DIGITAL TECHNOLOGY FUNDAMENTALS

Introduction

Audio was previously recorded, stored and reproduced by analog means and mediums. Now, however, due to the current advance in digital technology, audio can be stored and reproduced in a digital form. Such examples of this technology will be CD players, DAT players, digital consoles and digital samplers. For an audio engineer of the current times, it is vital that he/she must understand the underlying digital theories and concepts of these technologies that are available.

1. <u>Advantages of Digital Audio over Analog Audio</u>

Digital audio can exists in a non-linear form in comparison with analog audio. This non-linearity provides more flexibility in audio editing processes. *(Analogue editing compared with a DAW such as Protools)*

There is virtually no degradation of original signal when copying. A perfect copy of the original source can be made in the digital domain whereas in analog domain tape noise and hiss are added to the duplicates*. (Cassette dubbing compared to CD burning)*

Digital audio can be stored on more reliable media such as CDs, MDs and hard disks. These storage mediums have longer life expectancy than analogue storage mediums like tape do. (*CDs last longer than cassette tapes*)

Analog Audio	**Digital Audio**
♦ Obvious generation lost ♦ Noise added during copying ♦ No perfect copy can be made ♦ Can only be stored on limited analog medium ♦ Cannot be manipulated by computer	♦ No generation lost ♦ No noise added during copying ♦ Perfect copy can be made ♦ Can be stored on large number of digital medium ♦ Can be manipulated by computer

Analogue technologies have more circuitry that adds noise and distortion to the signal than digital technologies. Therefore digital technology has more dynamic range than analogue technology. (*16bit Digital Audio recorder has a dynamic range of 97.8dB - a value 30db above the noise figure for most conventional analog tape recorders.)*

2. <u>Binary numbers</u>

Binary numbers are used to represent data in the digital realm. In this system, there are only two states (a "0" state and a "1" state). Binary digits used in digital devices are called bits (short for Binary digITS). All numbers in the decimal system can be represented by the binary system. The decimal system is known as base 10 because there are 10 numbers (0-9) to represent all the figures in this system. The binary system is known as base 2 because it has only 2 numbers (0 and 1) to represent all the figures in its system.

Digital devices need a fixed length of binary numbers called a "word". Word length refers to the number of digits and is fixed in the design of a digital device *(00000001- 8-bit word length).* For example the Alesis XT20 uses 20-bit word length.

The following is a further illustration of the above-mentioned principle:

For systems with 2-bit word length there can be only four representations

00

01

10

11

For systems with 3-bit word length there will be eight representations

000

001

010

011

110

111

100

101

This can be calculated by the following formulae

Number of representations = 2 to the power of n (when n= word length)

Therefore a 16-bit system would have, 2 to the power of 16, which is equal to 65536 representations.

3. <u>Conversions of Binary numbers to Decimal and vice versa</u>

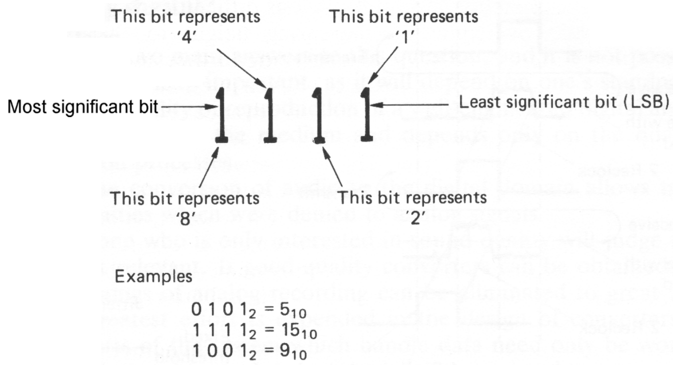
For example for a 5 digit word can represent all combination of numbers from 0 to 31 (2 to the power of 5) where:

00000 (Bin) = 0 (Dec)

00001 (Bin) = 1 (Dec)

11111 (Bin) = 31 (Dec)

