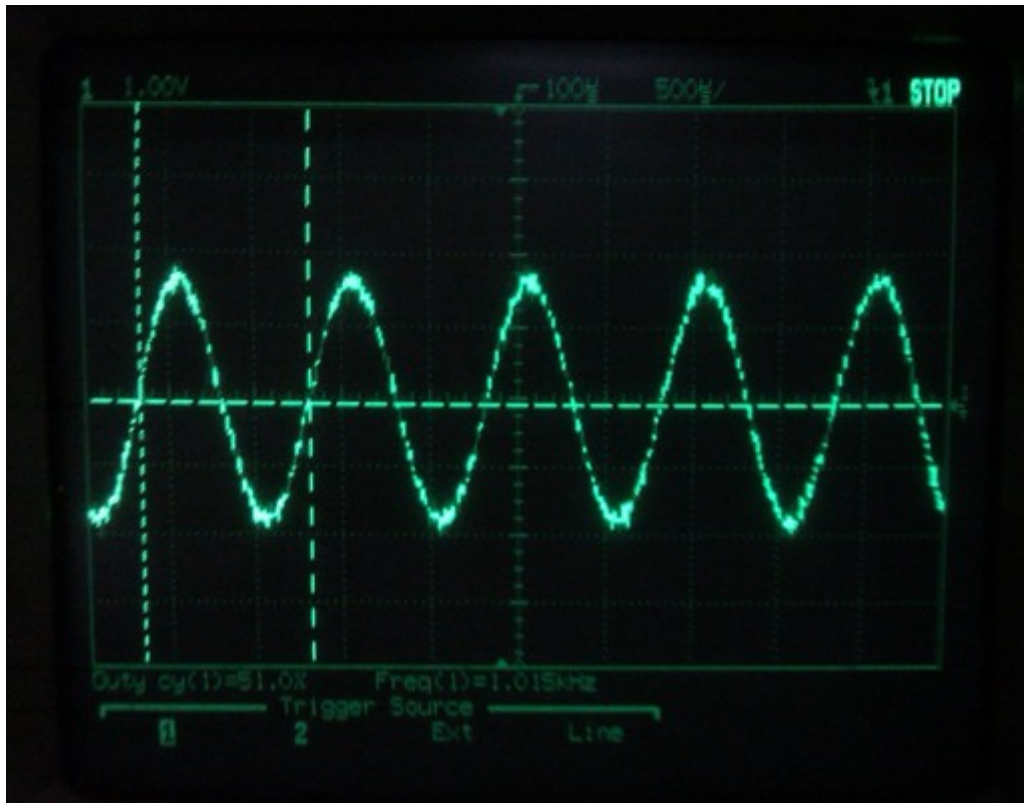


Sampling Theory



Swansea College of Art

Founded 1853

MMusTech Level 4
The Studio Engineer
Handout #3



PRIFYSGOL CYMRU
Y Drindod Dewi Sant
UNIVERSITY OF WALES
Trinity Saint David
SWANSEA - ABERTAWE

Introduction

This guide presents the principles of digital audio from an introductory level. The guide is not designed to give direct advice on digital audio, but more to help form an overall understanding of how sound is represented as digital information.

What is digital audio?

Digital data is a method of storing values in binary form. Binary is essentially the language of computers. If we want to use a computer to work with information from the real world (in our case sound) then this information must be presented in or converted to binary. All digital file formats store information in binary form, regardless of their type. Text documents, digital images and digital video files all contain information in binary. The file format that contains the binary and the way which the binary is encoded (stored) presents the binary data as a word processing document, a photo, a video and so on - usable files. Binary data itself is meaningless unless it is stored in the correct format, one that applies to the source of the binary data (sound, text and so on). For example audio binary information would be useless in a PDF format but far more usable when saved as an MP3.

Sound waves travel through the air by the changes of air pressure at different moments in time. The stronger the change in pressure then the resulting effect becomes louder. And as the speed of the air pressure changes becomes faster the perceived pitch becomes higher. These changes of amplitude (loudness) and frequency (pitch) are the two principle factors of sound. Other conditions such as the environment of the sound (such as the room) and the source of the sound help shape the amplitude and frequency to provide the sound's character, but essentially these two things are the defining components we are interested in. Analogue audio recordings, such as tape, capture continuous changes of sound during recording. For example sound pressure recorded through a microphone is converted to electrical voltage. The changes in voltage represent changes

in amplitude and frequency and is recorded onto a medium such as tape. Digital recording does a very similar thing, the main difference being that the recording is not continuous it happens in steps, at regular time frames. At each step the amplitude and frequency are recorded (known as sampling). The closer together that samples are recorded then the more accurate the recording becomes. When samples of amplitude and frequency information are captured are too far apart then essential information is missed. The reason that digital audio is not a continuous recording process, like analogue recording, is that it just isn't possible. However with the right technique enough information can be captured to naturally recreate the waveform of sound travelling through air. This information is stored into a digital file format, and the information therein can be used to reproduce the original acoustical sounds at a later time.

Binary form

Binary form is the decimal counting system that is used for storing digital data. Binary numbers are expressed in bits where a 'bit' is an abbreviation of the term 'binary digit'.

