Analog to Digital Conversion

How do we make the conversion from analog signals such as a human voice to the digital domain? And furthermore, why would we want to convert an analog signal to a digital one?

To answer the second question, first consider the problems of:

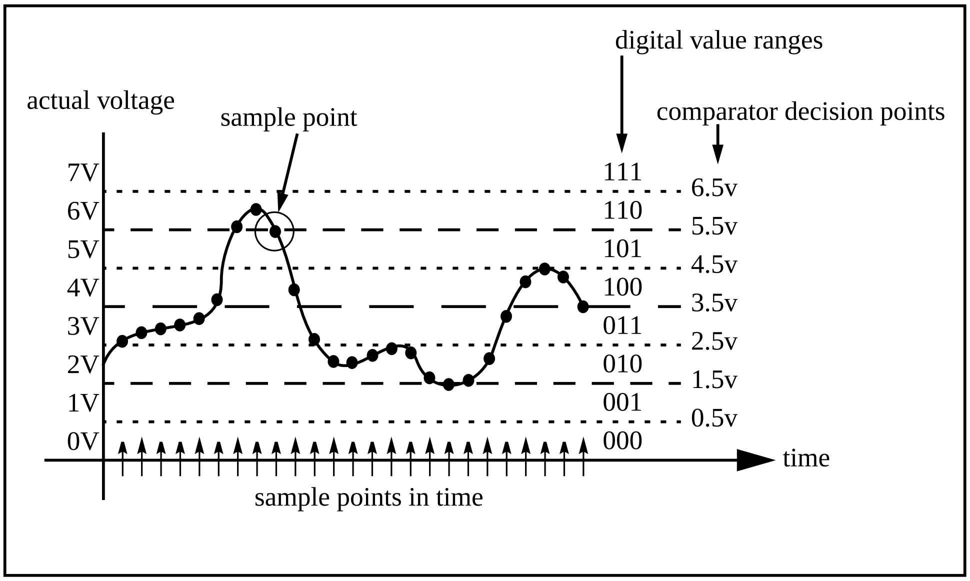
- storing analog data
- replicating or reconstructing analog data
- manipulation and processing of analog data

The storage of music in an analog fashion has been done for years with vinyl records and casette tapes. The storage media is large and is sensitive to many influences that degrade the fidelity (accuracy) of the data. Dust, warpage, and degradation of the physical media all contribute to a loss of accuracy over time or use. As these analog media are used repeatedly, data is actually lost or altered. We would like to have a way to infinitely reuse data and be able to store it in a small space and have it be relatively insensitive to environmental influences. Today, music can be stored digitally in CDs or non-volatile semiconductor memory. This storage media is thousands of times smaller than tapes or records. It is also relatively insensitive to heat, humidity and cannot wear out.

Replication of data on analog media is also problematic. For example, when vinyl records are pressed, the wearing of the master copy makes each copy slightly different. Tapes, when duplicated not only duplicate the audio content, but also copy the noise present on the original tape. If a musical tape is made from another musical tape, which was made from other copies of other tapes, eventually the level of hiss or noise approaches the music level. However when copies of digital data are made, each copy is exactly the same. Except for extreme situations, no noise or errors are introduced. Also, with digital techniques of error detection and correction, even if an error is introduced to the data, it is often detectable and correctable.

The ability to process digital data with a computer is a valuable advantage without equal in the analog domain. For example, data sent from deep space probes are carried on signals that are extremely weak. Using digital signal processing, the signals may be extracted from the received noise. Using computers we can also encode data so that it is much smaller in size. This process is utilized in the “zip” and “unzip” programs on personal computers.