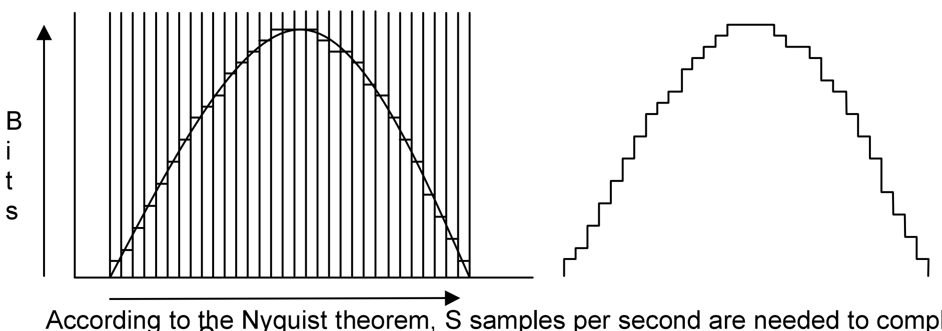
The first bit in a word is known as the Most Significant Bit (MSB) and the last bit is known as the Least Significant bit (LSB). One BYTE = eight bits and One NIBBLE = 4 bits.



The word length of a digital device becomes the measure of the "Resolution" of the system and with digital audio better resolution means better fidelity of reproduction. (*24 bit systems are superior to 16 bit systems)*

4. <u>Sampling</u>

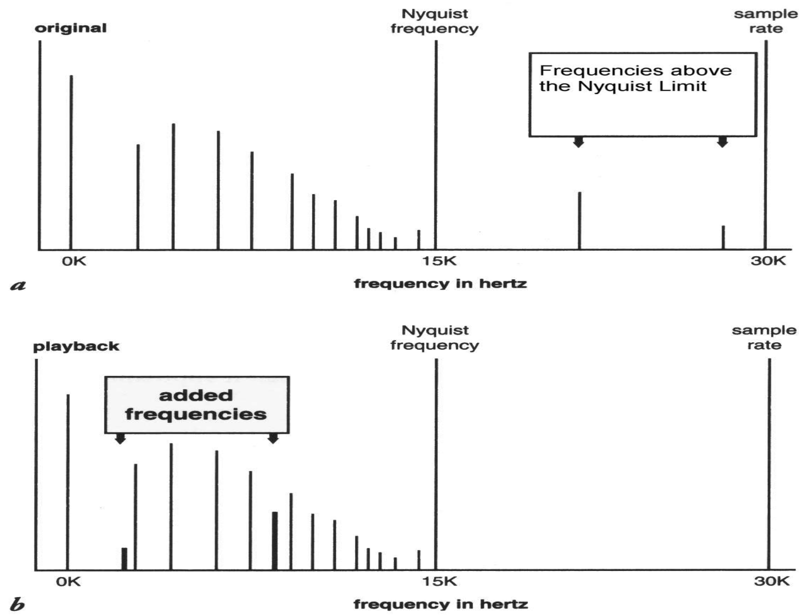
In the digital realm, recordings of analog waves are done through periodic sampling. That means when a sound wave is recorded, snap shots of the wave are taken at different instances. These snap shots are later examined and given a specific value (a binary number). This process is called discrete time sampling. The sampling rate of a digital system is defined as the number of snap shots or samples that are taken in one second. Therefore a device with a sample rate of 44.1kHz takes 44100 samples per second.



According to the Nyquist theorem, S samples per second are needed to completely represent a waveform with a bandwidth of S/2 Hz, i.e. we must sample at a rate which twice the highest through put frequency to achieve loss less sampling. Therefore for a bandwidth of 20Hz-20kHz, one must use a sampling frequency of at least 40 kHz. Therefore, it is a necessity to send the recording signal through a low pass filter before the sampling circuit to act in accordance with the Nyquist Theorem.

5. Aliasing

Alias frequencies are produced when the input signal is not filtered effectively to remove all the frequencies above half of the sampling frequency. This is because there are no longer adequate samples to represent the deviant high frequencies. The sample continues to produce samples at a fixed rate, outputting a stream of false information caused by deviant high frequencies. This false information takes the form of new descending frequencies, which were not present in the original audio signal. These are called alias or fold over frequencies. For example if S ids the sampling rate, and F is a frequency higher than half the sampling frequency, then a new frequency Fa is created where Fa = S-F. The solution to aliasing is to band limit the input signal at half the sampling frequency using a steep slope HPF called an anti-aliasing filter.

