FH MMA SALZBURG – AUDIO

ACOUSTIC BASICS AND DIGITAL AUDIO

TABLE OF CONTENTS

1. ACOUSTICS BASICS

1.1 SOUND AND WAVEFORMS

1.1.1 SOUND IN PHYSICS AND IN PSYCHOACOUSTICS

In physics, sound is a ***vibration*** that propagates from a ***sound source*** as an acoustic wave through an ***elastic transmission medium***, such as a gas (for example air), liquid (for example water) or solid.

While beyond the limits of our perception, sound in physics includes *infrasound* (below 20 Hz frequency) and *ultrasound* (beyond 20 000 Hz frequency) as well as sound below the *threshold of hearing* and beyond the *threshold of pain* (see § 1.2 and § 1.3). In Psychoacoustics (the branch of psychophysics involving the scientific study of sound perception), sound is the *reception of soundwaves through our hearing system* and their *perception by the brain*.

Within the context of Music Production and Sound Design, it is essential to understand the physiological limits of our hearing range, that are defined both in terms of frequency (20 – 20 000 Hz, see § 1.2) and dynamic range (0 – 137,5 dB SPL, see § 1.3). It is also very important to understand the perception of sound color (see § 1.4) and the perception of ambience effects (reverb, echo, delay, etc. see § 1.7).

The physiological limits and characteristics of our hearing also define the requirements for the audio equipment used, that should at least match or exceed these limits, as well as the recommended listening environment and sound pressure levels.

1.1.2 WAVEFORMS / SOUNDWAVES

A **waveform** is the visual representation of changes in the medium density plotted over time. The waveform *amplitude* is the strength of the displacement and relates to our perception of *loudness*, while the waveform *wavelength* (which is inversely proportional to its *frequency* in Hz) relates to our perception of *pitch*.

The purest of all soundwaves is the *sine wave*, which describes a smooth periodic oscillation and represents a single frequency (with no sound color or character – see § 1.4.1). All complex sounds can be reduced to their sine wave components called *partials* (see § 1.4.4).

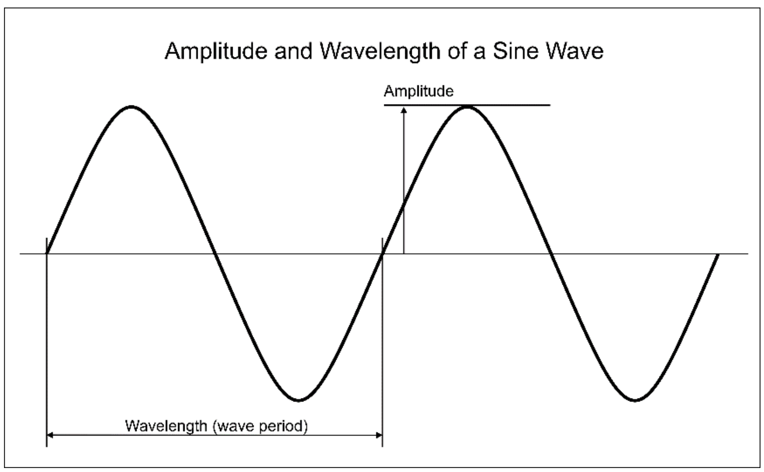


Figure 1: Amplitude and Wavelength of a Sine Wave

Soundwaves can be:

- ***periodic***, in which case they are classified as ***harmonic*** and have a clearly identifiable base frequency and pitch, as well as *harmonic partials* that are multiple integers of the base frequency.
- ***non-periodic***, in which case they are classified as ***inharmonic*** and feature *inharmonic partials* that are not multiple integers of the base frequency.
- ***random-like***, in which case they are classified as ***noise***, which is unpitched, indeterminate, uncontrolled or even random sound. *White noise* is the random sum of all frequencies.

Most pitched instruments (strings, woodwinds, brass, guitars, pianos, etc.) produce sounds that are mainly *harmonic* in character (but that can include some *inharmonic* or even *noise-like* components, like the breath noise of a flute, or the hammer impact of a piano). Most percussion instruments produce mainly *inharmonic* sounds, some are even mainly *noise-like* in character (for example the cymbals in a drumkit).

1.1.3 SOUNDWAVE PROPAGATION

In the absence of obstructions, soundwaves propagate spherically from the sound source as *longitudinal waves* also called *compression waves*. Longitudinal waves are waves of alternating pressure deviations from the equilibrium pressure (see § 1.3.2). The propagation speed of sound through air at sea level is about 340 m/s (see § 1.6).

On the planet's surface the propagation is *hemispherical*. In half or closed spaces, soundwaves encountering an obstacle (like a wall or ceiling) will generally be *reflected*, *refracted* or *absorbed*, depending on the angle of incidence, the material shape, and the material *absorption coefficient*. If soundwaves keep reflecting long enough within two or more surfaces or a closed space, the reflection pattern can be perceived as *reverberation* or *echo* (see § 1.7).

1.2 FREQUENCY AND BANDWIDTH

The **Hertz** is the unit for frequency (cycles per second) in all forms of sound, regardless of whether we are dealing with actual sound vibrations in the air (or another medium), with an analog audio signal (after being converted by a microphone) or with a digital audio signal (after being sampled by an analog-to-digital converter).

- **Hz** (Hertz) = unit of measure for frequency
- 1 Hz (Hertz) = 1 cycle per second
- 1 kHz = 1 000 Hz
- 1 MHz = 1 000 kHz
- 1 GHz = 1 000 MHz

The **human hearing frequency range** (or bandwidth) is usually considered to be 16 - 20 000 Hz. In practice, even for young and healthy individuals, the sensitivity drops rapidly beyond 16 000 Hz. With increasing age, the sensitivity in the high frequency range declines further. At age 50, the upper frequency limit us usually 12 000 – 14 000 Hz.

- **Infrasound** = frequencies below 16 Hz
- **Ultrasound** = frequencies above 20 kHz
- **1 octave** = doubling of the frequency in Hz
- **Human hearing frequency range** = about 10 octaves bandwidth (from 16 to 16 384 Hz there are exactly 10 octaves)

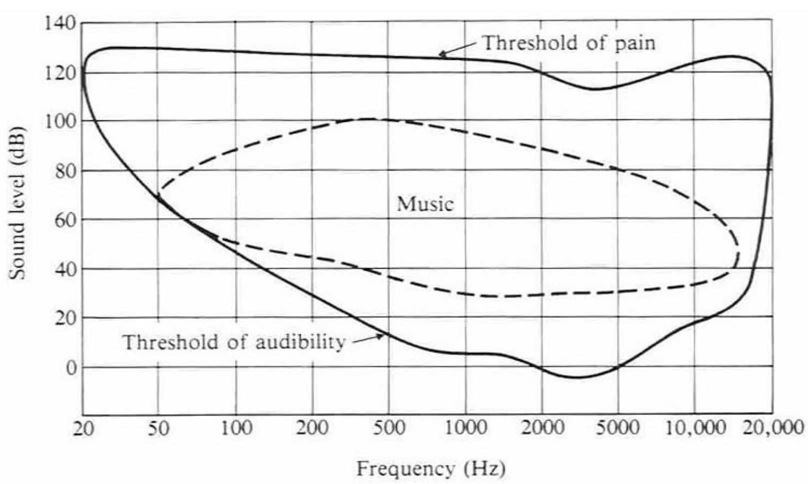


Figure 2: Hearing range, Threshold of Audibility and Threshold of Pain | from: *das Tonstudio Handbuch* - Hubert Henle

The **hearing frequency range in animals** can vary greatly, here are some examples:

- dog (60 – 45 000 Hz)
- cat (45 – 64 000 Hz)
- horse (55 – 33 500 Hz)
- mouse (1 000 – 91 000 Hz)
- bat (2 000 – 110 000 Hz)
- beluga whale (1 000 – 123 000 Hz)
- owl (200 – 12 000 Hz)
- chicken (125 – 2 000 Hz)

1.3 AMPLITUDE AND DYNAMIC RANGE

1.3.1 DECIBEL – POWER AND FIELD QUANTITIES

The **dB** (**decibel** = 1/10 of a Bel) is a logarithmic unit that expresses the ratio of two values of a physical quantity. In the context of audio this is either a *power/intensity* quantity, or a *field* quantity. As it just expresses a *ratio*, a dB is a *dimensionless unit*. However, if one of the two quantities is defined as a *reference value*, then the dB can express the *absolute level* of that physical quantity.

The dB uses a logarithmic scale in base 10. Two different scales are used when expressing a ratio in decibels, depending on the nature of the quantities: *power* or *field* (*root-power*).

When expressing a *power quantity* (for example: energy, acoustic intensity), the number of decibels is 10 times its $\log_{10}$.
This means a change in *power* by a factor of 10 corresponds to a 10 dB change in level.
The formula to translate sound *power* (intensity) (in W) to dB is:

- **$Li = 10 \times \log_{10} (I1 / I0)$ dB**

where "I0" is the reference intensity and "I1" the intensity being measured.

When expressing a *field (root-power) quantity* (for example: voltage, current, sound pressure), the number of decibels is 20 times its $\log_{10}$. This means a change in *amplitude* by a factor of 10 corresponds to a 20 dB change in level.
The formula to translate *field* (root-power) to dB is:

- **$Lp = 20 \times \log_{10} (U1 / U\ 0)$ dB**

where "U0" is the reference level (in the case of dB SPL: 0,00002 Pa) and "U1" the level being measured.

What is a Logarithm?
In mathematics, the logarithm is the inverse operation to exponentiation. That means the logarithm of a number is the exponent to which another fixed value, the base, must be raised to produce that number. Examples:

- the base 10 logarithm of 10 is 1
- the base 10 logarithm of 100 is 2, as $10^2 = 100$
- the base 10 logarithm of 1000 is 3, as $10^3 = 1000$
- the base 10 logarithm of 2 is 0,301, as 10 to the power of 0,301 = 2
- the base 10 logarithm of 4 is 0,602, as 10 to the power of 0,602 = 4

1.3.2 SOUND PRESSURE LEVEL

The **dB SPL** scale (*Sound Pressure Level* or *Acoustic Pressure*) is used to measure the amplitude of sound vibrations in the air. However, SPL measurements in other media (like water) are also possible.
The reference level is **0 dB SPL** corresponding to an air pressure deviation of 20 micropascals (0,00002 Pa) in air, or 1 micropascal in water. Due to difference reference levels, use of dB underwater can lead to confusion.

0 dB SPL is also the *Threshold of Hearing* (the quietest possible sound pressure level that can still be heard by human hearing), consequently most of the measurements in dB SPL are usually *positive values*, all the way up to the *Threshold of Pain* (137,5 dB SPL) and beyond. Both thresholds can vary with age, the condition of hearing and from person to person.

- **1 Pa** (Pascal) = unit of measure for *pressure* = 1 N/m² (1 Newton per square meter)
- Standard atmospheric pressure at sea level: 101 325 Pa = 1 013,25 millibar (1 millibar = 1 Hectopascal = 100 Pa)
- **dB SPL**: unit of measure for *Sound Pressure Level or Acoustic Pressure*
- **Threshold of Audibility** (or Threshold of Hearing): 0 dB SPL = 0,00002 Pa (pressure deviation)
- **Threshold of Pain**: 137,5 dB SPL = 150 Pa (pressure deviation); note: both thresholds vary with frequency
- **Human hearing dynamic range**: about 137,5 dB from the threshold of audibility to the threshold of pain

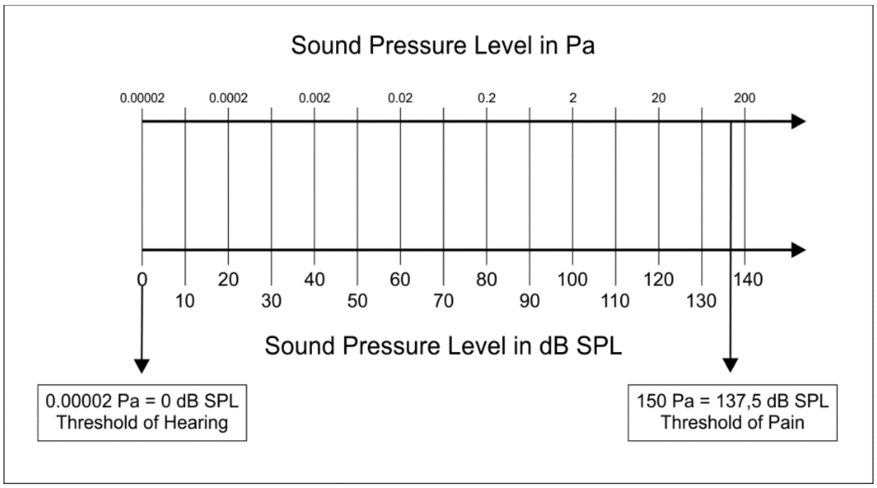


Figure 3: Sound pressure levels in Pascal (linear scale, Pa) and dB SPL (logarithmic scale).
A factor 10 in Pa corresponds to a 20 dB change.

