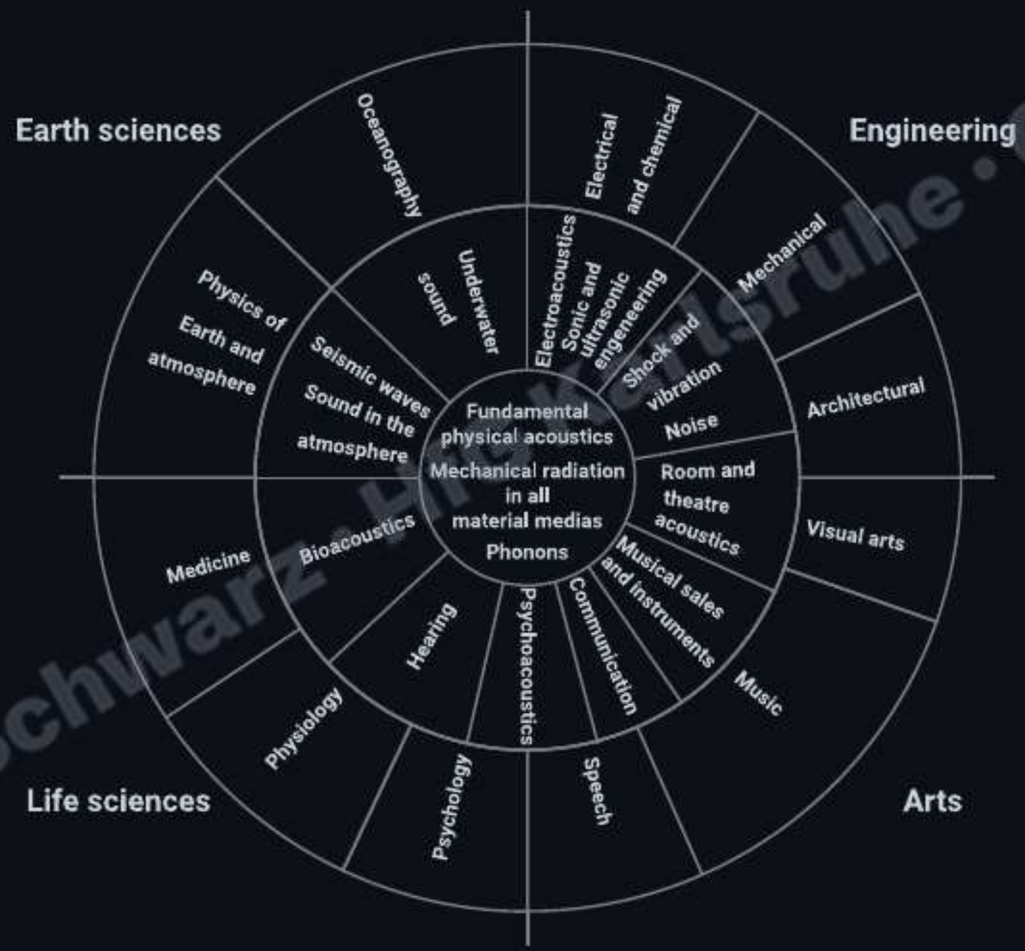
VIBRATION AND SOUND

The physical and perceptual nature of sound

A complex relationship between:

1. Physical disturbance in a medium and transfer of energy
2. Psycho-physical perception and sensory experience of the physical stimuli

Wheel of acoustics



(Beyer 300)

Sound (physics)

Etymology: Derived from Latin *sonare* (to sound)

Pressure or density variations in an elastic medium (e.g., air):

- particle displacement (e.g., air molecules)
- particle velocity

Elasticity and inertia

- **Elasticity:** The property of a material or medium that enables it to return to its original shape or equilibrium after being deformed, once the applied force is removed.
- **Inertia:** An object in motion remains in motion, and an object at rest stays at rest, unless acted upon by an external force (Newton's First Law of Motion).

Vibration and sound

Mechanical vibration is capable of producing sound, e.g.:

- strings (chordophones)
- membranes (membranophones)
- plates (struck idiophones)

Oscillation

A process that returns to the same state after repeating periods:

- periodic vibration or cyclical process
- number of occurrences of a repeating event per second
- measured in hertz (Hz)

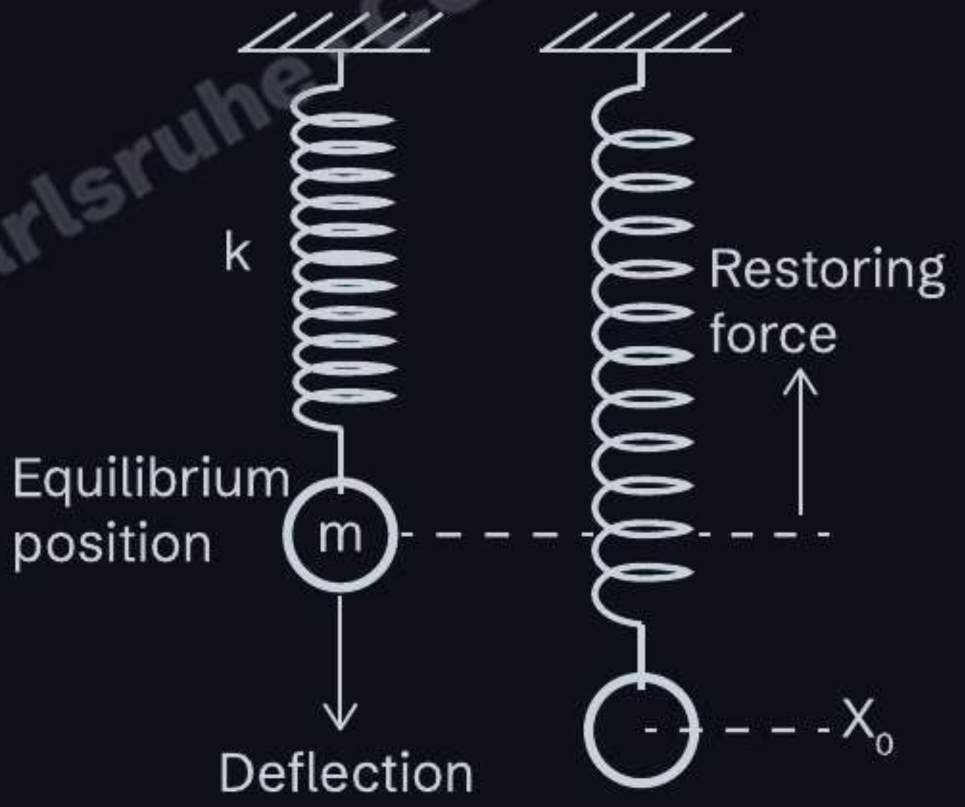
Case study: spring-mass system

Vertical back and forth movement of a mass on a spring:

- Newton's Second Law: $F = m\frac{dv}{dt} = ma$
- Hooke's Law (restoring force): $F = -kx$

constant:

- k spring constant
- m mass



Simple harmonic motion of a spring-mass system

- Newton's Second Law: $F = ma$
- Hooke's Law (restoring force): $F = -kx$

$$ma = -kx$$

variables:

