x

Sound Primer

Notes for Basic Electronic Music and Sound Recording Classes

Prepared by David Adamcyk (rev. May 2025)

1 Acoustics

1.1 Sound

• **Sound** is a disturbance that propagates in a medium. More specifically, sound is the result of the following chain of events:

An object vibrates (ex: violin body) -> the vibrations displace surrounding air molecules -> the displacement of air molecules propagates outwardly -> the displacement reaches a human ear and pushes its ear drum (i.e., the displacement applies pressure on the ear drum) -> the ear system converts the motions of the ear drum into brain signals.

• One has to distinguish between objective (measured by instruments) and subjective (experienced by human listening) properties of sound. Humans only hear sound when the vibrations occur in air at a rate of 20vibration/sec to 20,000vibration/sec. When vibrations are outside of this range, we do not perceive them as sound, though they can still be measured with instruments. We call the **20-20K** vibrations/sec range the **Audio Range.**

• Sound waves are **longitudinal** waves (medium moves in same direction as propagation).
  • See animation

• Sound can be represented in many ways. One very common approach is to plot the variations of **pressure in time,** which is known as the **waveform** presentation of sound:

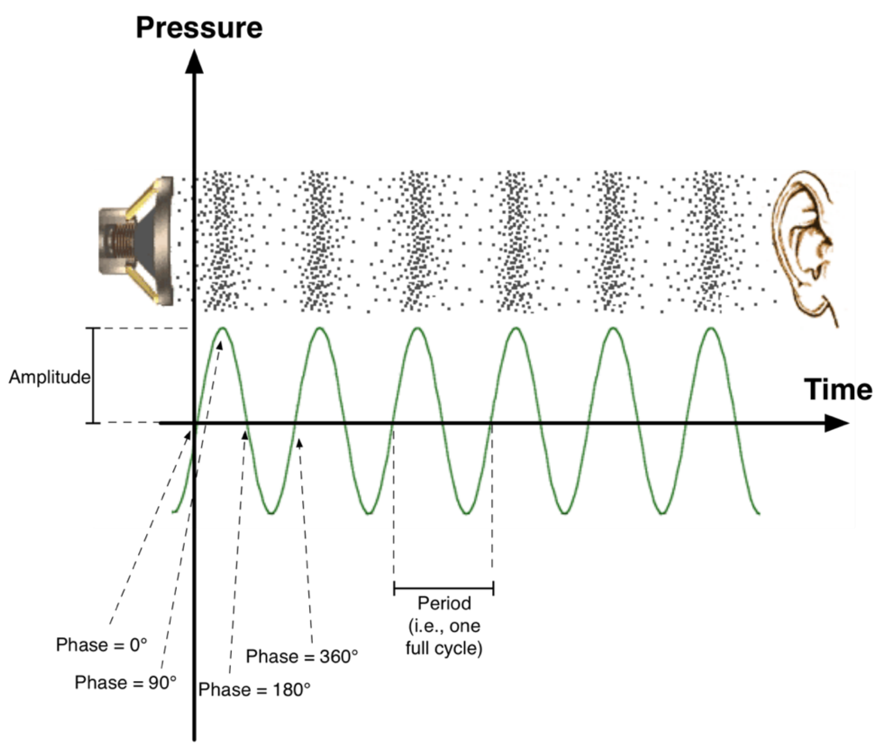


Figure 1 Waveform presentation of sound: variations of pressure in time

• **Frequency**: the number of complete wave periods (or cycles) per second. The unit is the Hertz, cycles/second. Notice that the frequency is a rate, whereas the period is an amount. Frequency is directly linked to what we perceive as **pitch**. For instance, if an object vibrates 440 times per second (i.e., 440Hz), we perceive its pitch as the A above middle C.

• **Amplitude**: the measure of wave's pressure. This is directly linked to **volume**: more pressure (greater amplitude) is louder; less pressure (smaller amplitude) is quieter.

• **Phase**: the fraction of the wave cycle that has elapsed relative to the origin, expressed in degrees.

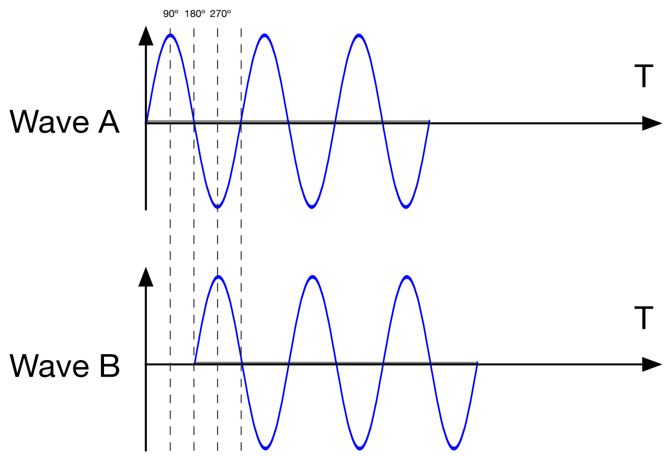


Figure 2 Two waves with same frequency and amplitude, but different phase.

Further Learning

• **Speed of sound:** Speed at which a sound wave propagates in a medium. It is dependent on the elasticity of the medium. In air, sound propagates at ~1125feet/sec. This =**~1foot per 1ms** (good approximation for time aligning speakers).

• **Wavelength**: the distance over which the wave's shape repeats.

• **Period (or cycle)**: Time interval required to do a complete cycle of vibration. Unit is seconds/cycle.

• **Simple harmonic motion**: Motion of an idealized mass on a spring (or of a simple pendulum). The mass on a spring (or simple pendulum) model is used to approximate how objects vibrate.

1.2 Sound Spectrum and Timbre

• Most objects that produce sound vibrate multiple ways at the same time; they have multiple **modes of vibrations**.

• Each mode of vibration produces a simple sound, known as a **sine tone (**because its waveform can be expressed with sine/cosine equations).

• The mode of vibration with the lowest frequency is generally known as the **fundamental**. Modes of vibration with higher frequencies are generally known as **partials**. The complete collection of partials of a sound is known as

• Partials in a sound can be viewed with a **spectrum analyzer**, which shows the **instantaneous spectrum** of a sound (or signal) as it varies in time (see **Error! Reference source not found.Error! Reference source not**

Figure 3 Spectrum analyzer view in Logic's Channel EQ showing instantaneous spectrum of a trumpet. Note that this is clearly a harmonic spectrum: the frequency of each partial (1000Hz, 1500Hz, 2000Hz, 3000Hz, etc.) is an integer multiple of the fundamental 500Hz.

found.).

• Certain objects (musical instruments, for example) vibrate in a way that the frequency of each successive mode of vibration (i.e., partial) is an **integer multiple of the fundamental**. This special order of partials is called a **harmonic series (or harmonic spectrum).** For example, if frequency of lowest mode of vibration is = f, then f2 = f*2; f3 = f*3; f4=f*4;...fn=f*n.

• The series of tones in a harmonic spectrum can be converted into musical notation (see **Error! Reference source not found.**):

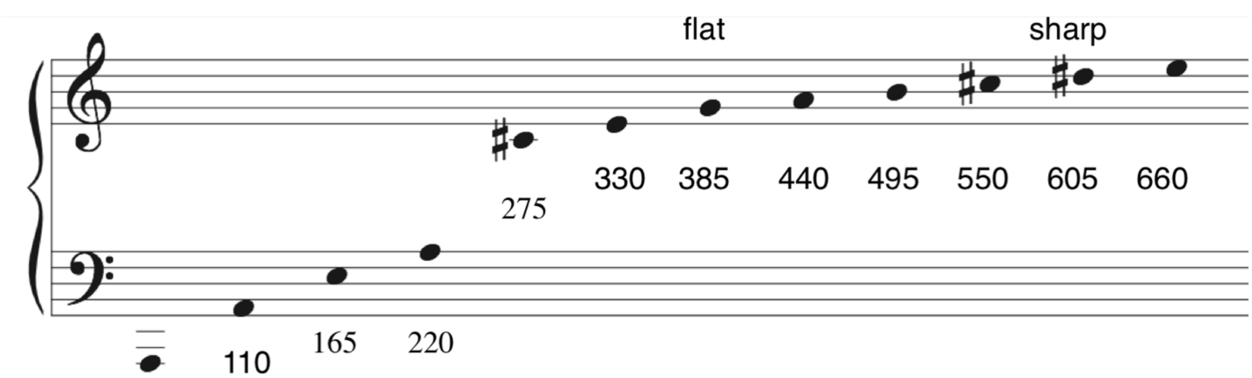


Figure 4 Example of a harmonic series starting on a low A (55Hz)

• The **perceived pitch** of a sound with a harmonic spectrum tends to be the frequency of the fundamental.

• When we hear a series of sine tones related by integer multiple of a fundamental, our brain automatically groups them together and perceives all the sine tones as one tone (i.e., a complex tone) with a specific **tone color, or timbre**. This process is called **spectral fusion**.

• For harmonic sounds, the spectrums are basically always the same, and it is the **variation in amplitude of each partial** that produce differences in timbre.

• Sounds whose partials are NOT integer multiples of a fundamental are said to be called **inharmonic**.

• Inharmonic sounds tend to sound less “fused” than harmonic sounds, meaning that we can hear individual partials more clearly.

• We tend to hear inharmonic sounds as “noisier”. As the order of the partials becomes increasingly random, we hear the result as increasingly noisy. The “noisiest” possible sound is **white noise**, which contains partials at all frequencies at equal amplitude.

• Partials in inharmonic sounds are NOT called harmonics, just partials.

• Examples of inharmonic spectrum: bells, gongs, tam tams, …most non-pitched percussion instruments.

••• Side Note

Sound is a dynamic phenomenon, and the spectrum of a sound is continually evolving as time passes. For example, the sound of a piano is both harmonic and inharmonic. The attack transient (the initial few milliseconds of sound) is very inharmonic and the spectrum varies substantially within a very short amount of time, while the sustain part of the sound (from which we determine the pitch) has a mostly harmonic spectrum which is relatively stable.

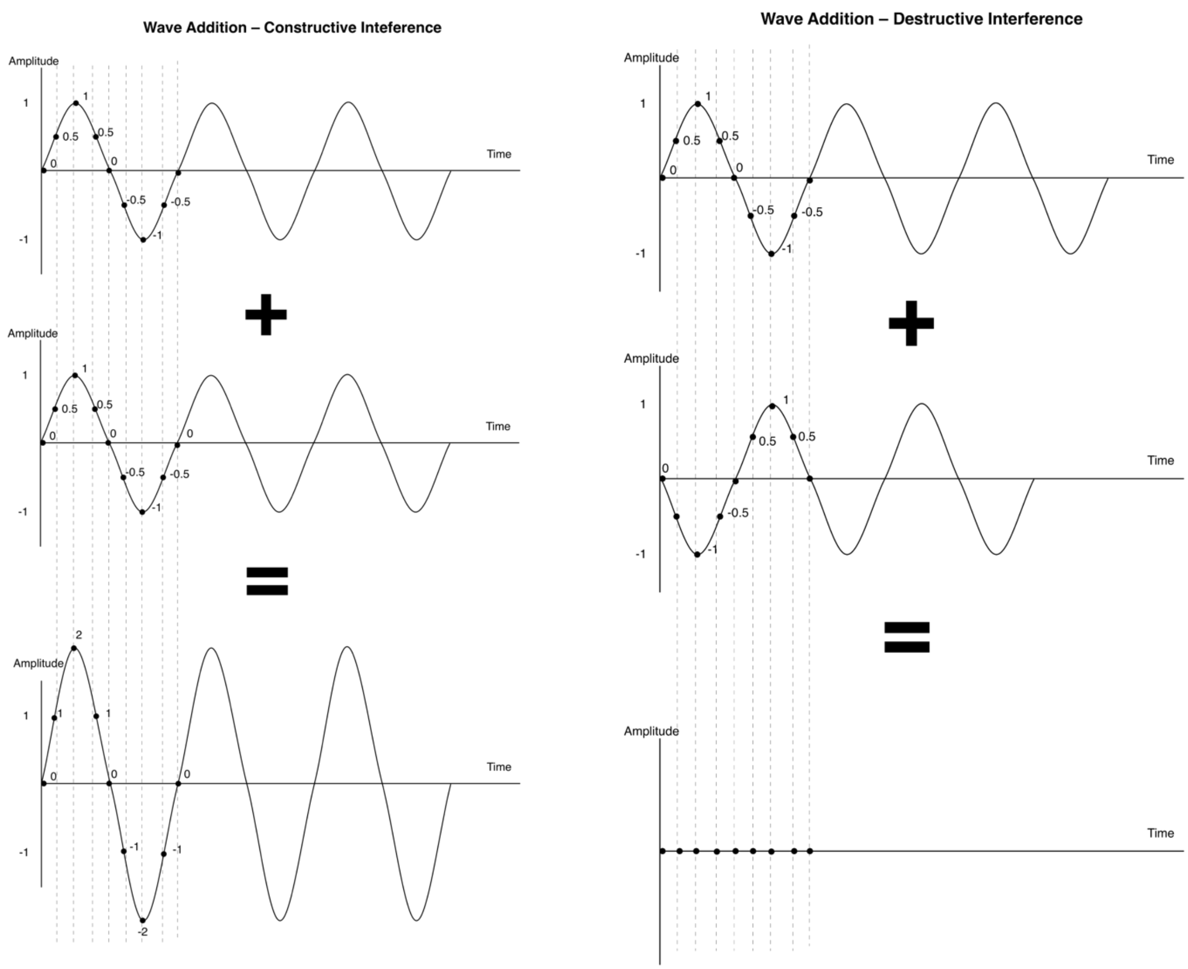
1.3 Combining Waves

1.3.1 Constructive and Destructive Interference

• When two waves meet in a medium (for example, when two sound sources play at the same time in a room), they will interact with one another: their amplitudes will add together. This results in **constructive** and **destructive interference**, where the amplitude of the resulting wave either increases or decreases.

• For two waves with exactly the same frequency, constructive interference occurs when two waves are perfectly in phase, and destructive interference occurs when two waves are perfectly out of phase (i.e., 180° out of phase).

• Example situations where waves are added together: when sound waves collide with one another in a room; audio mixing (for instance when two tracks play at the same time in Logic, they are summed at the output); additive synthesis, etc.



Resulting wave is amplified

Waves cancel each other out (-> phase cancellation or destructive interference)

Figure 5 Examples of constructive and destructive inteference

1.3.2 Beating

• Beating is a particular case of interference where adding together two waves with slightly different frequencies produces audible beats.

