Digital Audio

Prior to the advent of computers, sound recordings were made using exclusively analog systems, where the instantaneous signal amplitude was continuously conveyed by a voltage we could measure at any time. With computers using binary arithmetic, signals must be encoded from continuous analog representations into discrete digital measurements that can be stored as binary number streams, describing the signal amplitude at specific times. We therefore know the amplitude only at the times we make the conversion to digital measurements, referred to as samples. If we sample frequently enough, the information we store is fully adequate to describe the signal accurately even though we have periods of time in which the signal is not sampled. Motion pictures work similarly: we perceive continuous movement even though the displayed frames are only occurring 24 times a second. There are numerous potential shortcomings in the process of converting continuous analog voltages to sampled discrete binary measurements and these will determine the quality of sound reproduction we are able to deliver with digital audio technologies.

The initial and potentially most important aspect of digital audio is the conversion from analog to digital signal representation. Any error created at this stage will persist throughout the recording and playback process. Early analog-to-digital (A/D) converters had serious limitations that reduced the quality of the sound reproduced, giving digital audio a bad reputation among audiophiles. In the intervening years, most of the limitations have been corrected and modern digital recorders are greatly improved. We will investigate the improvements from the simplest digital audio systems to the newest digital technologies. In order to make sense of digital audio devices, we need first to understand how binary arithmetic is used to measure the signal and manipulate it digitally. Like the decimal system, the binary digits increase in weighting as we move leftward in the word from least significant bit (LSB) to most significant bit (MSB):

MSB   LSB

128	64	32	16	8	4	2	1

$$1 \quad 0 \quad 1 \quad 1 \quad 0 \quad 0 \quad 1 \quad 0_2 \; = 178_{10}$$

Fig. 1 Binary numbers

We are used to the decimal system, a number system based on tens. Each decimal digit represents a multiple of 10. In binary arithmetic, each digit represents a multiple of two. While computers use binary mathematics, we are able to program them using decimal numbers that are converted to binary by the software. When acquiring data from an analog audio signal however, the conversion produces binary numbers directly. Number systems using 8 and 16 as the base sometimes make handling digital information simpler and these systems, called octal and hexadecimal, are frequently encountered in digital systems. Octal representations are denoted by the subscript "8" and hexadecimal by "H":

$$1111111_2 = 177_8 = 7F_H = 127_{10}$$

Computers are frequently programmed using octal or hexadecimal numbers and registers read back from digital hardware often display the results in these forms. Both octal and hexadecimal are easily translated into binary should the need arise, while decimal is not as easily converted. (You'll feel so much younger if you say you're 28_H as opposed to 40_{10}!)

The process of measuring the signal amplitude is known as quantization. Quantization converts a continuously varying voltage into a series of discrete measured, or quantized, binary numbers. The digital representation of the analog signal is encoded using digital words, with the number of bits per word determining the number of

discrete amplitude values we can discern in the resulting data. The more bits per word, the finer the minimum level difference we can distinguish. The fine level discrimination is important to the dynamic range we can handle in our digital system, since the smallest level change we can encode is the analog level that corresponds to the difference between the least significant bit being set on or off. If we fix the maximum input signal amplitude, the minimum difference we can measure will determine the dynamic range we can provide.

In analog audio, we are accustomed to thinking of the low amplitude limits of the system being determined by the residual noise in the system. This is usually determined by thermal noise in resistances and the noise sources of the active devices employed. In digital audio, the low-level limitations are due to the error in measuring the small signal resulting from the finite number of bits in the word length we use to encode the signal. As the input signal level decreases, the error in the measurement relative to the signal amplitude increases. With large signals, many bits are available to describe the signal amplitude. As the signal level decreases, it approaches the voltage corresponding to the least significant bit in the converted word. The difference between the signal amplitude and the LSB voltage equivalent is the error in the conversion. When the signal falls below the LSB equivalent voltage, the output may even be zero. The signal to error ratio therefore decreases as the signal level decreases. If we consider the error to be a form of distortion, we find that digital audio distortion increases as the signal amplitude decreases, the opposite of what usually happens in analog audio where distortion increases as the signal amplitude increases. Of course, if the analog signal exceeds the upper limit of the converter there is high distortion, but the distortion does not increase until the limit is exceeded.