- a acceleration
- x displacement from equilibrium

Acceleration

$a = \frac{dv}{dt}$ first derivative of the velocity with respect to time

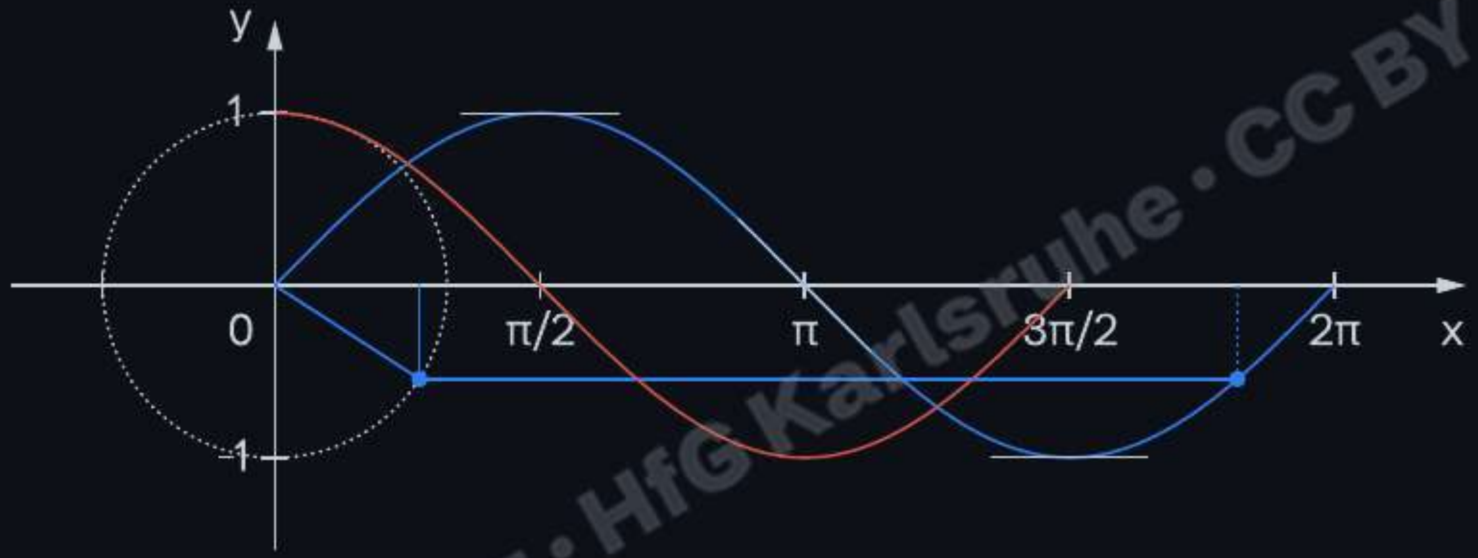
or

$a = \frac{d^2x}{dt^2}$ second derivative of the position with respect to time

$$\frac{d^2x}{dt^2} = -\frac{k}{m}x$$

(second derivative of the function is the function)

Sine and cosine



The gradient of the tangent equals the derivative of the function at the point where the curve and tangent line meet.

view in graphing calculator

Solving the differential equation

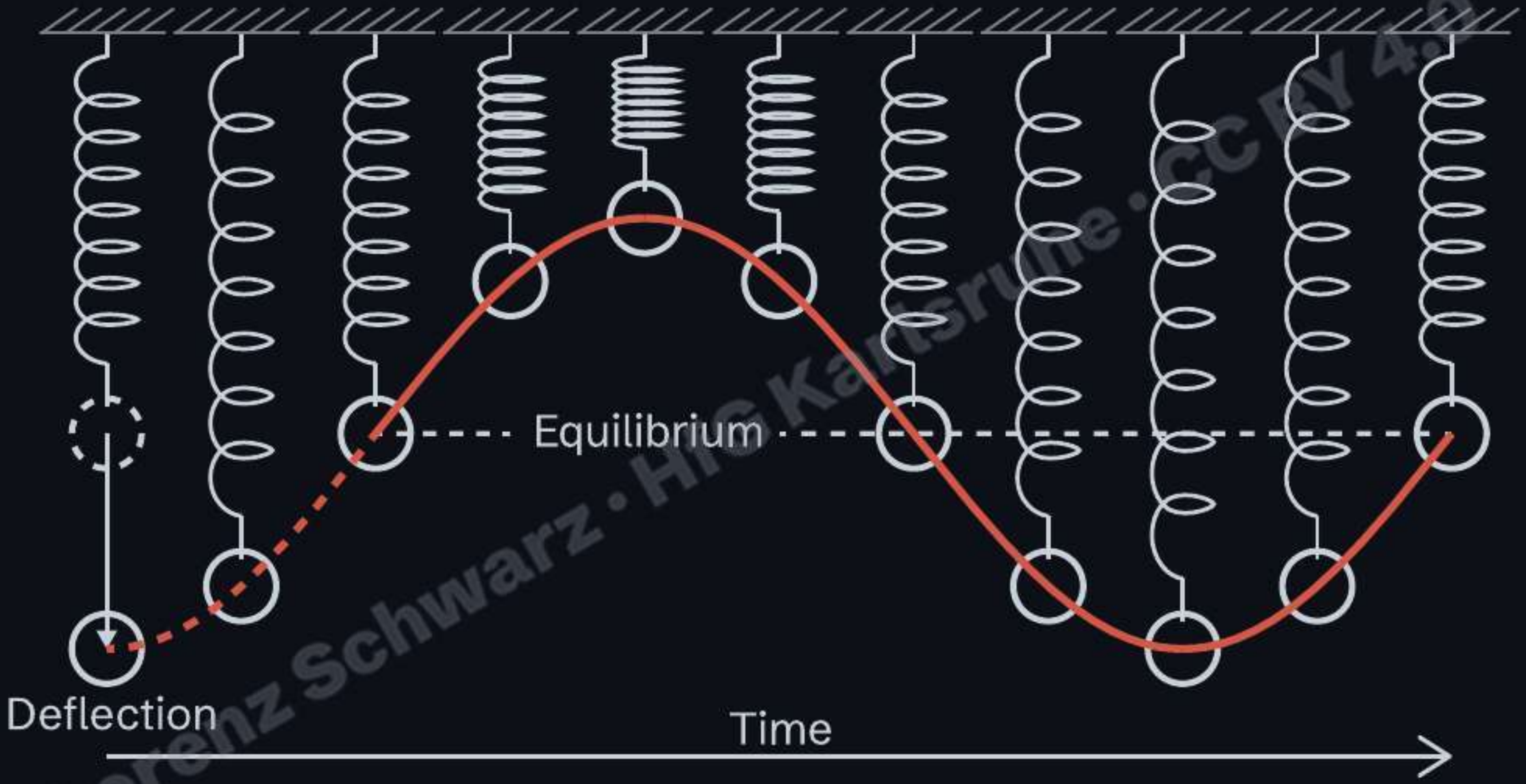
- first derivative of $\sin(\omega t)$ is $\omega \cos(\omega t)$
- and second derivative of $\sin(\omega t)$ → $-\omega^2 \sin(\omega t)$

$$-\omega^2 \sin(\omega t) = -\frac{k}{m} \sin(\omega t)$$

$$\omega = \sqrt{\frac{k}{m}}$$

$$2\pi f = \sqrt{\frac{k}{m}}$$

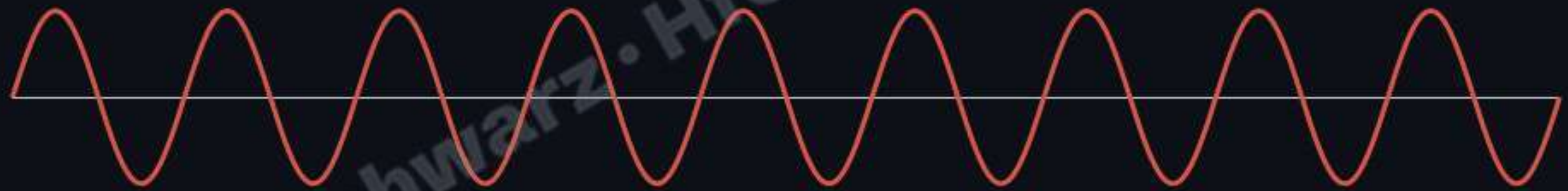
$$f = \frac{1}{2\pi} \sqrt{\frac{k}{m}}$$



Simple harmonic motion of a mass-spring system

Oscillation of a mass-spring system

The mathematics confirms what we observe: a mass on a spring oscillates sinusoidally. The sine wave is the fundamental pattern underlying all sound.



▶ Sine wave 400 Hz

Sine wave

Medieval Latin *sinus*, from Latin, curve

Displacement plotted against time describes a curved and symmetrical rise and fall with no abrupt changes:

- simplest periodic function
- describing periodic phenomena (vibration)
- "pure tone", because it has no other constituent frequencies.

Sine wave function

The time dependence of a harmonic motion is described by a sine (or cosine) oscillation whose argument is a linear function of time:

$$x(t) = A\sin(2\pi f t + \varphi) = A\sin(\omega t + \varphi)$$

- A peak amplitude (nonnegative)
- $\omega = 2\pi f$ angular frequency (radians/seconds and f in Hertz)
- t time (seconds)
- φ initial phase (radians)

→ *All complex oscillations can be related to the sine wave.*

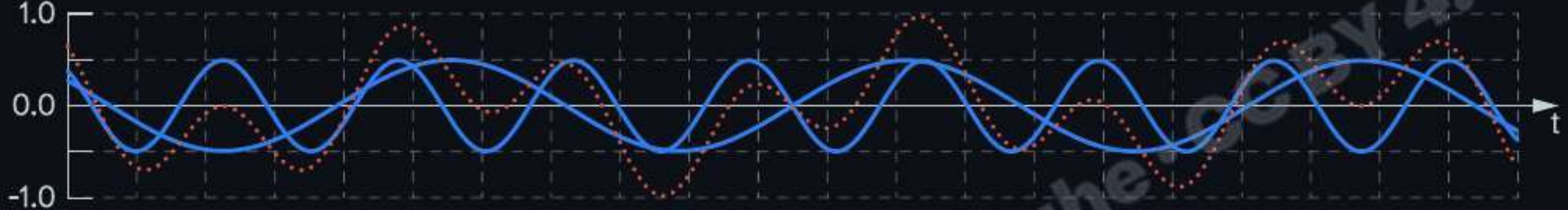
Superposition of sine waves

When a particle undergoes two or more simultaneous oscillatory movements in the same direction, the result is a combined oscillatory movement, determined by the sum of the individual oscillations.