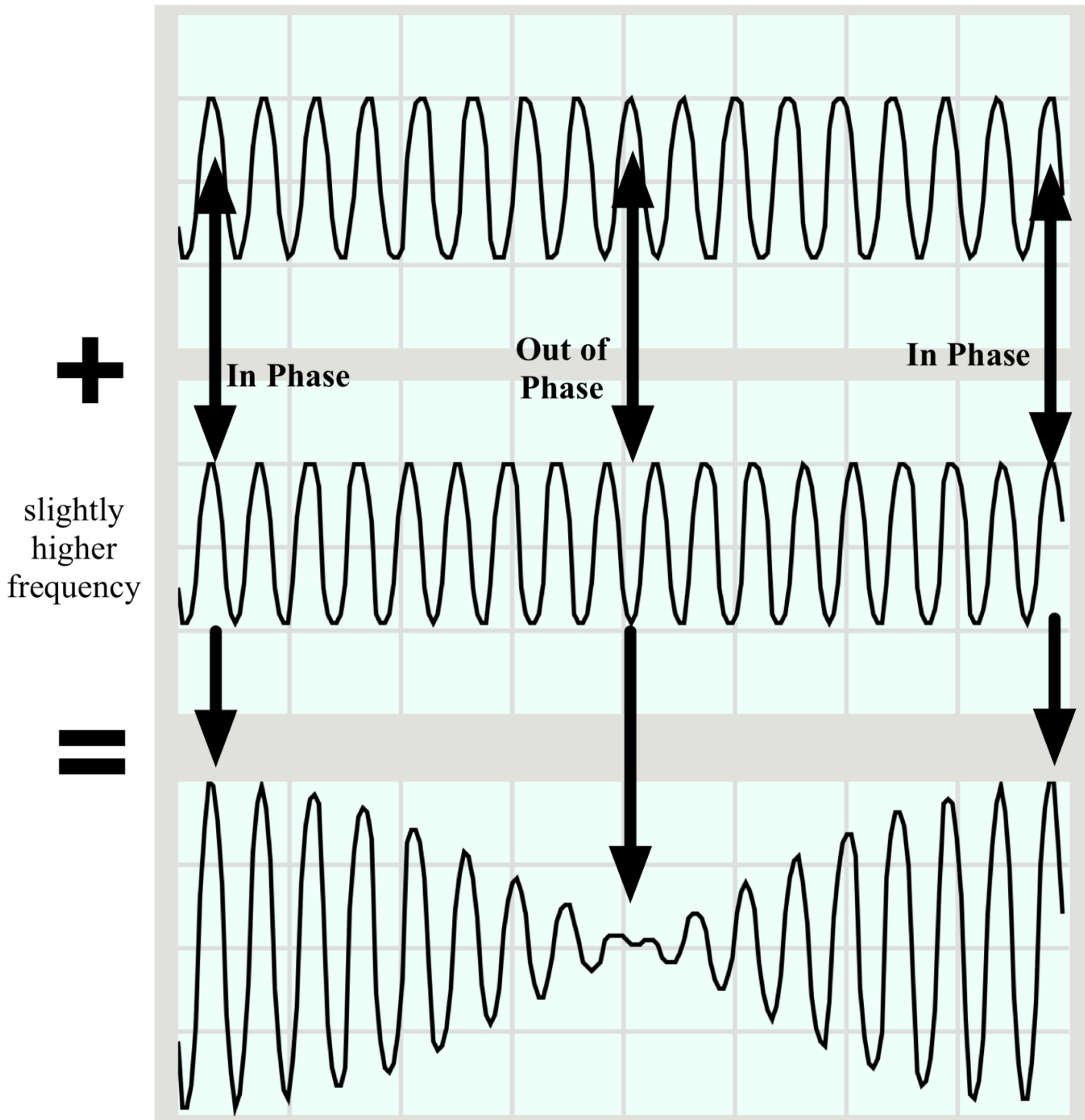


Figure 6 Example of beating

1.3.3 Multiplication:

• Action of multiplying the amplitude values of two waves at a given time.
• Example situations where waves are multiplied together: modulation techniques (ex: ring modulation, amplitude modulation, frequency modulation); amplitude envelope (ex: automation in logic); etc.

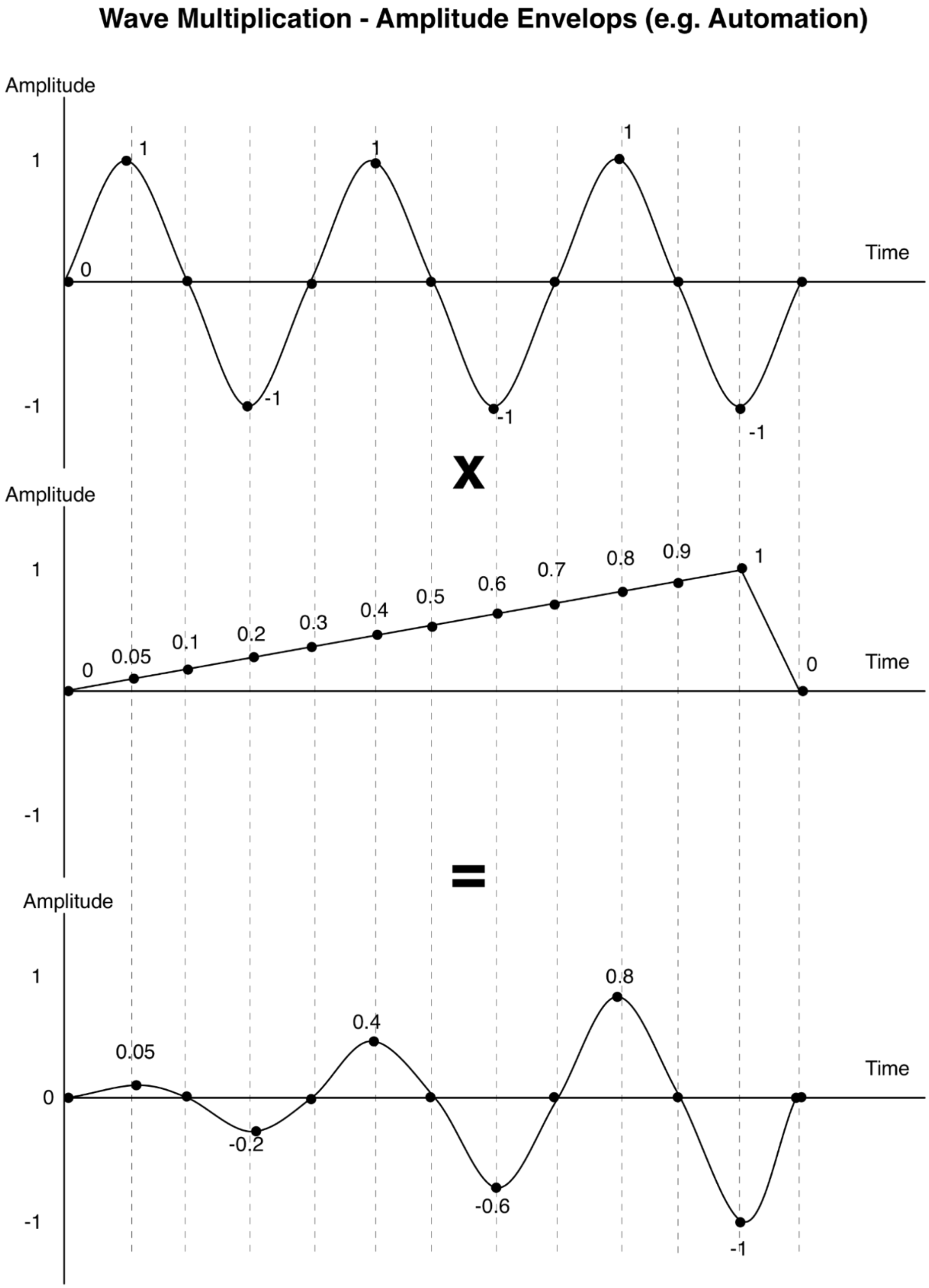
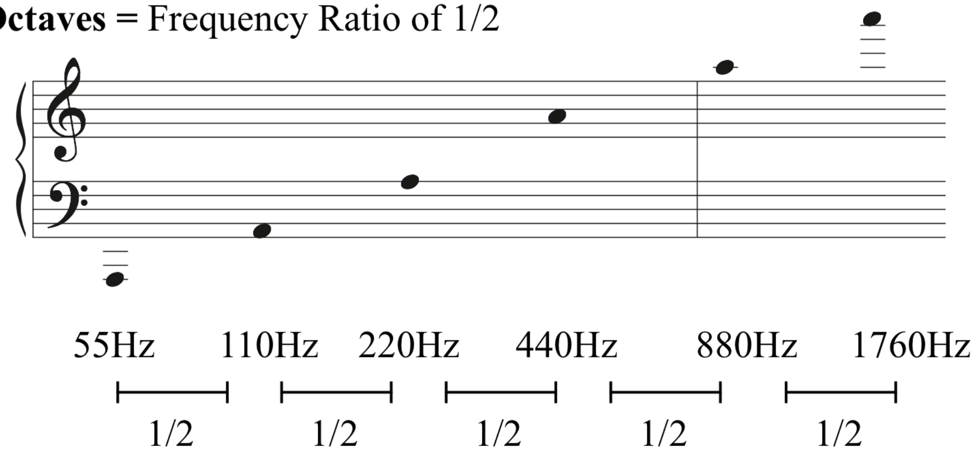


Figure 7 Creating a fade

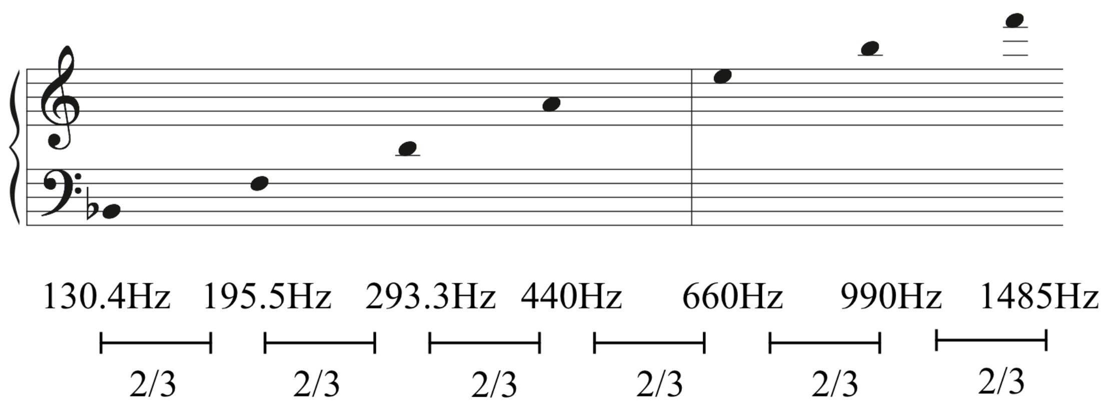
2 Linear vs Logarithmic

• In a linear scale, change between two values is based on the **difference** between the values:
• Ex: in the sequence 100, 200, 300, 400…. the difference between values is 100 (200-100=100; 300-200=100; 400-300=100; etc.)
• In a logarithmic scale, change between two values is based on the **ratios** between the values
• Ex: in the sequence 100, 200, 400, 800, 1600, the ratio between two consecutive values is 1/2 (100/200=1/2; 200/400=1/2; 400/800=1/2; 800/1600=1/2; etc.)
• Two aspects of sound use logarithmic scales: frequency (pitch) and amplitude (volume).

Octaves = Frequency Ratio of 1/2



Fifths = Frequency Ratio of 2/3 (for pure fifths, not equally tempered)



2.1 Frequency, Logarithmic Scales, and Musical Intervals

• Musical **intervals** are frequency ratios. Therefore, a sequence of frequencies related by octaves will be logarithmic.
• Example: octaves, for example, have a frequency ratio of ½, and fifths have a ratio of 2/3.

3 Decibels

• As mentioned in section 1, amplitude is directly linked to the perception of volume (the greater the amplitude of a wave, the greater its volume, and vice versa).
• In the world of recorded sound, the unit used to describe amplitude is generally the **decibel (dB)**
• A decibel is a **comparison** between a measurement and a reference measurement.
• Because the reference can change, there are **different types of decibels**: dBFS, dBSPL, dBVU, dBu, dBv, etc.
• The two types of decibels most encountered in modern sound recording are the dB FS (i.e., FULL SCALE), and dB SPL (i.e., SOUND PRESSURE LEVEL).

3.1 dB SPL

• In **dB SPL** (sound pressure level), the reference measurement is the sound pressure of the threshold of hearing (=0.00002 Pa).
• Other sound pressure measurements will always be **greater** than the reference, so dB SPL are always **positive**.
• dB SPL are generally used when describing everyday sounds, and are measured with a dB SPL meter (or phone app).

Table of sound levels L and corresponding sound pressure and sound intensity

Examples	Sound Pressure Level L_p dBSPL	Sound Pressure p N/m^2 = Pa	Sound Intensity I W/m^2
Jet aircraft, 50 m away	140	200	100
Threshold of pain	130	63.2	10
Threshold of discomfort	120	20	1
Chainsaw, 1m distance	110	6.3	0.1
Disco, 1 m from speaker	100	2	0.01
Diesel truck, 10 m away	90	0.63	0.001
Kerbside of busy road, 5 m	80	0.2	0.0001
Vacuum cleaner, distance 1 m	70	0.063	0.00001
Conversational speech, 1m	60	0.02	0.000001
Average home	50	0.0063	0.0000001
Quiet library	40	0.002	0.00000001
Quiet bedroom at night	30	0.00063	0.000000001
Background in TV studio	20	0.0002	0.0000000001
Rustling leaf	10	0.000063	0.00000000001
Threshold of hearing	0	0.00002	0.000000000001

Figure 8 - Examples of dB SPL (source: [Evergreen State College]). Note that the dB values are always **positive**.