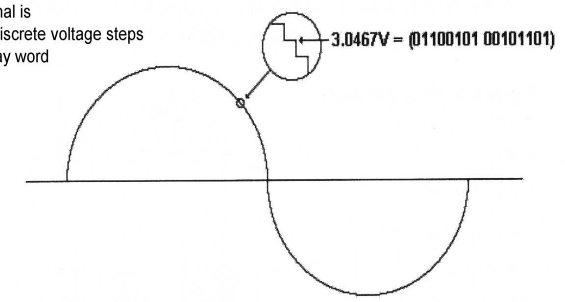


6. Quantization

Quantization is the measured amplitude of an analog signal at a discrete sample time. The accuracy of quantization is limited by the system's resolution, which is represented by the word length used to encode the signal, i.e. the number of bits in a signal such as 8,12, 16, 20, 24 bits. No matter how fine we make the resolution we cannot fully capture the full complexity of an analog waveform. For the current standards a 16-bit (65563 steps) representation is acceptable.

8-bit word	=	(nnnn nnnn)	=	256 steps
16-bit word	=	(nnnnnnnn nnnnnnnn)	=	65,536 steps
20-bit word	=	(nnnnnnnnnn nnnnnnnnnn)	=	1,048,576 steps
24-bit word	=	(nnnnnnnnnnnn nnnnnnnnnnnn)	=	16,777,216 steps

The amplitude of the analog signal is
is broken down into a series of discrete voltage steps
which are then assigned a binaray word



Quantization Error

This is the difference between actual analog value at the sample time and the chosen quantization intervals value i.e. the difference between the actual and measured values. Quantization error is limited to +- 1/2 interval at the sample time. At the system output, this error will be contained in the output signal. This error will sound like white noise that is heard together with the program material. However, there will be no noise in silent passages of the program. In addition, quantization noises changes with amplitude, becoming distortion at low levels.

Dither

Although quantization error occurs at a very low level, its presence must be considered in hi fidelity music. Particularly at low levels, the error becomes measurable distortion in reactive with the signal. To fix quantization errors, a low-level noise called dither is added to the audio signal before the sampling process. Dither randomises the effect of quantization error. Dither removes the distortion of quantization error and replaces it with low-level white noise.

S/E Ratio

Signal to error ratio is closely akin, although not identical to signal to noise ratio. Whereas signal to noise ratio is used to indicate the overall dynamic range of an analogue system, the signal to error ratio of a digital device indicates the degree of accuracy used when encoding a signal's dynamic range with regard to the step related effects of quantization.

The signal to error ratio of a certain digital device can be formulated as follows:

S/E = 6N + 1.8 (dB)

7. Pulse Code Modulation

A method of encoding digital audio information which uses a carrier wave in the form of a stream of pulses which represents the digital data. The original analog waveform is sampled and its amplitude quantized by the analog to digital (A/D) converter. Binary numbers are sent to the storage medium as a series of pulses representing amplitude. If two channels are to be sampled the PCM data may be multiplexed to form one data stream. The data is processed for error correction and stored. On playback the bit stream is decoded to recover back the original amplitude information at proper sample times and the analog waveform is reconstructed by the digital to analog converter (DAC).

8. Linear PCM recording section

Dither Generator

An analog noise signal is added to the analog signal coming from the line amplifier. The dither causes the audio signal to constantly move between quantization levels. The noise should resemble noise from analog systems, which is very easy on the ear. Gaussian white noise is often used.

Input Low Pass Filter

The analog signal is low-pass filtered by a very sharp cut-off filter to band limit the signal and its entire harmonic content to frequencies below half of the sampling frequency. The ideal LPF would have a "Brick wall" cut off, but this is very hard to achieve. In professional recorders with a sampling frequency of 48kHz, the input filters are usually designed for 20Hz-20kHz. Thus proving a guard band of 2kHz to ensure the attenuation is sufficient.

Sample and Hold

The S/H circuit time samples the analog waveform at a fixed periodic rate and holds the analog value until the A/DC outputs the corresponding digital word. Samples must be taken precisely at the correct time.

In audio digitization, time information is stored implicitly as samples taken at a fixed periodic rate, which is accomplished by the S/H circuit. An S/H circuit is essentially capacitor and a switch. Maintaining absolute time throughout a digital system is essential. Variations in absolute timing called jitter can create modulation noise.

