MUSC 208 Winter 2014 John Ellinger Carleton College

Digital Audio Fundamentals

Digital audio is a mix of mathematics, computer science, and physics. Sound waves are converted into streams of numbers that are processed by the computer and converted back into a sound wave.

A modern recording session uses both analog and digital hardware. The analog devices are the microphone and the speaker. The microphone converts sound waves into voltages and the speaker reverses the process converting voltages into sound waves. The digital devices are the ADC (Analog Digital Converter), an optional DSP (Digital Signal Processing) unit, and the DAC (Digital Analog Converter). The ADC, DSP, and DAC that are found within a modern computer are sufficient for all but the most critical audiophile recordings.



http://upload.wikimedia.org/wikipedia/commons/8/84/A-D-A_Flow.svg

Input

When sound waves hit the diaphragm of the microphone the diaphragm moves. As the diaphragm moves it generates very small voltage fluctuations. The voltages are so small they need to be amplified to be useable. This amplification is done either through a microphone preamplifier or a mixing board. When graphing an analog signal, the x axis represents time and the y axis represents amplitude. Analog signals are continuous in the mathematical sense that a y value exists for every x value.



ADC (Analog Digital Converter)

The analog signal then travels to the ADC to be converted into numbers the computer can process. In the past very expensive hardware devices were needed to convert the analog signal into a digital stream of numbers. Today the ADC is part of the computer. Before the microphone signal gets to the ADC it needs to be amplified and passed through a low pass anti-aliasing filter. The low pass filter is necessary because of a fundamental principle of digital audio, the Nyquist theorem.

Nyquist Theorem

Harry Nyquist described this theorem in a 1928 paper. Here is one definition of the Nyquist theorem from Wikipedia:

"In essence the theorem shows that an analog signal that has been sampled can be perfectly reconstructed from the samples if the sampling rate exceeds 2B samples per second, where B is the highest frequency in the original signal. If a signal contains a component at exactly B hertz, then samples spaced at exactly 1/(2B) seconds do not completely determine the signal."

http://en.wikipedia.org/wiki/Nyquist-Shannon sampling theorem

Intuitively it takes two points to sample one period of a sine wave.



Nyquist Rate

The Nyquist Rate is the minimum sampling rate needed to completely capture the highest frequencies that occur in the sound to be sampled. If the highest frequency in

the sound to be sampled is f, then the Nyquist rate is 2f. In practice the sampling rate is always somewhat higher than twice the highest frequency expected. The Nyquist Rate applies to both analog and digital signals. The Nyquist Frequency applies only to digital signals.

Nyquist Frequency

The Nyquist frequency is the frequency is equal to one half the sampling rate. Any frequencies in the signal before sampling that are higher than the Nyqust frequency will be appear as an aliased frequency in the samples. The Nyquist frequency is sometimes referred to as the foldover frequency because signals that exceed half the sampling rate are folded back (aliased) into the sampling rate as if the graph below was folded right top to bottom at the horizontal SR/2 line.



Aliasing

Any frequency f above the Nyquist Frequency will be heard at the alias pitch of SR - f. Frequencies from 0 Hz to the Nyquist frequency are heard at their true frequency. Frequencies from SR/2 Hz to SR Hz are heard as descending frequencies according to the formula SR - f. Frequencies from SR Hz to integer multiples of SR Hz rise and fall similar to the o-SR range. This graph illustrates what happens as frequencies exceed the Nyquist frequency.



The following picture shows you what would happen if the sample rate was 1000 Hz and the signal contained a frequency of 700Hz. The 700 Hz frequency would be fold over to the left of the Nyquist frequency (NF) and be heard as a tone of 300 Hz.



Negative Frequencies

According to the alias formula a frequency of 46100 Hz sampled at 44100 Hz will be aliased to a frequency of -2000 Hz. A negative frequency is a positive frequency phase shifted by 180°. You can't hear the difference.

After the low pass filter has removed signals above the Nyquist frequency the signal goes to the Analog Digital Converter.

Analog Digital Converter (ADC)

The Analog Digital Converter (ADC) converts the amplified voltage signal into a stream of numbers. The ADC determines the rate at which the numbers are produced (sampling rate) as well as the minimum and maximum numbers used to represent changes in amplitude (bit depth).

When graphing digital signals, the x axis represents time as uniformly spaced discrete samples and the y axis represents amplitude values at each sample time. The amplitude values in between sample times are unknown and undefined.