The dynamic range of a digital system is approximately equal to two raised to the power of the number of bits in the word. More accurately, the dynamic range of a digital system using n bits per word is:

$$SNR(dB) = 20\log(\sqrt{3/2} \times 2^n)$$

$$SNR(dB) = 6.02n+1.76$$

The energy in the error produced by the converter is a statistical probability function and the result must be calculated by integrating the product of the error and its probability. We can use the 20 log 2^n approximation to see roughly how word length affects the dynamic range of digital audio systems:

n Bits	Dynamic Range (dB)
2	12
4	24
6	36
8	48
10	60
12	72
14	84
16	96
18	108
20	120
24	144

In theory, the more bits used to encode sample words, the greater our confidence in the accuracy of the measurement and the better the sound quality. In reality, the physical process of conversion limits the real

accuracy predicted by theory. Nonetheless, the more bits we use the better the fidelity as long as the electronic circuitry used can perform with the required accuracy.

The other critical factor in our measurement of the analog signal is the regularity of the sampling. We assume that each sample is taken at precisely the same interval. Small deviations in the timing will result in sample error, since the signal is changing with time and measurements made at incorrect times will have different values than samples made at the correct time. Sampling requires that we have very stable timing reference available and that the circuitry is able to perform the conversion within the time allowed between samples. Variations in the sample time are known as jitter, a potential cause of error in the conversion process and in the transfer of digital signals between devices.

The Nyquist Theorem predicts that a continuous function, like an analog audio signal, can be exactly represented by a sampled discrete time sequence if the sample rate is greater than twice the highest frequency contained in the original signal. The theory assumes an infinite discrete time sequence, however, which is impossible to accomplish with real converters. There is therefore some error inherent in the analog-to-digital conversion process from a theoretical standpoint. Nonetheless, if the signal spectrum is limited to less than half the sample rate, we are able to reconstruct the original signal adequately.
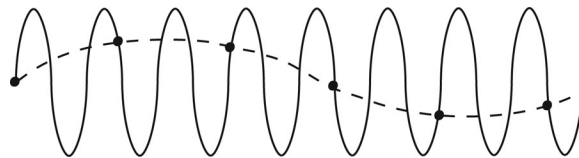

Fig. 2 A simple example of aliasing

In motion pictures, the result of sampling at too low a rate is obvious: the spokes of a rotating wheel appear to spin backwards when the wheel rotates more than one rotation between frames. The same phenomenon, known as aliasing, takes place in sampled audio. Instead of rotating backwards, the frequencies above half the sample rate are "folded over", creating audible frequencies below half the sample rate. This is called aliasing. The dots in Figure 1 mark the sample times, showing how the high frequency sine wave has undergone more than one cycle between samples. The resulting sine wave is of a lower frequency than either the signal or the sample rate. Alias frequencies are produced as sums and differences of the sample rate and the audio frequencies present above half the sample rate.

In order to prevent aliasing, analog low-pass filters are required at the A/D input to remove any frequencies greater than half the sample rate. For 44.1 kHz sample rates, this means frequencies above 22.05 kHz must be greatly attenuated. Analog filters capable of very sharp corner frequencies are quite difficult to implement, making such converters complicated. Each stage of the filter produces 6 dB/octave attenuation, so many stages are required to reduce the high frequency content above half the sample rate dramatically without reducing frequencies below the critical frequency. These filters may degrade the audio signal audibly, since such complex analog filters introduce phase and amplitude irregularities for frequencies well below the cutoff.

In theory, the sample rate must only be higher than twice the highest frequency present in the signal. How much higher it must be in practice is open to argument. The 44.1 kHz sample rate used in compact discs and many other digital devices was established at a time when digital storage space was limited, in both delivery media and recording devices. We are now able to handle much higher data rates and can use higher sample rates and longer words, but it is not clear how much is enough. The balance between high sample rates and the amount of data that needs to be stored is a tradeoff that must be decided by digital recording system users.

As the ability to convert from analog to digital representations has been refined, we find that high sample rates

and longer word lengths are available, now up to 192 kHz sample rates and 24 bit words. Higher sample rates mean that much less complex anti-alias filters are possible, reducing the potential for signal distortion. Higher speed data handling may improve sound quality by improving the time resolution important for transient signals. Further, the ability to perform digital filtering on the data after it is acquired reduces the complexity of analog filters needed without increasing the final amount of data stored. Any added data that must be stored and the higher data acquisition and transfer rates needed to handle the conversions are now easily accomplished even by inexpensive personal computer systems. SACD's, DVD's and proposed new digital delivery media have much higher data storage capabilities than compact discs. There are now viable alternatives to the compact disc's 44.1 kHz/16-bit sample rate for delivery media and users are able to apply the appropriate sample rate to any project. While the CD sample rate may be adequate for delivery of most music, higher rates and longer word lengths are useful in the recording studio as they provide more latitude in signal processing and mixing.

