing was the alternative to these clumsy devices. However laptop computers broke the back of the problem and now 250 GByte hard disk drives ($0.66/GB Feb 3, 2003) and 1 GB Compact Flash memory cards make the need for in situ processing of data less urgent.

Data telemetry in real time or in delayed transmission still presents a data bottleneck, particularly satellite data communication. This data throughput limit is partly the bandwidth of the channel and partly power limits in the transmitting instrument. For any technology, acoustic, radio, optical, at various range and noise environments, there is a cost of energy per bit transmitted. Even for optical fiber there is some cost although in observatories, much of the charm is the extreme economy of data transmission. There is still some need to be conservative in data quantity.

When data are presented in graphic or tabular form in publications or reports, their information content is much less than even 2 MBytes. If one knew in advance what processing was needed to reduce the data to that which would be published, one could save a lot of post-processing time. Obviously only a small part of the data in an experiment is published, the rest serving to prove the validity and develop the statistics of the sample and isolate the phenomena to report. Still it is a sign that one is largely in the observational rather than the measurement phase if one cannot do some data compression in situ. This reduction must be done sometime if it is to become understandable to an observer.

Problem Set - Noise and Sampling

1) What is the voltage noise of a 100K resistor at room temperature and a bandwidth of 10 MHz? Watch units.

2) An FM radio receiver may see an antenna with 300 ohm impedance. (This is reactive, not resistive, but a maximum power match might use a transformer to reflect the input transistor impedance to that value. Think of a 300 Ω noise source at the input.) A commercial FM station is ~ 100 MHz and the radio channel (not TV) is 30 kHz wide. What signal strength is needed to get 0 db S/N? [db = 20*$\log_{10}$(voltage ratio)]

3) A deep space probe at a range of 10 astronomical units drives its 2 meter diameter 900 MHz antenna with 10 watts. What is the signal received at a 10 meter diameter dish antenna on earth? What is a reasonable baud rate at room temperature? At liquid helium temperature?