Figure 9 - dB SPL Meter

3.2 dB FS

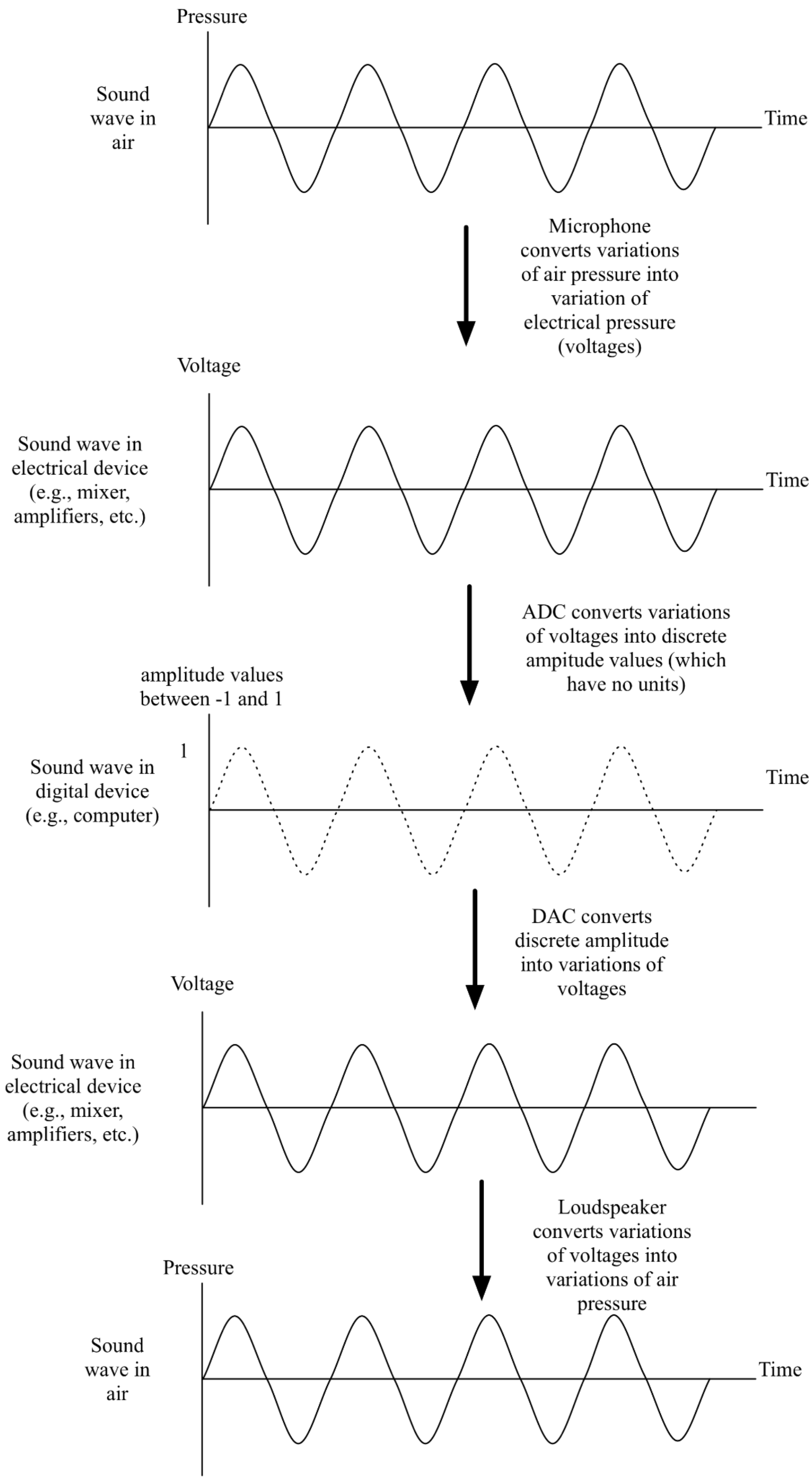
• In dB FS (i.e., full scale), the reference measurement is the maximum amplitude value in digital audio, which is 1.
• Other digital amplitude measurements will always be **smaller** than the reference, so dB FS are always **negative**.
• dB FS are exclusively used with digital audio devices (ADC, DAC, software, digital mixers, etc.).

Figure 10 - Use of dB FS in Logic Pro. Note that the dB values are always **negative**.

Further Learning
• Just like frequency, amplitude (sound pressure) is also experienced logarithmically.
• For example, a sound wave whose pressure increases from 1 Pa to 2 Pa (Pascales, unit of pressure) and then increase from 2 Pa to 4 Pa would be perceived as a gradual crescendo, not a sudden boost (in both cases the pressure ratios are 1/2); both increases are perceived as the same relative amount of change (just like we perceive two different octaves as the same relative amount of change of frequency)
• To express amplitude values, we use **decibels**. Decibels are a ratio; you compare two amounts.
• The mathematical definition of a decibel is dB = 20log10(A1/A2)
• Different dB scales exist, and each scale has its own reference (i.e., a fixed value to which all other values are compared). In the equation above, A2 is the reference.
• In the dB SPL scale, the reference is always the threshold of hearing (0.00002 Pa). In this scale, dBs are always positive, because the amplitudes being measured are always larger than the reference.
• Example: what is the level in dB SPL of a typical conversational speech? According to the chart below, a vacuum at 1m has a pressure of 0.063 Pa, so…
dB = 20log10(0.063[vaccum at 1m]/0.00002 [threshold of hearing] = 70dB
• Example #2: 20log10(0.5/1) = -6dB
• In the dB FS scale (FS means Full Scale), which is used exclusively in digital audio, the reference is 1 [maximum possible amplitude in digital audio]. With dB FS, the dBs are always negative, because the amplitude being measured is always smaller than the reference.
• Good to remember: doubling amplitude= 6 dB increase; halving of amplitude=6dB decrease.

4 Digitizing Sound

4.1 General Overview of the Complete Process



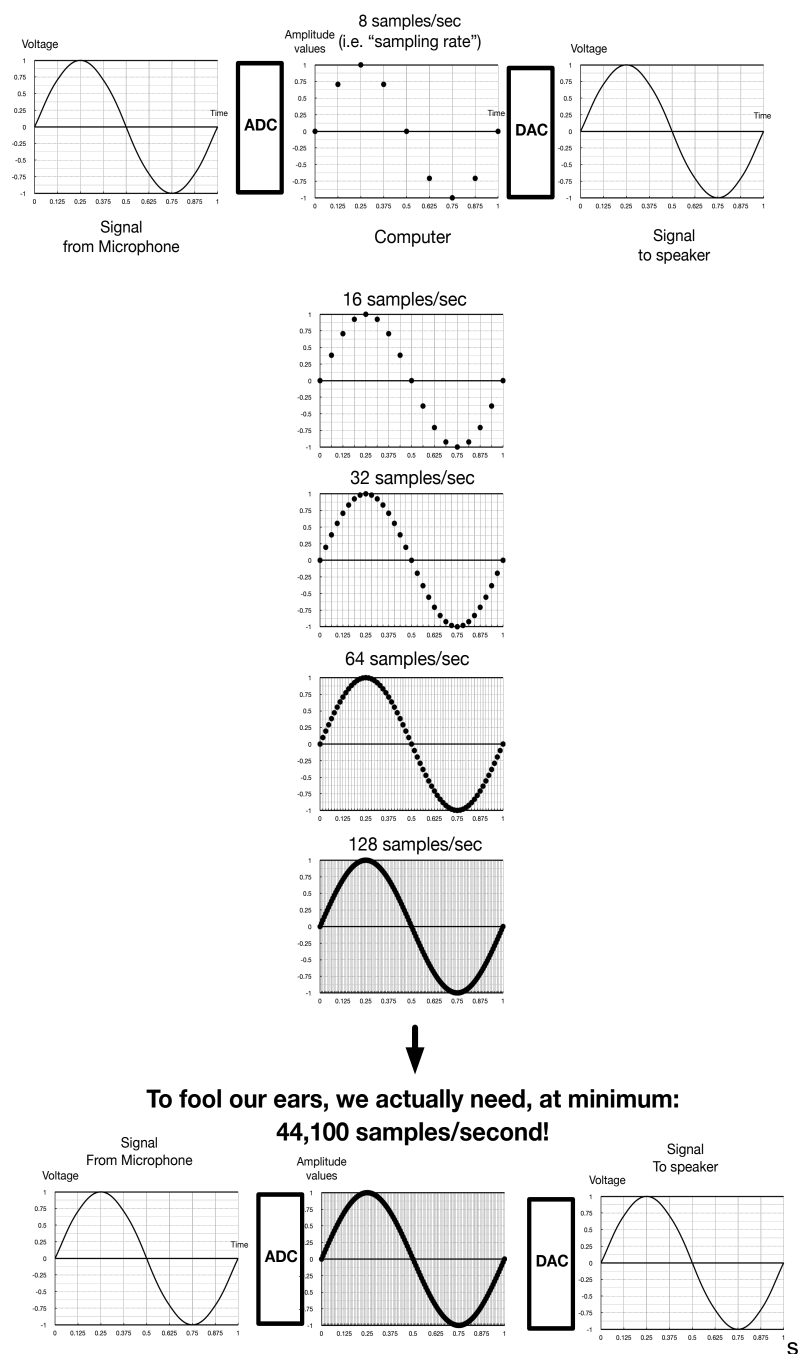
4.2 Definitions

• **ADC (Analog to Digital Converter)**: Device that converts a waveform expressed by an analog (i.e., continuous) electrical signal (usually expressed in volts) into a stream of 1s and 0s.

• **DAC (Digital to Analog Converter)**: Device that converts a stream of 1s and 0s representing an audio waveform into an analog electrical signal (usually expressed in volts).

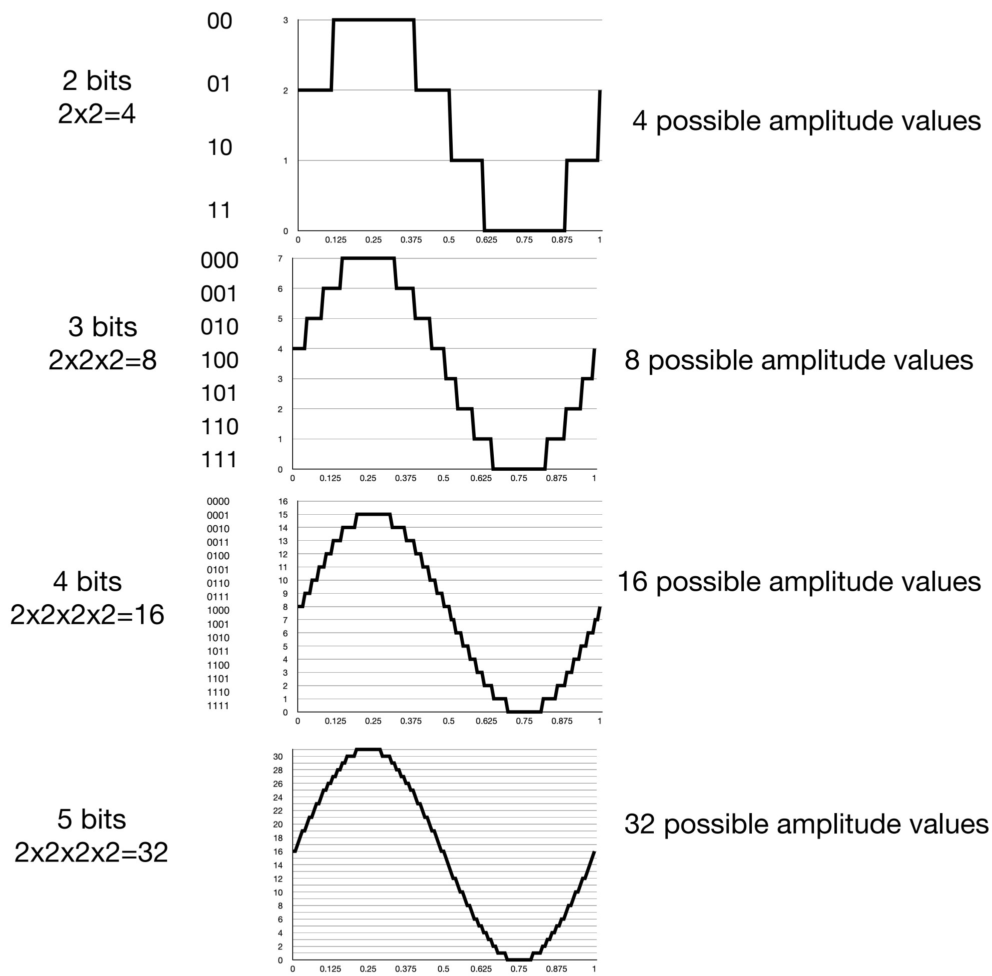
• **Sampling Rate:** Rate that specifies how many amplitude measurements per second are made by an ADC. In general, a higher sampling rate implies higher sound quality since the digital representation of the sound is more accurate.

Sample Rate



• **Bit Depth**: Expression of the precision of each amplitude measurement made by the ADC hardware. The more bits of resolution, the greater the accuracy. A bit is defined as one binary digit, that is, either the value 0 or the value 1. Thus, when only one bit is used to represent the value of an amplitude measurement, the amplitude must be either 0 or 1. If two bits are used to represent amplitude values, then four possible amplitude values can be represented (0, 1, 2, and 3). More generally, the number of amplitude values that can be represented by a given number of bits is 2**N where N is the number of bits. Using a large number of bits per sample allows the full amplitude range of an ADC to be partitioned into a large number of very small steps. In general, a higher bit depth implies higher sound quality.