While it's tempting to think that the above signal must have been a sine wave...



it could also have been a very jagged wave because we don't know what happened between the sample points.



Sample Rate

The number of samples taken per second is called the sampling rate. The higher the sample rate the more closely the digital sound will match the analog signal.

Audio CD Sampling Rate

Audio CD's are sampled at a rate of 44,100 samples per second. The sampling frequency is 44100 Hz or 44.1 samples every millisecond. The sampling period is 1/44100 or 0.00002267 second. The audio CD sampling rate will capture frequencies up to the Nyquist frequency, 22050 Hz, well above the range of human hearing.

The following plots show the effect of sampling a one second one Hz sine wave at different sample rates. You can see from the plots that the more samples per second, the more accurate the sine wave.





64 Samples Per Second



Bit Depth

Bit depth determines the minimum and maximum range of numbers and that represent the amplitude of the signal. The greater the bit depth, the more gradations there are between loud and soft passages. The bit depth of an audio CD is 16 which means

amplitude values can range from zero to $2^{16} = 65,536$ possible values. In practice half the values are positive and half are negative shifting the range from 2^{-15} to 2^{15} , or ±32.767. Bit depths of 24 are also used which represents $2^{24} = 16,777,216$ values that range from ±8,388,607. The resulting sample values are further normalized to the range -1.0 to +1.0 used in the DSP unit. These plots show result of sampling a sine wave at various bit depths.

A bit depth of 4 can represent $2^4 = 16$ values and has an amplitude range from -7 to +7.



A bit depth of 6 can represent $2^6 = 32$ values and has an amplitude range from-31 to +31.



A bit depth of 7 can represent $2^7 = 128$ values and has an amplitude range from -63 to

m208w2014



A bit depth of 8 can represent $2^8 = 256$ values and has an amplitude range from -127 to +127.



Audio Storage Requirements

The CD sample rate of 44,100 samples per second and the bit depth is 16 or two bytes for each sample. Stereo sound uses two channels left and right making 88,200 total samples. One minute of stereo will use 88200 samples per second * 60 seconds * 2 bytes per sample = 10,584,000 bytes. That's 10 Megabytes per minute. An audio CD can hold about 640 Mb, or about an hour's worth of music. Increasing either the sampling rate or the bit depth will further increase the size needed to store the data.

Sample Rate	In Ads	Bit Depth	Number Bytes	One minute of mono in Megabytes	Audio CDs needed for one hour stereo
44100	44.1K	16	2	5168	0.9
44100	44.1K	24	3	7752	1.4
48000	48K	16	2	5625	1
48000	48K	24	3	8438	1.5
88000	88K	16	2	10313	1.9
88000	88k	24	3	15469	2.8
96000	96K	16	2	11250	2.1
96000	96K	24	3	16875	3.1
192000	192K	16	2	22500	4.1
192000	192K	24	3	33750	6.2

Prefixes

These prefixes refer to numerical quantities. For example a 1 gigahertz computer's CPU is timed with a clock running in nanoseconds. A slow digital audio recorder can record 44.1K samples every second. I purchased my first computer hard drive in 1987, a 1 Mb (Mega-byte) drive that cost \$1000. I recently purchased a 3 Tb (Tera-byte) hard drive for \$169.

Prefix	Value	Power of 10	Abbreviation
Tera	1,000,000,000,000	10 ¹²	т
Giga	1,000,000,000	10 ⁹	G
Mega	1,000,000	10 ⁶	М
Kilo	1,000	10^{3}	К
Deci	0.1	10^{-1}	d
Centi	0.01	10^{-2}	С
Milli	0.001	10 ⁻³	m
Micro	0.000001	10 ⁻⁶	μ
Nano	0.00000001	10 ⁻⁹	n
Pico	0.00000000001	10 ⁻¹²	р

DSP Unit

After the signal leaves the ADC it may undergo further Digital Signal Processing (DSP). Today DSP effects are done inside the computer with specialized software packages. There are many free effects available for download on the internet. Many of these effects are in the VST or AU format which can be loaded as plug-ins in most of todays audio software. Common DSP effects amplify the sound, change the duration or pitch, add reverb, emphasize or attenuate selected frequencies, or emulate expensive hardware devices of the past. DSP effects can range from subtle enhancement to wild distortion.