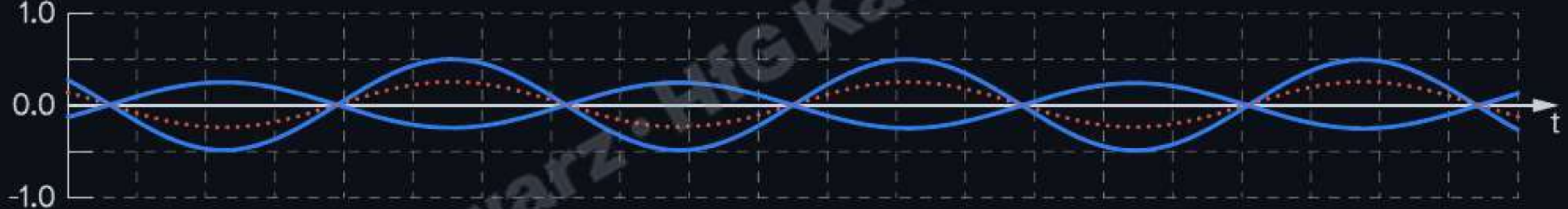
$$x(t) = A_1 \cos(\omega_1 t) + A_2 \cos(\omega_2 t + \varphi)$$

view in graphing calculator

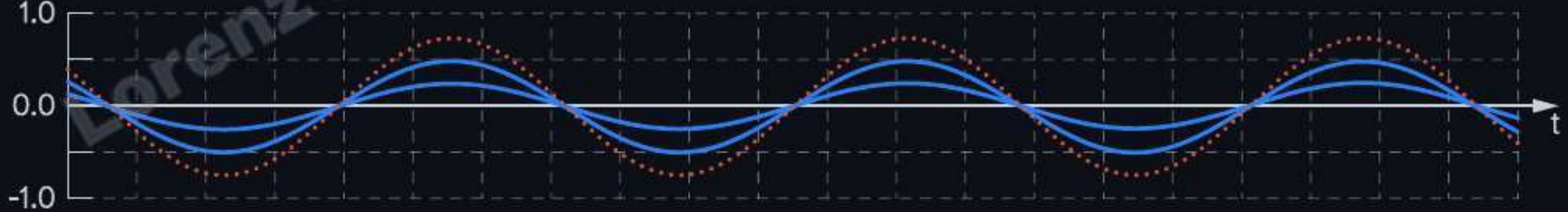
1. Two different frequencies, same amplitude



2. Same frequency, opposite phase, one at half amplitude (partial cancellation)



3. Same frequency, same phase, one at half amplitude (constructive)



Oscillation and pressure waves

The spring-mass system showed periodic oscillations.
Other mechanical systems like strings or speakers create periodic displacements.

In air, this displacement creates:

- Compression (molecules pushed together)
- Rarefaction (molecules spread apart)

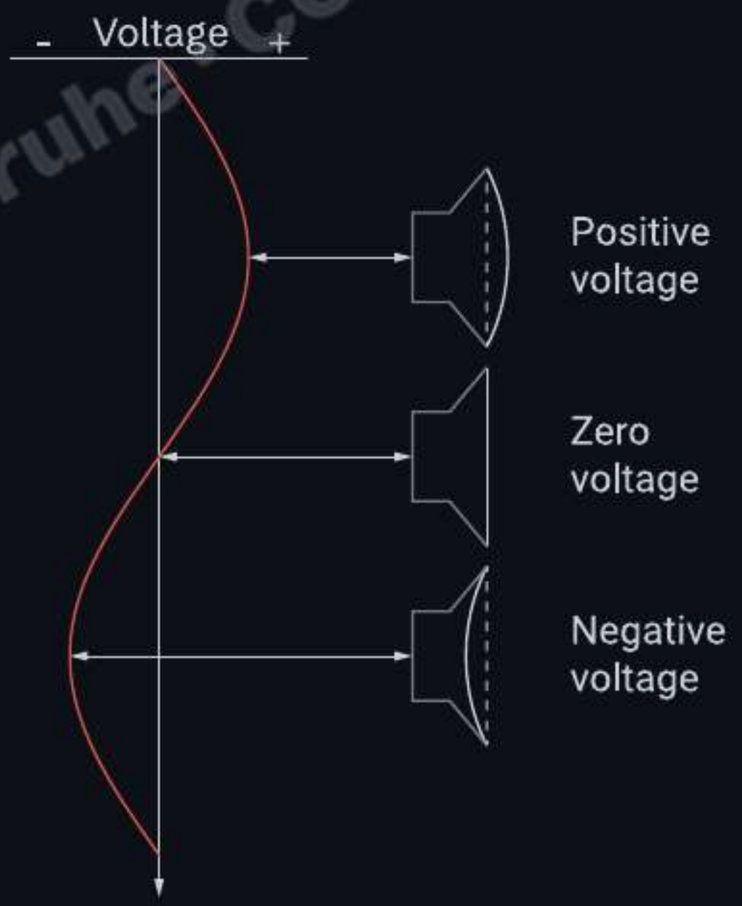
→ *These periodic displacements (pressure variations) propagate as sound waves.*

Back and forth movement of a speaker

The electrical audio signal causes the diaphragm of the speaker to move in an analogous manner:

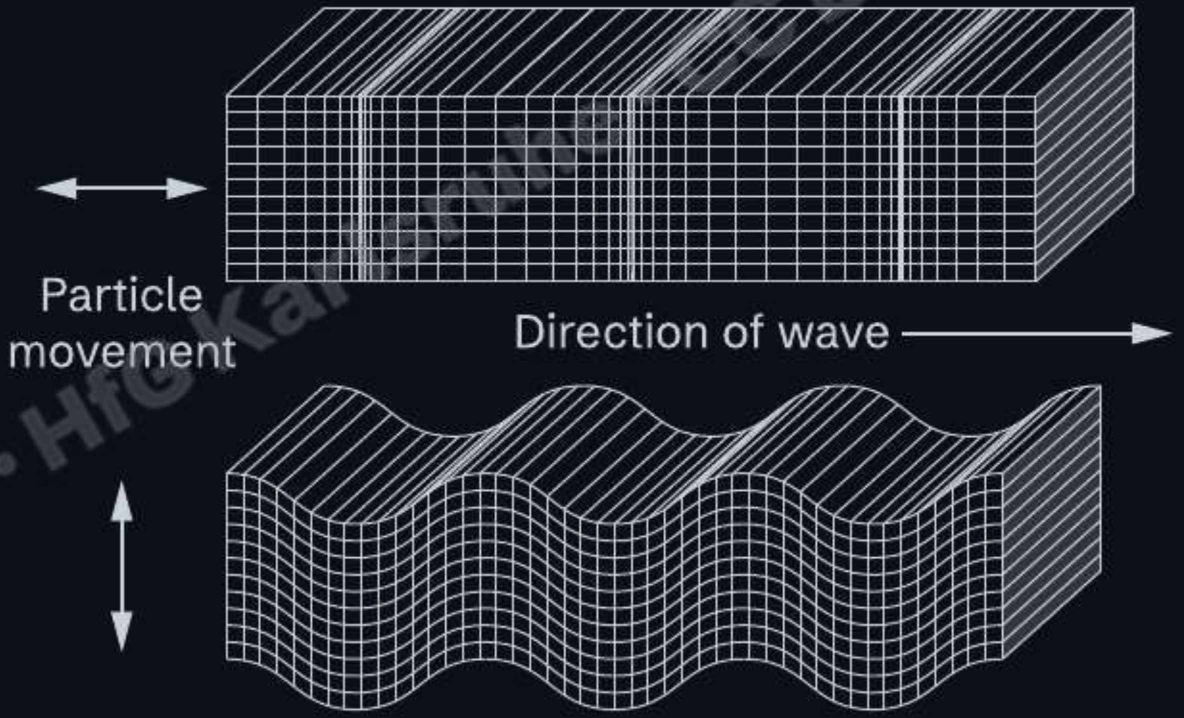
- When it moves forward, it compresses the air particles in front of it.
- When it moves backward, it creates a region of lower pressure.