1.3.3 AMPLITUDE IN ANALOG AUDIO SIGNALS

The amplitude in **analog audio signals** (electrical voltage – for example sound that has been captured by a microphone or generated by a synthesizer) is measured in **dBu** (reference level 775 mV) or in **dBV** (reference level 1 V). Popular standards are +4 dBu (for professional audio gear) and -10 dBV (for consumer equipment).

- **dBu** = unit of measurement for the *amplitude* of analog audio signals (electrical voltage)
  reference level: **0 dBu = 0,775 V** (775 mV)
  A popular calibration level for *professional audio equipment* is +4 dBu
  (the "0" on the VU meter equals in fact to an amplitude of +4 dBu)

- **dBV** = unit of measurement for the *amplitude* of analog audio signals (electrical voltage)
  reference level: **0 dBV = 1 V** (1 000 mV)
  A popular calibration level for *consumer audio equipment* is **–10 dBV**
  (the "0" on the meter equals in fact to an amplitude of –10 dBV)

Figure 4: Millennia TCL 2 compressor with VU-meters– VST plugin modelled by brainworx

1.3.4 AMPLITUDE IN DIGITAL AUDIO SIGNALS

The amplitude in **digital audio signals** such as PCM (Pulse Code Modulation) is measured in **dBFS** (decibel relative to Full-Scale) and the reference level is 0 dBFS. A *Full-Scale Signal* is the maximum possible value expressed in bits for a digital audio signal before clipping occurs. Therefore, all peak amplitude measurements smaller than the maximum are *negative values* in dBFS. A positive dBFS value means the signal is “clipping” (being truncated, as it is exceeding the maximum available amplitude).

- **dBFS** = unit of measurement for the *amplitude* of digital audio signals
  reference level: **0 dBFS** = Full Scale Signal

Examples of dB calculations

- How much difference in dB is there between an electrical signal of 2V amplitude, compared to the reference level 0 dBu? $20 \times \log_{10} (2 / 0{,}775) = 8{,}234$ dB
- How much difference in dB is there between an electrical signal of 2V amplitude, compared to the reference level 0 dBV? $20 \times \log_{10} (2 / 1) = 6{,}020$ dB
- How much difference in dB is there between the reference levels 0 dBV and 0 dBu? $20 \times \log_{10} (1 / 0{,}775) = 2{,}21$ dB

Remember: the difference in level between -10 dBV and +4 dBu is [14 - 2,21 = 11,79 dB] and not 14 dB!

1.3.5 PEAK VS RMS AMPLITUDE

The **Peak Amplitude** (Pk) is the maximum excursion of the wave from the zero or equilibrium point (the *equilibrium pressure* in the case of actual sound waves). While in symmetrical waveforms like a sine wave the positive and negative peaks are identical, this is not necessarily true for every audio signal.

The **Peak-to-Peak Amplitude** (Pk-Pk) is the measurement from a negative peak to a positive peak. In the case of a sine wave, the peak-to-peak value is exactly twice the peak value, because the negative and positive phases of the waveform are symmetrical, but this is not necessarily true for every audio signal.

The **Root Mean Square Amplitude** (RMS) is the *square root* of the *mean over time* of the *square values* of the waveform. In the case of the sine wave, the RMS value is 0,707 times the peak value. The RMS value is proportional to the area under the curve.

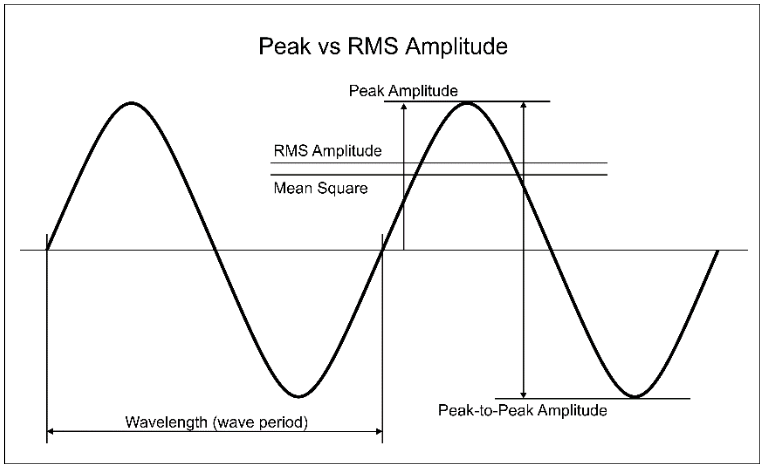


Figure 5: Peak vs RMS Level

If the negative peaks are rectified (flipped to positive) and the area under the resulting curve averaged to a constant level, that level would be proportional to the RMS value:

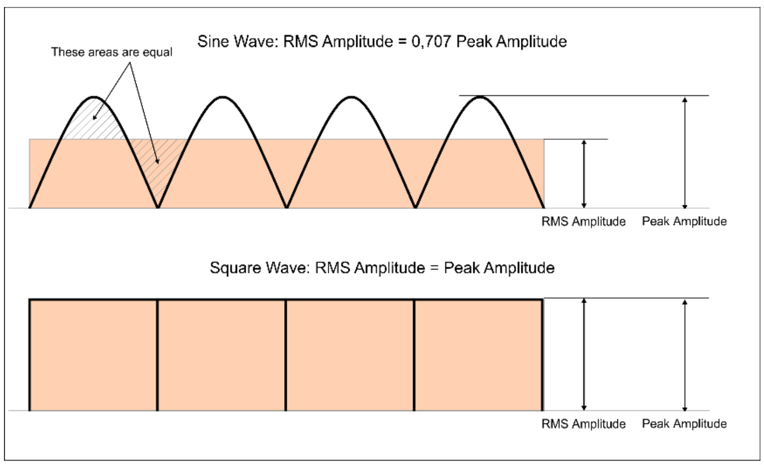


Figure 6: Comparison of RMS and Peak Amplitude in a Sine and Square Wave

In a square wave the RMS value is identical with the peak value, and this is true for square waves only.

The RMS value of a vibration signal is an important measure of its amplitude. As mentioned before, it is numerically equal to the square root of the average of the squared value of amplitude. To calculate this value, the instantaneous amplitude values of the waveform must be squared, and these squared values averaged over a certain length of time. This time interval must be at least one period of the wave to arrive at the correct value. The squared values are all positive, and thus so is their average. Then the square root of this average value is extracted to get the RMS value.

RMS values are important because our perception of loudness depends mainly on the overall amount of energy in a sound wave (which is well approximated by the RMS amplitude value) and less on the short peak amplitude values.

1.4 SOUND SPECTRUM AND COLOR

1.4.1 SINE WAVE

A stated in § 1.1, a ***sine wave*** describes a smooth periodic oscillation and represents a single frequency (with no sound color or character). For example, a sine tone at 100 Hz appears as a single frequency spike in a spectrum analyzer (Freq. vs. Energy graph):

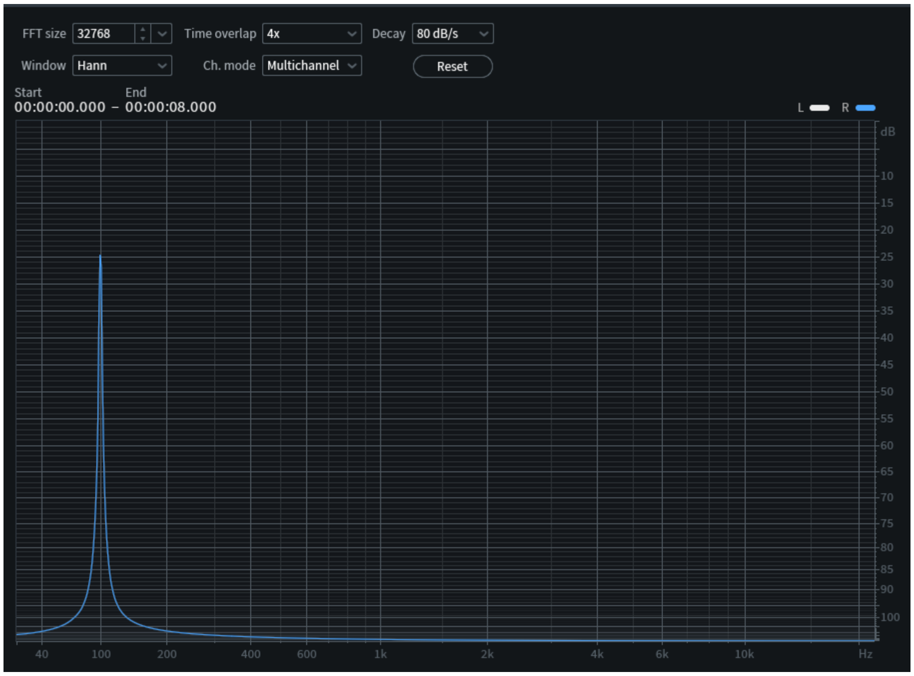


Figure 7: Spectral analysis of a 100 Hz sine wave

Very few objects or instruments produce sounds close to a sine wave spectrum: for example, a tuning fork, or the glass harmonica. However, even in these instruments other spectral components are present. A ***pure sine wave*** is a mathematical function and from a sound perspective it can only be generated electronically.

Most acoustic instruments, as well as most synthesizers, produce sounds that are a combination of several sine wave components called ***partials***, as well as ***noise*** (random-like) components. The variations in amplitude and frequency of the partial tones, as well as the differences in dynamic changes over time, define a complex *dynamic sound spectrum*, which we perceive as *sound color and character*.

All complex sounds can be reduced to their sine wave components called partials – see § 1.4.4.

1.4.2 HARMONIC AND INHARMONIC SOUNDS

Most acoustic pitched instruments (including piano, strings, woodwinds, brass, guitars, etc.) as well as tuned percussion instruments (marimba, xylophone, vibraphone, celesta, glockenspiel etc.) generate sounds that are mainly ***harmonic*** in character. Harmonic sounds have a clearly identifiable base frequency and pitch, and their partials are *multiple integers* of the base frequency. Their waveform is substantially ***periodic*** – after each period of the base frequency the waveforms repeats with little or no variation.
The sound of these instruments, however, usually includes also *inharmonic* or even *noise-like* components, like the breath noise of a flute, or the hammer impact of a piano.

Most untuned percussion instruments (drums, bass drum, snare, toms, cymbals, gongs, bongos, congas, timbales, talking drum, taiko, etc.) generate sounds that are mainly ***inharmonic*** in character, some are even mainly ***noise-like*** in character (for example the cymbals in a drumkit). Their waveform is ***non-periodic.***

1.4.3 STANDARD WAVEFORMS AND NOISE TYPES

Among the most popular waveforms used in analog subtractive synthesizers are the ***triangle***, the ***sawtooth*** and the ***square*** wave. These are all *harmonic* waveforms. Their partial energy decreases with increasing multiplier frequency. While the sawtooth contains both odd and even partials, both the triangle and the square wave contain only odd partials. The triangle however has a much darker sound character, as the upper harmonics are *proportional to the inverse square* of the partial number, while in sawtooth and square waves they are just *proportional to the inverse* of the partial number.