After any optional DSP processing, the signal is almost ready to be played but first it needs to be passed through the Digital Analog Converter.

Digital Audio Converter (DAC)

The DAC converts the processed digital signal back into a an analog signal. Sometimes DSP effects add unwanted frequencies above the Nyquist frequency that need to be filtered out before playback. The signal is sent through another low pass filter.

Low Pass, Reconstruction Filter

This Low Pass filter removes those unwanted frequencies and is sometimes called a smoothing filter. The signal is once again an analog signal that can be sent to the output device.

Output

This sampled, processed, smoothed, and reconstructed analog signal can finally be played through speakers or headphones.

Bels and Decibel

A Bel is a sound intensity measurement named after Alexander Graham Bell, the inventor of the telephone. The Bel scale is a logarithmic scale whose units are powers of ten. One Bel is 10^1 and 4 Bels is 10^4 . The Bel scale measures ratios between the lowest intensity sound we can just barely hear (10^0) and the highest intensity sound we can tolerate before the sound becomes painful (10^{12}) . The lowest intensity is referred to as the threshold of hearing. The highest intensity is referred to as the threshold of pain. The ratio between them is 12 Bels or one trillion to one. A decibel is 1/10 of a Bel. In the decibel scale the the ratio of the threshold of pain over the threshold of hearing is 120 decibels or 120 dB. The B is capitalized in honor of Alexander Graham Bell. Bels and decibels have no physical units, they are simply numbers that express a ratio of how much louder or softer one sound is to another. A 10 dB difference between sounds is a 10 times increase in intensity. A 40 dB difference between two sounds is an intensity difference of 10,000 (10^4). Whether the two sounds were 10 dB and 50 dB or 70 dB and 110 dB there is a 40 dB difference.

Positive dB's represent an increase in volume (gain) and negative dB's represent a decrease (attenuation) in volume. Every 10 decibel change represents a power of ten change in sound intensity. For example, there is an 80 dB difference between the softest symphonic music (20 dB) and loudest symphonic music (100 db). That's an intensity

difference of one hundred million, 10^8 .

Powers of Decibels **Positive Magnitude** Description 10 10¹⁶ 160 1,000,000,000,000,000,000,000,000 Space Shuttle 10¹⁵ 150 1,000,000,000,000,000,000,000 10¹⁴ 140 1,000,000,000,000,000,000 Jet takeoff

This table shows some relative decibel levels.

10 ¹³	130	1,000,000,000,000,000	
10 ¹²	120	1,000,000,000,000	Threshold of pain Amplified rock band
10 ¹¹	110	100,000,000,000	
10 ¹⁰	100	10,000,000,000	Loudest symphonic music
10 ⁹	90	1,000,000,000	
10 ⁸	80	100,000,000	Vacuum Cleaner
10 ⁷	70	10,000,000	
10 ⁶	60	1,000,000	Conversation
10 ⁵	50	100,000	
10 ⁴	40	10,000	
10 ³	30	1000	
10 ²	20	100	Whispering Softest symphonic music
10 ¹	10	10	
10 [°]	1	1	Threshold of hearing

Digital audio often reverses the decibel scale making o dB the loudest sound that can be accurately produced by the hardware without distortion. Softer sounds are measured as negative decibels below zero. Software decibel scales often use a portion of the o dB to 120 dB range and may choose an arbitrary value for the o dB point.

Powers of 10	Decibels	Magnitude	Magnitude Squared Power Audio Amplitude	A Possible Software dB Scale
10°	0	1	1.000000	10
10 ⁻¹	-10	0.1	0.316228	0
10^{-2}	-20	0.01	0.100000	-10
10^{-3}	-30	0.001	0.031623	-20
10^{-4}	-40	0.0001	0.010000	-30
10^{-5}	-50	0.00001	0.003162	-40
10 ⁻⁶	-60	0.000001	0.001000	-50
10 ⁻⁷	-70	0.0000001	0.000316	-60

10^{-8}	-80	0.00000001	0.000100	-70
10 ⁻⁹	-90	0.00000001	0.000032	-80
10^{-10}	-100	0.000000001	0.000010	-90
10 ⁻¹¹	-110	0.0000000001	0.000003	-100
10 ⁻¹²	-120	0.000000000001	0.000001	-110

This Logic Pro dB scale goes from 0 dB down to -60 dB. The 0.0 dB setting on the volume fader on the right corresponds to -11 dB on the dB scale.