Bit Depth



To fool our ears and minimize noise, we need, at minimum:
16 bits (= 65,536 possible amplitude values)

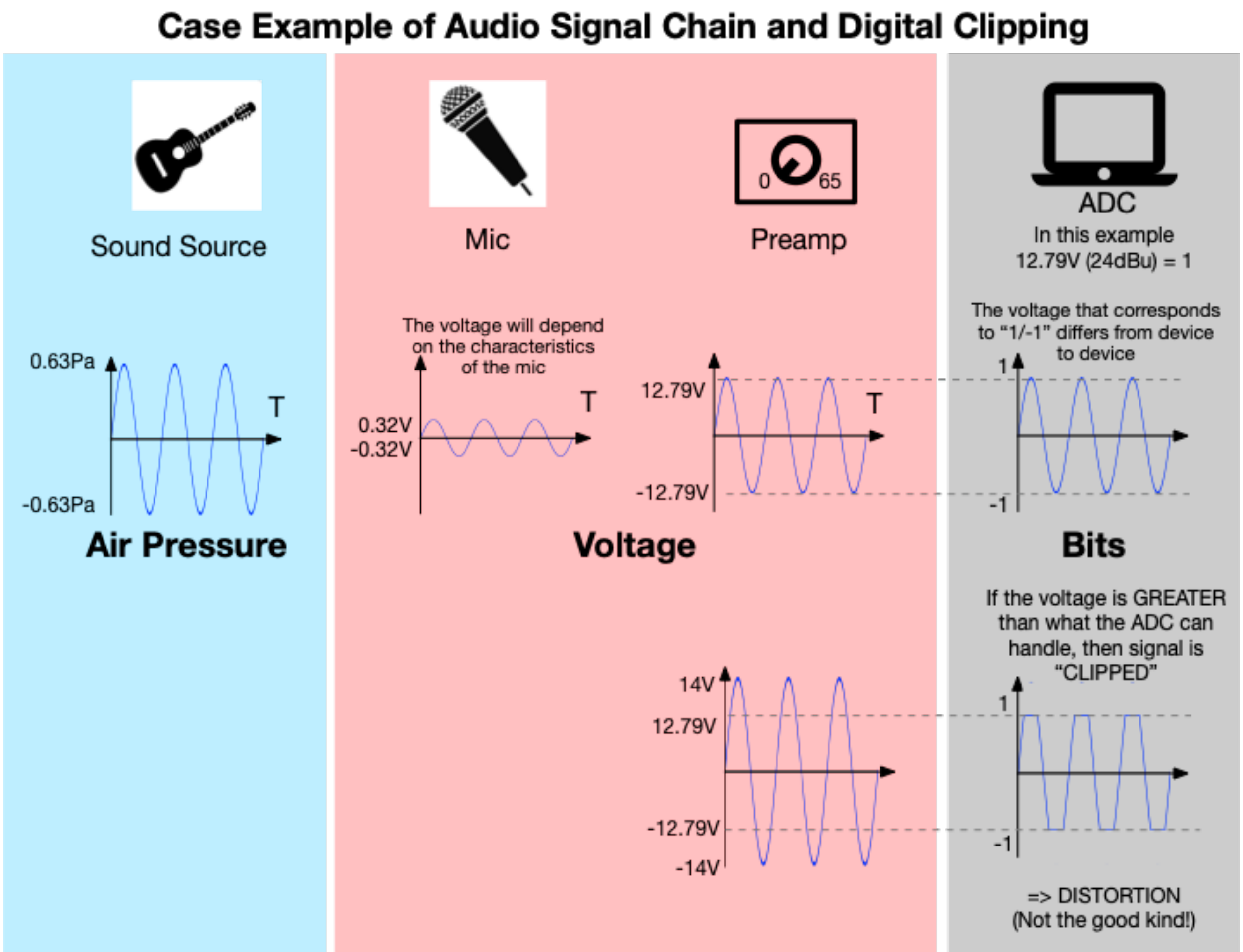
However, to maximize sound quality,
ALL YOUR PROJECTS SHOULD BE RECORDED AT:
24 bits (= 16,777,216 possible amplitude values)

• **Common Audio Resolutions:**
CD Audio: 44.1KHz SR, 16bits
Sound for video: 48KHz, 24 bits
High Resolution audio: 96KHz, 24 bits

• **Common Audio File Formats:**
AIFF, WAV, CAF

4.3 Clipping

In digital audio, all possible values of a given bit depth are scaled down to the range of 1 to -1. Thus, digital waveforms can never go above 1 or below -1. If a waveform goes beyond 1 to -1, then the values that are out of range are "clipped" to 1 or -1, resulting is a distortion of the waveform and a very unpleasant sound:



5 <u>Metering</u>

Meters on an audio devices or software show the amplitude of a signal. They are useful to know how strong or weak the signal is, and thus help make appropriate gain level decisions. There are generally two types of meters: peak and RMS.

• **Peak** meter: meter that shows a signal's exact level, from peak to peak. Generally, digital devices use peak meters (the meters in Logic's mixer are peak meters). As an overlying rule, **always use Peak meters when setting preamp levels when recording digitally**.

• **RMS** meter (Root Mean Square): meter that show an average of the signal over a given time window (also called the RMS-size), which varies between 50-500ms depending on how the meter was designed.

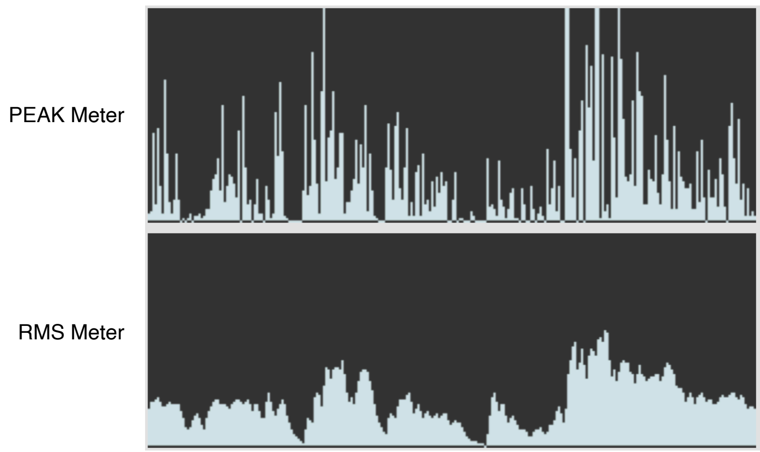
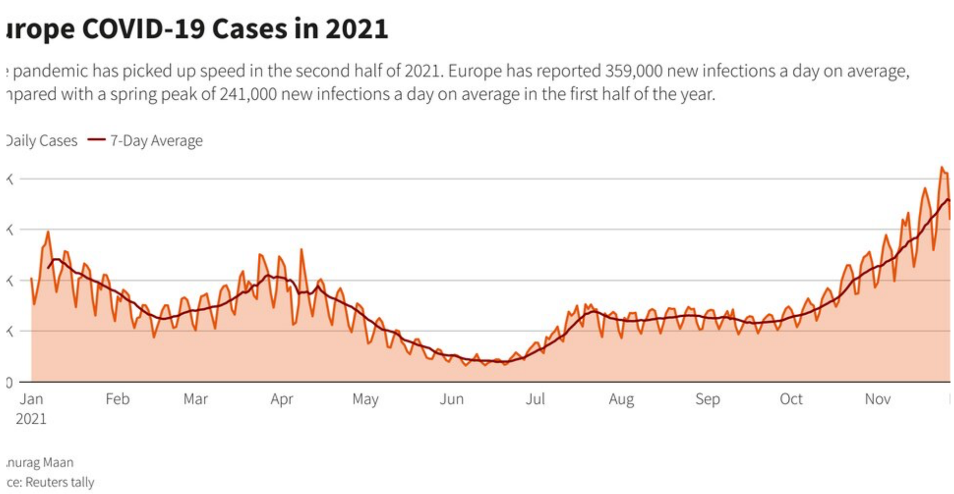


Figure 11 - Visualization of a PEAK meter vs an RMS meter

Figure 12 - Logic's Level Meter plugin, which can show both PEAK and RMS levels

The difference between Peak and RMS meters can be compared a familiar recent event: the day-to-day COVID cases (i.e., PEAK), versus the seven-day average COVID cases (i.e., RMS).



6 Basic Digital Signal Processing (DSP)

6.1 Filters

A filter is a hardware device or a software program that cuts, attenuates or boosts certain frequencies.

6.1.1 Low Cut Filter (a.k.a., High Pass Filter) & High Cut Filter (a.k.a., Low Pass Filter)

Filter which cuts out partials that are above or below a certain frequency.

Figure 6-1 – Example of low-cut filter (a.k.a., high-pass filter).

Figure 6-2 – Example of high cut filter (a.k.a., low-pass filter). Same as low-cut, except reversed.

<u>**Common Parameters**</u>

Cutoff frequency (fc): The frequency after which other frequencies are cut (technically, is the point where the level transmitted by the filter drops by 3dB).

Slope (a.k.a., roll-off): The rate at which the attenuation increases beyond the cut-off frequency, expressed in dB/octaves. With a steep slope, frequencies are filtered out quickly beyond the cutoff frequency (and vice versa for a gentle slope).

<u>**Typical Usage Examples**</u>

• <u>Removal of low rumbles (from ventilation, mic handling, etc.) and bleed from nearby low frequency sound sources</u>: low cut filter with cutoff frequency set just below the lowest frequency of a sound source. Note that here, the filter does not typically alter the sound source, it simply removes unwanted low frequency sound.

• <u>Filter sweeps</u>: Automated typically high cut filter that sweeps upwards to create an energy build-up.

6.1.2 • Band filter (a.k.a, peak or notch filter).

Filter that boosts or attenuates a band of frequencies. This band is defined by three parameters:

Figure 6-3 – Two of examples of band filters. On the left, a band filter with negative gain attenuates a band of frequencies; on the right, a band filter with positive gain boosts a band of frequencies.

Common Parameters

Center frequency (Fc) of the filter, which controls the location of the filter band on the frequency spectrum;

Q (i.e., "quality"): The control of how broad or narrow the band is.

Gain: The amount of boost or attenuation of the band.

Typical Usage Examples

• Boosting the 2-5kHz region with a wide band to make speech more intelligible.

• Removing a ground loop tone with a high Q and high attenuation band asset at 60Hz.

6.1.3 • Shelf filter.

Similar to a low/high-pass filter, but the filtering levels off at a specific frequency.

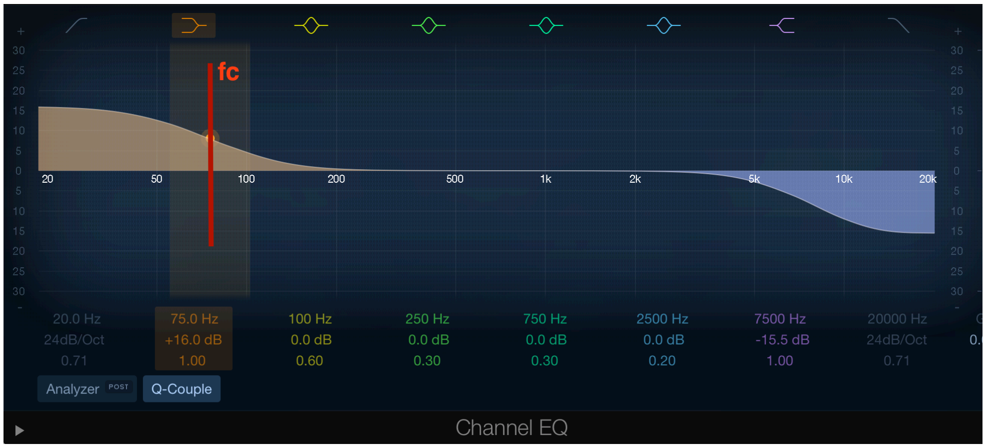


Figure 6-4 – Example of a low-shelf and high-self filter

Common Parameters

Cutoff frequency (fc): The frequency after which other frequencies are attenuated (technically, is the point where the level transmitted by the filter drops by 3dB).

Q (i.e., "quality"): The control of how broad or narrow the band is.

Gain: The amount of boost or attenuation of the band.

Typical Usage Examples

• Attenuating the low frequencies created from **proximity effect** (i.e., the natural tendency for low frequencies to be boosted when a directional microphone is placed very close to a sound source).

6.2 Delay based effects

6.2.1 Delay (a.k.a echo):

Delay is the process of "delaying" a signal. In other words, it's a device that takes an audio signal, stores it, and plays it back after a specified **delay time**. *In digital audio, this is accomplished through the use of buffers (holding places for digital data) which are continuously updated so that they always store the most recently received signal*. Delay lines are often called "**taps**", as a reference to plumbing: an audio tap diverts an audio signal the same way a faucet diverts water. Most delay-based effects allow for **feedback**: the delayed signal is sent back into the delay. If the delayed signal is not attenuated before being sent back into the buffer, an infinite feedback loop will be created. If the delayed signal is attenuated before being sent back into the buffer, then the delayed signal will progressively fade-out with each iteration. Certain delay units of plug-ins have multiple taps (called **multi-tap delays,** like Delay Designer in Logic), each with its own delay time, feedback level, and filter.

Many common effects are simply delay lines with specific delay times:

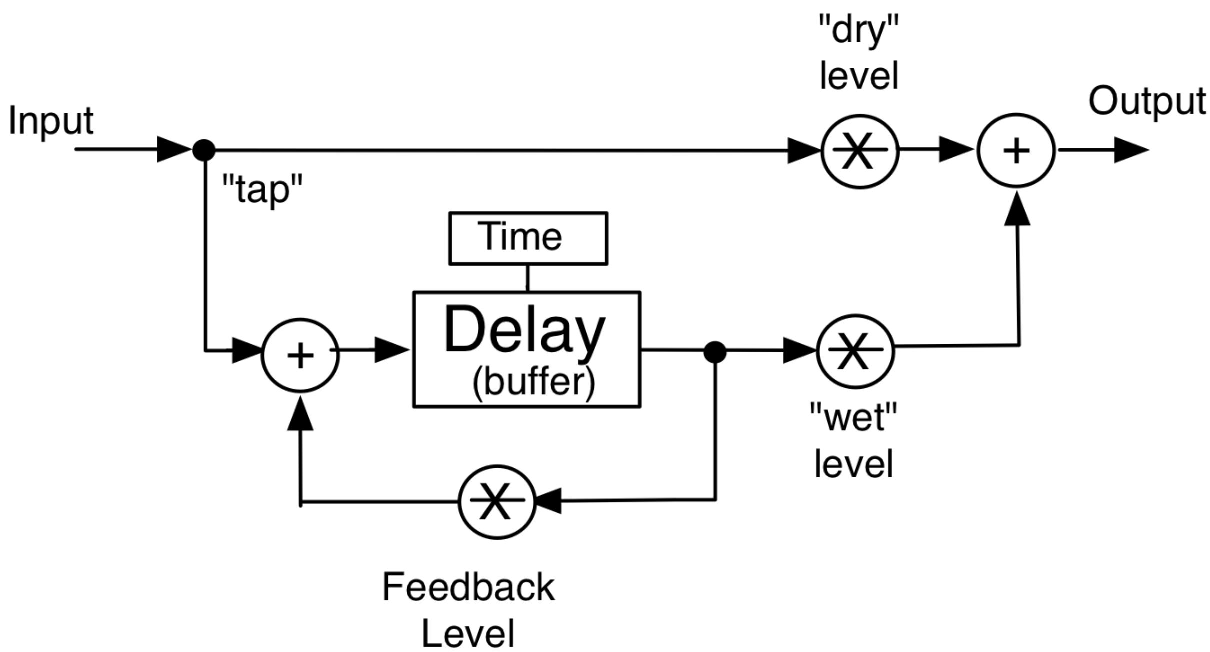


Figure 6-5

6.2.2 Flanger:

A delay-based signal processing technique, which consists of combining a signal with a delayed version of itself while dynamically varying the delay time. The variation of the delay time is typically done with an LFO (low frequency oscillator). Usually, flangers use a range of short delay times, between 0-20ms. Because the delay times are so short, the combination of the signal with its delayed version will cause certain partials to cancel out, resulting **in comb filtering** (a filtering effect which exhibits peaks and valleys equally spaced in frequency). This results in the flanger's characteristic 'swooshing' effect.