	Partials / Harmonics	Partial Energy
Triangle	Contains only odd partials	Proportional to the inverse square of the partial number
Square Wave	Contains only odd partials	Proportional to the inverse of the partial number
Sawtooth / Ramp	Contains all odd and even partials	Proportional to the inverse of the partial number

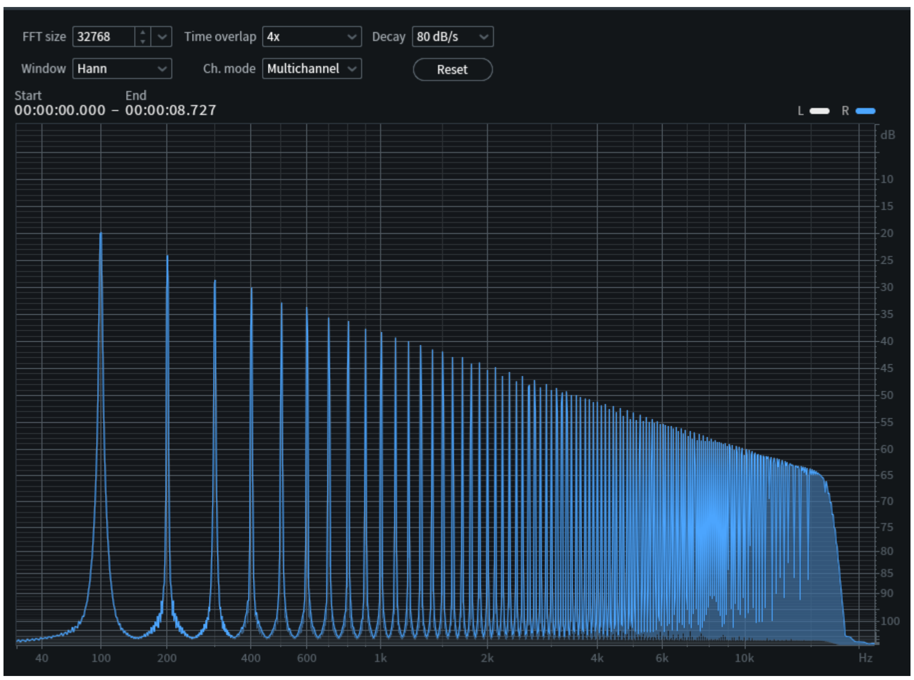


Figure 8: Spectral analysis of a 100 Hz sawtooth wave – each spike represents a sine wave harmonic partial

If no spectral component can be clearly identifiable, sound is classified as ***noise***, which is often described as unpitched, indeterminate, uncontrolled or even random sound. Wide spectrum random-like noise can be classified as white, pink or brown depending on the distribution of the spectral energy:

	Spectral Energy
White Noise	Constant power across the complete audible spectrum
Square Wave	The power drops -3 dB /octave with increasing frequency
Sawtooth / Ramp	The power drops -6 dB /octave with increasing frequency

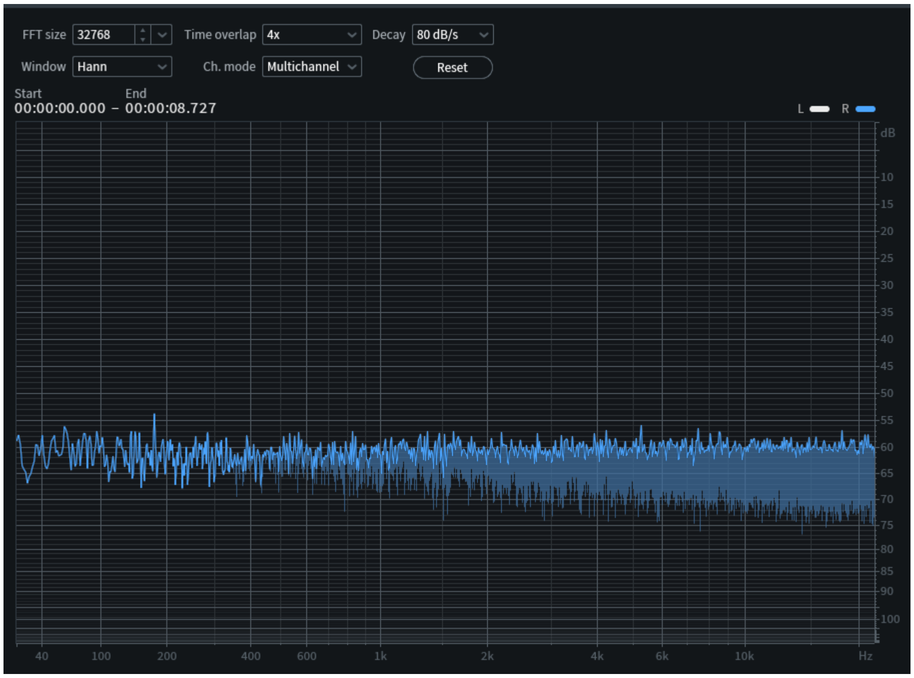


Figure 9: Spectral analysis of white noise, showing equal energy distribution across the complete sound spectrum

1.4.4 PARTIALS AND HARMONICS

The sine wave components of a complex sound are called ***partials***.

In harmonic sounds partials are *multiple integers* of the base frequency and are also called ***harmonic partials***. The frequency of the harmonic partials is equal to their progressive number multiplied by the base frequency.

In inharmonic sounds the partials are *not multiple integers* of the base frequency and are referred to as ***inharmonic partials.***

The terms *Fundamental Tone* (for the 1st Harmonic) and *Overtones* (the 1st Overtone = the 2nd Harmonic and so on) are also used but can lead to confusion because their progressive number does not correspond to their frequency multiplier, like for harmonic partials.

These are the first 8 harmonic components of a harmonic sound with base frequency 100 Hz:

Progressive Intervals	Interval relative to Fundamental Tone	Overtones	Harmonic Partials	Multiplier	Frequency
Base frequency	Fundamental Tone	Fundamental Tone	1st Harmonic	1 x 100 =	100 Hz
octave	octave (P8)	1st Overtone	2nd Harmonic	2 x 100 =	200 Hz
just perfect fifth	P8 + just perfect fifth (P5)	2nd Overtone	3rd Harmonic	3 x 100 =	300 Hz
just perfect fourth	2P8	3rd Overtone	4th Harmonic	4 x 100 =	400 Hz
just major third	2P8 + just major third (M3)	4th Overtone	5th Harmonic	5 x 100 =	500 Hz
just minor third	2P8 + P5	5th Overtone	6th Harmonic	6 x 100 =	600 Hz
septimal minor third	2P8 + septimal min. seventh (m7)	6th Overtone	7th Harmonic	7 x 100 =	700 Hz
septimal major second	3P8	7th Overtone	8th Harmonic	8 x 100 =	800 Hz

How can we tell instruments so easily apart if a piano, a violin, a flute, a trumpet playing a 100 Hz note share the same harmonics (frequency-wise)?

While the frequency of the harmonics is substantially identical between these instruments, their *amplitude* and *amplitude variations over time* and the resulting *dynamic frequency spectrum* are quite different. Most instruments are characterized by specific ranges in the spectrum that are emphasized, while others are less prominent. These ranges are called ***formants*** and remain relatively constant even when different notes are played on the instruments.

Formants appear as “mountains” (emphasized ranges) and “valleys” (weaker ranges) in the sound spectrum:

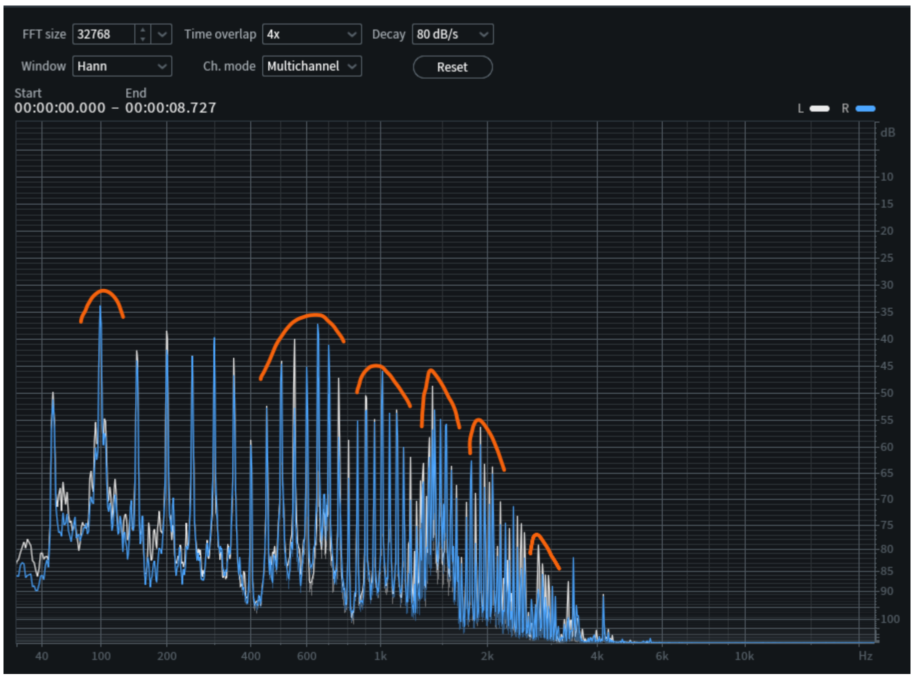


Figure 10: spectral analysis of a 50 Hz piano sound, with characteristic formants clearly visible

The formants, together with other sound characteristics, including the *envelope* (dynamic variation of the sound over time, defined by attack, decay and release time, as well as the sustain level in sustain-type sounds), *additional non-harmonic* and *noise components* all contribute to the precise identification of different instruments and sound sources.

Consider also that the lower the base frequency is, the more harmonics are within our hearing range. This is one of the reasons why bass sounds are often perceived as “richer” and “more interesting” than high pitched ones: due to the higher number of partials, bass sounds offer a much wider range of possible sound colors and characters.

Base frequency		Number of partials in hearing range
20 Hz	20 000 / 20 =	1 000
100 Hz	20 000 / 100 =	200
200 Hz	20 000 / 200 =	100
1 000 Hz	20 000 / 1 000 =	20
2 000 Hz	20 000 / 2 000 =	10
10 000 Hz	20 000 / 10 000 =	2

1.5 RELATION BETWEEN PERCEIVED DYNAMIC AND FREQUENCY RANGE

The **Equal-Loudness Contours** (ISO 226:2003 – earlier known as **Fletcher/Munson Curves**) are a measurement of *subjective loudness perception*, plotted for different dynamic ranges and across the complete frequency range.

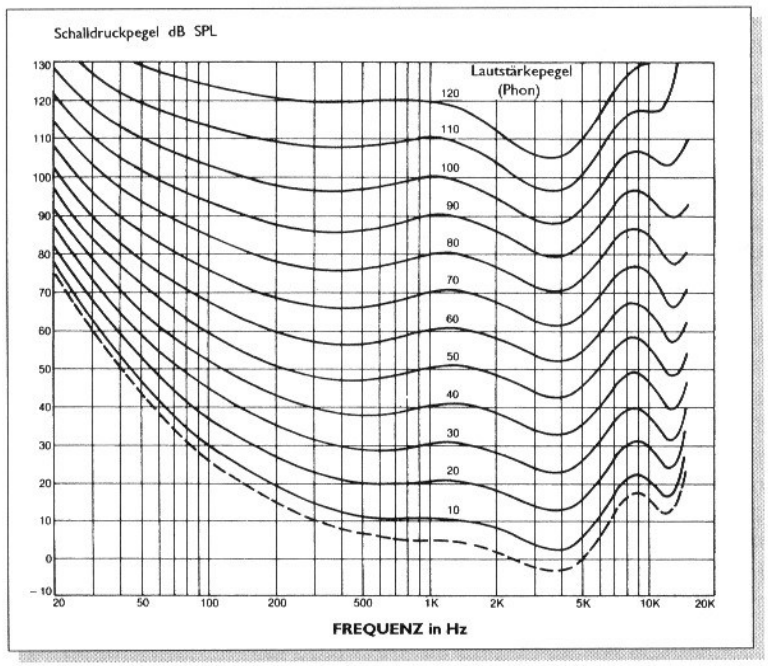


Abb. 1.4: Fletcher/Munson-Diagramm

Figure 11: The Equal Loudness Contours (Fletcher/Munson Curves) | from: *das Tonstudio Handbuch* - Hubert Henle

The **Phon** is the unit of measure for the *perceived loudness*; 1 Phon corresponds to 1 dB at 1 000 Hz in the dB SPL scale; it does not correspond to the dB SPL scale at any other frequency; see also the *Fletcher/Munson diagram* for reference.