Analog-to-digital converters may be designed in several ways. Originally, digital devices used multi-bit converters, converting each word by comparing the signal to a digital-to-analog converter's output and continually adjusting the digital word that sets the D/A converter's output until the two voltages agree. This is known as successive approximation conversion (or "guess-and-check"). This type of converter requires one iteration for each bit in the data word. Newer A/D converters use sigma-delta conversion, where the signal amplitude is compared with its previous value at a very high rate and a single bit is used to tell if the latest measurement is larger than the previous one or not. The individual differences are then summed to produce the digital word. The latest system of A/D conversion uses a single bit converter at extremely high sample rate, like the sigma-delta converter, except that the data stream is used without conversion to multi-bit words and is known as direct stream digital (DSD). Each system has strengths and weaknesses that contribute to the perceived sound quality and ease of data manipulation.

A/D converters ideally produce output words that are linearly related to the input voltages they measure. In some cases, however, there are deviations from linearity, especially in the low order bits. Some codes may be missing or the voltage steps may not be exactly equal. These faults limit the accuracy of the conversion and produce audible distortion of low-level signals. Changes in the temperature of the converter circuitry can contribute to drift over time. With careful design, these errors can be minimized. Most modern converters produce acceptable linearity over a wide amplitude range, although 24 bit converters are not likely to produce real 144 dB dynamic ranges, in part because the analog circuitry cannot produce such low noise levels at room temperature.

Multi-bit converters are widely used in digital audio devices. These digital systems employ coding schemes that make use of a multi-bit digital word to encode the measured analog signal amplitude. The coding system used is most often pulse code modulation (PCM), in which the bits' states, on or off, code for one and zero in the binary digital word. Other methods can use the width of a digital pulse to encode amplitude in a system known as pulse width modulation (PWM) or the distance between pulses known as pulse position modulation (PPM). The advantage of the PCM system is that it can easily be used for calculations by computer logic circuitry. The newer direct stream digital data cannot as easily be manipulated although it may be stored.

The process of converting the stored and processed digital data back to analog is handled by a digital-to-analog converter (DAC). This is the reverse of the A/D converter and involves much the same technology. Like the A/D converter, filtering is necessary however in the case of the D/A converter, it is to remove multiples of the spectrum folded around multiples of the sample rate. These unwanted frequency bands are called images and the DAC filter is known as an anti-imaging filter. The images are created by the sample pulse modulating the signal which creating multiples of the original spectrum at multiples of the sample rate. While these are ultrasonic and not audible to the ear, they would potentially overload analog electronics following the DAC, creating audible distortion. Digital-to-analog converters are also susceptible to jitter like A/D converters, so

solid clocking is required to the DAC.

Another issue with D/A converters relates to what happens as the digital word feeding the converter changes from sample to sample. The output of the DAC is always connected to the analog output, so any glitches that occur as the digital word changes would be sent to the output. To prevent this, a sample-and-hold (S/H) circuit is used. The S/H circuit samples the analog value at the DAC after any glitches subside and stores it on a capacitor until the next value is ready. This is an analog sampler, but it is still a form of sampling. The width of the sample-and-hold pulse affects the frequency response of the system and it can be used to tune the output frequency response of the converter to offset other aspects of the system that may alter the response. This is known as the aperture effect.

Both A/D and D/A converters can be improved by running at much higher sample rates than the input quantizer or D/A converter and using digital signal processing to restore the desired output sample rate. This is known as oversampling and its main advantage is the ability to use digital filters instead of steep analog filters to remove aliasing and imaging. In the case of A/D conversions, the quantizer can be run at much higher sample rates than the final sample rate and the data averaged to produce the lower sample rate, a process known as decimation. At the D/A end, analog values at times between stored samples can be created using a process known as interpolation. Interpolation filters work by inserting evenly spaced zeros between samples and multiplying the samples by a series of coefficients derived from polynomials selected by statistical predictions of analog signal behavior.

