Inside Digital Audio

Before the development of high-speed low-cost digital computers and associated electronic circuits, all recording and manipulation of sound was done with 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 that 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 amplitude measurement (quantization) must be extremely accurate. But even if these processes were 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 (modulating) 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 "fold-over" 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 that 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 feasible. 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 more than 40,000 samples/second. This means that for 1 second of stereo sound, we will have more than 80,000 numbers.

This assumes that there are no signal components present above 20,000 Hz. How do we guarantee that?

Anti-Alias Filters

In order to prevent aliasing, we must remove any frequencies above half the sampling rate from the signal, which is done by filtering the signal. If we want 20 kHz frequencies to be reproduced, we must filter very sharply above this frequency. Unfortunately, a simple filter cannot remove frequencies near 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 that is passed through. 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 while it is digitized, reducing the need for sharp analog filters. These procedures come under the heading of **oversampling**, a technique that allows high-speed sampling without increasing the amount of data to be stored. Even in these systems some analog filtering is 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 demands on the analog filters are further relaxed.

Real A/D converters

Most of the trouble with digital systems is involved in the transition from analog to digital and back again. The process must be perfectly accurate and very fast in order to adequately capture and reproduce the analog signal. Recent developments in integrated circuit manufacturing have led to great improvements in the conversion process and significant improvement in the sound of the average digital audio system. Much of the improvement has dealt with increased accuracy of the conversion process by using more bits: while many systems store only 16 bit words, the converters may quantize to 18 or even 20 bits. This helps because most of the distortion produced in the A/D conversion results from the nonlinearities associated with the lowest-order bits in the converter. There is an inherent limit to the accuracy of quantization determined by the least significant bit: there is always 1/2 LSB uncertainty due to the size of the smallest signal quantizable. If we want 16-bit linearity, we can use an 18-bit converter and disregard the two lowest-order bits, thereby increasing 16 bit linearity. The use of modern 24-bit converters guarantees linearity to at least 20 bits and many systems now store 24-bit audio data directly.

The simplest common A/D converter is known as the **successive approximation register** (SAR). This circuit works by setting a register and converting the value to analog via a D/A converter. The analog output from the register is continuously compared with the input voltage while the register is repeatedly incremented or decremented by control logic until the two are equal, at which time the digital number in the register is used as the value of the analog input voltage. In general, n D/A conversions must occur per sample period in an n-bit converter, so the clock must run at nearly 1 MHz in a one channel 16-bit converter. In older systems, the stereo channels were multiplexed through a single A/D converter that had to run at double speed. This also resulted in a small time shift between the channels. In order to guarantee that the analog input voltage doesn't change while the conversion is taking place, the analog voltage must be held constant while the A/D process takes place. This is done with a circuit called a **sample-and-hold**. It stores a single voltage value on a capacitor while the conversion proceeds and then reads in a new value for the next conversion. Using dual sample-and-hold circuits clocked together, the small time delay between multiplexed channels could be eliminated.

A newer type of A/D converter is known as a 1-bit or sigma-delta type converter, now used for Direct Stream Digital (DSD) recording as well as in the majority of PCM A/D converters. This converter works by very rapidly comparing the difference between the input and summed previous samples taken at a very high rate (oversampled), quantized as either 0 (difference is positive) or 1 (difference is negative). If the difference is 0, the pulse alternates between 1 and 0. The series of pulses can be used to generate the multi-bit codes actually used for data storage, or in the case of DSD, the single-bit bit-stream itself is stored. This technique trades the accuracy and linearity of a one-bit quantizer for the greatly increased sample rate necessary. An additional benefit of this type of converter is the ability to use digital signal processing to shape the noise generated in the process in such a way as to place it predominantly in high frequencies which are inaudible and easily removed by simple analog filtering.

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 that 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 re-creation 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, like A/D converters, 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 that can increase the linearity of the output and continued improvements in the digital circuits that perform the conversions are contributing to better overall performance. Converters capable of sampling to 24-bit resolution at 96 kHz sample rates are now becoming standard and have significantly reduced the inaccuracies of the process.

The output of the D/A converter is connected to the outside world at all times. Since transitions between sequential words may cause glitches as the registers change value, these glitches would be fed to the output of the device and enter the rest of the audio chain. To prevent this, we use the same sample-and-hold circuitry we employed in the A/D circuit to hold the previous output value until the next value is ready to be sent out. By tailoring the sample-and-hold pulse timing, some signal conditioning can be accomplished resulting in improved high-frequency performance.

Dither

While analog audio produces a constantly varying voltage or current, digital audio produces a non-continuous list of numbers. The maximum size of the numbers will determine the dynamic range of the system, since the smallest signal possible will result from the lowest order bit (LSB or least significant bit) changing from 0 to 1. The D/A converter will decode this change as a small voltage shift, which will be the smallest change the system can produce. The difference between this voltage and the voltage encoded by the largest number possible (all bits 1's) will become the dynamic range.

This leads to one of the major differences between analog and digital audio: as the signal level increases, an analog system tends to produce more distortion as overload is approached. A digital system will introduce no distortion until its dynamic range is exceeded, at which point it produces prodigious distortion. As the signal becomes smaller, an analog system produces less distortion until the noise floor begins to mask the signal, at which point the signal-to-noise ratio is low, but harmonic distortion of the signal does not increase. With low amplitude signals, a digital system produces increasing distortion because there are insufficient bits available to accurately measure the small signal changes.

There is a difference in the type of interference at low signal levels between analog and digital audio systems. Analog systems suffer from the noise generated by electronic circuitry. This noise is composed white noise, which has equal power at every frequency, pink noise, which has equal power in each octave, and other noise sources all of which are uncorrelated with the signal. It is the "hiss" like a constant ocean roar with which we are so familiar. The low-level noise generated by digital systems is different: it is correlated with the signal because it is really a form of signal distortion produced by quantizing errors and has a distinctly unpleasant sound: something like a "gritty" or "grainy" sound. Most listeners find digital noise to be less acceptable than analog white noise. Do we need to use noise reduction with digital systems as we found necessary on analog recorders? Well, the answer is yes, sort of. As we will see, there are digital techniques that may be used to reduce the undesirable low-level distortions without processing the signal as we did for analog noise reduction.

The solution to the dilemma of quantization error is counter-intuitive: by adding a small amount of wideband random noise, the signal correlated conversion error can be improved. This is process is known as dithering and it is applied any time the quantization bit depth is changed. While we do hear the added noise, it has the effect of de-correlating the conversion error with the signal and increasing the effective dynamic range of the system. Very low level signals become clearly audible, although in the presence of the added noise.