The **isophon** (= equal loudness) contours show what sound pressure level (in dB SPL) is required to produce a certain loudness sensation (in Phon) at different frequencies.

Analysis of the isophon contours shows that *our perception of loudness changes with the frequency*: we are generally less sensitive to low (less than 100-200 Hz) and very high frequencies (more than 10 kHz), and most sensitive in the range between 2 and 5 kHz (which is very important for speech recognition).

It also shows that *our perception of loudness changes across its dynamic range*: our hearing "frequency response" is less linear at very low listening levels, and more linear at mid to high listening levels (which is why 85 dB SPL are usually used as reference for mixing and audio production).

Example:
A subjective loudness of 90 Phon requires 90 dB SPL at 1 kHz, only 80 dB SPL at 4 kHz, but as much as 110 dB SPL at 30 Hz!

Even the difference in dB required for the subjective perception of "double as loud" is not constant: at 1 000 Hz, about 10 dB are required, but at very low frequencies variations of just 6 dB may produce the same effect.

The non-linearity of our audio perception is more pronounced at very low listening levels: this is why home and car stereo systems often have a *loudness* switch, which boosts low and high frequencies to compensate for the ear non-linearity when listening at very low levels.

When mixing and mastering, it is important to work at reference listening levels, typically around 80-85 dB SPL: at this level there are less deviations in loudness perception at the different frequencies (we hear “more linear”).

Some engineers prefer to mix and master at lower levels to avoid listening fatigue (around 70-75 dB SPL), switching at the reference level (or even louder listening levels, in the case of certain music styles) only for short periods of time.
Listening from the next room or in noisy environments like a car (non-ideal conditions) can also be very helpful, to check if the most important elements in a mix are still clearly hearable.

1.6 SOUND SPEED AND TRANSMISSION

Figure 12: The Lockheed SR-71 Blackbird, the fastest jet aircraft ever built (in service between 1966 and 1999), that could travel up to Mach 3,3 (= 3,3 times the speed of sound, about 3.600 km/h)

1.6.1 SPEED OF SOUND, WAVELENGTH AND FREQUENCY

The speed of sound depends mainly on the *medium density*. The air density at sea level depends on temperature, therefore the speed of sound in air is mainly *temperature dependent*.

Speed of sound (propagation in Earth atmosphere)

- at sea level, 20° C temperature: 343,8 m/s (approx. 340 m/s)
- at sea level, 0° C temperature: 331,8 m/s

Higher altitudes translate to lower air temperatures and consequently lower speed of sound. For example, Mach 1.0 (1x speed of sound) at sea level is corresponds to 1225 km/h, Mach 1.0 at 9144 m altitude corresponds to 1091 km/h.

Wavelength (in m) = c / f (where “c” is the speed of sound in m/s, and “f” the frequency in Hz); for example: the wavelength of 440 Hz = 340 / 440 = 0.772 m

Frequency (in Hz) = c / L (where “c” is the speed of sound in m/s, and “L” the wavelength in m); for example: the frequency of a sound with 6 m wavelength = 340 / 6 = 56,6 Hz

Quick reference: sound travels 340 m in 1 sec; 34 m in 100 ms; 3,4 m in 10 ms; **34 cm in 1 ms**
The average distance between our ears is about 17 cm, therefore the **maximum delay between right and left ear** for an incident wave is **0,5 ms**

Speed of sound propagation at 0° C through different media:

- Air: 331 m/s
- Water: 1485 m/s
- Copper: 3710 m/s
- Iron: 5100 m/s
- Wood: 3000-4000 m/s
- Glass: 5000 m/s
- Hard rubber: 1500 m/s

1.6.2 SOUND TRANSMISSION IN VACUUM

If there is no elastic medium between the sound source and the receiving end (a microphone, or our ears), like in the vacuum of outer space, no sound transmission can occur. The sound effects featured in most science-fiction movies and series such as Star Trek and Star Wars (lasers, blasters, phasers, railguns, explosions, engine and warp sounds, etc.) are pure fiction. The sound designers know there would be no sound in space, but they are requested to add sound effects nevertheless "to add an emotional layer" to the movie.

While there is no sound transmission in vacuum, structural sound transmission can still occur by direct contact between objects. For example, an astronaut using an electric screwdriver while repairing a satellite can hear the noise from the tool and the screw transmitted directly through his/her suit and body. In *Gravity* you can see and hear exactly that.

While more an exception than the rule, there are a few science fiction movies or series with very realistic rendering of sound in space (or lack thereof). Here some examples:

2001 – A Space Odyssey (1968 – directed by Stanley Kubrick).

Especially the sequences aboard and outside the Discovery spaceship (inside service pods and on spacewalks) on the way to Jupiter feature very realistic sound design.

This movie also features a very realistic rendition of weightlessness as well as apparent gravity inside a centrifuge aboard the Discovery.

Obviously "realistic sound" does not apply to the *Beyond Infinite* psychedelic sequence.

Gravity (2013 – directed by Alfonso Cuarón).

Right from the opening scene, very realistic rendering of structural sound transmission during a spacewalk followed by (spoiler alert) "silent explosions" and destruction of nearby spaceship and objects in Low Earth Orbit.

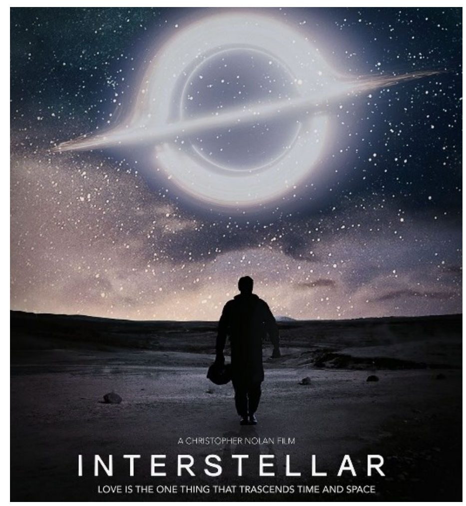
This opening scene proves in my opinion that NOT using sound effects where you are not supposed to hear anything can, in fact, make a scene like this not only more realistic, but also more terrifying.

The Expanse (2015 – 2020 Syfy TV Series).

While "sounds in space" are featured to make combat and action scenes more dramatic (railguns, point defense cannons, missiles, rocket engines and explosions), they are processed in a way that it feels almost like being underwater; this is achieved using a low pass filter and almost no reverb and somehow helps suggesting that … you should not really be hearing this.

Sounds as heard from the perspective of the characters inside their spaceships or spacesuits during spacewalks is generally rendered in a very realistic way

In The Expanse the "laws of physics" are respected more than in the majority of other SF movies or series to date: for example, there is no artificial gravity. Apparent gravity in space can only be achieved by rotation (for example using a large centrifuge, like in Medina Station) or linear accelleration (a spaceship accelerating or decelerating).



Interstellar (2014 – directed by Christopher Nolan).

This movie generally features a very realistic rendition of sound in space – watch for example the silent explosion aboard the Endurance and the following "docking scene" between Lander shuttle and the Endurance.

Interstellar also features the most scientifically accurate visual rendition of a giant black hole with superheated rotating accretion disc of matter, that was realized with custom visual effects based on the actual mathematical equations that describe how the extreme gravity of the black hole not only bends space, but also the path of light rays, making objects behind the black hole visible.

The last part of the movie is questionable, as it drops scientific rigor in favor of metaphysics and philosophical themes.

1.7 SOUND REFLECTION, REFRACTION, ABSORPTION, REVERB AND ECHO

The **soundwave propagation in a free field** (an ideal open space with no boundaries and no reflections) for a sound source that is not moving is *spherical* in shape.

On the planet's surface, in absence of obstruction, the propagation is *hemispherical*. In half or closed spaces, soundwaves encountering an obstacle (like a wall or ceiling) will generally be *reflected*, *refracted* or *absorbed*, depending on the angle of incidence, the material shape, and the material *absorption coefficient*. If soundwaves keep reflecting long enough within two surfaces or a closed space, the reflection pattern can be heard as *reverberation* or *echo* or a combination of both.

1.7.1 REFLECTION, ABSORBTION AND TRANSMISSION OF SOUND

When a soundwave encounters a boundary, part of the wave energy undergoes reflection, part is absorbed by the material (the sound energy is converted into heat) and part undergoes transmission across the boundary. The amount of reflection depends on the difference in density, the rigidity as well as the thickness of the boundary. Generally, soundwaves are reflected "mirror-like" by a smooth hard surface: the *angle of incidence* is equal to the *angle of reflection*.

A hard material like concrete has a very high density and extreme rigidity. Subsequently a sound wave is almost completely reflected (up to 99%) by a concrete boundary, while very little energy is transmitted through.

Sound absorbing materials (like acoustic foam, glasswool, rockwool) cause mechanical losses via conversion of part of the sound energy into heat, resulting in significant acoustic attenuation.

The *absorption coefficients* can vary greatly depending on the material as well as the frequency range. A coefficient of 1.00 equals to 100% absorption and a coefficient of 0.00 equals to 100% reflection. Here are some examples:

Materials	Absorption coefficients by frequency (Hz)				
	125	**250**	**500**	**1,000**	**2,000**
Acoustic tile (ceiling)	.80	.90	.90	.95	.90
Brick	.03	.03	.03	.04	.05
Carpet over concrete	.08	.25	.60	.70	.72
Heavy curtains	.15	.35	.55	.75	.70
Marble	.01	.01	.01	.01	.02
Painted concrete	.10	.05	.06	.07	.09
Plaster on concrete	.10	.10	.08	.05	.05
Plywood on studs	.30	.20	.15	.10	.09
Smooth concrete	.01	.01	.01	.02	.02
Wood floor	.15	.11	.10	.07	.06

For more examples of material absorption coefficients see 4 of Recording Studio Design.
To learn more about sound proofing, sound transmission class and Rw see §1 of the same document.

- https://www.digitalnaturalsound.com/images/stories/fh_mma_courses/pdf/mg_studio_design.pdf

1.7.2 REFRACTION AND DIFFUSION

Soundwaves encountering an obstacle like a pillar or column tend to "bend" around the obstacle, in a similar way as water waves behave when encountering an obstacle. This phenomenon is called ***Refraction***. Objects that could cause refraction (like monitor screens) should not be placed on the sound path between studio speakers and the listening position, as they can impair the perception of spatial stereo as well as cause coloration of the sound due to comb filtering (overlapping of waveforms with short time-delays). Likewise, objects (like a music stand) should not be placed in the path between an instrument being recorded and the microphone.

Diffusion, acoustically speaking, is the spreading of sound energy evenly in a given environment. A perfectly diffusive sound space has the goal of achieving a smooth reverberation with constant reverberation time and pattern from any listening position, dense reflections, and no flutter-echo.

A ***diffusor*** is an acoustic element designed to effectively reduce distinct echoes and reflections while still leaving a live sounding space. Compared to a reflective surface, which will cause most of the energy to be reflected off at an angle equal to the angle of incidence, a diffusor will cause the sound energy to be radiated in many directions, hence leading to a more diffusive acoustic space. The spatial diffusion causes also a spread in time of subsequent reflections. Diffusors can aid sound diffusion, but this is not why they are used in many cases; they are more often used to remove coloration and echoes.