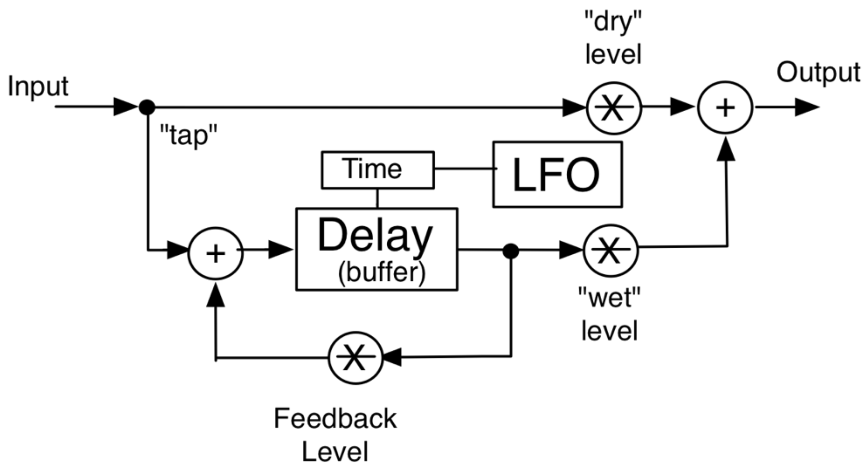


Figure 6-6

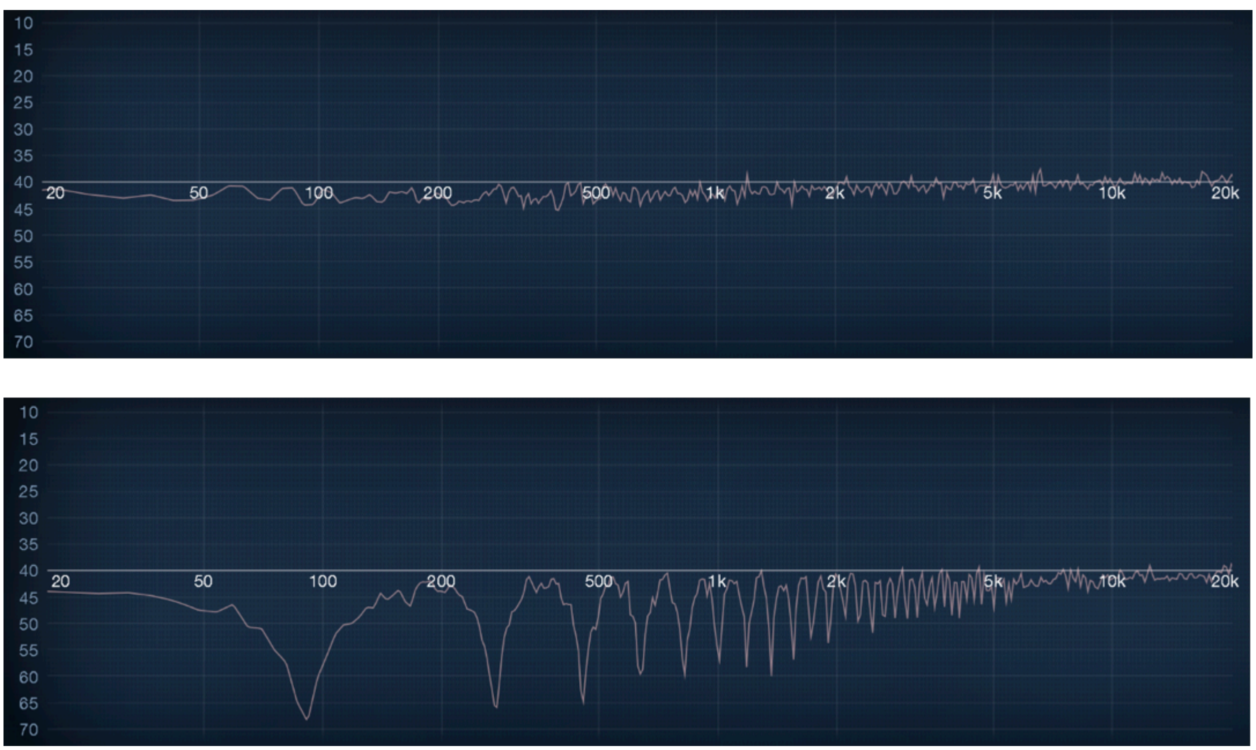


Figure 6-7 – Example of Comb Filtering. Here, a noise signal is delayed by 5ms and combined with itself. Note the resulting comb filtering.

6.2.3 Chorus:

"The creation from a single voice of the percept of multiple voices in unison" (Dodge p.74). In practice, this is basically the same effect as the flanger, except that the delay times are longer, i.e. ~10-50ms. Because the range of delay times is larger, comb filtering will be less audible, and the delayed signal starts to pitch shift. As the delay times increase, the pitch lowers; and as the delay times decrease, the pitch raises (just like in a Doppler effect).

6.3 Modulation

Modulation is the process of modifying a signal (or a parameter of a signal, like the amplitude) with another signal (the "**modulator**"). In audio, the modulator can be any signal but in most cases is a low frequency sine wave (a.k.a. an **LFO**, for Low Frequency Oscillator)

• **Ring modulation**: An audio process where the input signal and the modulator are simply multiplied together. When the frequency of the modulator is between 0Hz and ~20Hz, the effect is known as **tremolo** (i.e. we hear clear pulsations). When the frequency of the modulator is above ~20Hz, then new frequencies are generated around each frequency present in the input signal. For a modulator with a frequency of f_m and a carrier with a frequency of f_c, a ring-modulated signal will contain frequencies of ($f_m - f_c$) and ($f_m + f_c$). Note that the original signal is <u>not</u> present in the output.

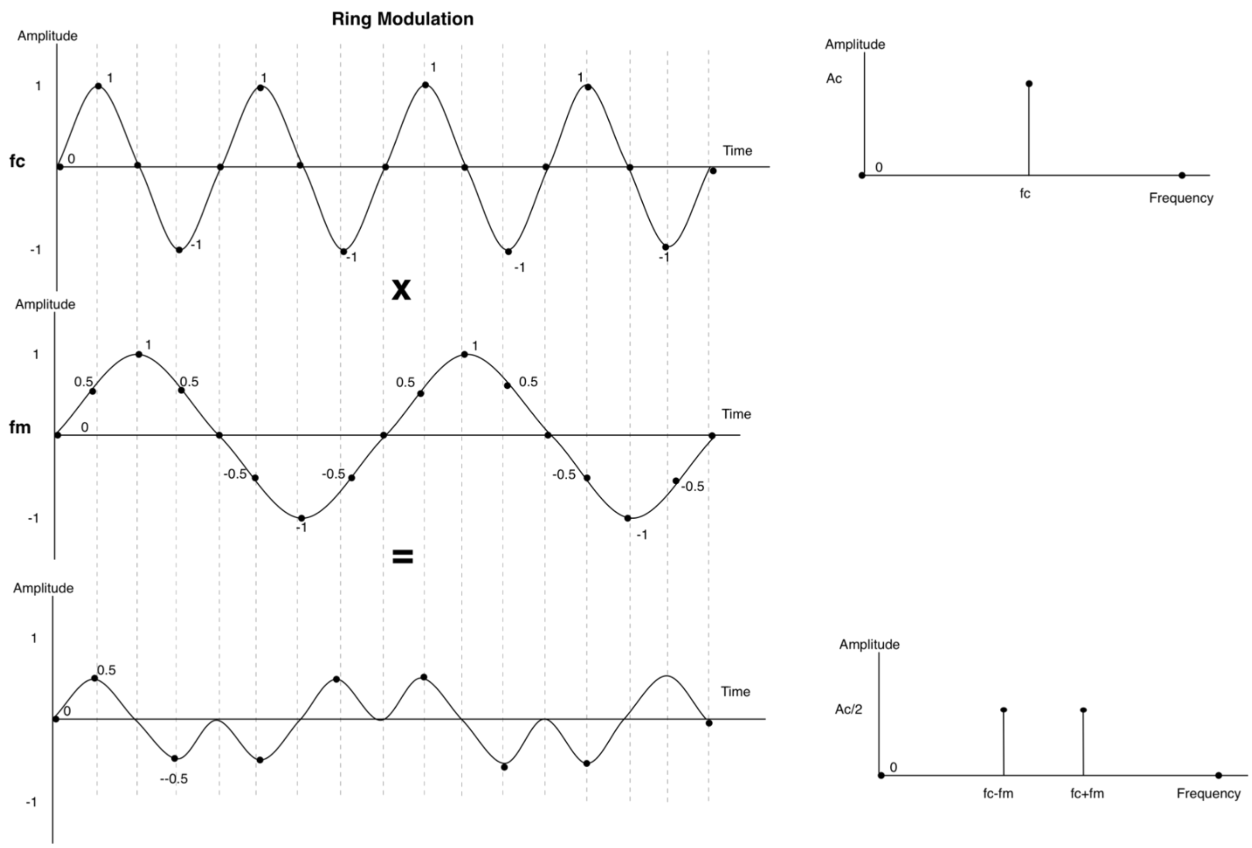


Figure 6-8

6.4 Distortion:

Distortion is the modification of an audio signal during its transmission, amplification, or other processing. In the context of musical instrument amplification, it refers to various forms of clipping (i.e. the truncation of the part of a signal that exceeds certain limits). Distortion in this context has different names: fuzz, overdrive, clipping, etc. In digital audio, distortion can refer to analog distortion emulation software (for example distortion plug-ins), but can also refer to a class of sound synthesis techniques in which a simple waveform is modified by a controlled amount. Examples of this are frequency modulation (i.e. modulating the frequency of a carrier by the frequency of a modulator) and non-linear waveshaping (i.e. substituting incoming amplitude values with others according to a wave table).

6.5 Dynamic Control

6.5.1 Compression.

Compression is a form of automatic volume control. A compressor basically “squeezes” an audio signal to reduce the parts with the greatest amplitude. More precisely, it lowers the amplitude of a signal by a given amount whenever it rises above a set **threshold**.

The **ratio** indicates how much the signal will be reduced, and the **attack** and **release** times dictate how fast the compressor reacts to an incoming signal once that signal crosses the threshold. The attack time is the time it takes to reach maximum compression after the signal crosses the threshold, and the release time is the time it takes for compression to stop after the signal returns below the threshold. Both the attack and release times are typically indicated in milliseconds.

The main use of a compressor is to reduce the amplitude of only the parts of a signal that are too high, while leaving the other parts intact. Once the high parts have been reduced, then the overall level of the signal can be raised, thereby increasing the overall perceived volume. In other words, compressors are useful tools to control the **dynamic range** of a signal (i.e. the difference in level between the lowest and the highest amplitude of a signal). More compression will create less dynamic range.

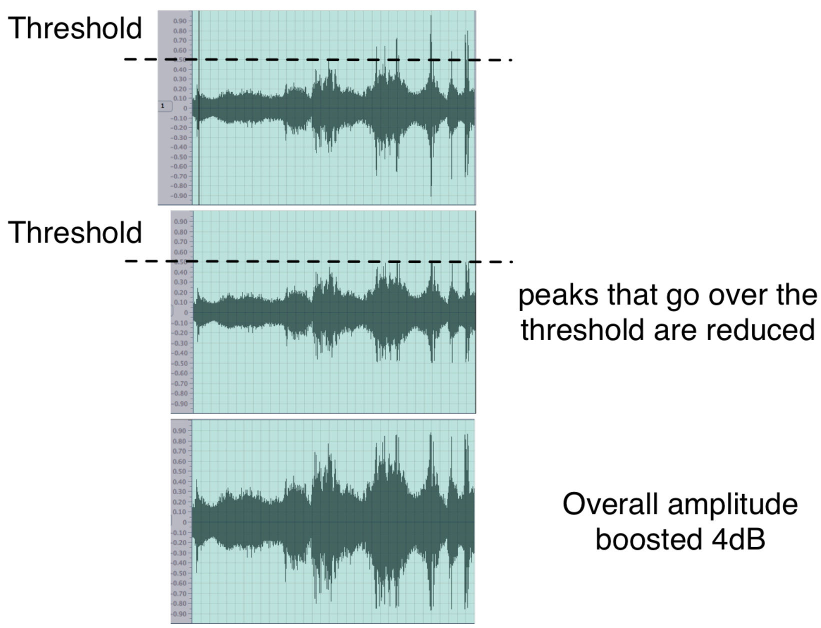


Figure 6-9 Example of effect of compression on overall dynamic range of a signal.

__Common Parameters__

Threshold: Level at which the compressor engages (i.e., starts to lower the incoming signal)

Ratio: The amount of level reduction, expressed as a ration of input level over threshold/output level over threshold.

Attack: Time required for the compressor to fully reduce the level at the given ratio. Lower time indicates a faster reacting compressor, and vice versa.