As opposed to decimal numbers (known as base 10) which use the ten digits 0 - 9, binary (known as base 2) only uses two digits, 0 and 1.

It is highly efficient in computer systems, as only two numerals are needed in ascertaining an electronic system's primary concern, which is whether a state is ON or OFF.

Using binary's two digits to express decimal values is expressed in the following table.

Binary Form					Decimal Form
16	8	4	2	1	
				0	0
				1	1
			1	0	2
			1	1	3
		1	0	0	4
		1	0	1	5
		1	1	0	6
		1	1	1	7
	1	0	0	0	8
1	0	0	0	0	16
1	1	1	1	1	31

This table shows how the 1's and 0's that make up binary expressions can be used to represent decimal numbers. Each binary column relates to a decimal value starting from 1 upwards (1, 2, 4, 8, 16, 24 and so on). For example a decimal value of 6 is equal to the binary value *110*, and 31 is expressed as binary *1111*.

Negative integers can also be expressed in binary form, commonly using an extra left most bit to indicate whether '1' the number is positive or '0' whether a number is negative. All digital information is stored in binary, and in digital audio files, this refers to the values of the recorded changes in electrical voltage, or amplitude. Frequency is derived from the relationships between the values.

Analogue to Digital Conversion

The Audio Signal Path

In order for binary data to be heard correctly it must be converted back to changes in air pressure via changes in voltage for headphones or loudspeakers. The diagram below illustrates the signal flow of sound from an analogue input through a digital system and back to an acoustic output.

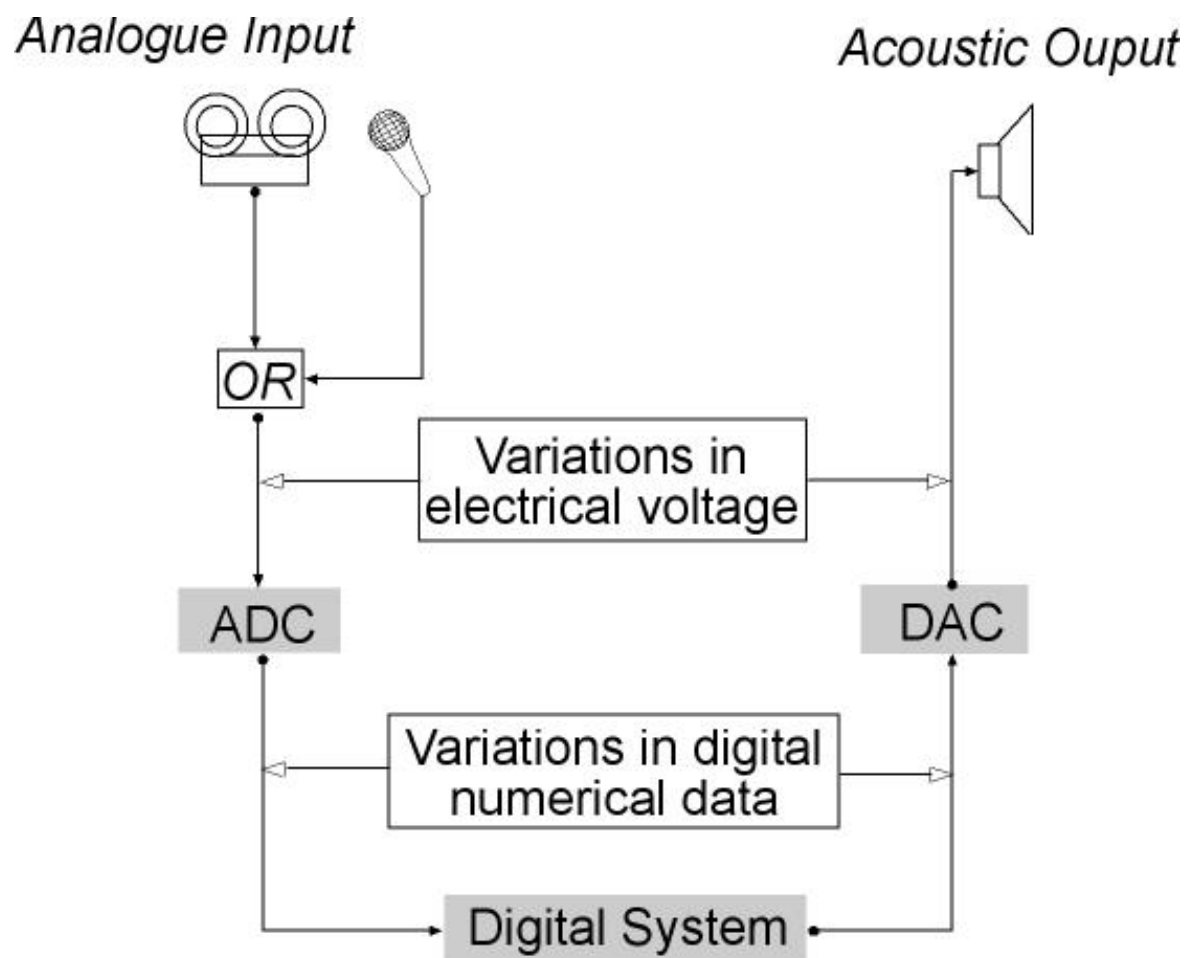


Diagram 1. Signal flow of recording audio into a digital system. The diagram shows how analogue audio is converted to digital information for capture and then the process is reversed for listening back to the sound