1.7.3 REVERBERATION

Reverberation, in psychoacoustics and acoustics, is a persistence of sound after the sound is produced. A ***reverb*** is created when a sound or signal is reflected causing numerous reflections to build up and then decay as the sound is absorbed by the surfaces of objects in the space – which include the room boundaries (walls, ceiling, floor) as well as furniture, people, and air. This is most noticeable when the sound source stops emitting energy, but the reflections continue, their amplitude decreasing, until zero is reached.

Reverberation is *frequency dependent*: the length of the decay, or reverberation time, is not constant across the complete frequency range. Different acoustic spaces may have very different reverberation time depending on their purpose. For example, an orchestra concert hall may have a reverb time of about 1,5 to 3 seconds, while a speech recording booth might have a rev time of just 0,1 sec. of rev

In comparison to a distinct *echo*, that is detectable at a minimum of 50 to 100 ms after the previous sound, reverberation is the occurrence of reflections that arrive in a sequence of less than approximately 50 ms. As time passes, the amplitude of the reflections gradually reduces to non-noticeable levels. Reverberation is not limited to indoor spaces as it can also occur in outdoor environments such as forests, stadiums, arenas, etc.

Reverberation time **RT_{60}** is usually stated as a decay time and is measured in seconds. Decay time is the time it takes the reverb signal to drop 60 dB below the original sound. If the decay rate (in dB /sec) is constant, it is sufficient to measure a drop of 20 dB and multiply the time by 3, or a drop of 30 dB and multiply the time by 2. These are the so-called RT_{20} and RT_{30} measurement methods.

The concept of reverberation time implicitly supposes that the decay rate of the sound is exponential, so that the sound level diminishes regularly, at a constant rate of so many dB per second. However, some acoustic spaces – depending on the disposition of reflective, dispersive and absorbing surfaces – might behave differently, causing *non-linear* decay curves.

Non-linear reverbs can also be generated electronically, the most well-known example being the **gated reverb** effect that became very popular in the 80s. The gated reverb applied on drums was discovered almost “by accident” in 1979, while Phil Collins (guest drummer for the album) and Hugh Padgham (the producer) were playing around with the compression and gate settings of the Townhouse Studios new SSL 400 console (which for the first time included dynamic controls on each channel). What they came up with was the drum sound you can now hear on the song “Intruder”, from Peter Gabriel’s 3rd studio album:

- https://www.youtube.com/watch?v=xvAmj3k3Imc

In the following years, gated reverb drum sounds appeared on countless other pop and rock productions.

1.7.4 ECHO

In acoustics, ***echo*** is a reflection of sound waves that arrives at the listener with a delay after the direct sound. The delay is directly proportional to the distance of the reflecting surface from the source and the listener.

A true echo is a single reflection of the sound source. Unlike the reverb, the echo reflection is very distinct (not diffuse) and occurs for instance when a highly reflective surface (like the rocky sides of two opposing mountains) bounces the soundwaves back. In the case of two opposite reflection surfaces, the soundwave might bounce back and forth, causing several distinct echo reflections.

Other examples are the echo produced by the bottom of a well, by a building, or by the opposing walls of a long corridor.

1.8 SOUND POWER, SOUND PRESSURE LEVEL AND LOUDNESS

Figure 13: High-end Hi-Fi system featuring Infinity K7 4-way loudspeakers

The **Sound Power** or **Intensity** is a measure of the *sound energy that passes through a given area each second*

Energy per second is measured in **Watt** (1 W = 1 Joule per second)

Intensity is related to the sound pressure amplitude: specifically, *the energy in a wave is proportional to the square of the pressure amplitude.*

Formula: $\mathbf{I = P^2}$, where “I” is the sound power (intensity), and “P” the sound pressure amplitude

Examples:

- double sound pressure = 4 times the power (intensity)
- 1/2 sound pressure level = 1/4 the power (intensity)

The formula to translate *sound power* (intensity) (in W) to dB is: $\mathbf{Li = 10 \times \log_{10} (I1 / I0)}$
where “I0” is the reference intensity and “I1” the intensity being measured

The formula to translate *sound pressure level* (in Pa) to dB is: $\mathbf{Lp = 20 \times \log_{10} (U1 / U\ 0)}$
where “U0” is the reference level and “U1” the level being measured

- The **sound power** (intensity) *doubles every 3,01 dB* [10 x log10 (2 / 1) = 3,01 dB]
- The **sound pressure level** (in Pa) *doubles every 6,02 dB* [20 x log10 (2 / 1) = 6,02 dB]
- The **perceived loudness** (in Phon) *doubles about every 10 dB*

Remember that perceived loudness is frequency dependent and varies quite a lot between individuals (see Equal-Loudness Contours!)

Examples:

- 10 times the power (in W) = +10 dB SPL = 3,162 times the SPL in Pa, but is just perceived as *double as loud*
- 100 times the power (in W) = +20 dB SPL = 10 times the SPL in Pa, but is just perceived as *4 times as loud*
- 1000 times the power (in W) = +30 dB SPL = 31,62 times the SPL in Pa, but is just perceived as *8 times as loud*

1.9 SOUND POWER, SOUND PRESSURE LEVEL AND DISTANCE

In a free field (an ideal open space with no boundaries and no reflections), *the sound pressure level is inversely proportional to the distance from the sound source,*

Formula: **p = 1/d** (p = sound pressure, d= distance ratio)

Examples:

▪ double distance	= 1/2 sound pressure level	-6,02 dB
▪ 4 times the distance	= 1/4 sound pressure level	-12,04 dB
▪ 1/2 the distance	= 1/0.5 = double sound pressure level	+6,02 dB

In a free field, *the sound power (intensity) is inversely proportional to the square of the distance from the sound source*

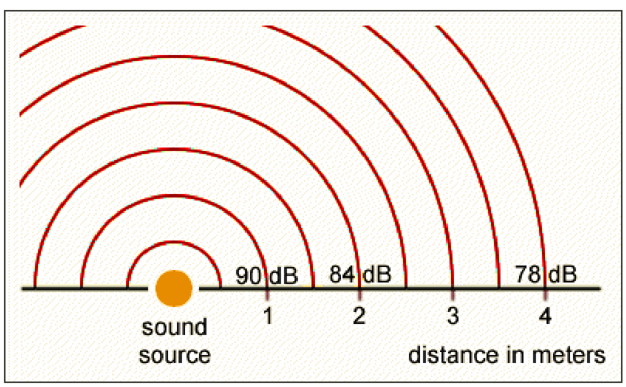
Formula**: i = 1/d²** (i = intensity, d = distance ratio)

Think it like this: the sound waves carry energy; doubling the distance, this energy is spread on an area that is 4 times as large.

Examples:

▪ double distance	= 1 / (2 x 2)	= 1/4 power in W	-6,02 dB
▪ 4 times the distance	= 1 / (4 x 4)	= 1/16 power in W	-12,04 dB
▪ 1/2 the distance	= 1 / (0.5 x 0.5)	= 4 times the power in W	+6,02 dB

When dealing with **sound amplification systems** (amplifiers and loudspeakers), it is important to know how to calculate the power requirements, to get the desired sound pressure level at the listening position. Of course, there are also other aspects to consider, such as room size, absorption coefficients, reverberation time, etc.



Example:

A loudspeaker has an *efficiency* of 90 dB SPL /W at 1 m distance; what sound pressure level will it produce at 32 m distance, in a free field?

p = 1/d = 1/32; to calculate the ratio in dB: 20 x log10 (1/32) = 30,10 dB

90 - 30,10 = 59,90 dB. Therefore, at 32 m distance this loudspeaker just produces about 60 dB SPL

How much more power is required to still produce 90 dB SPL at 32 m distance?

For each additional 10 dB SPL we need 10 times the power (in W); for +30 dB SPL we need 1 000 times the power!

If this loudspeaker needed 1 W to produce 90 dB SPL at 1 m, we are going to need 1 000 W power to produce 120 dB SPL at 1 m, corresponding to about 90 dB SPL at 32 m distance.

2 DIGITAL AUDIO RECORDING AND PROCESSING

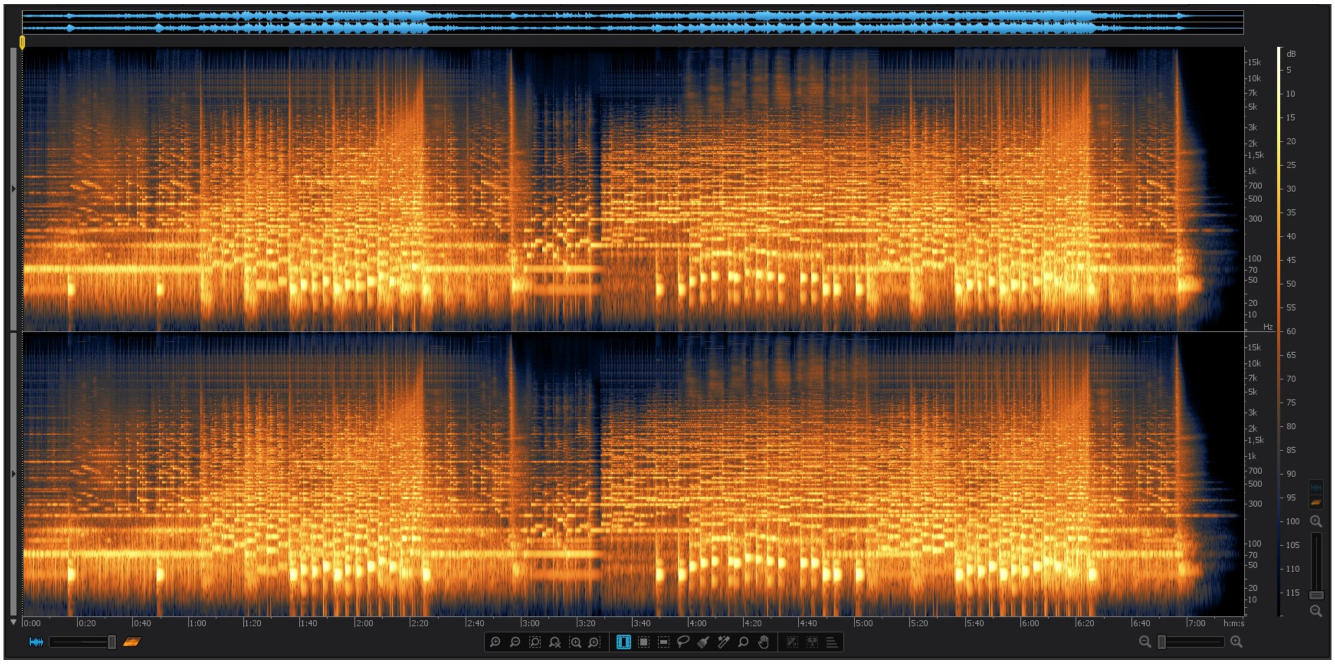


Figure 14: Spectral view of a stereo music track in iZotope Studio RX
Amplitude plotted from black (silence), through blue (low level) and orange (medium level), up to yellow (high level)

2.1 AUDIO SAMPLING

2.1.1 BASICS

Audio Sampling: the information of a band-limited analog audio signal is reduced from *continuous-time* (non-quantized state) to *discrete time*, (quantized state), resulting in a *finite amount of digital* (= numerical) *information*.

The component of an audio interface that samples and records the analog signal input is called **ADC**: analog-to-digital converter.

This **digital information** can be *processed in real time* (by DSP systems or general-purpose computer CPUs) or *stored in digital form* on a *magnetic* medium (DAT, ADAT tape, Hard Disk), *optical* medium (CD/DVD-R) or *solid-state* memory (Solid State Drive, RAM) for later processing.

One of the main advantages of audio sampling is that digital audio can be easily stored and further processed without the loss of quality or additional noise and distortion typically associated with analog audio processing.