→ *These alternating compressions and rarefactions propagate through the air as sound waves.*



Sound wave propagation

- Sound is transmitted as longitudinal waves (compression waves) through *gases* and *liquids*.
- It can be transmitted as both longitudinal and transverse waves through *solids*.



Longitudinal wave (top) and transverse wave (below)

Transverse and longitudinal waves

Transverse wave:

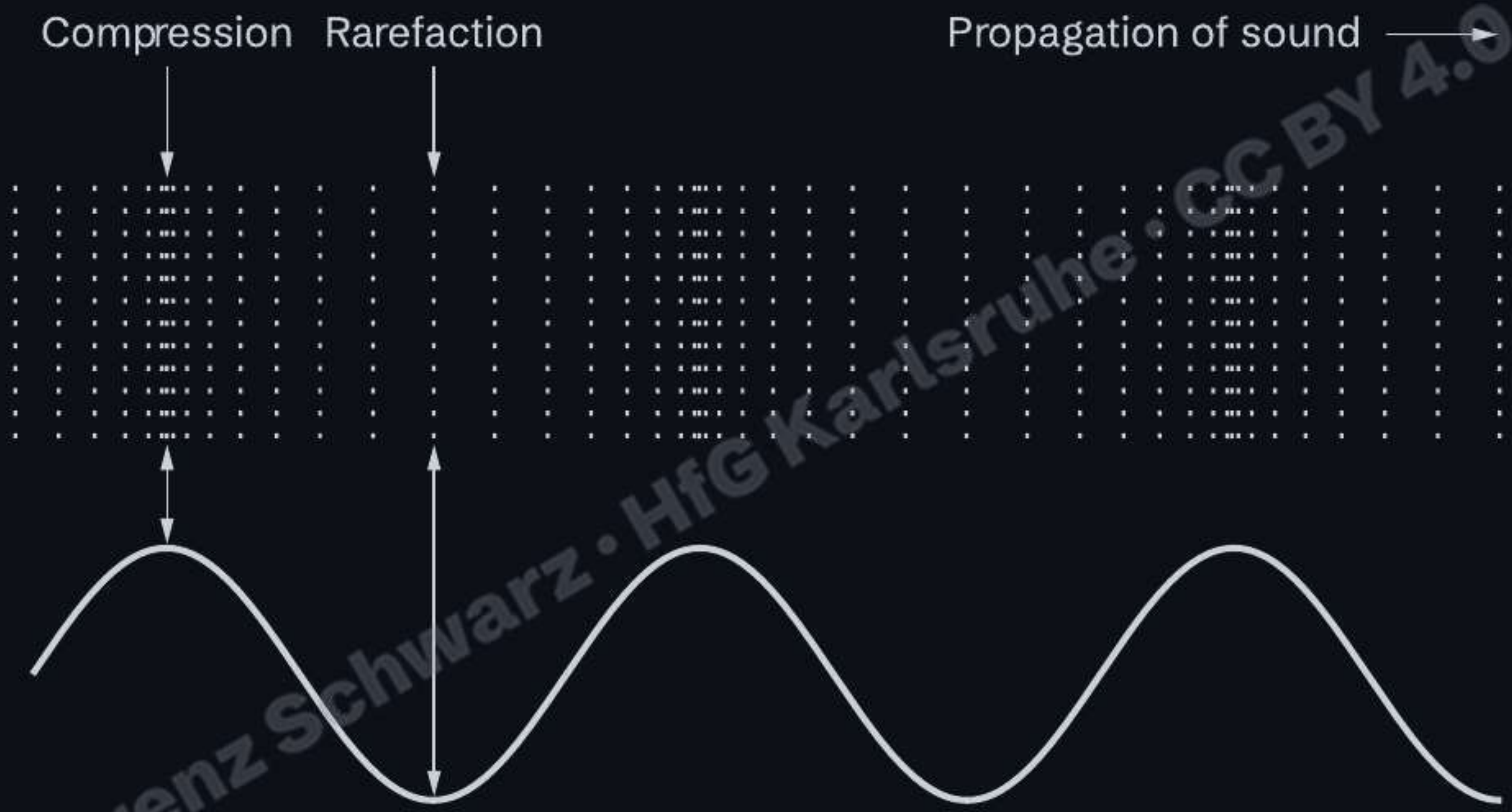
- particles move perpendicular to the direction of the wave

Longitudinal wave:

- particles move parallel to the direction of the wave

→ *Longitudinal waves are considered for airborne sound.*

view in graphing calculator



Longitudinal waves are also called compression waves.

Quantifying sound in space

- **Field quantities** (at a point in space):
  - Sound pressure (Pa) — pressure deviation from ambient atmospheric pressure
  - Particle velocity (m/s) — velocity of particle oscillation around equilibrium
- **Energy quantities** (rate of energy transfer):
  - Intensity (W/m²) — energy flow per unit area
  - Power (W) — total energy radiated from source

→ *Impedance (Pa·s/m) links pressure and velocity as their ratio.*

Sound pressure p (sound field quantity)

Sound pressure is a property of the sound field at a specific point in space.

It represents variations in air pressure (local compressions and rarefactions) caused by sound waves, typically measured with a microphone, relative to the ambient (static) atmospheric pressure.

$$p_{total} = p_{stat} + p$$

- p = time-varying pressure
- p_{stat} = static pressure
- p in pascals (Pa) = N/m^2

Sound pressure level (SPL)

Sound pressure level (L_p) expresses sound pressure on a logarithmic scale in decibels:

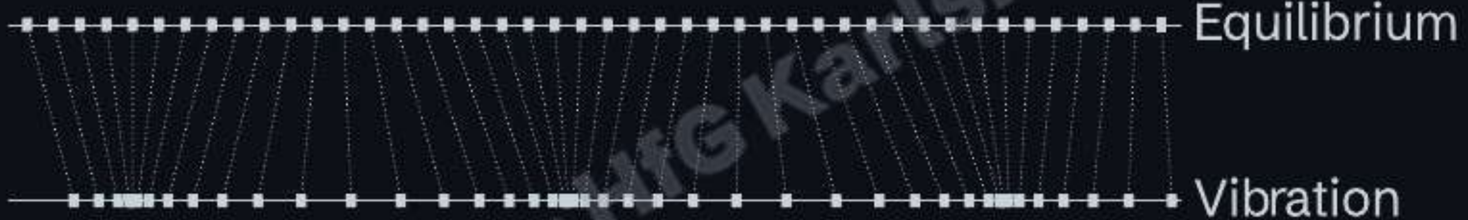
$$L_p = 20\log_{10}\left(\frac{p}{p_0}\right)\,\mathrm{dB\ SPL}$$

p — measured sound pressure (Pa)
p_0 — reference sound pressure

- Reference: $p_0 = 20\,\mu\mathrm{Pa} = 0\,\mathrm{dB\ SPL}$ (threshold of human hearing at 1 kHz)
- Pain threshold: $p \approx 20\text{–}60\,\mathrm{Pa} \approx 120\text{–}130\,\mathrm{dB\ SPL}$

Particle velocity v

Particle velocity is the speed of the particles vibrating around their rest position (equilibrium).



→ *Particle velocity must not be confused with the speed of sound.*

Sound power

Sound is a form of energy:

- Property of the sound source, equal to the total power emitted by that source in all directions.