Decibels To Amplitude

Roads lists the decibel formula on page 39 as:

$$dB = 10 \log(\frac{level}{referenceLevel})$$

That's correct as long as we're measuring the intensity or sound pressure levels between two sounds. It's not correct when comparing two different amplitude readings; for example voltage levels from a microphone, or amplitude values in a digital audio waveform editor. Digital audio amplitude values are real numbers between 0 and 1.0. The reference amplitude 1, the maximum amplitude possible. Because voltage squared relates to power the formula to use when calculating amplitude is:

$$dB = 20\log(amplitude) = 10\log(\frac{amplitude^2}{1})$$

As long as amplitudes never exceed 1.0, all dB readings will be negative except for amplitude 1.0 which is 0 dB.

This chart shows decibel values and their amplitude equivalents.

Decibels	Amplitude
10	3.162278
0	1.000000
-10	0.316228
-20	0.100000
-30	0.031623
-40	0.010000
-50	0.003162
-60	0.001000
-70	0.000316
-80	0.000100
-90	0.000032
-100	0.000010
-110	0.000003

Amplitude To Decibels

The formula for converting amplitude to decibels is:

amplitude =
$$10^{\frac{dB}{20}}$$

This chart shows amplitude values and their decibel equivalents.

Amplitude	Decibels
1	0.0
0.9	-0.9
0.8	-1.9
0.7	-3.1
0.6	-4.4
0.5	-6.0
0.4	-8.0
0.3	-10.5
0.2	-14.0

m208w2014

0.1	-20.0
0.01	-40.0
0.001	-60.0

Dynamic Range

The decibel formula can be used to calculate the dynamic range of sound for a given bit depth. The formula is:

dynamic Range in dB for n bits $\approx 6 * n = 20 * \log(2^n) = n * 20 * \log(2) = 6.0206$

Bit Depth	± Amplitude Range	Dynamic Range dB
8	127	48
12	2,047	72
16	32,767	96
20	524,287	120
24	8,388,607	144
32	2,147,483,647	192

It should be apparent from this table that a bit depth of 32 is overkill, and why 24 bit resolution is sufficient for all professional recording systems.

Equal Loudness Contours

Different frequencies at the same dB level may not be perceived as the same volume when we hear them. The dB level low frequency sounds need to be raised match the same apparent volume as a higher frequency sound. The lines on this chart represent isophons or sounds with the same perceived loudness. Frequency is displayed on the X axis, and decibels on the Y axis. The phon lines are named at the reference frequency of 1000 Hz. Sounds below the threshold line are inaudible. Using the 80 phon line, the chart shows that a 100 Hz sound at 92 dB will sound as loud as a 1000 Hz sound at 81 dB.



http://en.wikipedia.org/wiki/File:Lindos1.svg

Noise Induced Hearing Loss

"Sound pressure is measured in decibels (dB). Like a temperature scale, the decibel scale goes below zero. The average person can hear sounds down to about 0 dB, the level of rustling leaves. Some people with very good hearing can hear sounds down to -15 dB. If a sound reaches 85 dB or stronger, it can cause permanent damage to your hearing. The amount of time you listen to a sound affects how much damage it will cause. The quieter the sound, the longer you can listen to it safely. If the sound is very quiet, it will not cause damage even if you listen to it for a very long time; however, exposure to some common sounds can cause permanent damage. With extended exposure, noises that reach a decibel level of 85 can cause permanent damage to the hair cells in the inner ear, leading to hearing loss. Many common sounds may be louder than you think...

• A typical conversation occurs at 60 dB – not loud enough to cause damage.

• A bulldozer that is idling (note that this is idling, not actively bulldozing) is loud enough at 85 dB that it can cause permanent damage after only 1 work day (8 hours).

• When listening to a personal music system with stock earphones at a maximum volume, the sound generated can reach a level of over 100 dBA, loud enough to begin causing permanent damage after just 15 minutes per day!

• A clap of thunder from a nearby storm (120 dB) or a gunshot (140-190 dB, depending on weapon), can both cause immediate damage." <u>http://www.dangerousdecibels.org/education/information-center/noise-inducedhearing-loss/</u>

Decibel (Loudness) Comparison Chart

http://www.gcaudio.com/resources/howtos/loudness.html

The Top 10 Loudest Noises

http://listverse.com/2007/11/30/top-10-loudest-noises/