During **playback,** the process is inverted: the digital information is used to reconstruct the wave function of the original analog audio signal; provided sampling rate and quantization are sufficiently accurate (for example, for CD quality 16-bit per sample and 44,1 kHz sampling rate), the reconstructed analog audio signal will be virtually identical to the original.

The component of an audio interface that converts the digital information back into an analog audio signal at the output is called **DAC**: digital-to-analog converter.

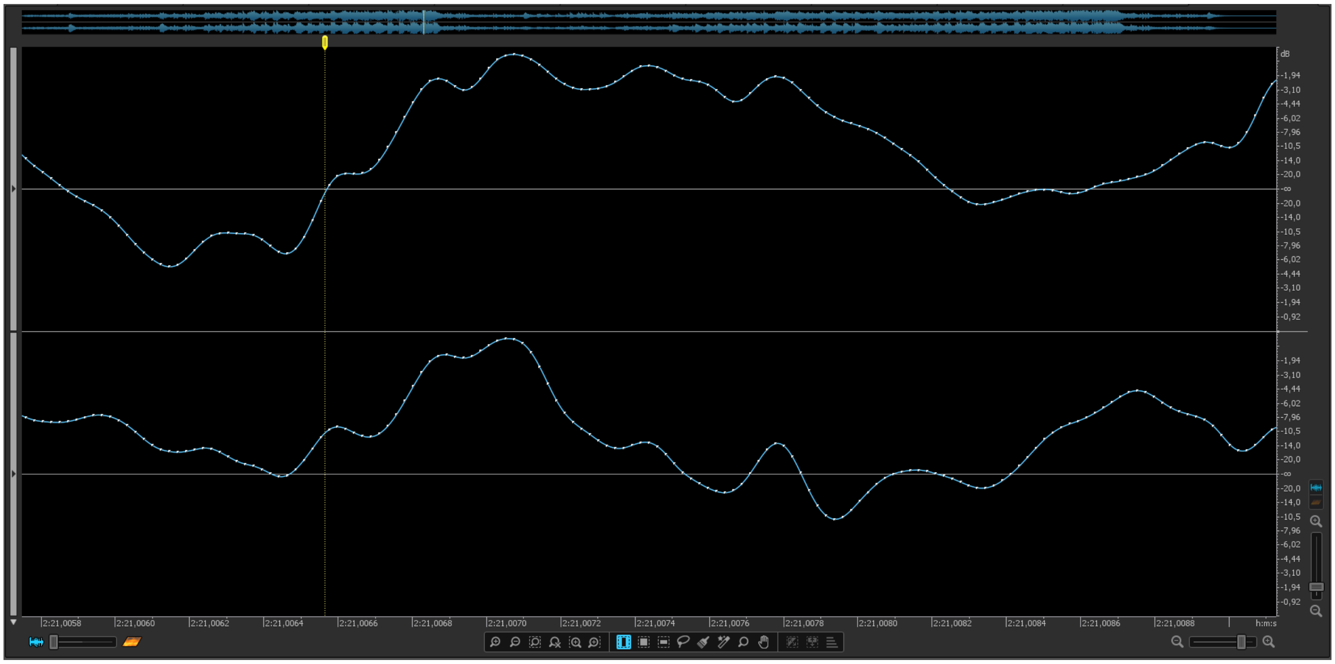


Figure 15: Waveform visualization of a stereo audio track in iZotope Studio RX,
showing individual samples (white dots) and the reconstructed wave function (in blue)

2.1.2 MAIN PARAMETERS

The quality of a recorded digital audio signal depends mainly on two factors:

The **Sampling Frequency** or **Sampling Rate** (in Hz) defines how many samples (discrete amplitude measurements) per sec. of the analog signal are being taken (*temporal accuracy*). The **Frequency Bandwidth** depends on the sampling rate and is always less than ½ the sampling rate (see Nyquist Theorem under).

The **Quantization Precision** (bit-depth) defines the accuracy of the amplitude measurement of each sample (*amplitude accuracy*). The available **Dynamic Range** (difference between the loudest and the quietest possible signals) and the **S/N Ratio** (signal-to-noise ratio) depend on the Quantization Precision.

2.1.3 NYQUIST-SHANNON SAMPLING THEOREM

The **Nyquist-Shannon Sampling Theorem** (named after Harry Nyquist and Claude Shannon) establishes a sufficient condition for a sample rate that permits a discrete sequence of samples to capture all the information from a continuous-time signal of finite bandwidth.

The sampling theorem introduces the concept of a sample rate that is sufficient for perfect fidelity for the class of functions that are band-limited to a given bandwidth, such that no actual information is lost in the sampling process. It expresses the sufficient sample rate in terms of the bandwidth for the class of functions. The theorem also leads to a formula for perfectly reconstructing the original continuous-time function from the samples.

The sampling rate (in Hz) must be at least 2 times the desired bandwidth (in Hz); in other words: to accurately sample and reproduce a given frequency it is necessary to have at least one sample per positive and one sample per negative phase of the wave cycle.

- **Nyquist frequency** = ½ the sampling rate of a discrete signal processing system
- **Nyquist Rate** (2 B) = twice the bandwidth of a bandlimited function (or in this case audio signal) = Minimum Sampling Rate
- **Nyquist Criterion**: $f_s > 2B$ = the Sampling Rate must be more than double the required signal bandwidth.

When the Nyquist Criterion is not met, a condition called **aliasing** occurs, in form of signal components not present in the original signal, and not related to it harmonically. This should not be confused with **harmonic distortion**, which can, in moderate amounts, be considered a pleasant side-effect of analog processing.

Aliasing happens when signals above the Nyquist frequency enter the ADC without proper "anti-aliasing filtering" and get "mirrored" around the Nyquist frequency itself, appearing back into the hearable freq. range.

Example:

in a 44 100 sampling rate system, the Nyquist frequency is 22 050 Hz; without filtering, a signal of 30 000 Hz would be mirrored at 22 050 - (30 000 - 22 050) = 14 100 Hz

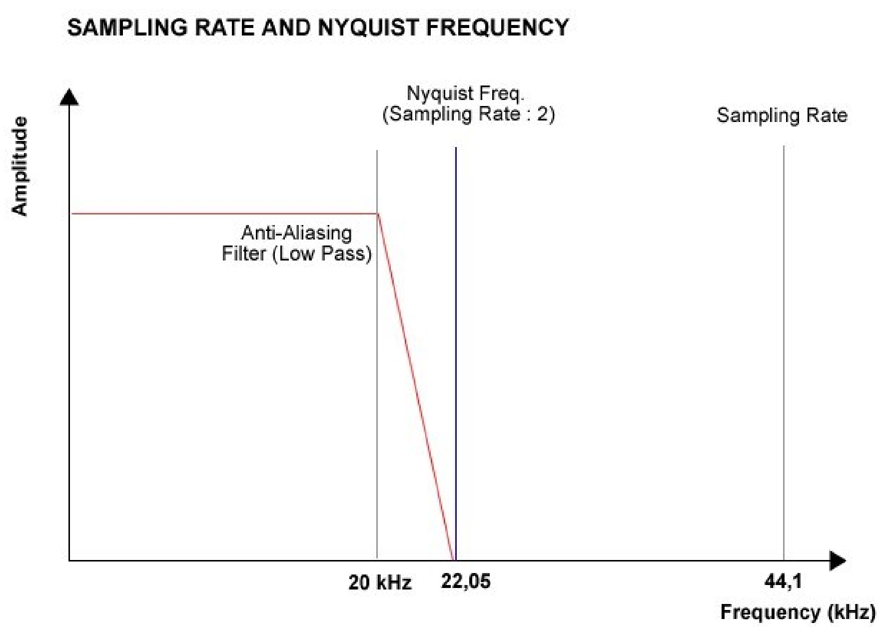


Figure 16: Sampling Rate and Nyquist Frequency

An **anti-aliasing filter** is a low pass filter with a razor-sharp slope; as practical filters cannot be manufactured with “infinite slope”, the sampling rate must be a bit higher than 2B (twice the bandwidth) to effectively remove all frequencies above the Nyquist, without affecting the desired bandwidth. This is the Nyquist Criterion: $f_s > 2B$

Example:

In a typical *CD quality* recording (44 100 Hz sampling rate) the anti-aliasing filter must not affect signals under 20 000 Hz, but must filter all signals above 22 050 Hz to avoid aliasing, leaving just 2 050 Hz "transition band" to go from "full pass" to "full cut".

A similar (inverse) process occurs in the DAC: after DA conversion, an analog low pass filter removes all undesired artifacts generated by the sampling process over 20 000 Hz

Oversampling is the process of sampling a signal at a sampling frequency significantly higher than the Nyquist Rate. Modern ADC and DAC often work at 64 or 128 x oversampling (for example 128 x 44 100 = 5 644 800 Hz = 5,644 MHz). Oversampling can make it easier to realize analog anti-aliasing filters. Without oversampling it is very difficult to implement high-quality filters with the sharp cutoff necessary to maximize the use of the available bandwidth, without exceeding the Nyquist Frequency. By increasing the bandwidth of the sampling system, design constraints for the *analog anti-aliasing filter* may be relaxed.

After conversion, a *digital low pass filter* is used (which is easier and cheaper to implement than a comparable analog anti-aliasing filter and can also be very accurate); finally, the signal is *downsampled* to the desired rate (for example 44 100 Hz) before further processing or storage on a medium. A similar inverted process occurs at sample playback.

2.2 DIGITAL AUDIO – SIGNAL BANDWIDTH AND DYNAMIC RANGE

2.2.1 SAMPLING RATE AND AUDIO BANDWIDTH

The available **audio bandwidth** (the frequency range of audio, that can be digitally generated or digitally recorded) depends primarily on the **sampling rate**, but also on the design of the ADC and DAC units, which might implement different types of anti-aliasing filters, resulting in different response in the frequency range beyond 20 kHz.

According to the Nyquist Theorem, the sampling rate of a digital audio recording must be *at least* double than the required bandwidth.

Example: A digital recording with *44,1 kHz sampling rate* allows for a *5 – 20 500 Hz bandwidth*

Here are some standard sampling rate formats, the Nyquist frequency (1/s SR) and the resulting available bandwidth to be expected on a modern audio interface (here RME Fireface UCX):

Sampling Rate (Hz)	Remarks	Nyquist Frequency (Hz)	Available Bandwidth (Hz) (approx.)
44 100	CD-Audio Standard	22 050	5 - 20 500
48 000	Video and Cinema Standard	24 000	5 - 22 500
88 200	HR Audio (less common)	44 100	5 - 41 000
96 000	Used for HR-Audio Productions or advanced Sound Design	48 000	5 - 45 000
176 400	HR Audio (less common)	88 200	5 - 82 000
192 000	Used for HR-Audio Productions or advanced Sound Design	96 000	5 - 90 000

2.2.2 QUANTIZATION PRECISION AND DYNAMIC RANGE

The **dynamic range** depends on the **quantization precision** (bit-depth). A higher bit-depth results in a finer resolution in the sampling process (as each sample can be measured with higher precision), which translates in improved S/N Ratio and available dynamic range.

Formula: $\text{SQNR} = 20 \log_{10} (2^{Q}) \approx 6{,}02 \cdot Q$ dB

bit-depth	Max Positive Value	Max Negative Value		Available Dynamic Range
8	+ 127	- 128	8 x 6,02	**48,16 dB**
16	+ 32 767	- 32 768	16 x 6,02	**96,33 dB**
24	+ 8 388 607	- 8 388 608	24 x 6,02	**144,49 dB**

Thumb Rule: you get roughly 6 dB of additional dynamic range for each extra bit of precision.

The 24-bit signal theoretical dynamic range (144,49 dB) can only be achieved in the digital domain; the effective dynamic range of audio recordings depends on the quality of the ADC units, that includes analog components.

Good quality 24-bit ADC/DAC units currently manage between 110 and 128 dB dynamic range.