1. A transducer (which converts one form of energy to another) such as a tape machine or a microphone converts air pressure variations into an electrical signal (voltage).
2. An analogue-to-digital converter (ADC) converts the signal into digital numerical data by repeatedly measuring the signal, with regards to changes in voltage.
3. The numerical data is passed to a digital system and analysed, stored or manipulated.
4. The digital system creates a stream of output values, either from the stored values, manipulated values of the input or as a continuous real time stream of the input source.
5. A digital-to-analogue converter (DAC) converts the output of the digital system to variations in electrical voltage.
6. An acoustic output, such as a loudspeaker device, converts voltage changes to air pressure variations (audible sound).

Sampling

Sampling is the method of converting an analogue signal to digital data. A useful analogy is how a series of photographs or images (like a flip book) can mimic the perception of motion when they are viewed one after another in quick succession. In a similar way digital sampling does the same thing, it takes snapshots of information at quick and regular intervals in an attempt to recreate a constant waveform.

An acoustic waveform is a continuous line of information. When viewed extremely close up there are no breaks in the waveform, even when distortion or noise is present. This continuous line is represented numerically by a digital system at specified points in time. Subsequently, the quality of the digital reading increases alongside the number of samples taken.

Diagram 2a (overleaf) shows an analogue waveform with sample points marked out. Changes in the signal occur on the Y-axis against time on the X-axis. The sample values are captured and are shown in diagram 2b relative to the time captured. Once a reading is taken, this value is stored and held until the next sample. This is known as sample-and-hold and is common in most digital audio systems. The effect can be seen in Diagram 2c. A digital system uses a process known as anti-aliasing to smooth the curves between the sample steps, to help recreate the original waveform as seen in diagram 2d. In practice this process is done by the analogue to digital converter (ADC) within the audio interface or sound card. To be able to listen to the newly created digital information the reverse process is required, and this is handled by the digital to analogue converter (DAC).

The reading of these values occurs at a fixed rate by a sample clock within the ADC and is known as the sampling frequency, i.e. once every .25ms, expressed as 4000Hz ($f = 1/t$). This is also known as pulse code modulation (PCM) sampling.