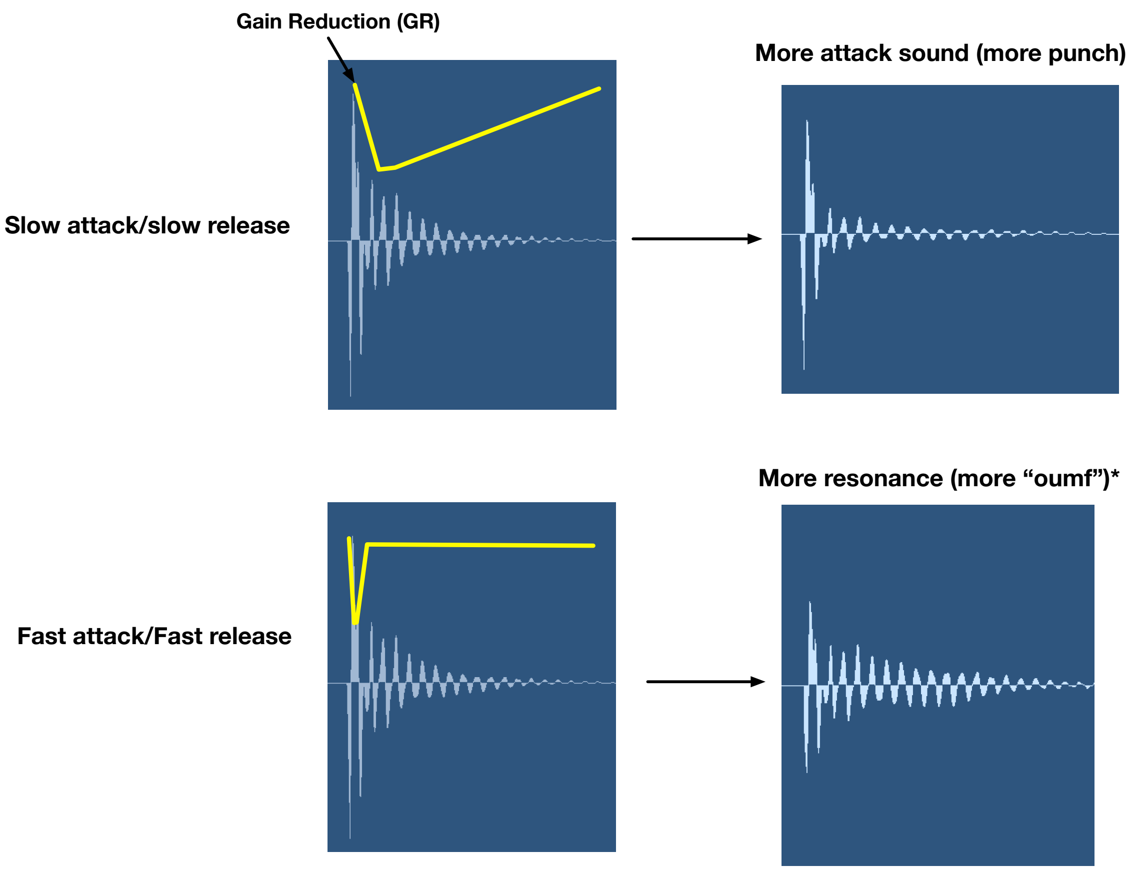
Release: Time required for the compressor to return to the uncompressed level at the given ratio. Lower time indicates a faster reacting compressor, and vice versa.

Makeup Gain: Amount of gain added to the signal after it has been compressed.

<u>Typical Usage Examples of Compressors</u>

• Leveling a signal: evening out the dynamic range of a particular instrument, voice, sub-mix, or entire mix, so that it doesn't become too quiet or too loud.

• Shaping the transients of signals by carefully setting the attack and release times as below.



***Note that 4dB of make up gain was added**

Further Learning

The **ratio** is usually expressed as *x*:1, which means that when the input signal is *x*dBs above the threshold, the output will only be 1dB above the threshold. Let's look at a few examples to understand:

<u>Example 1</u>: A compressor's threshold is set to -20dB and the ratio is 2:1. The incoming signal is -18dB. What would the compressor output?

-18dB is 2dB above the -20dB threshold, so only 1dB is let through (->2:1 ratio)
which means -20dB + 1 dB = **-19dB**

<u>Example 2</u>: The compressor has the same threshold settings, but the incoming signal is -14dB. What would the compressor output?

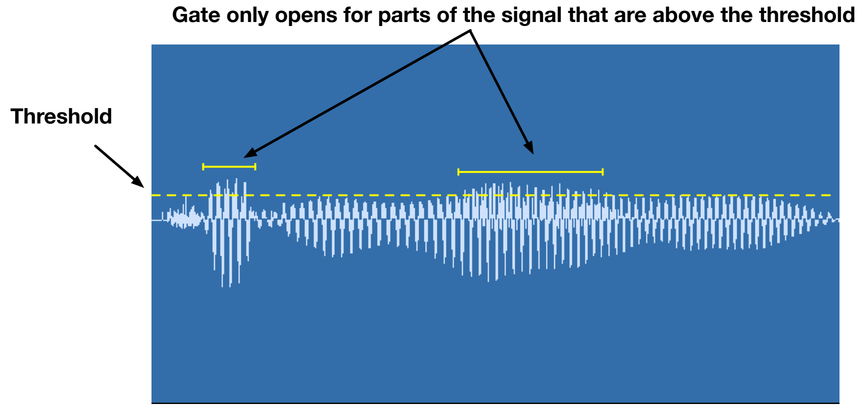
-14dB is 6dB above the -20dB threshold, so 3dB are let through (6:3 = 2:1)
which means -20dB + 3dB = **-17dB**

6.5.2 Limiting

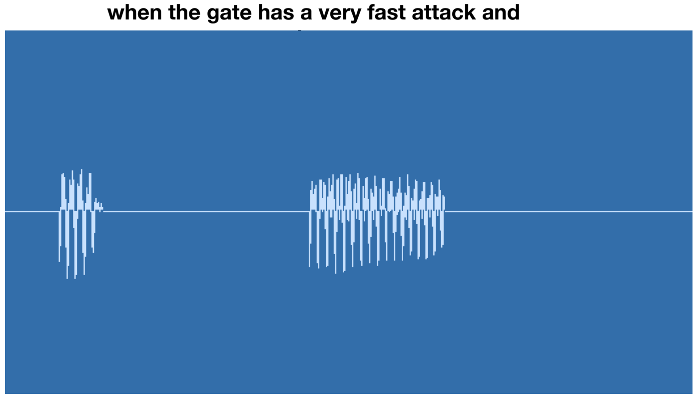
While a compressor proportionally reduces the signal when it exceeds the threshold, a **limiter** reduces any peak above the threshold to the threshold level, effectively limiting the signal to this level. Typically, it is said that a limiter has an infinitely high ratio (∞:1).

6.5.3 Gating

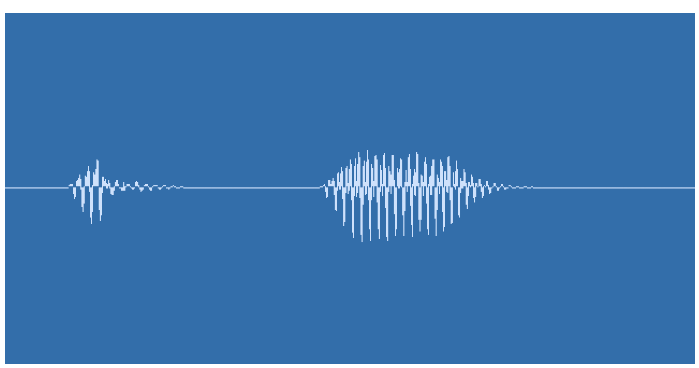
Whereas a compressor lowers the level when the signal is louder than the threshold, a **gate** lowers the signal level whenever it falls below the threshold. Louder sounds pass through unchanged, but softer sounds, such as ambient noise or the decay of a sustained instrument, are cut off. Noise gates are often used to eliminate low-level noise or hum from an audio signal (source: Logic manual).



Which results in sharp (almost clicking) sounds when the gate has a very fast attack and



Or more gentle sounds when the gate has a slower attack and release



7 Fast Fourier Transform (FFT) based Analysis/Re-synthesis

An FFT is a fast (computationally efficient) way to decompose an audio signal into its sinusoidal components. It is based on Fourier's theory from 1822 which stated that complicated vibrations could be analyzed as a sum of many simultaneous simple signals. The FFT as we know it is accredited to Cooley and Tukey (1965) who implemented a way for computers to quickly perform a Fourier analysis instead of the computationally heavy Discrete Fourier transforms (DFT) used until then.

Explanation: FFTs are performed on small overlapping **windows** (i.e. segments) of an audio signal, the size of which is calculated in samples (**window size**). One way to understand an FFT is to compare it to a large bank of narrow band-pass filters (or peak filters) distributed equally to cover a large frequency spectrum. As a signal goes through the bank of filters, only the filters which correspond to a specific frequency present in the signal will be excited. Hence, once we know which filters were excited, then we also know which frequencies were present in the signal. The precision of the analysis depends on how many filter are used: more filters=better precision. In an actual FFT, there are no band-pass filters but rather there are **bins** (i.e. 'frequency containers'). The number of bins used by the FFT is called the **FFT size** (which is directly related to the window size). To obtain a total picture of the audio signal, each FFT is chained to the next, and the complete chain is converted to a graphical representation known as a **sonogram**.

A fade-in/fade-out is operated on each window before the FFT is calculated. This is to avoid clicks which would result from each window not starting and ending at a zero crossing. The contour of the fade-in/fade-out curve is known as the **window type** and some have specific names (ex: Blackman, Hanning, Hamming, etc.). The amount of overlap between windows is given by the **window step**.

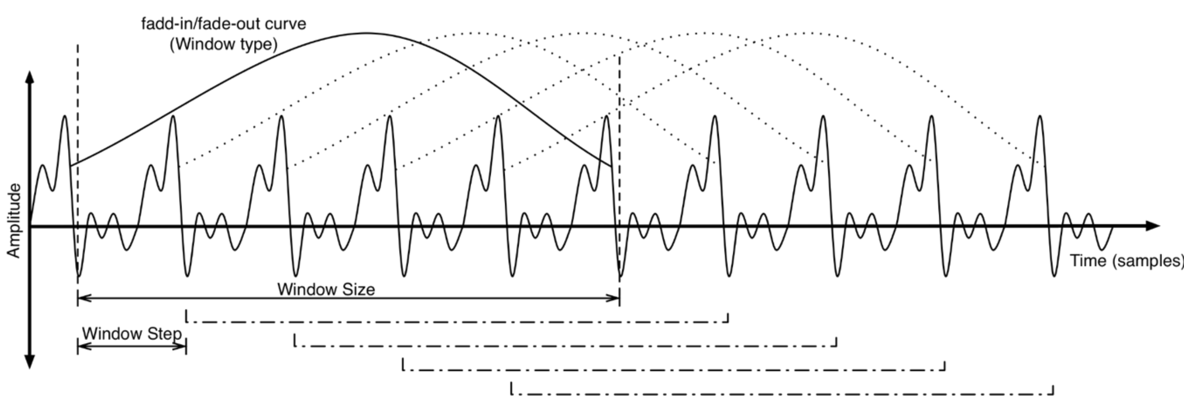


Figure 7-1

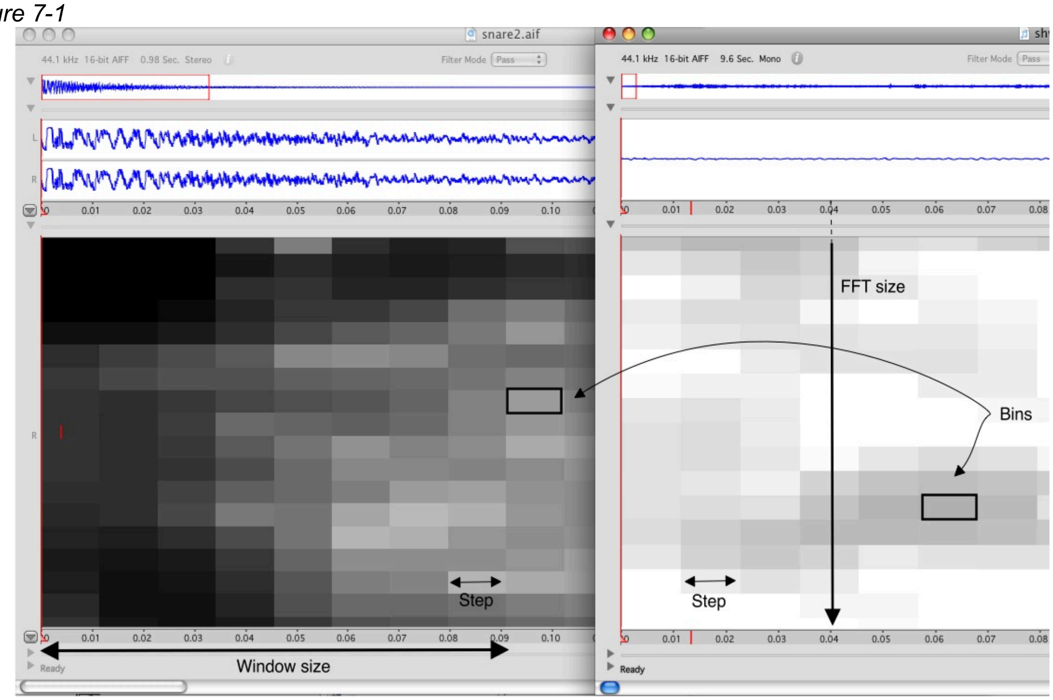


Figure 7-2

NOTE: In general, a bigger window will produce greater frequency resolution, but poorer time resolution; whereas a smaller window will produce greater time resolution, but poorer frequency resolution.

<u>Formulas</u>: **Window sizes are calculated in number of samples**. For an FFT to accurately decompose a sound into its sinusoidal components, the window must contain enough samples to represent at least 5 periods of the lowest frequency present in the sound.

• For a given sampling rate (SR), the following formulas can be used to calculate the window size (WS), or the lowest frequency (f) that can be accurately represented:

WS= (SR/f)*5

f= (SR*5)/WS

For example, with:

SR = 44.1 kHz
f = 60 Hz (so that one can hear the lowest C of the cello)

WS = (44100 Hz / 60 Hz) * 5
WS = 3675 samples

OR, with:

SR = 44.1 kHz
WS = 1000 samples

F = (44100 Hz * 5) / 1000
F = 220.5 Hz

8 Reverberation

Reverberation is the product of sound reflecting off of all the surfaces present in a space. As listeners, the sound we hear is actually always colored by reverb since reflected sound waves collide with one another resulting in constructive or destructive interference. Reverb is therefore an important factor to consider when working on any type of audio project.