→ *Sound power is neither dependent on room nor distance*

Sound intensity I (sound energy quantity)

- Sound intensity is acoustical power per unit area (W/m^2).
- Sound intensity level (SIL) is its logarithmic representation (dB).

$$L_I = 10\log_{10}(\frac{I}{I_0})$$

Reference sound intensity for the auditory threshold (at 1000Hz):

$$I_0 = 10^{-12} W/m^2$$

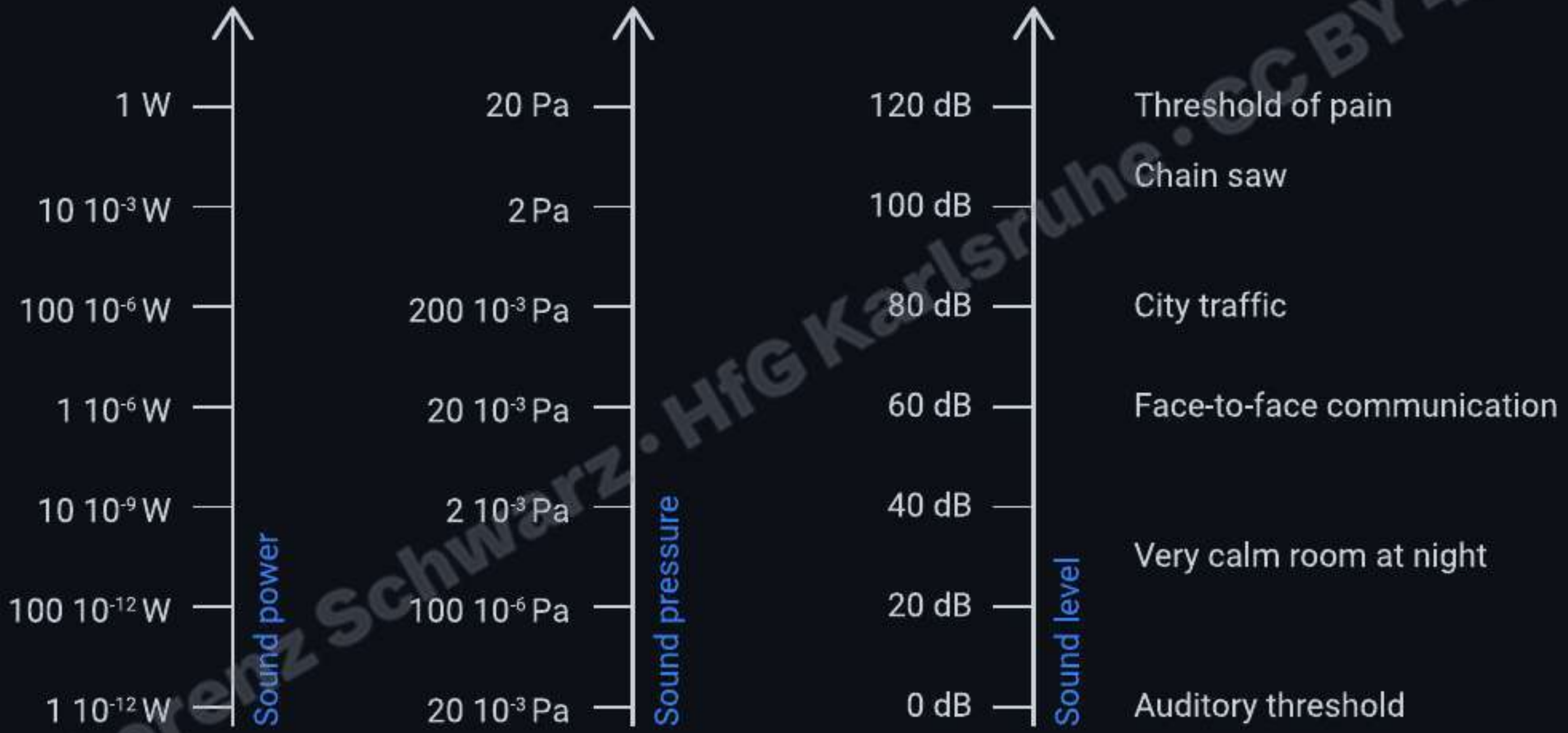
Impedance Z

Acoustic Impedance (Z) is the ratio of sound pressure (p) to particle velocity (v) in a sound wave:

$$Z = \frac{p}{v}$$

Specific Acoustic Impedance (Z_0):
For a plane wave or in the far field, the specific acoustic impedance for air at standard temperature and pressure is approximated as:

$$Z_0 \approx 413\ \mathrm{Pa{\cdot}s/m} \approx 413\ \frac{\mathrm{kg}}{\mathrm{m}^2 \cdot \mathrm{s}}$$


1 W
$10\ 10^{-3}$ W
$100\ 10^{-6}$ W
$1\ 10^{-6}$ W
$10\ 10^{-9}$ W
$100\ 10^{-12}$ W
$1\ 10^{-12}$ W
Sound power
20 Pa
2 Pa
$200\ 10^{-3}$ Pa
$20\ 10^{-3}$ Pa
$2\ 10^{-3}$ Pa
$100\ 10^{-6}$ Pa
$20\ 10^{-3}$ Pa
Sound pressure
120 dB
100 dB
80 dB
60 dB
40 dB
20 dB
0 dB
Sound level
Threshold of pain
Chain saw
City traffic
Face-to-face communication
Very calm room at night
Auditory threshold

Sound intensity and sound pressure level

Sound pressure level L_p:

$$L_p = 20 \log_{10} \left(\frac{p}{p_0} \right)$$

Sound intensity level L_I:

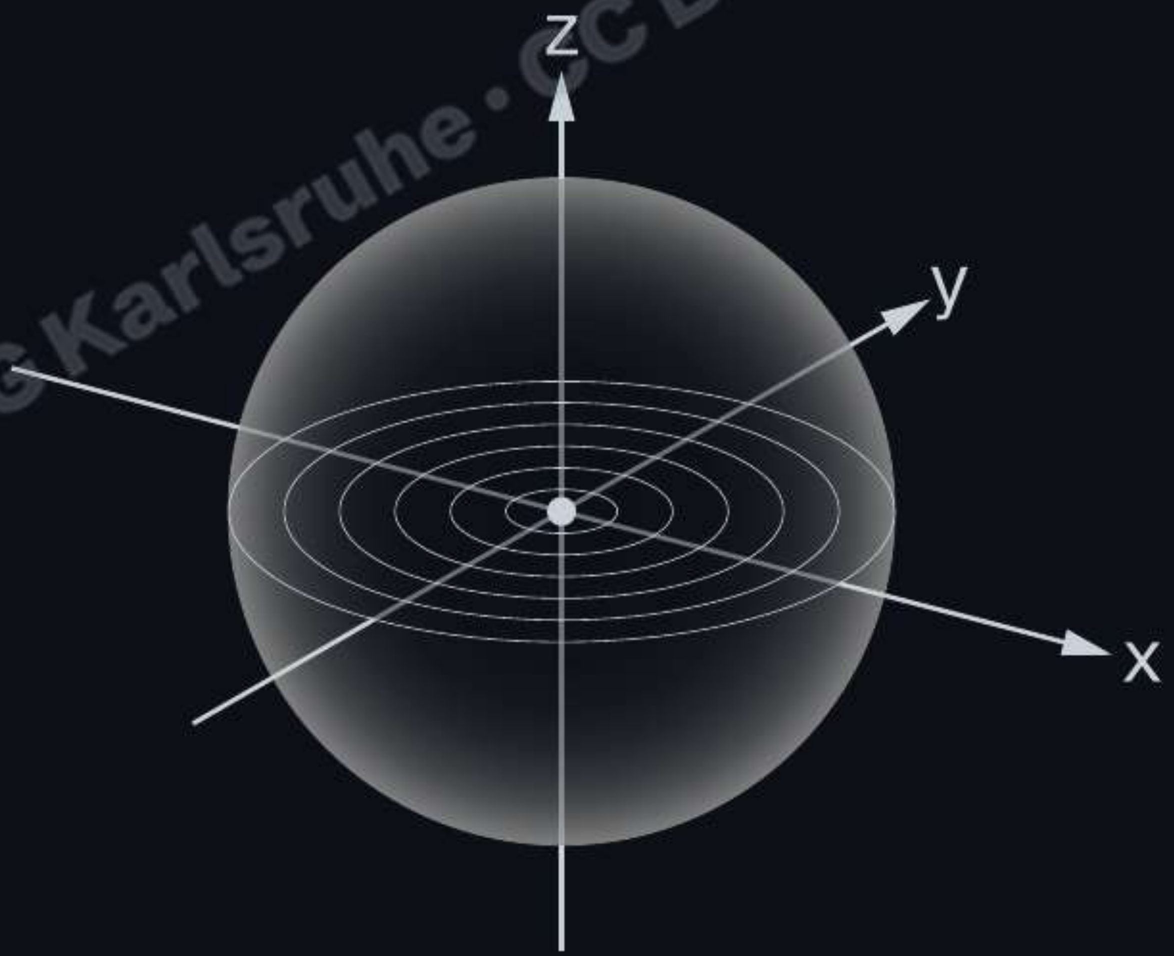
$$L_I = 10 \log_{10} \left(\frac{I}{I_0} \right)$$

