Introduction to Digital Audio

Before the development of high-speed low-cost digital computers and associated electronic circuits, all recording and manipulation of sound was done by analog circuits. These circuits functioned by creating an electrical analog of sound, storing and processing that analog replica. The analog signal was continuous, meaning that no matter when one looked at the signal, there was a voltage or current value which represented the instantaneous value of the signal. In order to make use of digital representations of the sound, it was necessary to convert the continuous analog representation into a stream of numbers. Since it takes time to make the conversion from analog into digital, it is not possible to create a continuous digital representation of the sound. Furthermore, unless we can use infinitely large numbers, it is also impossible to measure every possible voltage level with complete precision. So we must trade the continuity of analog systems for the advantages of digital signal representation. The arguments still rage about which system is more desirable, so it is up to the engineer to decide for him/herself which system of audio representation they prefer for a given application.

A/D conversion:

Since no true digital microphones exist, the analog microphone signal must be converted into a digital representation. This process is known as **analog-to-digital (A/D) conversion**. It is this process and the complimentary conversion back to analog (D/A conversion) that impart many of the shortcomings to digital audio. In order to faithfully recreate the original signal, both the sampling rate and measurement (quantization) must be extremely accurate.

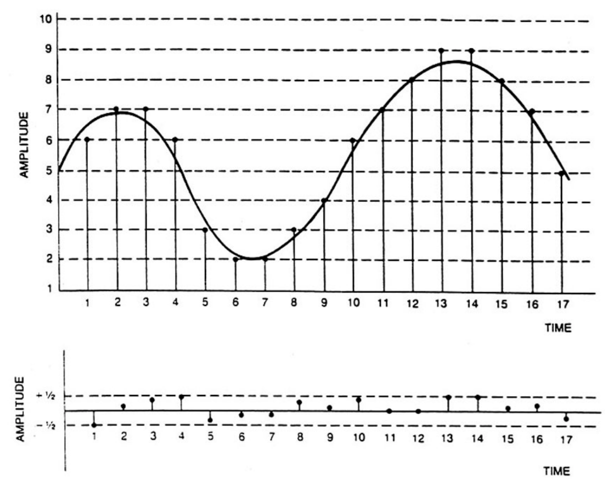

Figure 1. Error in sample measurement

In converting analog signals to digital representations, we use binary digital words to store the measured signal amplitude at intervals determined by the sample rate. The more bits (binary digits) the words contain, the more resolution we can provide: the bit length of the word determines the dynamic range we can accurately describe. Some current systems use 16 bits, as in CDs and DAT recorders, while many newer systems use 24-bit words. In digital electronics, each bit is coded by a single transistor, which can be either on or off, thus conveying the two possible digital states. Each bit encodes a binary weighting value like the decimal system: a 1's bit, a 2's bit, a 4's bit, etc. These digital words are often displayed in the form of the contents of digital electronic registers, groups of transistors wired to do the type of weighting necessary to perform binary mathematics.

128	64	32	16	8	4	2	1	
0	1	0	1	1	0	1	0	= 90

Figure 2 Weighted registers

Even if the digital measurement was perfect, there is still a significant problem associated with the conversion process: the signal must not contain any frequencies higher than half the sampling rate. This is a result of the way the audio signal interacts with the sampling frequency. In a way, the process of sampling is equivalent to combining the sampling signal with the audio signal, generating sum and difference frequencies. If the sample rate is low, the difference frequencies are audible and constitute a form of distortion: the so-called “foldover” distortion or “aliasing”. In order to minimize aliasing, the sample rate should be much higher than the highest frequency present in the signal. This can be accomplished either by increasing the sample rate or by limiting the signal bandwidth. Since increasing the sample rate increases the amount of data which must be stored, it is preferable to keep the sample rate as low as possible. Limiting the bandwidth will decrease the high-frequency content of the signal, also not desirable. We must therefore find a balance between sample rate and signal bandwidth. Of course, as the speed and capacity of digital systems continues to improve, the use of higher sample rates and longer word lengths is relaxing many of the constraints originally adopted to make digital recording feasable. The mathematics of sampling dictates that we sample at a rate higher than twice the highest frequency present in the signal (the Nyquist Theorem). If we want the entire audio bandwidth of 20 to 20,000 Hz, we must sample at greater than 40,000 Hz. This means that for 1 second of stereo sound, we will have more than 80,000 numbers. But this assumes that there are no signal components present above 20,000 Hz.

How do we guarantee this?

Anti-Alias Filters:

In order to prevent aliasing, we must remove any frequencies above half the sampling rate from the signal. This is done by filtering the signal. Since we want 20 kHz frequencies to be reproduced, we must filter very sharply above this frequency. Unfortunately, a simple filter cannot remove frequencies near to the 20 kHz cutoff very well. We must use a very sharp, complicated filter to remove the unwanted frequencies without also losing some frequencies inside the audio bandwidth. These filters are known as “brick-wall” filters because they cut off so rapidly above their corner frequency. It is not unusual to find 12 pole filters employed as anti-aliasing filters. As you might imagine, the design of these critical filters is very complicated. It is not possible to filter a signal so heavily without making some changes to the signal which is passed through. Often, the transient response suffers audibly as the complex filter responds to the driving signal. These filters are responsible for much of the criticism of digital audio, especially for the so-called harshness of early digital recorders. In recent years, the speed of computer chips has increased to the point where sophisticated mathematical processes can be applied to digitized signals which remove the unwanted frequencies from the signal after it is digitized, reducing the need for sharp analog filters. These procedures come under the heading of **oversampling**, a technique which allows high-speed sampling without increasing the amount of data to be stored. Even in these systems some analog filtering is still required, but simple filters with just a few poles are used, thereby reducing the deleterious effects of brick-wall filters. And as sample rates extend to 96 kHz and even 192 kHz, the requirement for analog filters is further relaxed.

D/A conversion:

Once the signal is stored digitally, it must eventually be converted back to analog for us to hear it. This process is essentially the reverse of sampling: the numbers are converted back to analog voltages. But since we have samples at discrete times, what should the output voltage do between the samples? What it should do is what the original signal did, but unfortunately, we no longer know what that was. One solution is to hold the last value until the next value is converted. This results in a staircase-like output which was not present in the original signal. The more bits we use for quantization, the smaller the step size, but we will always have some step-like distortion of the original signal. Again, we must filter the output to remove the effects of quantization.

Increasing the number of bits we use to quantize the analog signal results in a more accurate recreation of the input, but we pay for the increase in accuracy by storing more data. This also means that the data path between internal chips and memory must have more separate lines, since each bit is handled in parallel through its own data line. This leads to greater complexity in the circuitry required to perform the conversion. Another problem with D/A converters is that they are not completely linear in the lowest order bits. A 16 bit D/A converter may only be linear to 14 bits, thus introducing another kind of distortion. Fortunately, there are methods which can increase the linearity of the output. And continued improvement of the digital circuits which perform the conversions are contributing to better overall performance. Converters capable of sampling to 24-bit resolution at 96kHz and even 192 kHz sample rates are now becoming standard and reducing the inaccuracies of the processes.