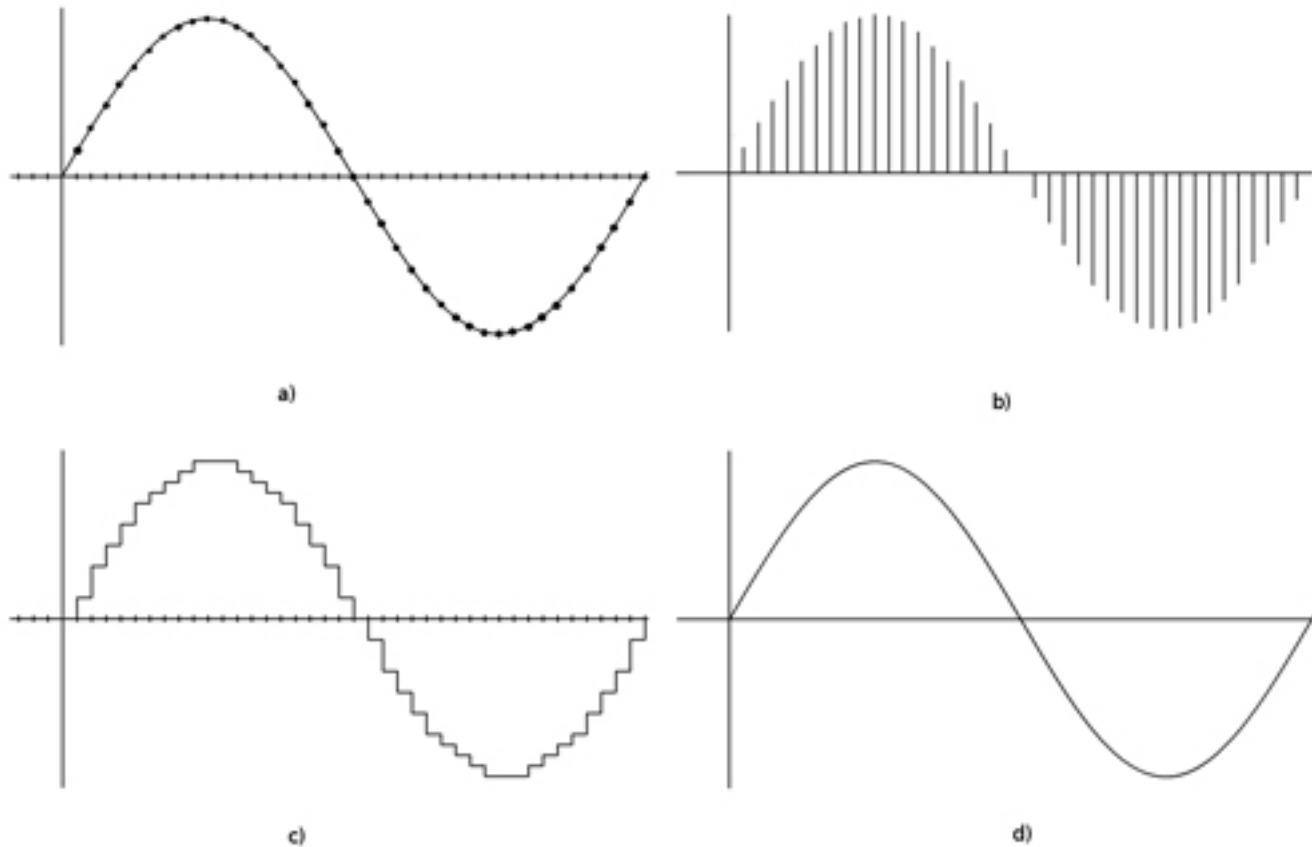


Diagram 2. Diagram 2a shows a sound wave with sample points at regular intervals in time from right to left on the x-axis. The sample values shown in Diagram 2b create the digital representation of the wave file shown in Diagram 2c. A low pass filter is used upon output to smooth the edges of the new digital wave, which represent added high frequencies, with the resultant wave shown in Diagram 2d.

Audio sampling is based on the Nyquist-Shannon sampling theorem, which states that for accurate signal reconstruction the sampling frequency should be at least twice the bandwidth of the source signal. Therefore, if a source signal has a bandwidth of 0 - 1KHz, then the sampling rate needs to be at least 2KHz.

Sample Rates

The quality of reproducing an analogue signal is increased with a higher sampling rate, as more information is captured. However, the question of which sample rate is most effective to preserve optimum quality and accuracy is dependent on a number of factors. With constant advances in technology, the storage of large amounts of data generated from high sample rate recordings is becoming less of a concern. On the other hand, when digitising audio from low quality media sources (for example, reformatting a low quality digital file), a higher sample rate may not be necessary. If storage space is of no concern and optimum quality is required, then it is advisable to digitise with the highest sample rate available, however this issue is often a compromise.

The maximum frequency audible to the human ear is approximately 20KHz (this decreases considerably with age). Therefore, in keeping with the Nyquist-Shannon theorem the sampling rate needed to capture all of the audio in this range needs to be approximately 40KHz, although it should be pointed out that many recordings may not have content that contains frequencies at the higher or lower end of this range). Due to the practicalities of building these electrical systems the most commonly used baseline sample rate is 44.1KHz (which also stems from early video digitisation systems),. However electronic systems in practice behave differently than the ideal theory of principles and there is a strong argument that optimum audio fidelity is often not being achieved. Studies have shown that methods of modern post-production can often utilise frequencies higher than the human audible range to benefit the audible content, but in practice this is widely debated and varies between recordings. As a result higher sample rates are being used more within pro-audio systems. 48KHz, 96KHz, and 192KHz are becoming more commonly used as the CD is starting to be replaced by media capable of supporting high sample rates.

Even though we may not hear frequencies above the threshold of hearing, they can still modulate frequencies within the audible range, and thus affect the sound. High sample rates also allow for the flaws in anti-aliasing filters (mentioned

later) not to affect high frequencies within the audible range due to imperfections of electronic systems in practice. It is considered that a sample rate of 44.1KHz is adequate for most work at a consumer level but a higher rate with a minimum of 48KHz is appropriate for more accurate results of digitisation, especially when archiving material.