→ *Human hearing primarily responds to sound pressure.*

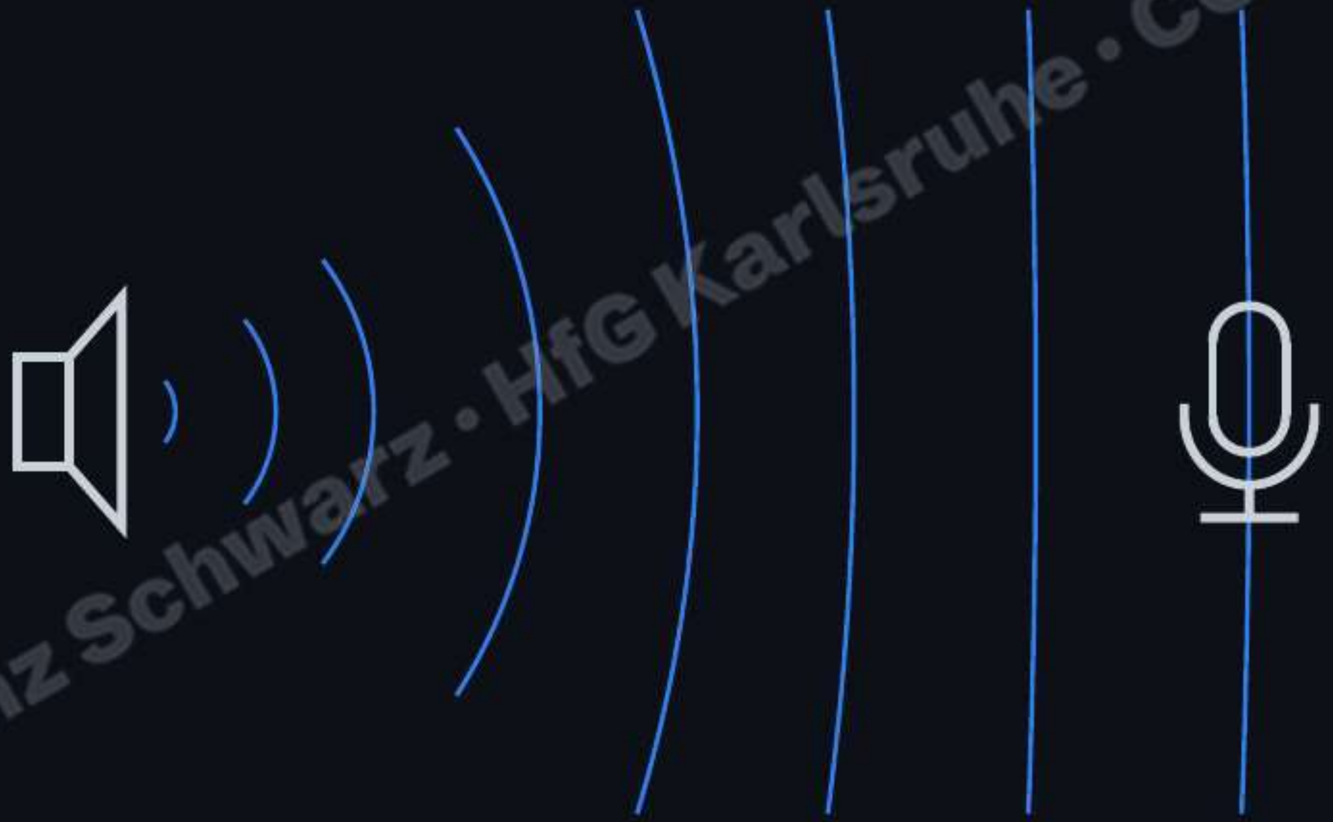
Spatial behavior

An idealized wave that radiates uniformly in all directions from a single point source in 3D space and attenuates with distance:

- The acoustic field variables depend only on the radial coordinate (r) and time (t).
- Surface area of spherical wavefront: $A = 4\pi r^2$



Wave propagation: spherical (short distance), plane (long distance)



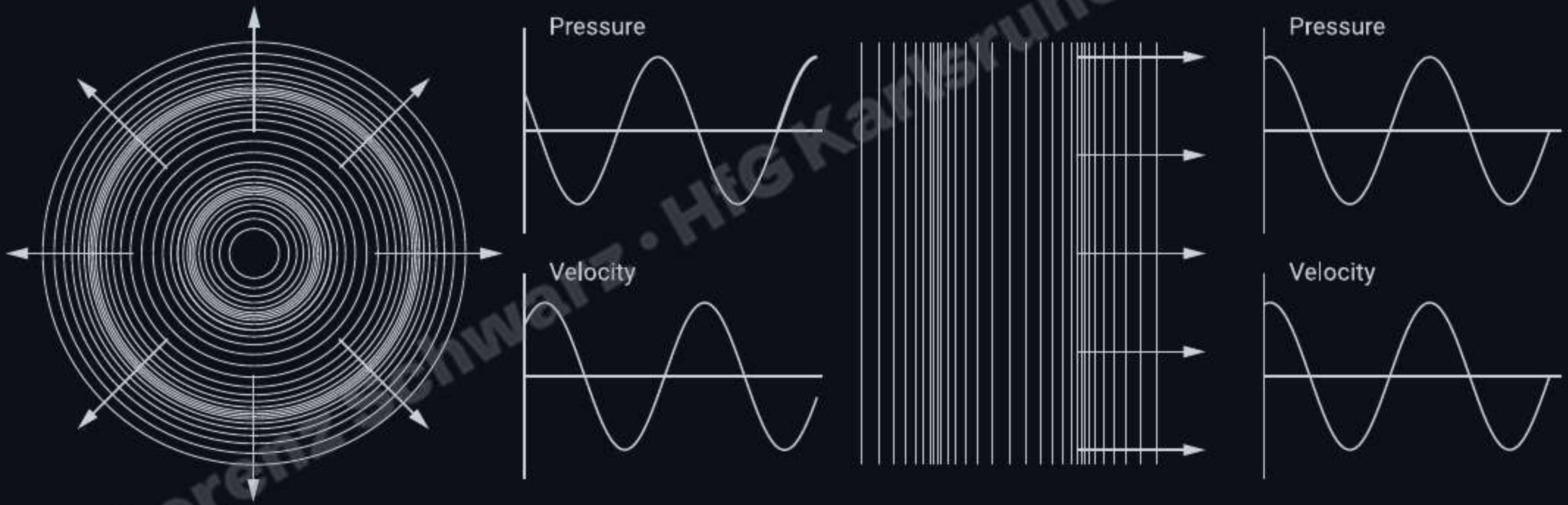
Acoustic fields and their properties

- **Near Field:** A property of the source; the region close to the sound source where the sound pressure and particle velocity are not proportional (non-linear behavior).
- **Far Field:** A property of the source; the region farther from the sound source where sound waves are proportional to the inverse of the distance (linear behavior).
- **Free Field:** A property of the environment; sound propagates without reflections or obstructions.
- **Diffuse Field:** A property of the environment; sound energy is uniformly distributed due to multiple reflections.

Near and far field

- Near field:
  - For point sources, the near field is often approximated as $r < \lambda$
  - Particle velocity shows strong deviations in the near field.
  - (Where λ = wavelength, distance over which wave repeats)
- Far field:
  - Ratio of sound pressure and particle velocity is constant (in phase).
  - The curvature of the wavefront becomes plane.
  - Sound pressure approximately follows inverse-distance behavior