Whatever system is used to acquire the digital data, we are faced with the same dilemma we encounter in the analog recorder when it comes to storing the information. While computer memories may store the data temporarily, most computers use dynamic random access memory (DRAM) chips that lose the data when the power is removed. For long-term storage, magnetic media are used for digital recording as well as for analog. Since digital data require only two states and not the complete linearity demanded by analog recordings, there is a difference in the way the process is applied. Digital magnetic recorders use saturation recording, leaving all magnetic domains polarized in one direction or the other with no intermediate levels required, so the bias current used in analog recording is unnecessary in digital recorders. The density is quite high in digital recorders and that introduces some problems not encountered in analog recorders. The number of bits per unit area of medium is limited in the longitudinal recording method used for analog record, although it is sufficient for the demands of analog recording. Digital magnetic media may benefit from closer packing of domains, which is achieved by using perpendicular recording, in which the domain magnetic fields are magnetized perpendicular to the medium surface instead of along the surface as in analog recorders.

Digital data stored magnetically requires only two discernable states for binary information. This is easily accomplished by magnetizing domains fully in one or the other polarity. While this avoids the non-linear region of the H-M curve, it introduces another problem, the interference between closely occurring bits. If the write head and the medium are not capable of altering the magnetic polarities as rapidly as the bits are changing, the magnetization from the previous bit will affect the next bit. This causes the data to be altered, since the overlap makes discriminating between ones and zeros unclear. This inter-symbol interference limits the data density that can be stored.

We have several options for storing digital data, including dedicated devices using tape or discs as media and general-purpose personal computers with added interfaces to acquire and convert analog audio. The high data density required for storing digital data made early digital recorders quite complicated, requiring rotating magnetic head recorders designed for video recording or using high tape speeds with stationary heads requiring more than one data track for each audio channel to provide enough bandwidth. While the more complicated recorders are still in use, they've largely been replaced by computers as the preferred storage device for digital audio recordings. The low price and wide availability of large, fast disc drives has spurred a move to the PC as

the digital audio recorder of choice, especially since the computer can take on the functions of editing, mixing, processing, and storing the entire project in a single device.

The evolution of digital recorders has been rapid. Rotary-head modular digital multitrack recorders and stereo DAT recorders enjoyed only a few years of widespread use before recordists moved to the general-purpose computer as the preferred platform for digital recording. These machines bridged the gap between high-cost stationary head professional digital recorders like the Sony DASH and Mitsubishi Pro-Digi systems and analog multitrack. The Alesis ADAT and TASCAM DTRS machines used video tape, cheap and readily available, to provide inexpensive access to digital recording for a wide range of users. These machines, while initially inexpensive, suffered from their complexity when head wear and transport malfunctions required difficult repair and diagnosis procedures. Yamaha produced the DMR/DRU series of recorders that used stationary heads and proprietary tape cassettes to deliver 20-bit 8-channel recording in the early 1990's, but they were expensive relative to the ADAT and DTRS machines and never caught on. None of these tape-based systems survived the move to computer-based systems and all have been phased out or will soon be retired. While tape provides the advantage of removable media, the large size of hard drives and the availability of plug-and-play computer interfaces for storage media has diminished the attractiveness of tape-based digital recorders.

The ability to use inexpensive, mass-produced personal computers for digital audio recording and mixing has greatly expanded the accessibility of these tools. The addition of a FireWire or USB audio interface and some software is all that is required to create a digital studio entirely within the PC. This has had a dramatic effect on the recording studio and the music business in general. Essentially the entire recording studio may now be contained in a single piece of equipment, with the ability to recall the entire project and studio configuration in a few seconds. The advantages of digital audio are hard to ignore, even for those dedicated to the analog studio paradigm.

Using personal computers for audio recording has introduced a new set of difficulties. Each operating system and hardware platform requires different software and there are differences in the bus structures and interface buses available that complicate the choice of peripheral audio interfaces. Input/output buses include FireWire and USB high speed serial interfaces, both of which are possible choices for connecting multi-channel A/D and D/A modules to the PC to provide audio access. The software for recording interacts with the operating system to access these audio channels and may do so with differing speed capabilities on different computers. With the main choices for personal computer systems, Macintosh and Windows/Intel, both types of interface are supported but different recording programs are required and the performance of the audio interfaces may differ due to differences in the hardware and software employed in the particular computer used. The recording engineer must now have some knowledge of the internal workings of their computer. If something goes wrong with the recording system, it becomes necessary to isolate the problem by troubleshooting a complicated series of interactions between software, computer and peripheral hardware that may not be well documented. Each manufacturer provides information about their part of the system, but no one company is responsible for the entire system, leaving the user to deal with the problem.