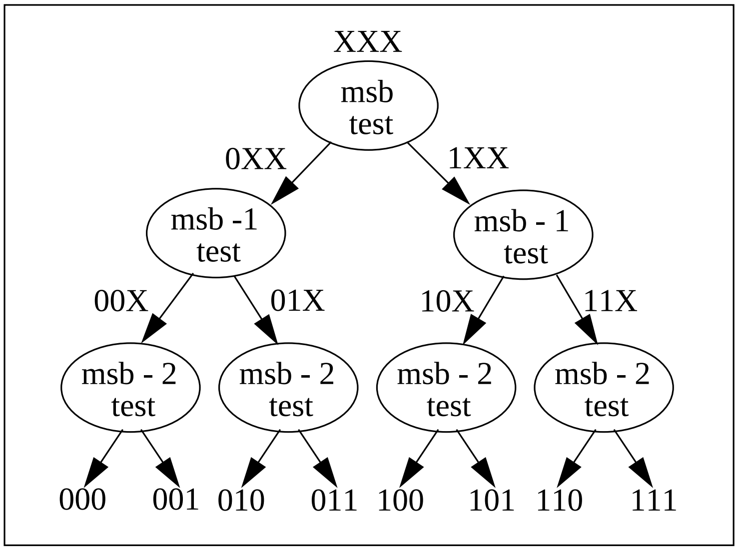
Given all these advantages for digital data, let’s answer the original question; how do we convert analog signals to a digital representation? The waveform below might be representative of a voice signal in a telephone system. The conversion of analog to digital is done by “sampling” the analog waveform at discrete times. Sampling means to simply take a voltage reading of the signal at discrete moments in time. A series of samples is converted to a set of sampled voltages.



A signal range is chosen for the digital representation of the analog waveform. In the example shown, we use three binary (base 2) bits to represent eight possible voltage levels (2^3=8 levels). The range chosen must be able to cover all the possible input levels. For a compact disk recording where high fidelity and dynamic range are needed we would use 16 bits, while a telephone circuit can get by with only 8 bits.

The bit values sampled are representations of the voltage sampled at a single point in time. A binary encoding is used to represent the voltage level. Each digit represents a power of 2. For example 101_2 is equivalent to $(1 * 2^2) + (0 * 2^1) + (1 * 2^0) = 5_{10}$. In the example above, waveform voltage samples above 4.5 volts and below 5.5 volts are considered to be 5 volts. The difference between the actual voltage and the quantized voltage, is called the quantization error. If the waveform is music, quantization error is experienced by our ears as distortion.

One way to decide the value of a sample is to check if the input voltage is above a level determined by repeatedly bisecting the voltage level range. The method is called successive approximation. The steps in successive approximation are shown below.



To determine the value of a sample, the question is asked “is the sample point above or below the 3.5V decision point for bit 2, the most significant bit (MSB)?” If so, bit 2 is set to “1”. Else, the bit is set to “0”. The question is asked again within the half range that the sample was found in but to determine the value of bit 1. As with the MSB, if the sample level is above the 5.5 volt level we choose a “1”, else a “0”.

This process can be repeated as many times as we have bits to represent our signal. Each time we do a comparison, we get another bit of information and more accurate estimate of the true signal level. Modern A to D converters have up to 24 bits of resolution. (1 part in 16 million!)

When we have a finite number of bits to represent the real valued input signal, we introduce *quantization error*. This error or “noise” is the approximation error made in the conversion decisions. Without infinite bit resolution, we can only approximate the real voltage level.

How fast should we sample the input data? What if we sampled the input signal at 1/2 the rate in our example? We would loose a lot of information about the waveform especially where it is changing quickly. So, how fast is fast enough? One of the most important rules for accurate sampling is called the *Nyquist Theorem*, which states that the highest frequency which can be accurately represented is one-half of the sampling rate.

For example, the telephone system digitizes our voices for digital transmission. The frequency spectrum that carries the bulk of intelligence in a human voice is confined to 300-3000hz. Thus with an upper limit of 3000hz, the sampling rate is chosen as 8000 samples per second. For a audio CD with frequency components to 20khz, the sampling rate is set to 44.1Khz. The sampling rate is usually in excess of the minimum required to allow for non-ideal circuits.

If we don’t sample fast enough, bad things happen. The problem is referred to as *aliasing*. It manifests itself as unmusical tones in the incorrectly digitized material. Because of this, A/D converters must use lowpass filtering to remove all signals above the Nyquist frequency. Of course, it also means that in order to get high-fidelity sound, we have to take a lot of samples.