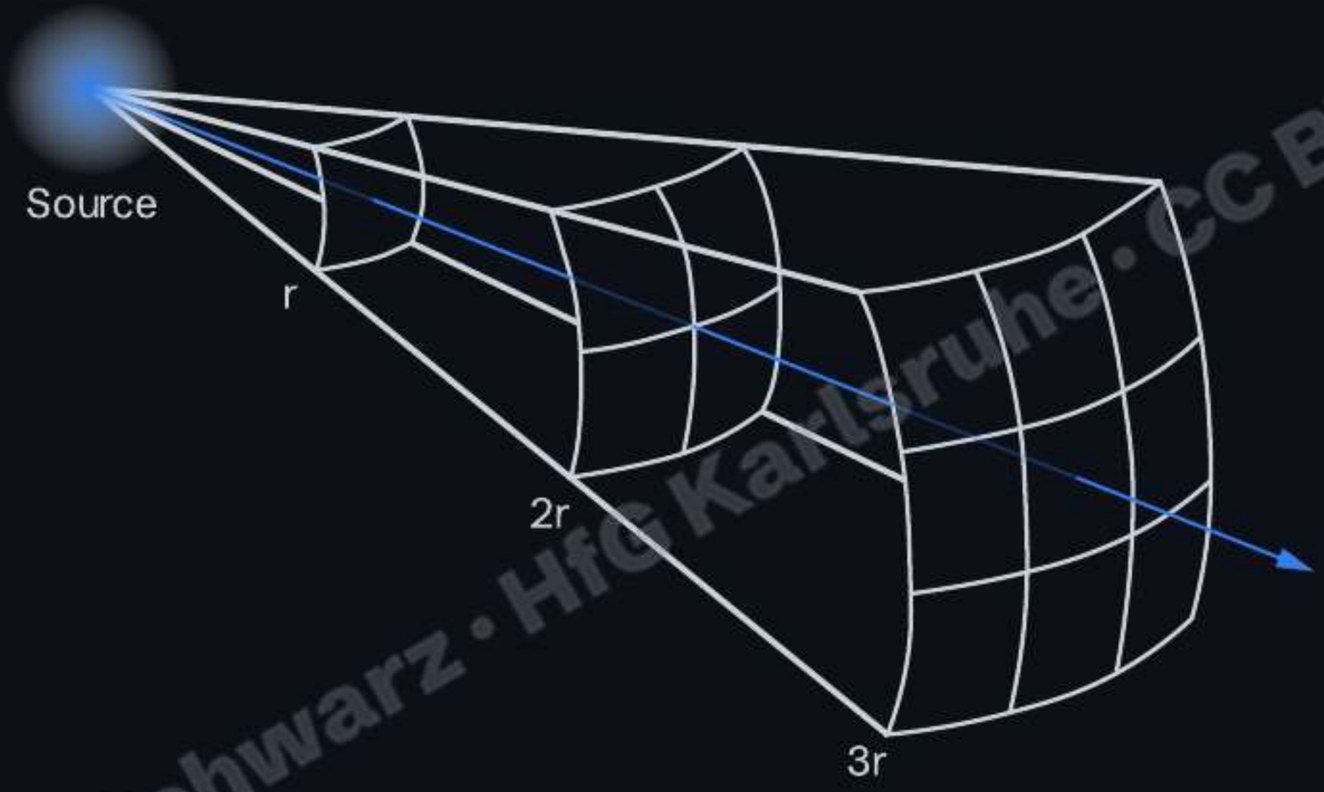
Spherical wave and plane wave



Free field

Region where sound propagates without any interference from reflective surfaces, obstacles, or boundaries, resulting in no reverberation or echo (only direct sound).

→ *Sound is attenuated according to the inverse-square law.*



Sound gets weaker as the distance from the sound source increases. (Doubling the radius increases the surface area of a spherical wavefront by a factor of four.)

Sound propagation with distance

In free-field conditions, sound level decreases as sound energy spreads over a larger area with distance.

- **Sound intensity:** Doubling distance decreases the level by about 6 dB (intensity level)

$$I \propto \frac{1}{r^2}$$

- **Sound pressure:** Doubling distance decreases the level by about 6 dB SPL

$$p \propto \frac{1}{r}$$

→ *Both intensity level and sound pressure level drop by about 6 dB per distance doubling*

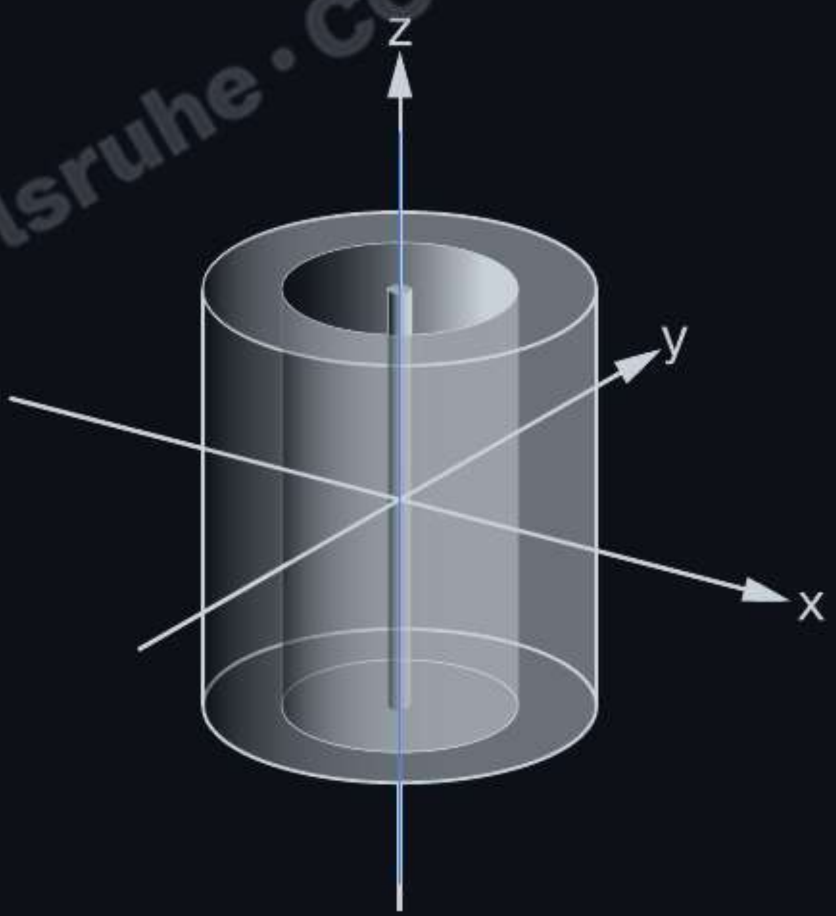
Line source

Cylindrical wavefront radiating from a one-dimensional line source (no vertical dispersion).