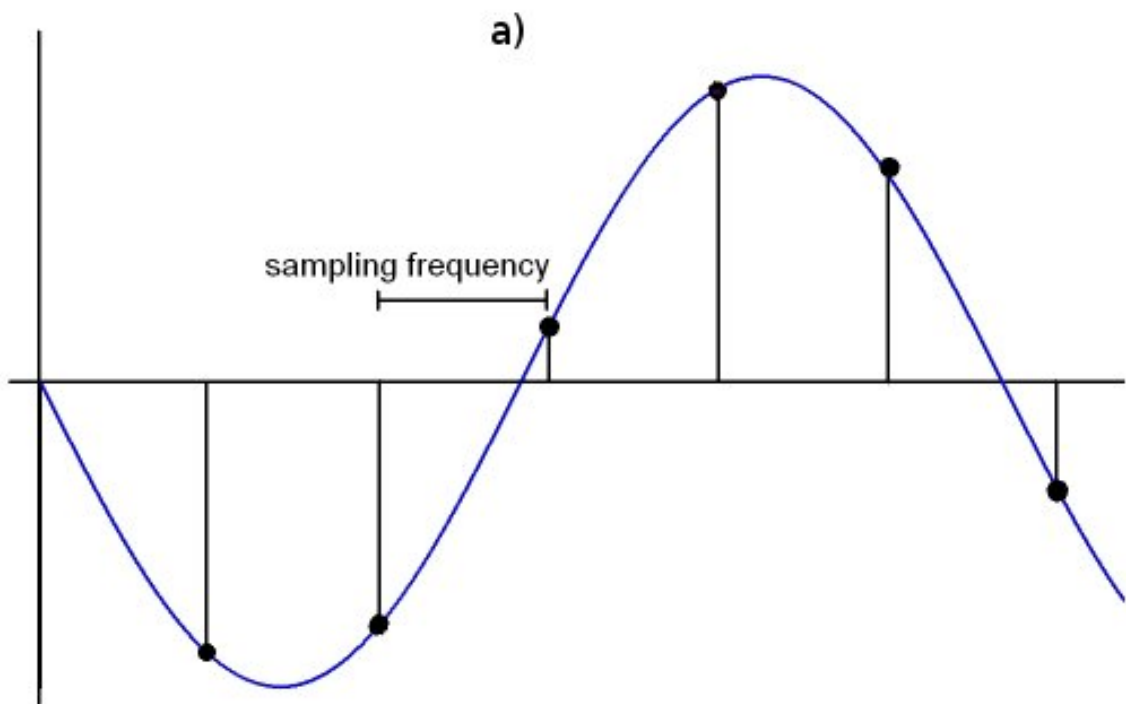
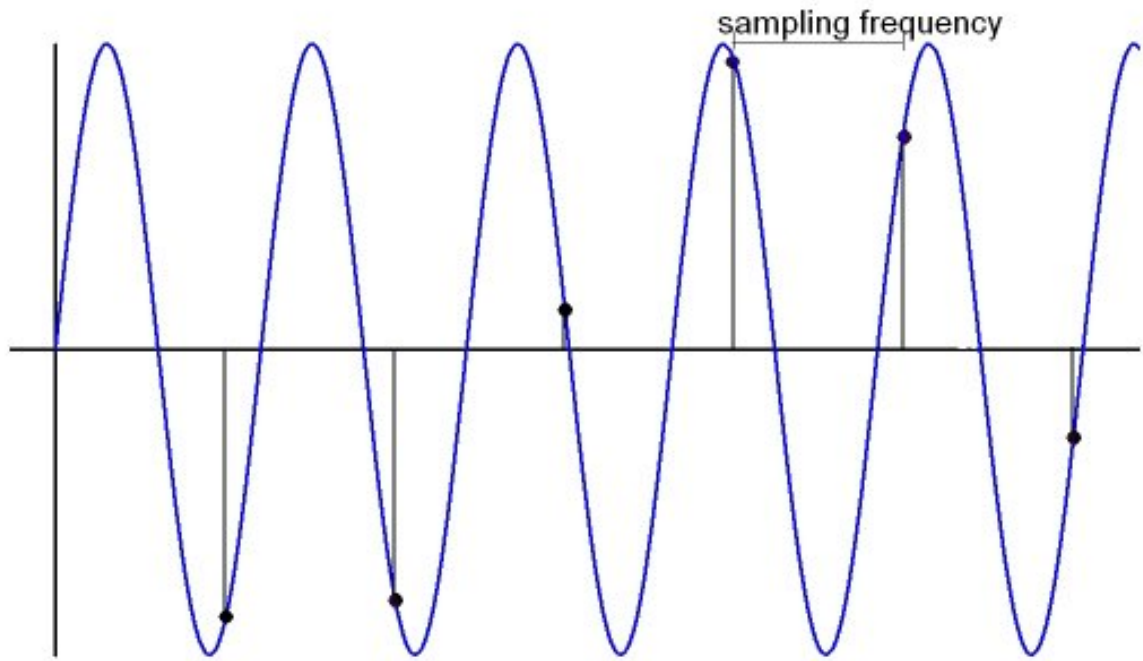
Aliasing

One of the main reasons the sampling theorem needs to be adhered to is to avoid the effects of aliasing. When the sample rate is less than twice the frequency of the maximum frequency of the source material, a new alias frequency is created from the sample points (see Diagram 3 below), which is added to the signal. The effect is known as aliasing. Aliasing occurs when the lower alias frequency produced is in the audible range. Multiple frequencies above this threshold create further alias', which also become blended in with the original signal.

Providing the input frequency is higher than half of the sampling frequency, the approximate new alias frequency can be derived from:

Alias frequency = sampling frequency - input frequency

For example, an input frequency of 26Khz sampled at a rate of 40Kz, produces an alias frequency of 14Khz (40Khz - 26KHz).



b)

Diagram 3. 3a shows the sample points on an source waveform before analogue-to-digital conversion. As the input signal is greater than half of the sampling frequency the original input signal is filtered out producing a lower alias frequency shown in Diagram 3b. The frequency of the waveform is measured by the time it takes to

complete one full cycle. In the diagram the sampling frequency is not fast enough to capture the information needed to replicate the source waveform.

To reduce the effect of aliasing, a low-pass filter is used at the input stage of an ADC that removes any frequencies higher than half the sample rate. This is referred to as an anti-aliasing filter, the effects of which are shown in Diagram 2. Although a low-pass filter will theoretically remove aliasing, in practice problems occur due to the properties of electronic components that potentially degrade the quality of the signal once anti-aliased.

Quantisation

Quantisation errors occur during sampling where the areas between samples hold no data, as opposed to an analogue signal which is a constant stream of values. Each digital sample is rounded up to the nearest value which in turn deviates from the original source signal. The result of this is audible artefacts, known as quantisation noise.