Analog to Digital conversion

This is the most critical component of the entire system. The circuit must determine which quantization increment is closest to the analog waveform's current value, and output a binary number specifying that level. This is done in less than 20 microseconds. In a 16-bit linear PCM system each of the 65 536 increments must be evenly spaced throughout the amplitude range so that even the LSBs in the resulting word are meaningful. Thus the speed and accuracy are the key requirements for an A/D converter.

8.5 Record processing

After conversion several operations must take place prior to storage:

Multiplexing - Digital audio channel data is processed in a single stream. However the A/DC outputs parallel data, i.e. entire words. The multiplexer converts this parallel data to serial data.

Data coding - Raw channel is properly encoded to facilitate storage and later recovery. Several types of coding are applied to modify or supplement the original data. A synchronisation code is a fixed pattern of bits provided to identify the beginning of each word as it occurs in the bit stream. Address codes are added to identify location of data in the recording. Other specifications such as sampling frequency, table of contents, copyright information, even Time Code can be added.

9. Error protection, correction and concealment

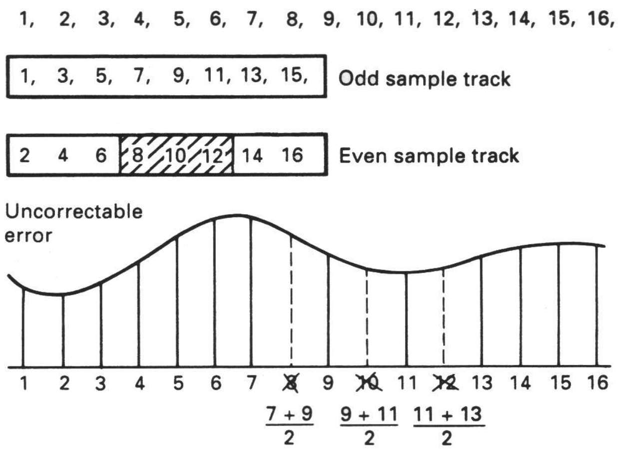
As anyone familiar with analog recording will know, magnetic tape is an imperfect medium. It suffers from noise and dropouts, which in analog recording are audible. In a digital recording of binary data, a bit is either correct or wrong, with no intermediate stage. Small amounts of noise are rejected, but inevitably, infrequent noise impulses cause some individual bits to be in error (bit errors). Dropouts can cause a larger number of bits in one place to be in error. An error of this kind is called a burst error. Whatever the medium and whatever the nature of the mechanism responsible, data are either recovered correctly, or suffer some combination of bit errors and burst errors. In Compact Disc, random errors can be caused by imperfections in the moulding process, whereas burst errors are due to contamination or scratching of the CD surface.

The audibility of a bit error depends upon which bit of the sample is involved. If the LSB of one sample was in error in a loud passage of music, the effect would be totally masked and no one could detect it. Conversely, if the MSB of one sample was in error in a quiet passage, no one could fail to notice the resulting loud transient. Clearly a means is needed to render errors from the medium inaudible. This is the purpose of error correction.

In binary, a bit has only two states. If it is wrong, it is only necessary to reverse the state and it must be right. Thus the correction process is trivial and perfect. The main difficulty is in identifying the bits, which are in error. This is done by coding the data by adding redundant bits. Adding redundancy is not confined to digital technology: airliners have several engines and cars have twin braking systems. Clearly the more failures which have to be handled, the more redundancy is needed. If a four-engined airliner is designed to fly normally with one engine failed, three of the engines have enough power to reach cruise speed, and the fourth one is redundant. The amount of redundancy is equal to the amount of failure, which can be handled. In the case of the failure of two engines, the plane can still fly, but it must slow down; this is graceful degradation. Clearly the chances of a two-engine failure on the same flight are remote.

In digital audio, the amount of error, which can be corrected, is proportional to the amount of redundancy; the samples are returned to exactly their original value. Consequently corrected samples are inaudible. If the amount of error exceeds the amount of redundancy, correction is not possible, and, in order to allow graceful degradation, concealment will be used. Concealment is a process where the value of a missing sample is estimated from those nearby. The estimated sample value is not necessarily exactly the same as the original, and so under some circumstances concealment can be audible, especially if it is frequent. However, in a well-designed system, concealments occur with negligible frequency unless there is an actual fault or problem.