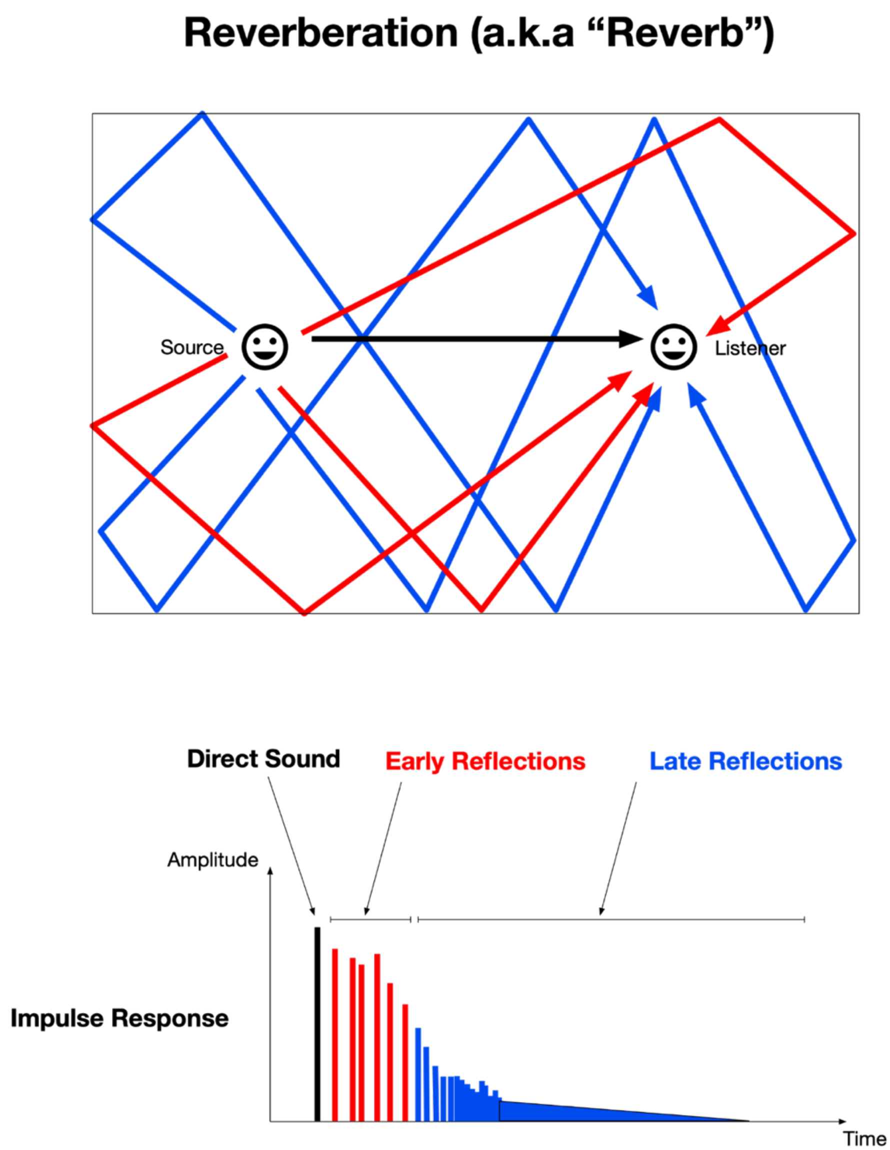


Figure 8-1 – Various attributes of reveberation

8.1 Common Attributes of Reverb

• The reflections that arrive right after the direct sound until about 50ms have elapsed are called **early reflections**. If early reflections are too evenly spaced, they may color the sound to the point of sounding metallic.

• **Late reflections** are reflections that arrive about 50ms after the direct sound.

• **Reverberation time (a.k.a., decay)**: The amount of time for the level of all reflections to drop by 60dB compared to the level of the initial direct sound (the sound may not be necessarily audible for this long, depending on room noise, other sounds, etc.).

  • Reverberation time depends on the volume of the room and the type of reflective surfaces. Large volumes tend to have longer reverberation times, and highly reflective surfaces (hard, solid, non-porous) tend to increase reverberation time, while absorptive surfaces reduce reverberation time.

- Reverberation time is not constant throughout frequency spectrum. High frequencies tend to be absorbed (by various materials found and through air absorption) with greater ease than lower frequencies. This is often reproduced by applying a low-pass filter on the reverb tail.

• **Onset delay (a.k.a pre-delay):** Time it takes for the first reflection to arrive after the direct sound. A long delay (>50 msec) gives an echo effect. A very short delay (<5 msec) gives the feeling of a very small space. Normal is between 10 to 20 msecs

• **Reverb density**: amount and level of the reflections in the reverb tail.

8.2 Types of reverb devices

In the studio, reverb is simulated with the use of devices (hardware or software) that produce reflections patterns. Two basic types of reverb devices exist:

• **Algorithmic reverb** attempt to reproduce the reflection patterns through the use of mathematical algorithms that more efficiently create delay-type effects. These reverbs can be extremely realistic or sonically impressive. Examples include Lexicon 480L, TC Electronic M3000, Bricasti M7, Valhalla Room, and Chroma Verb.

• I**mpulse Response reverb** add reflections to another sound through a process known as convolution. Convolution is a complex mathematical operation that "blends" one function with another" (cross-synthesis would be another example of convolution). These devices, also known as 'convolution reverbs', are generally regarded as the most realistic option since they imprint onto a sound all the non-linear subtleties of a space. Examples include Altiverb and Space Designer.

9 Auditory Localization

"Auditory localization is the human perception of the placement of a sound source" (Dodge, 308)

9.1 Types of Auditory Localization Cues:

• **Interaural time differences (ITD)**: Delay between the arrival of a sound at one ear and at the other. If sound is directly in front, ITD=0. ITD cues becomes less precise as the sound source moves to the side of the listener. ITD cues are most effective between 270 to 500 Hz, but don't work at all above 1400 Hz.

• Interaural intensity differences (IID): Difference in the intensity level of a sound at both ears. Due to the head partially dampening the sound at the ear on the opposite side of the sound source. If sound is directly in front, IID=0. Below 500 Hz, there are no significant IID cues, but become gradually more important as frequency increases. Above 1400 Hz, IIDs are the most important means of sound localization.

NOTE: For low sounds (spectral energy below about 270 Hz), both ITD and IID are ineffective. This is why stereo systems will only have one subwoofer, and it does not really matter where it's placed in the room.

• Head-related filtering (measured as Head-Related Transfer Functions, or HRTFs): Expression of the filtering imparted to a sound by the shape of the head and the pinnae (the external part of the human ear). Their effect is more pronounce at frequencies greater than 4kHz.

******Side Note**
"Listeners estimate [location] almost entirely from the attack portion of a sound" (Dodge p.311). Hence, sounds most easily located in space will be iterative rather than sustained (i.e. lots of attacks), and will have considerable frequency content between about 500 Hz (C5) and 1400 Hz (F6)–where both ITDs and IID are prominent.

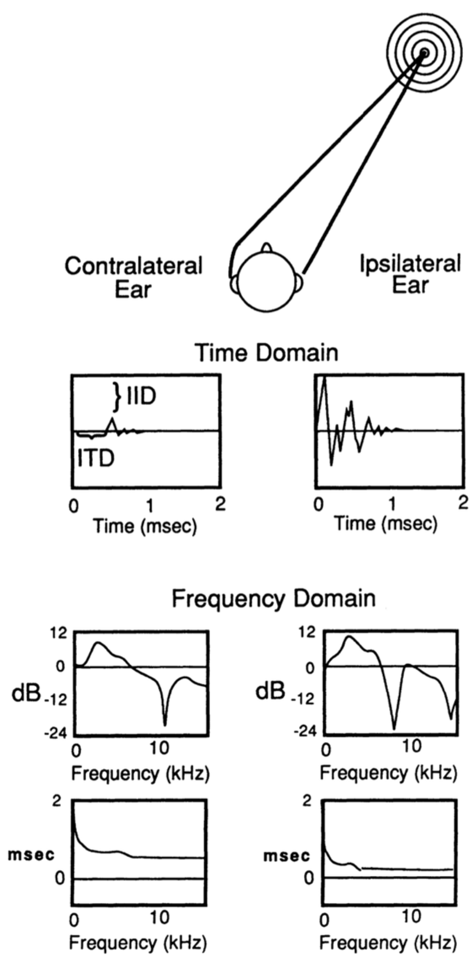


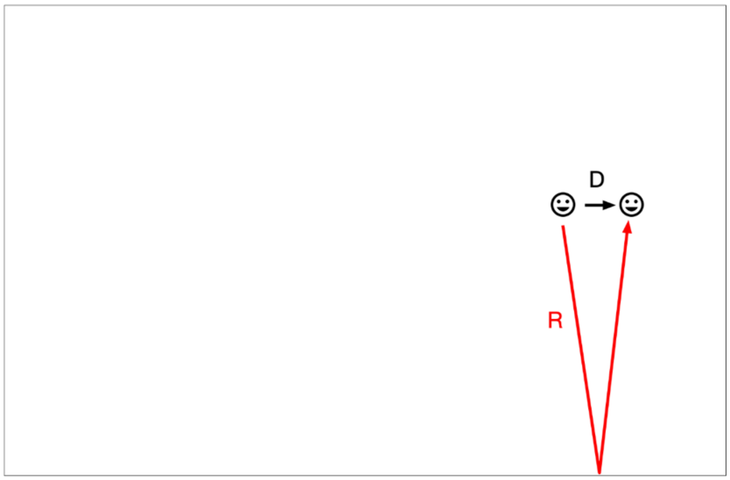
Figure 25 – Illustration of ITDs and IIDs (src.: Kendall, Gary S. A 3-D Sound Primer: Directional Hearing and Stereo Reproduction. Computer Music Journal, Vol. 19, No. 4, 26.)

9.2 Distance cues

• Intensity of the sound. Most influential cue when hearing a sound for the first time.
• Ratio of reverberated to direct sound (**R/D ratio**). As distance increases, the ratio gets smaller (more reverb). The **critical distance** is defined by an R/D ratio of 1. Beyond this point, it becomes progressively harder to discern the distance of a sound as the amount of reverb is greater than the amount of direct sound.
• Amount of high frequency energy in the sound.

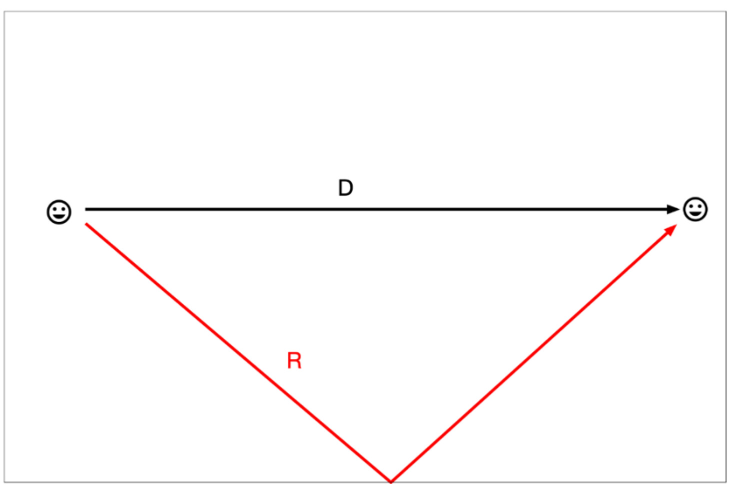
Source Sounds Close

D much greater than R



Source Sounds Far

D a little greater or equal than R



10 Signal Types

Coming soon

11 Cable Types

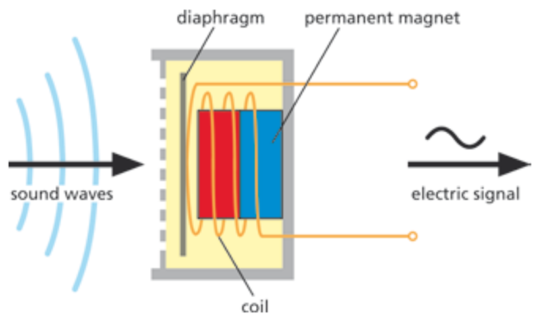
Coming soon

12 Microphones

12.1 Types:

• **Dynamic**: Operates via a diaphragm attached to a thin coil wrapped around a magnet. The magnetic flux of the coil moving in an electromagnetic field creates a voltage.

- Pros: rugged, simple construction therefore generally lower cost, can handle very high sound pressure levels (good for drums, trumpets, etc)
- Cons: not very sensitive (can be noisy for quiet sounds), generally weaker high frequency response compared to condenser mics.
- examples: Shure SM57 and SM58, Sennheiser 421, Beyer M88, EV RE20

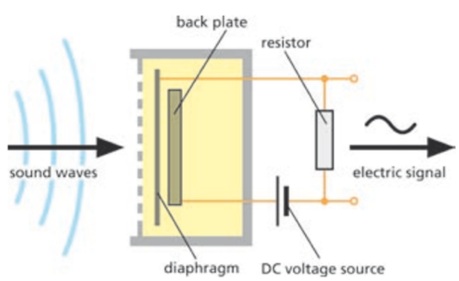


Dynamic microphone

Figure 26 – Sideview of dynamic microphone (src.: Holmco Speaker Microphones Catalogue)

• **Condenser**: Works on the principle of a capacitor (which has two plates): the diaphragm acts as one plate of a capacitor, and the vibrations change the distance between the two plates. An electrical current, called **phantom power (abbreviated as 48v)**, charges one of the two plates of the capacitor, and as the diaphragm moves the voltage changes.