In digital systems if the input signal is silent, then the result is digital silence; a sample of value 0. However, with a sinusoidal wave input the rounding up of values creates an audible effect, the granulation effect, especially as the waves decay to zero. This noise is similar to the hiss of quiet or silent tape being played.

The more accurate the sample reading then the less need there is for quantisation. The bit depth (see below) is the range of accuracy the sample can be read at.

Dither

Dither is a common technique used to reduce quantisation noise. Dither adds an element of very low-level noise to a signal prior to analogue-to-digital conversion. This noise forces the quantisation process to jump between adjacent levels at random. The audible effect is known as *soft-landing* as the grittiness is converted into a soft pool of low-level noise, which is far more natural on the human ear.

Dither is not generally considered necessary at higher bit depths, such as 20-bit and above, as the human ear is not as susceptible to quantisation errors at this quality. However when converting to 16-bit and lower, dither becomes necessary to maintain high audible quality.

Digital distortion (Clipping)

When sampling analogue audio care needs to be taken to the input gain level of the analogue source going into the digital system. Gain is the amplitude level of the signal that is fed into the ADC. In digital systems the available range of dynamic levels (loud and quiet) are fixed. Any signals above the maximum distort, and the result is a loss of sound quality and audible distortion, which is far less forgiving than analogue distortion.

Imagine forcing a pie into a cylindrical pipe whose diameter was much shorter than the width of the pie. In order to successfully force the pie into the pipe, edges of the pie would break off flattening each side. If you were to examine the pie as it came out of the other side of the pipe it would be fair to say that it wouldn't resemble the original pie, before it entered the pipe.

The volume ranges of analogue and digital systems are not matched, and this needs to be addressed before recording audio or digitising media. Let's imagine you were to play an analogue audio file into a digital system. Let's also imagine that the digital system has an input range of 1 - 10. Every time the volume of the analogue file goes above what the digital system perceives as 10, the digital system would run out of available range to capture the sound properly. The result is that the signal gets clipped at the maximum level and is heard as distortion. This effect can be seen in Diagram 4 where the input signal's amplitude is increased half way along the time line. The solution in this case is to turn down the volume of the analogue input signal so that it never goes above 10.

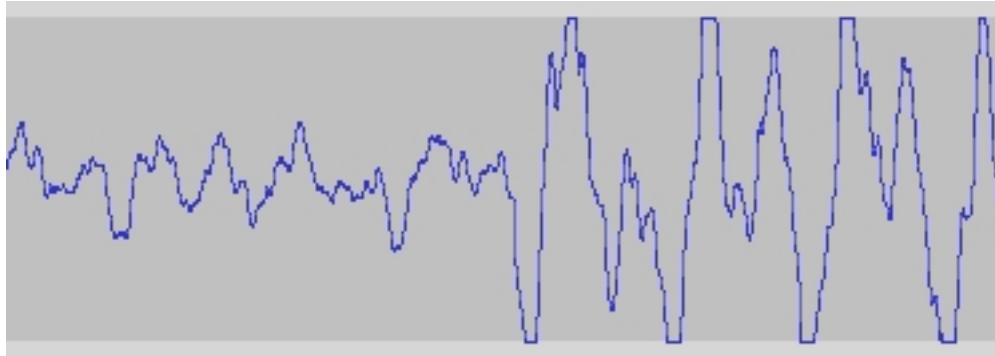


Diagram 4. Clipping occurs during the second half of the waveform where the peaks and troughs of the wave are harshly rounded off.

Digital clipping is not easy to undo. When a file is clipped the wave shape is altered and would need to be repaired back to the original form of the analogue file. If this data is lost through clipping then it can be very complicated to render it back to its original form. As shown in diagram 4 an analogue input signal becomes clipped when sampled. The resultant waveform is shown with the distorted elements visible in the peaks and troughs of the wave.