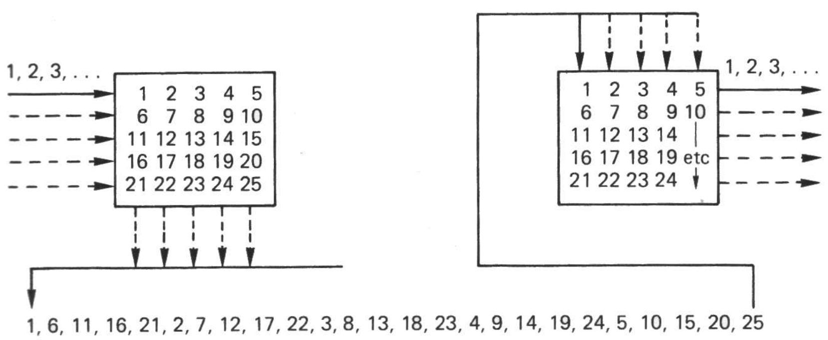
Concealment is made possible by rearranging or shuffling the sample sequence prior to recording. This is shown in the following diagram.



In cases where error correction is inadequate, concealment can be used provided that the samples have been ordered appropriately in the recording. Odd and even samples are recorded in different places as shown here. As a result an uncorrectable error causes incorrect samples to occur singly, between correct samples. Therefore an approxiamtion of the existing values are taken via the process of interpolation.

In cases where the error correction is inadequate, concealment can be used provided that the samples have been ordered appropriately in the recording. Odd and even samples are recorded in different places as shown here. As a result an uncorrectable error causes incorrect samples to occur singly, between correct samples. In the example shown, sample 8 is incorrect, but samples 7 and 9 are unaffected and an approximation to the value of sample 8 can be had by taking the average value of the two. This interpolated value is substituted for the incorrect value. Odd-numbered samples are separated from even-numbered samples prior to recording. The odd and even sets of samples may be recorded in different places, so that an uncorrectable burst error only affects one set. On replay, the samples are recombined into their natural sequence, and the error is now split up so that it results in every other sample being lost. The waveform is now described half as often, but can still be reproduced with some loss of accuracy. This is better than not being reproduced at all even if it is not perfect. Almost all digital recorders use such an odd/even shuffle for concealment. Clearly if any errors are fully correctable, the shuffle is a waste of time; it is only needed if correction is not possible.

In high-density recorders, more data are lost in a given-sized dropout. adding redundancy equal to the size of a dropout to every code is inefficient. The following figure shows that the efficiency of the system can be raised using interleaving.



Sequential samples from the ADC are assembled into codes, but these are not recorded in their natural sequence. A number of sequential codes are assembled along rows in a memory. When the memory is full, it is copied to the medium by reading down columns. On replay, the samples need to be de-interleaved to return them to their natural sequence. This is done by writing samples from tape into a memory in columns, and when it is full, the memory is read in rows. Samples read from the memory are now in their original sequence so there is no effect on the recording. However, if a burst error occurs on the medium, it will damage sequential samples in a vertical direction in the de-interleave memory. When the memory is read, a single large error is broken down into a number of small errors whose size is exactly equal to the correcting power of the codes and the correction is performed with maximum efficiency.

Interleave, de-interleave, time compression and timebase correction processes cause delay and this is evident in the time taken before audio emerges after starting a digital machine. Confidence replay takes place later than the distance between record and replay heads would indicate. In DASH-format recorders, confidence replay is about one-tenth of a second behind the input. Synchronous recording requires new techniques to overcome the effect of the delays.

The presence of an error-correction system means that the audio quality is independent of the tape/head quality within limits. There is no point in trying to assess the health of a machine by listening to it, as this will not reveal whether the error rate is normal or within a whisker of failure. The only useful procedure is to monitor the frequency with which errors are being corrected, and to compare it with normal figures. Professional digita

Some people claim to be able to hear error correction and misguidedly conclude that the above theory is flawed. Not all digital audio machines l audio equipment should have an error rate display. are properly engineered, however, and if the DAC shares a common power supply with the error-correction logic, a burst of errors will raise the current taken by the logic, which in turn loads the power supply and interferes with the operation of the DAC. The effect is harder to eliminate in small battery-powered machines where space for screening and decoupling components is hard to find, but it is only a matter of design; there is no flaw in the theory.