- Pros: very sensitive (good for quiet sounds), excellent response to high frequencies
- Cons: less rugged, higher cost (though not so much today), can more easily distort on loud sounds.
- Examples: AKG C-414, Neumann KM184, Shure KSM147, Rode NT1A, Rode NT55,



Condenser microphone

Figure 27 - Sideview of condenser microphone (src.: Holmco Speaker Microphones Catalogue)

• **Ribbon**: Works on same principle as dynamic microphone, but the diaphragm is a thin aluminum ribbon.
  - • Pros: although large, can pick up a lot of high-frequency detail without sounding as harsh as condenser microphones; don't require phantom power.
  - • Cons: the most fragile of microphones
  - • Examples: RCA 44 and the AEA R44

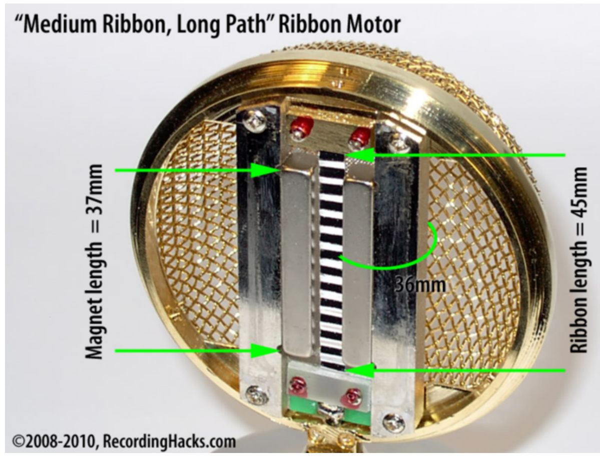


Figure 28 - Ribbon mic construction (src: http://recordinghacks.com/2008/11/01/chinese-ribbon-microphone-designs/)

• **Contact (piezo)**: Contact mics only pickup surface vibrations by using two discs (one of brass, the other of ceramic) sandwiched together. When stuck to a plane surface (a wood table for example), the surface's vibrations will deform the discs which creates a voltage due to the crystalline structure of the ceramic.
  - • Pros: only pickup surface vibrations (i.e., they don't' pickup air pressure), cheap, possibly DYI
  - • Cons: generally noisy and need lots of gain
  - • Example: Piezo discs, Schertler Dyn-Uni, AKG 411

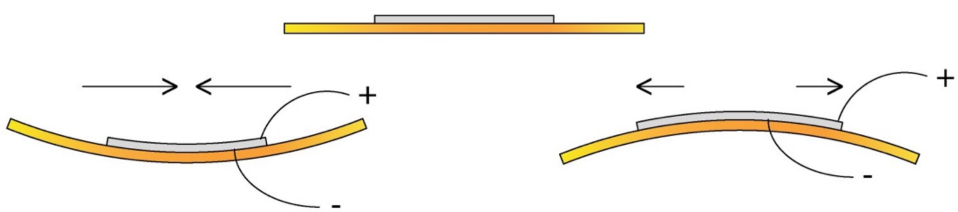


Figure 29 -Bending of ceramic disc creates voltage in piezo device (i.e., contact mic). Src.: Piezoelectricity - Wikipedia

• **DI (Direct Input)**: Not a microphone per se, but rather a devices that converts an unbalanced, high-impendence instrument signal (such as a guitar or a piezo contact mic) to a balanced, low-impendence microphone signal. Active DIs require phantom power; passive DIs do not required phantom.
  - • Pros: creates a balanced signal, maintains full frequency spectrum of instrument
  - • Cons: none
  - • Examples: Radial ProD2, Radial JDI,

Figure 30 - DI box example (src: https://www.radialeng.com/product/prod2)

12.2 Directional Patterns (a.k.a. polar pattern or directionality)

The directional pattern of a microphone determines its sensitivity to sound, relative to angle. In other words, some patterns are better than others for picking up sound coming from certain directions. Directional patterns are displayed as a graphic, commonly referred to as a microphone's polar pattern.

The following is a list of the three basic polar patterns and their features, plus derivatives of the cardioid pattern.

• **Omnidirectional**: equal sensitivity at all angles (360° microphone)

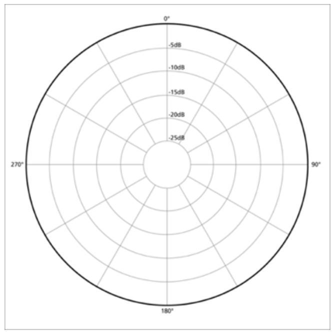


Figure 12-6

• **Bidirectional (a.k.a. figure-eight)**: equal sensitivity at 0° and 180°, no sensitivity at 90° and 270°

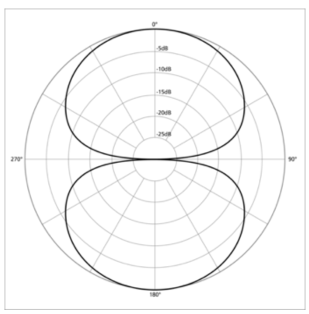


Figure 12-7

• **Unidirectional**: most sensitive to sound source at 0°, least sensitivity at 180°.

Progressing from super-cardioid to shut-gun, pickup on the sides decreases while pickup from the rear increases. These derivatives of the cardioid pattern may improve separation when recording multiple sources with multiple microphones (i.e. less "bleeding").

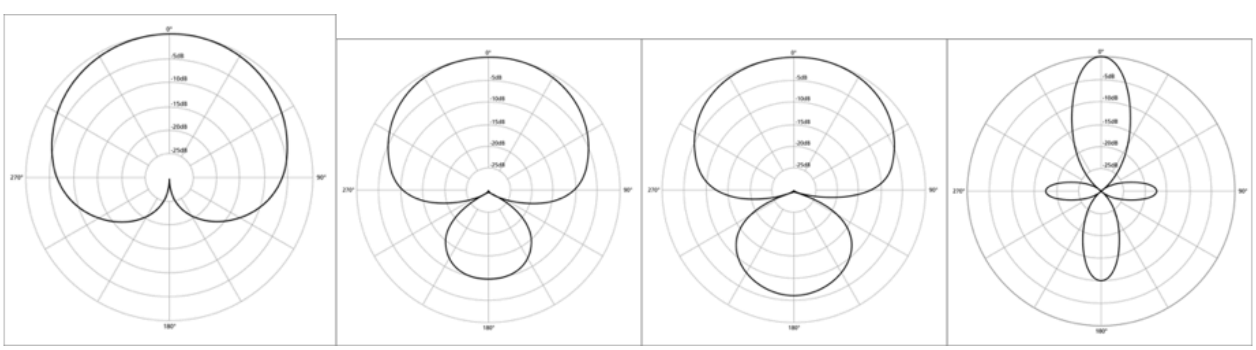


Figure 12-8 | From left to right: Cardioid, Super-Cardioid, Hyper-Cardioid, Shot Gun

****Side Note**
Polar patterns are not constant throughout the entire frequency range of a microphone. Most microphones tend to be less directional in the lower frequencies, and increasingly directional in the upper frequencies. The example below, of an omnidirectional microphone, is quite typical.

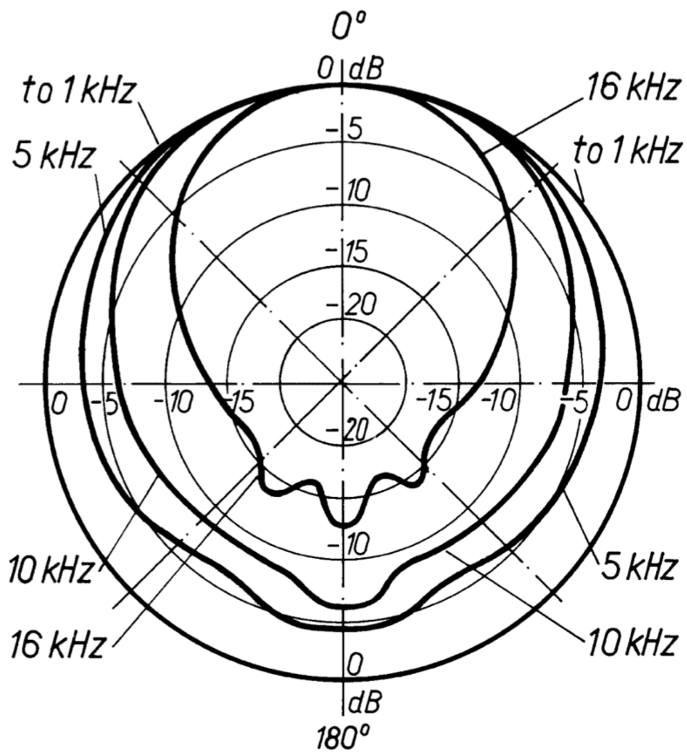


Figure 34 – Typical omnidirectional polar pattern (src.: Boré,Gerhart and Stephan Peus. Microphones for Studio and Home-Recording Applications Operation Principles and Type Examples. Neumann, Berlin, p21.)

13 <u>Recording</u>

13.1 Mono vs. Stereo

• A mono recording uses one microphone. This gives a flat sound without any special imaging. This can be very useful for projects where instruments will be recorded, analyzed, and the sound processed and later spatialized.
• A stereo recording uses at least two microphones and thus creates a much better sound image. As a consequence, stereo takes twice the computer memory and processing, and stereo recordings can also be harder to mix in a multi-track project since constructive/destructive interference can drastically change the sound of the recording when both channels are summed together (when panning, for example).

13.2 Stereo Mic'ing Techniques

Stereo recording attempts to emulate the way humans hear. In other words, a stereo recording will contain, at least in part, the information we need to locate sounds in space: inter-aural time differences (ITD), inter-aural intensity differences (IID), and in the case of binaural recordings, head-related filtering (these are further discussed in section 9). There are many ways to record in stereo, but the following four techniques are the most common:

• Coincident pair (a.k.a XY pair): Two cardioid microphones are at the same place, typically pointing at an angle between 90° and 135° to each other. Captures IIDs only.

- Pros: Mono compatible since it eliminates phase cancellation problems. Sound very focused in the center.
- Cons: Stereo less precise than other techniques.

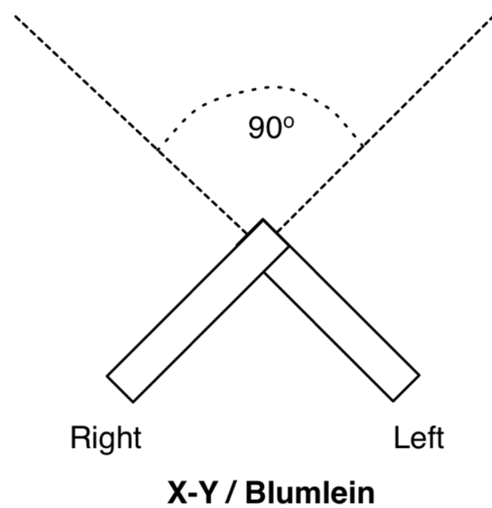


Figure 13-1

• **Near-coincident pair**: Two cardioid microphones typically placed 10-25 centimeters apart at an angle of between 90º-110º. Captures both IIDs and ITDs.

- Pros: More precise stereo image than XY.
- Cons: Mono compatibility less good than XY.
- Common varieties: ORTF (17 cm apart, 110º), NOS (25cm apart, 90º)

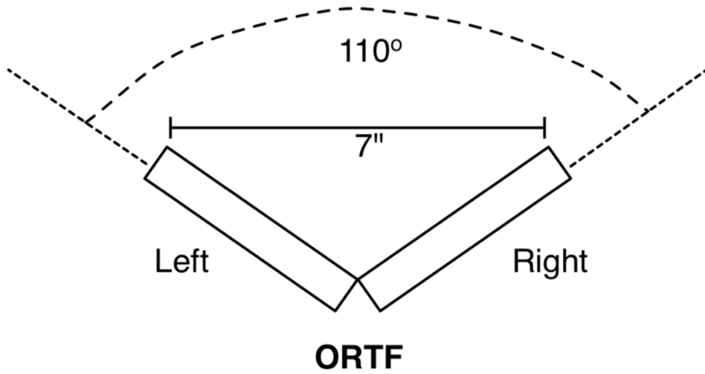


Figure 13-2

• **Spaced pair (a.k.a AB)**: Two omni microphones placed in parallel 20” to 10’ apart. The distance depends on the size of the sound source. Captures only ITDs.

- Pros: Very simple setup. Provides very wide sound field.
- Cons: Mono compatibility is poor. Because it captures a lot of room sound, the technique is only as good as the room you’re recording in.

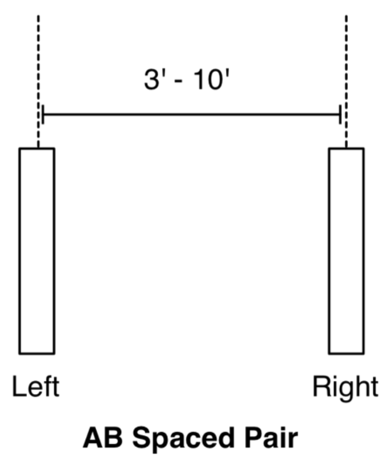


Figure 13-3

• **Binaural (a.k.a dummy head)**: Two very small omni microphones are places inside the ears of a dummy head. Captures IIDs, ITDs, and head-related filtering.

- Pros: Extremely realistic sound reproduction, even for sound in back of the head.
- Cons: Only works if listening to the recording with headphones; complicated setup

13.3 Effects of Source Distance from Microphone while Recording

• Mic distance affects the high frequency content of a sound source (a closer mic may make a source sound more bright yet “unnatural”, and a farther sounds less bright but more “natural”)
• Mic distance may increase bass frequency response of cardioid mics via the “proximity effect” (see Shure sm57 beta spec sheet)
• Mic distance has an effect on the amplitude of the signal
    • see Inverse Square Law -> -6dB level decrease as distance from mic to source is doubled

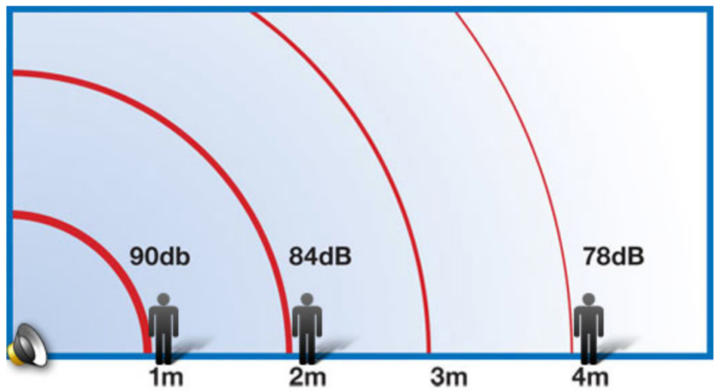


Figure 38 – Illustration of inverse square law (src.: Extron website, https://www.extron.com/calculators/inverse-square-law/?tab=tools)

• Mic distance has an effect on the presence of the room in a recording (reflections, D vs R)

Bibliography

Benson, K. Blair, editor. Audio Engineering Handbook. New York: McGraw-hill, 1988
Dodge, Charles and Thomas A. Jerse. Computer music. United States: Schirmer, 1985

Roads, Curtis (1996). The Computer Music Tutorial. MIT Press.

Unspecified author. What is reverb? Found in Sound on Sound online: http://www.soundonsound.com/sos/may00/articles/reverb.htm. Accessed March 26th.

Unspecified author. What is reverb? Found in Sound on Sound online: http://www.soundonsound.com/sos/may00/articles/reverb.htm. Accessed March 26th. 2010.

Bridget Shield, Trevor Cox. Concert Hall Acoustics: Art and Science. http://www.acoustics.salford.ac.uk/acoustics_info/concert_hall_acoustics/?content=testing. Accessed March 26th, 2010.

Unspecified author. 'Space-Designer' in Logic Studio: Instrumenta and Effect Documentation. 2007.