2.2.3 QUANTIZATION NOISE AND DITHERING

At the lowest boundaries of the available dynamic range, the rounding errors caused by insufficient accuracy in the measurement of each sample become audible in form of digital noise. This can become a problem in very quiet passages of a classical recording, or in the fade-out part of a song.

To reduce the noise introduced by quantization errors, a process called ***dithering*** can be used: this adds a small amount of random-like noise (called *dither*), that masks the quantization errors. To further reduce the impact of the noise added by dithering, additional ***noise-shaping*** can be also applied, which basically applies an EQ curve to the dithering noise, moving most of the noise components outside the frequency ranges we are most sensitive to. This is a typical dithering + noise shaping EQ curve:

Figure 17: iZotope Studio RX - Dithering with Noise Shaping Curve (Resampling)

This process (dithering with or without noise-shaping) is usually applied at the very end of the mastering process, when the high-resolution original master (for example 96 kHz 24-bit) is exported at a lower sampling rate and bit precision (for example 44,1 kHz 16-bit for CD Audio). It should be applied only *once*, as repeating the process can raise the level of dithering noise to the point it becomes audible.

2.2.4 AMPLITUDE MEASUREMENT – DECIBEL FS

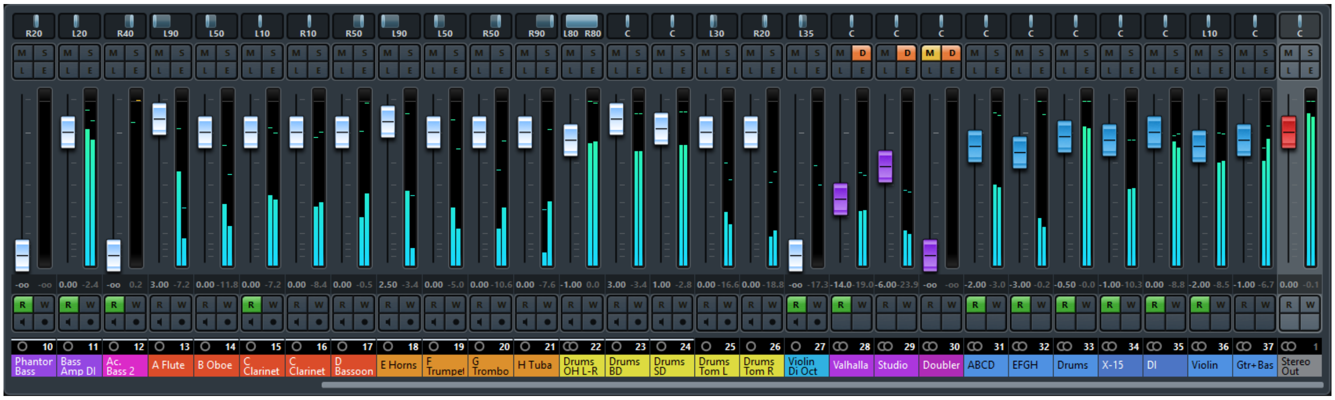


Figure 18: Digital mixer in Cubase 8.5 showing amplitude peak values in dBFS

The amplitude of digital audio signals is measured in **dBFS** (see § 1.3.4 for details).

2.2.5 CLIPPING

If the amplitude of a digital audio signal exceeds "0 dBFS", **clipping** occurs. Clipping is the abrupt truncation of the waveform, causing distortion in form of undesired and harsh harmonics. Clipping does not occur when **32-bit float** processing or file formats are used. The audio engine of Digital Audio Workstations such as *Cubase*, *Nuendo* and *Logic* work internally in 32-bit float mode. A mixdown or master can also be exported to a 32-bit float WAV file format.

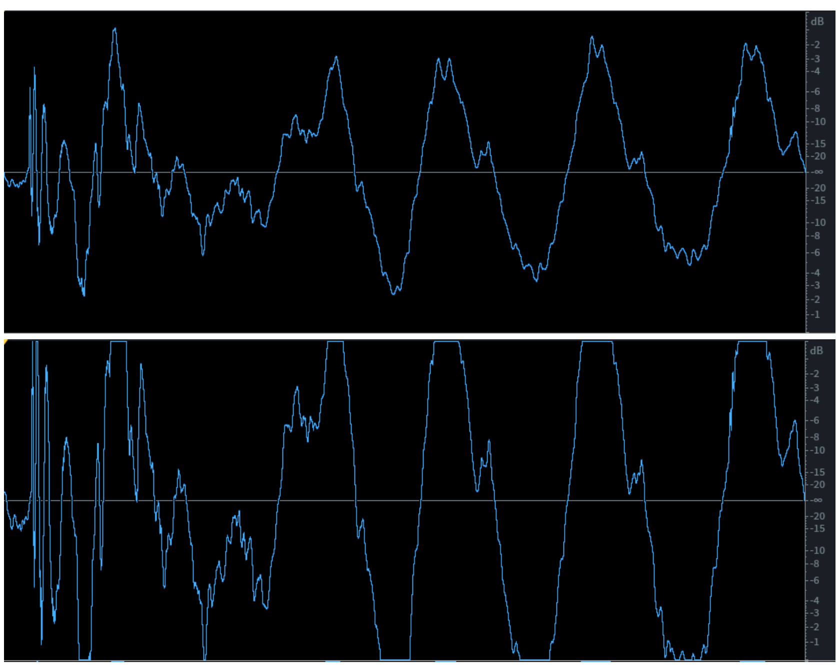


Figure 19: Top: properly encoded audio signal. Bottom: the same audio signal, boosted + 6 dB and severely clipped

Even in systems supporting **32-bit float resolution** (where theoretically the full 24-bit resolution of the signal is maintained throughout the signal path), the signal sent to the hardware output should never exceed 0 dBFS, or clipping occurs in the DA converter, causing distortion.

To avoid clipping of the DA stage in a HDR systems supporting internally *32-bit float resolution*, it is usually enough to reduce the level of the master output fader, unless clipping occurs before in some plugin that does not support the 32-bit float format.

Good quality 24-bit ADC/DAC units currently manage between 110 and 128 dB dynamic range.

2.3 HIGH RESOLUTION AUDIO RECORDING AND PROCESSING

2.3.1 QUANTIZATION PRECISION AND DYNAMIC RANGE

Using 24-bit quantization instead of 16-bit provides significant improvements in dynamic range and noise floor, making it the de facto standard for professional digital recording and production.
Theoretically, 16-bit audio offers about 96 dB of dynamic range, while 24-bit extends this to approximately 144 dB—an increase of 48 dB (since each additional bit adds ~6 dB). This extra headroom reduces quantization noise, allowing for cleaner recordings, especially of quiet passages, and more flexibility during mixing without introducing audible artifacts.

2.3.2 SAMPLE RATE: THE DEBATE

The benefits of higher sample rates (e.g., 96 kHz or 192 kHz) over standard rates like 44.1 kHz or 48 kHz remain debated for general listening and basic production.
Some engineers argue that higher rates improve transient accuracy, but the relevant high-frequency components lie in the ultrasonic range (>20 kHz), beyond human hearing.
Higher sample rates do provide advantages during signal processing, particularly for EQ in regions near or above 20 kHz, as they allow filter responses to maintain symmetry without abrupt cutoff artifacts. However, most modern digital equalizers handle this internally through oversampling, mitigating limitations in band-limited signals.

2.3.3 KEY ADVANTAGES IN ADVANCED SOUND DESIGN

One area where higher sample rates offer clear, practical benefits is advanced sound design, especially when creating exaggerated, larger-than-life effects. Sound designers often pitch-shift source material downward to make elements (e.g., vehicles, creatures, or impacts) sound bigger and more imposing. For instance, simulating a tyrannosaur roar might combine a lion's roar with an elephant's growl, then transpose the result down in pitch.

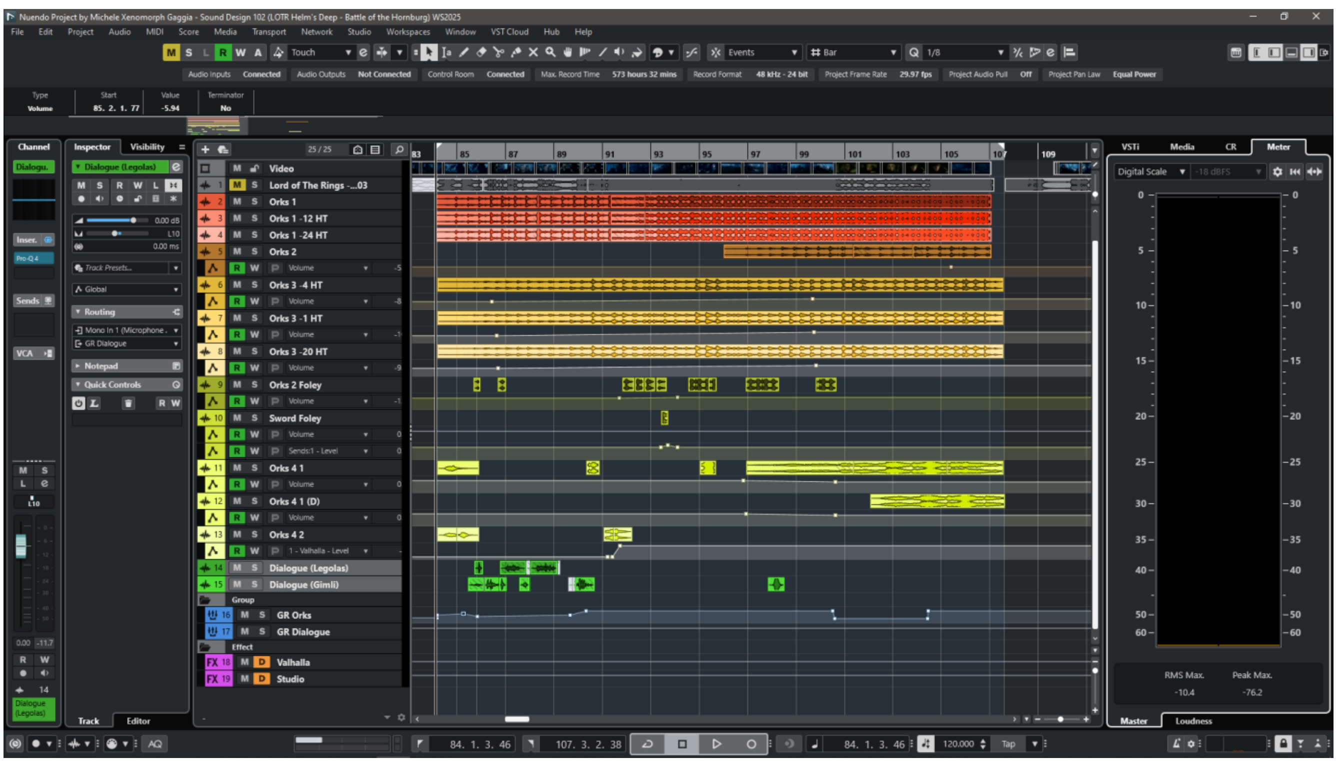
- At a standard 48 kHz sample rate (available bandwidth: 22,5 kHz), pitching down one octave halves the effective bandwidth to ~11 kHz, resulting in a duller, muffled sound.
- Recording at 96 kHz (available bandwidth: 45 kHz) preserves full audible bandwidth (22,5 kHz) even after a one-octave downward shift, as previously ultrasonic components shift into the hearable range, enriching the spectrum.
- At 192 kHz (available bandwidth: 90 kHz), up to two octaves of downward transposition can maintain high-frequency content without significant loss.

Techniques like this are widely used in film and game sound design.

This is how we recreated the sound effects of the *Lord Of The Rings – Battle of the Hornburg* opening sequence in our class, using common objects (pens, tablets, keys, metal bottles) and our voices (grunts, screams).

The raw recording (performed at 192 kHz) was pitched down 1 and 2 octaves to simulate heavy spears, shields, metal armor and weapons. Human growls and screams were pitched down 1 octave to simulate the terrifying Uruk-Hai orks. The very high sampling rate made it possible to apply extreme pitch shifting, without significant loss of audio bandwidth.