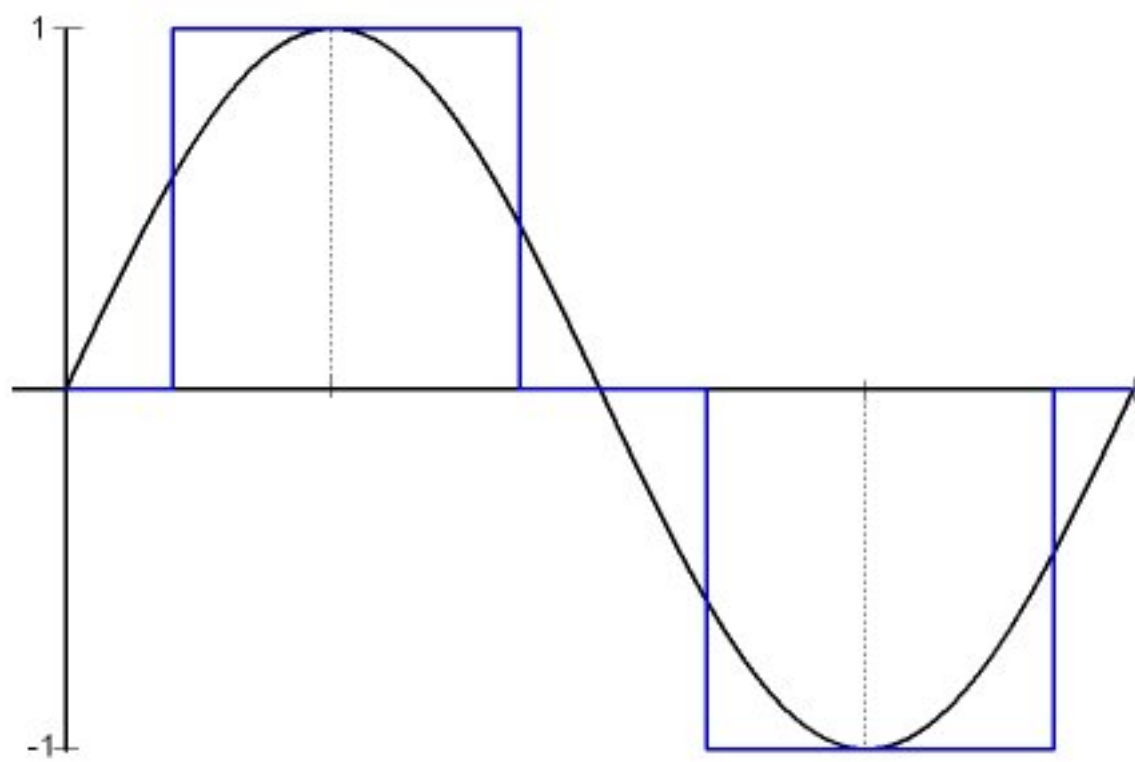
To avoid digital clipping, gain levels should always be set before digitisation so that the maximum output level of the analogue input is well within the limits of the range of the DAC.

Bit Depth

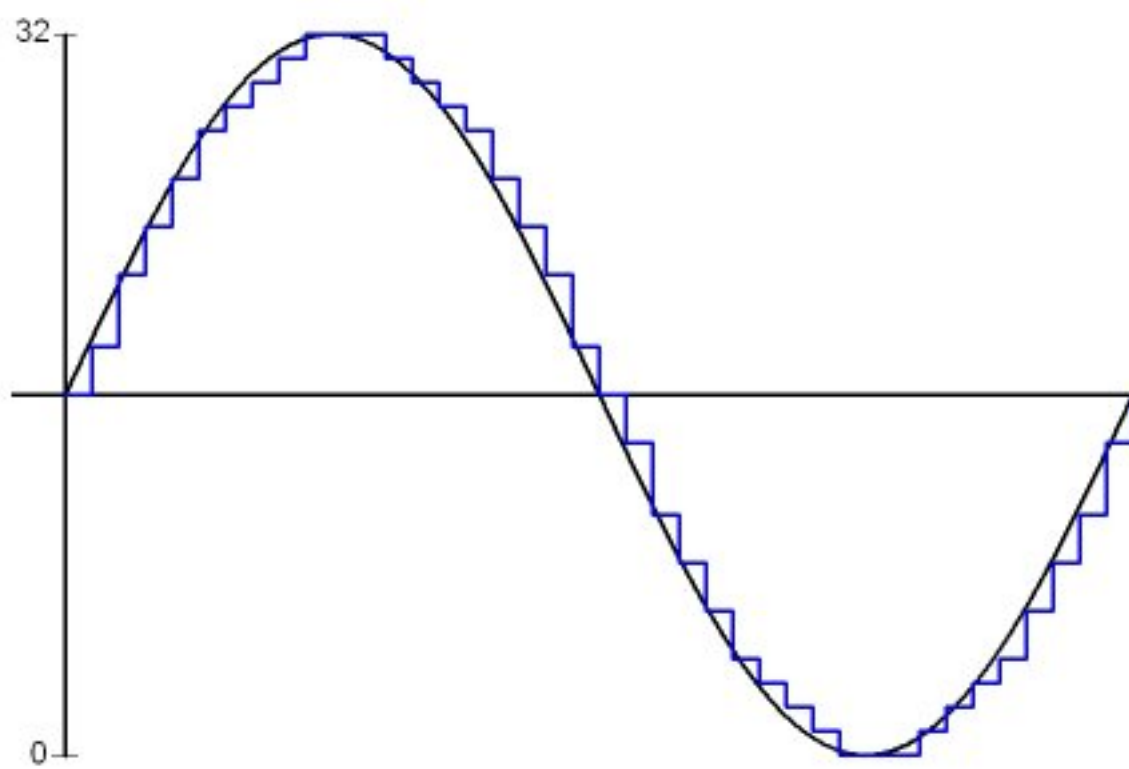
The bit-depth (or sample-width) determines the accuracy of the amplitude measurements per sample, i.e. the range of values available per sample.

A higher bit-depth allows for greater audible perception in the subtle changes in loudness - the dynamic range. This is a measurement of the ratio of the loudest available undistorted noise and the quietest available. 16-bit sampling is the standard baseline for consumer level audio (used in CD audio). However, 24-bit sampling offers a larger dynamic range and is becoming more commonly used in production and digitisation. Similar to the sample rate, a higher bit depth demands greater computational speed from the ADC. The maximum available will vary depending on the equipment in use.