Error protection and correction are provided so that the effect of storage defects is minimised. The data is processed before storage by adding parity bits and check codes, both of which are redundant data created from the original data to help detect and correct errors. Finally interleaving is employed in which data is scattered to various locations on the recording medium.

9.1 Modulation process.

This is the final electronic manipulation of the audio data before storage. Binary code is not recorded directly, rather a modulated code in the form of a modulation waveform which is stored and which represents the bit stream.

In the Binary bit stream there is really no way to directly distinguish between the individual bits. A series of ones and zeros would forma static signal upon playback and timing information would be lost. Additionally it is inefficient to store binary code directly onto the medium. It would take too much storage space.

Typically, in the modulation process, it is the transition from one level to another rather than the amplitude levels, which represents the information on the medium. Various modulation codes have been designed:

Non-return to zero code (NRZ) - 1s and 0s are represented directly as high and low levels. Used only where synchronisation is externally generated such as video tape recordings.

Non-return to Zero Inverted code (NRZI) - only 1s are denoted with amplitude transitions. Thus any flux change in the magnetic medium indicates a 1.

Modified Frequency modulation (MFM) - Sometimes called miller code.

Eight to Fourteen modulation (MFM) - Used for CD storage. Blocks of 8 bits are translated into blocks of 14-bits using a look-up table. Each one represents a transition in the medium, which in a CD would mean physical presence of a pit-edge.

10. Digital Audio Reproduction

Digital audio reproduction processes can be compared to the reverse of the digital audio recording processes.

Demodulation Circuits

A preamp is required to boost the low-level signal coming off the tape heads. The waveform is very distorted and only transitions between levels have survived corresponding to the original recorded signal. A waveform shaper circuit is used to identify the transitions and reconstruct the modulation code. The modulated data, whatever its code (EFM, MFM, etc) is then demodulated to NRZ code, that is, a simple code which amplitude information represents the binary information.

Reproduction Processing

The production processing circuits are concerned with minimising the effects of data storage. They accomplish the buffering of data to minimise the effects of mechanical variations and transport problems (e.g. timing variations, tape stretch, bad head alignment) in the medium. These timing variations will cause jitter. The reproduction processing circuits also perform error correction and demulitplexing.

1. The circuits firstly de-interleave the data and assemble it in the correct order.
2. The data is read into a buffer whose output occurs at an accurately controlled rate thus ensuring precise timing data and nullifying any jitter caused by mechanical variations in the medium.
3. Using redundancy techniques such as parity and checksums, the data is checked for errors. When error is too extensive for recovery, error compensation techniques are used to conceal the error. In extreme cases the signal will be momentarily switched off.
4. The demulitplexer reconverts the serial data to its parallel form. This circuit takes single bits and outputs whole simultaneous words.

11. Digital to Analog Conversion

The D/AC is the most critical element in the reproduction system - determining how accurately the digitized signal will be restored to the analog domain. A DAC accepts input digital word and converts it into an output analog voltage or current.

Output sample and hold

When the DAC switches from one output voltage to another, false voltage variations such as switching glitches can occur which will produce audible distortion. The output circuit acquires a voltage from the DAC only when the circuit has reached a stable output condition. The S/H circuit holds correct voltage during the intervals when the DAC switches from samples

Hence false glitches are avoided by the S/H circuitry. It operates like a gate removing false voltages from the analog stream and like a timing buffer, re-clocking the precise flow of voltages. Its output is a precise "staircase" analog signal, which resembles the output of its counterpart in the recording conversion.

Output Low-Pass Filter and Oversampling

This "anti-imaging" LPF has a design very similar to the input "anti-aliasing" filter. Oversampling techniques are used in conjunction with this filter.

Oversampling has the effect of further reducing inter modulation and other forms of distortion. Whenever Oversampling is employed, the effective sampling rate of a signal-processing block is multiplied by a certain factor - commonly ranging between 12 and 128 times the original rate. This significant increase in the sample rate is accomplished by interpolating the original sample times. This technique, in effect, makes educated guesses as to where sample levels would fall at the new sample time and generate an equivalent digital word for that level.