A few additional effects were used: EQ (mainly bass-boost), compression, and reverb; on the master chain: master compressor, multiband limiter and true peak limiter. Some parameters were automated: level, panorama, and reverb send (varying the reverb amount, to simulate close-up and far-away sound events, like when hearing the far away horde from inside the Hornburg cave).



2.4 DAW – PROJECT AUDIO AND VIDEO SETTINGS

2.4.1 DAW – PROJECT SETUP

This is the **Project Setup** dialog box in Cubase/Nuendo. It includes settings for project duration, frame rate, time displays, record file format, and other project settings (pan law, max volume, etc.).

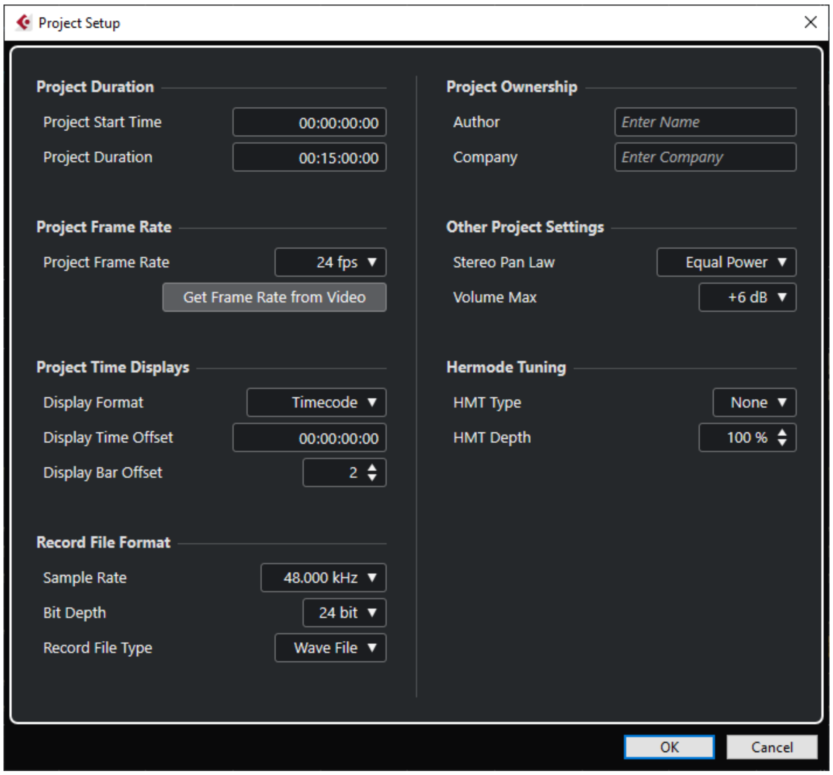


Figure 20: Cubase Project Setup Settings - Audio and Video Formats

2.4.2 DAW – PROJECT AUDIO SETTINGS

When working in a DAW, it is very important to set the correct project audio format properties. There are usually options for **sample rate**, **bit-depth**, **file format** (WAV, AIFF) and additional options for time display and project duration.

Typical recording formats and project settings are:

- 24-bit 44.1 kHz — for a CD production, the final master will be dithered to 16-bit
- 24-bit 48 kHz — for most audiovisual productions

  24-bit 88,2 / 96 kHz — double sample rate, high resolution audio
- 24-bit 176,4 / 192 kHz — quad sample rate, high resolution audio

WAV (Waveform Audio File Format) is the standard audio file format on Windows systems, while AIFF (Audio Interchange File Format) is standard on MacOS systems. Both can be used on both OS platforms.

2.4.3 DAW – RECORDING LEVEL AND HEADROOM

During a recording sound check, the input gain for each channel/instrument should be adjusted so that even the loudest passages can be recorded leaving a few extra dB of *headroom* (the difference in dB from the loudest peak being recorded and clipping at 0 dBFS). It is common practice to leave about 6 dB headroom during the sound check (meaning the loudest peak of the recorded signal is at -6 dBFS): this way louder peaks that might occur during the actual performance (when musicians might play with a higher energy than during sound check) can still be recorded without the risk of clipping.

However, in classical recordings it is likely that the average recording level will be around -18 dBFS. *Pianissmo* passages in a classical recording might be as quiet as -36 dBFS.

As discussed earlier (chapter 2.2), each bit precision translates to about 6 dB dynamic range. This means that when we are recording a -36 dBFS (RMS) signal in 16-bit audio file format we are left with a signal encoded with just 10-bit precision [16 – (36 : 6) = 10], that corresponds to about 60 dB of residual dynamic range. In this context, quantization noise and distortion become very audible and the quality of the recording is just comparable to that of a 1980s microcassette!

If you are recording the same -36 dBFS signal in 24-bit audio file format, you are still encoding this signal with 18-bit precision and 108 dB dynamic range [24 – (36 : 6) = 18].

Therefore, your recording file format should always be set at 24-bit depth.

> Remember that 24-bit is not "just 50% better" than 16-bit: it is 256 times more accurate, as each extra bit doubles the quantizing precision and adds 6 dB of dynamic range.

2.4.4 DAW – PROJECT VIDEO SETTINGS

When setting up a project that includes a video (for example for a trailer, commercial, etc.) it is important to set your DAW project to the correct **frame rate**. Typical frame rate settings are:

- Film, High-Definition Video, BluRay — 24 FPS
  48 FPS (double rate, i.e. The Hobbit)
- PAL (European Video Standard) — 25 FPS
  25P = 25 progressive frames /s
  50i = 50 interlaced fields /s
- NTSC (US Video Standard) — 29,97 FPS
  29,97P = 29,97 progressive frames /s
  59,94i = 59,94 interlaced fields /s
- Computer Video, Smartphones — 30 FPS
  60 FPS

> What is the difference between frames and fields?
> - 1 frame = full screen content (odd plus even horizontal lines)
> - 1 field = only the odd, or only the even horizontal lines = half screen content

2.4.5 DAW – OTHER PROJECT SETTINGS

In Cubase/Nuendo, these additional project settings are available:

- Timecode settings — project duration and timecode offset
- Time display format — bars and beats, seconds, timecode, samples
- Stereo pan law — Equal Power, 0 dB, -3 dB, -4,5 dB, -6 dB
- Volume fader max volume — +6 dB, +12 dB

2.5 DAW – SYSTEM LATENCY

The total **system latency** of your DAW (from audio input through processing and audio output) is the sum of AD conversion, input audio buffer latency, internal processing latency (some plugins like a look-ahead limiter require this), output audio buffer latency and DA conversion.

The typical **conversion time** @ 44,1 kHz sampling frequency required by a modern ADC or DAC with 128x oversampling is about 1 ms. This is an example of the total system latency using an RME Fireface UFX interface at different buffer settings:

Buffer Settings (samples @ sampling rate)	1024 @ 48 kHz 2048 @ 96 kHz 4096 @ 192 kHz	512 @ 48 kHz 1024 @ 96 kHz 2048 @ 192 kHz	256 @ 48 kHz 512 @ 96 kHz 1024 @ 192 kHz	128 @ 48 kHz 256 @ 96 kHz 512 @ 192 kHz	64 @ 48 kHz 128 @ 96 kHz 256 @ 192 kHz	32 @ 48 kHz 96 @ 96 kHz
AD conversion	1 ms	1 ms	1 ms	1 ms	1 ms	1 ms
Input Audio Buffer (rounded)	21–22 ms	10–11 ms	5–6 ms	2,8–3,2 ms	1,6–1,9 ms	1,2–1,6 ms
Internal Processing	variable	variable	variable	variable	variable	variable
Output Audio Buffer (rounded)	22–23 ms	11–12 ms	6–7 ms	3,5–4 ms	2,2–2,8 ms	1,8–2,4 ms
DA conversion	1 ms	1 ms	1 ms	1 ms	1 ms	1 ms
Total System Latency	**45–47 ms**	**23–25 ms**	**13–15 ms**	**8,3–9,2 ms**	**5,8–5,7 ms**	**5–6 ms**

Please note that low buffer settings can only be used as long the CPU audio load is very low (25% or less).

With increasing CPU load (due to VSTi instruments or audio plugins), drop-outs will happen at the lowest buffer settings, therefore switching to higher buffer settings (with resulting higher latency) is usually recommended.

During the recording session it is recommended to avoid using high-quality audio plugins to keep the CPU load as low as possible. If the CPU load is moderate, most audio interfaces can provide drop-out free operation with the audio buffers set at 128 to 256 samples @ 48 kHz, which corresponds to a I/O latency of 8–15 ms. Modern high-performance computers might even manage 64 samples @ 48 kHz.

Important: Cubase and Nuendo include a function called **Constrained Delay Compensation** (CDC). This can reduce the latency of the whole project to almost zero, but it will also disable all the effects that require plugin delay compensation (like for example a look-ahead limiter).

Later, in the mixing phase, the buffer size and resulting latency can be increased without side-effects (no real-time monitoring is required). With a larger buffer size (for example 1024 samples), it is possible to reach a much higher CPU load, using whatever high-quality audio plugins or VSTi that might be required for the production, without risking audio drop-outs.

Latencies up to 6 ms are excellent and up to 13 ms are still acceptable when playing software instruments (VSTi). Please note that in this case, only the output buffer latency and DA conversion times are relevant.
I/O latencies of 13–15 ms are however on the upper limit when offering monitoring to musicians or singers. Rhythmical instruments (like drums or bass) might require even shorter latencies for an accurate control of the groove feeling and performance (5–8 ms).

2.5.1 DAW – DIRECT MONITORING

Some audio interfaces (including the RME Fireface UFX) offer *zero latency monitoring* (the option is called **Direct Monitoring** in Cubase / Nuendo audio settings). When this mode is activated, the signal at the hardware input (what is being recorded) is rerouted directly to the hardware output (listening/monitoring) without being processed by the audio driver. This way it is possible to monitor a signal with virtually no latency (just the AD/DA latency, which is less than 2 ms), completely bypassing the internal DAW processing; however no DAW plugin can be used in this mode.

This is an excellent choice if you provide real-time monitoring to a drummer or other instruments that require very tight timing.

2.5.2 DAW – STORAGE USAGE

To calculate HD or SSD storage space usage recording at different resolutions (stereo, multitrack, 16 and 24 bit):
Storage space in Bytes/min = [Bytes of quantization per channel] x [n. of audio channels] x [sampling freq. in Hz] x 60 sec

- 16-bit, stereo, 44.1 kHz (*CD quality*): 2 Bytes x 2 Ch. x 44100 x 60 = 10 584 000 Bytes/min = 10,09 MB/min
- 24-bit, 8 track, 96 kHz: 3 Bytes x 8 Ch. x 96000 x 60 = 138 240 000 Bytes/min = 131,83 MB/min

Remember: for HD manufactures, 10 584 000 equals to 10,58 MB, but for your Operating System this is just 10,093 MB

1 KB = 1024 (and not 1000) Bytes; 1 MB = 1024 KB; 1 GB = 1024 MB; 1 TB = 1024 GB; etc.

Disk manufactures want to advertise larger disk capacities, so they sell a hard disk as a 120 GB capable device, while the effective capacity as seen by the OS and programs is only 111,7 GB.

RECOMMENDED LITERATURE

- EVEREST, F. Alton: *Master Handbook of Acoustics* – McGraw-Hill Education TAB;
  6th edition, 8 Dec 2014 (ASIN: B00O2A7GYW)
- HENLE, Hubert: *das Tonstudio Handbuch* – GC Carstensen 2001 (ISBN 3-910098-19-3)
- HÖMBERG, Martin: *Recording Basics* – PPV Medien 2002 (ISBN 3-932275-21-7)

WEBSITE

- http://www.digitalnaturalsound.com/fh-multimediaart/audio.html
- www.digitalnaturalsound.com or www.dns-studios.com > FH | MultiMediaArt > Audio