Some systems use lower bit-depths where quality is less of an issue and there are smaller computational demands on the system in use. For example audio in telephony uses 8-bit conversion as there is not a need for high audio fidelity, and to keep the costs down of manufacturing the components being used to process the audio.



a)



b)

Diagram 5. The accuracy of 2-bit quantisation (Diagram 5a), with three levels of amplitude resolution compared to 5-bit quantisation (Diagram 5b), with thirty-two levels of amplitude resolution. Shown against the black input waveform. In both examples the same sampling frequency is used (X- axis), but the number of available sampling values differs; 3 vs. 32 (Y-axis).

CD quality is 44.1Hz sampling rate and 16-bit. This was the standard laid out in the 1980s and was cost effective and adequate for the time. For spoken word and most musical content 16-bit sampling is considered to be of adequate quality. However, as standards and quality of digital equipment increases, higher quality through a greater dynamic can be achieved with 24-processing.

Bit Rate

Bit rate (which is easily confused with bit depth) refers to the transfer of data, such as downloading audio files via the internet. Compressing a digital file (making the file smaller) is often useful for sharing files, either for downloading or streaming. This is due to the larger size of uncompressed audio files and the extra time it can take to stream or download on slower, and sometimes even on fast network connections.

The bit rate is the expression of the amount of information that is stored per second. It is dependent upon the type of codec used to create the file (such as an MP3 audio file) and which, if any, data compression scheme is applied. A simple equation can be used to calculate the bit rate of a file when the following values are known.

Bit rate = (bit depth) x (sampling rate) x (number of channels)
For example, a 24 bit file sampled at 44.1Khz in stereo (2 channels) has a bit rate of $24 \times 44100 \times 2 = 2116800 \text{ bits/sec}$ or 21168 kbits/sec

In practical terms most sound editing software has a function to convert a file to a different bit rate. The options are often presented as common standardised bit rates that differ in quality. For example the MP3 codec compresses data into 5 common bit rates ranging from poor to good in quality and relate to different file sizes.

Storage of digital files

Digital audio files can be highly taxing on storage space, especially when created with high sample rates and bit depth. It is advisable to always take into account the size of files you may be working with to optimise your available storage capacity. A larger file size will also impact on the processing demands of a digital system especially when conducting analysis and modification.

For example, an uncompressed audio CD, 44.1KHz sample rate, 16bits (2 bytes per sample), 2 channels (stereo), lasting 45 minutes can be expressed as:

$$44100 \times 2 \times 2 \times 45 \times 60 = 476280000 \text{ bytes} \approx 454.2\text{MB}$$

Alternatively uncompressed audio stored on a digital multi-track recorder, 96KHz sample rate, 24bits (3 bytes per sample), 48 channels, lasting 45 minutes can be expressed as:

$$96000 \times 3 \times 48 \times 45 \times 60 = 37324800000 \approx 34.76\text{GB}$$

Discussion

Opinions of audio sound quality and fidelity have long been split in the audio industries due to the subjective nature of quantifying such matters. There are lots of arguments that technology still lags behind the theory of perfect digitisation, and many argue that digital sound has much harsher characteristics than analogue sound on the human ear. However, with the recent advances in modern technology and the need for long-term digital storage, digital audio processing has been at the forefront of creative and practical development, and has even sprung into its own art forms. It is generally considered that with high quality equipment, which is configured correctly, digital conversion of analogue audio is so transparent that all the necessary information from an analogue audio file can be transferred into the digital domain.

The most important component in the conversion process to the resultant quality of replication is the ADC. It is important to note that even though manufacturers claim to have similar specified equipment available, some ADC greatly outperform others when it comes to the internal processes of the converters which capture the information. A general rule of thumb to go on is the more expensive the converter then the better quality unit you will receive, although this is not always the case and nearly all systems are known to leave imprints of certain audio characteristics.

One thing that technology cannot replace in many ways is the subjectivity of the human ear, the most critical tool when working with sound. So when questions arise with regards to fidelity and quality, aside from standards of practice laid out, the most reliable tool is the trained ears of the listener.

Glossary

Bandwidth

The range of available frequencies for a specified signal or system. For example, a source file containing frequencies from 100Hz to 15KHz has a 14.9KHz bandwidth between the two frequencies.

Decibel

The decibel is a measurement of the relationship between sound intensity and perceived loudness by the listener.

PCM

Pulse Code Modulation (PCM).
In terms of digital audio files a linear PCM scheme stores binary integer values for each sample.

Sample-and-hold

A technique used in electronics when interfacing with real world inputs. A sample is taken with most commonly the voltage reading being stored within a capacitor. A switch disconnects the capacitor from the input, and thus the sampled value is held. At the next period of the sampling frequency the switch is opened, allowing for the input value to be read again then closed, repeating the process.

Sampling theorem

Often referred to the Nyquist or the Nyquist-Shannon theorem. Originates from work and research done in 1928 and 1933 in telecommunications and information systems.

*Document adapted from:

<http://www.jiscdigitalmedia.ac.uk/guide/an-introduction-to-digital-audio>