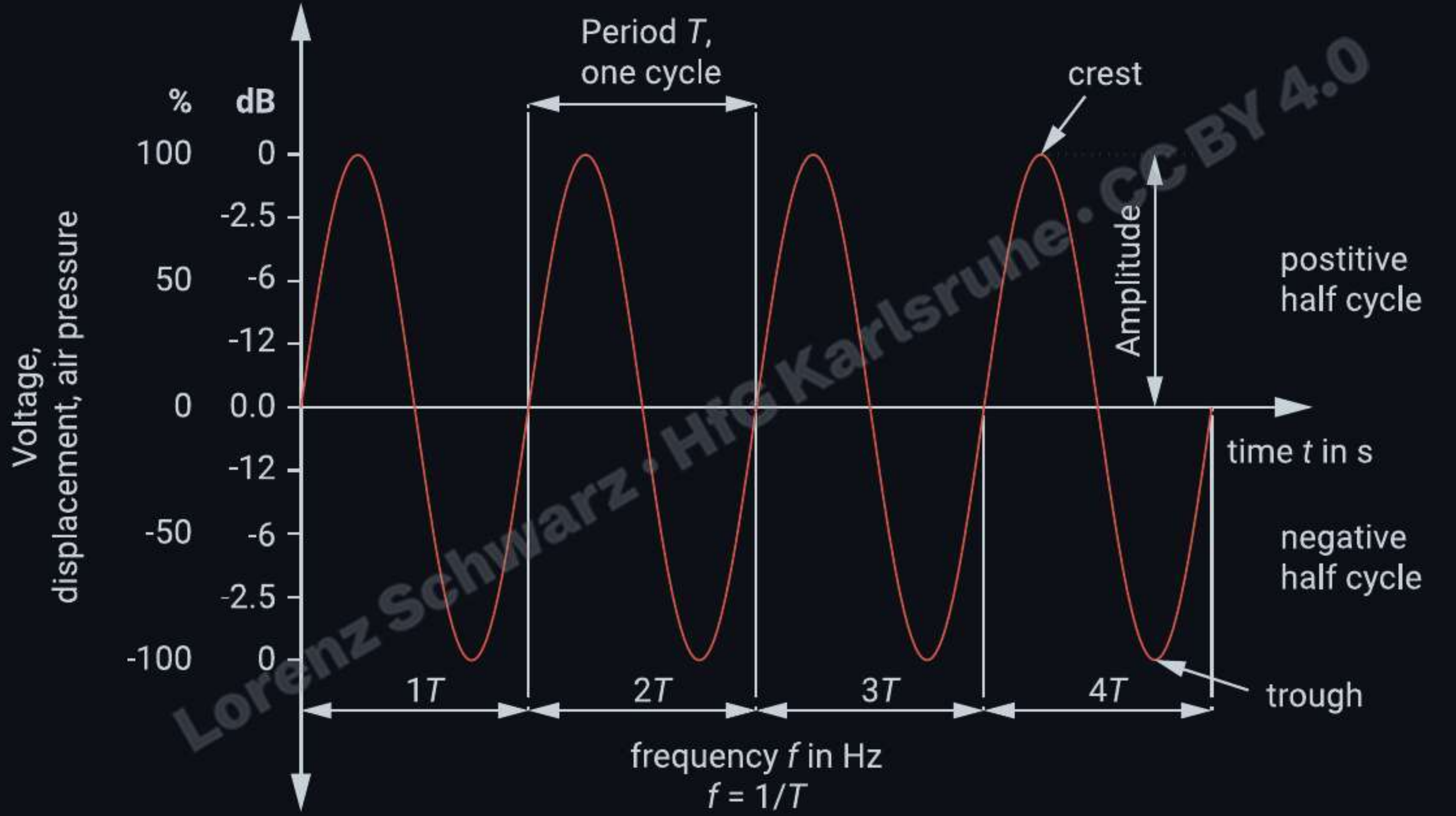
$A_{cylinder} = 2\pi\ r\ h$

Line source attenuates with the inverse of distance (1/r), which is a decrease of approximately -3 dB

Applications: Sound reinforcement situations (as much energy as possible for the audience, e.g., line arrays)



WAVE PROPERTIES

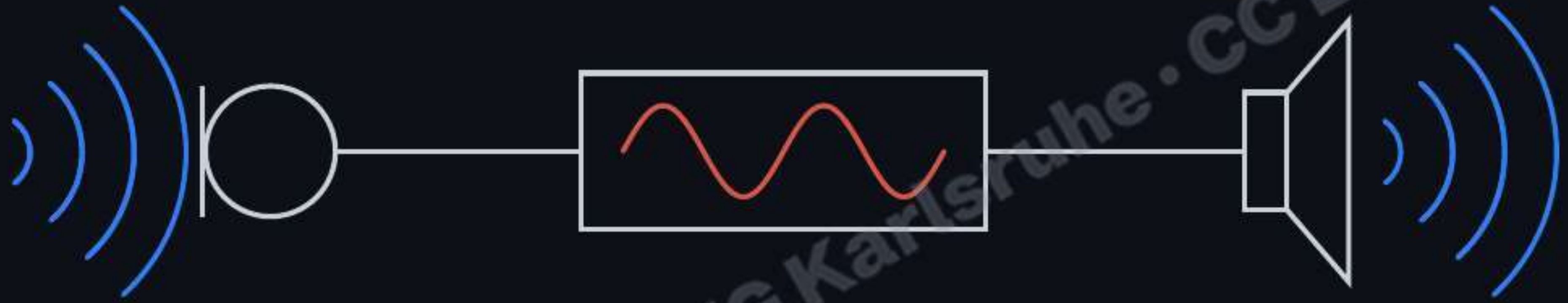


Period T, one cycle
crest
%
dB
100
0
-2.5
50
-6
-12
0
0.0
-12
-50
-6
-2.5
-100
0
Voltage, displacement, air pressure
Amplitude
postitive half cycle
time t in s
negative half cycle
1T
2T
3T
4T
trough
frequency f in Hz
f = 1/T

Audio signals (electrical representation)

A transducer (e.g., microphone) converts the varying air pressure of a sound wave into a continuous electrical signal, which is proportional to the sound pressure variations.

→ *Electrical sound signals in this form are analogous to the sound pressure levels.*



Acoustic Sound - Electrical Signal - Acoustic Sound

Electronic sound and audio signals

Electronic audio signals are variations in electrical energy (changes in voltage or current), that correspond to sound pressure variations.

→ *Audio signals can be converted into audible sound through a speaker or other transducers.*

Wave properties

The shape of the graph of a periodic function can be described using the following terms:

1. Amplitude $\hat{u}$ - The maximum instantaneous value of the wave.
2. Period T - The time interval after which the wave repeats.
3. Frequency f - The reciprocal of the period (oscillations per second).
4. Phase angle φ - Describes the offset or difference between two sine waves.
5. Wavelength λ - The distance over which the wave repeats (spatial period).

Related:
Zero crossing - Point where the wave crosses zero.

Amplitude

Amplitude describes the maximum variation of a periodic signal (such as air pressure, displacement, or voltage) within a single period:

- Maximum instantaneous value of the variable.
- Maximum distance between the resting position (equilibrium) and the point of maximum displacement.
- Perceived as the loudness of sound in the context of audio signals.

→ *Amplitude has no influence on frequency, wavelength, phase, period of time, and speed of sound.*

Amplitude and energy transfer in waves

Amplitude relates to the wave's energy:

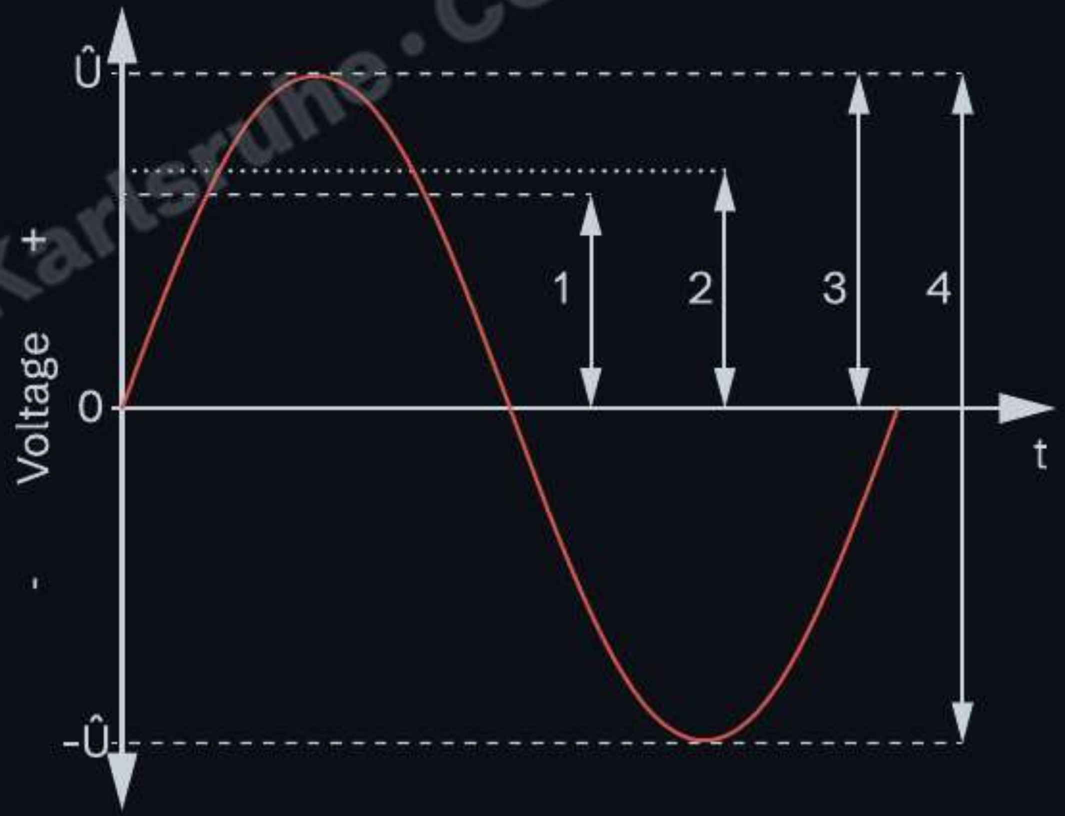
- Higher amplitude corresponds to greater energy transfer in the wave

This energy can be converted into work, heat, or other forms depending on the medium and interaction.

Amplitude

For a sinusoidal waveform:

1. Average rectified (ARV) $\frac{2\hat{u}}{\pi}$
2. Root mean square amplitude (RMS): $\frac{\hat{u}}{\sqrt{2}}$ (equivalent value of constant direct current)
3. Peak amplitude or semi-amplitude: $\hat{u}$ (maximum distance between resting position (equilibrium) and maximum displacement)
4. Peak to peak amplitude: $2\hat{u}$ (between maximum and minimum)



Root mean square (RMS)

The average value of a sine wave over a full cycle is zero (positive and negative halves cancel). RMS solves this by squaring values first, making them all positive:

RMS represents the *effective value*: the equivalent DC level that delivers the same power.

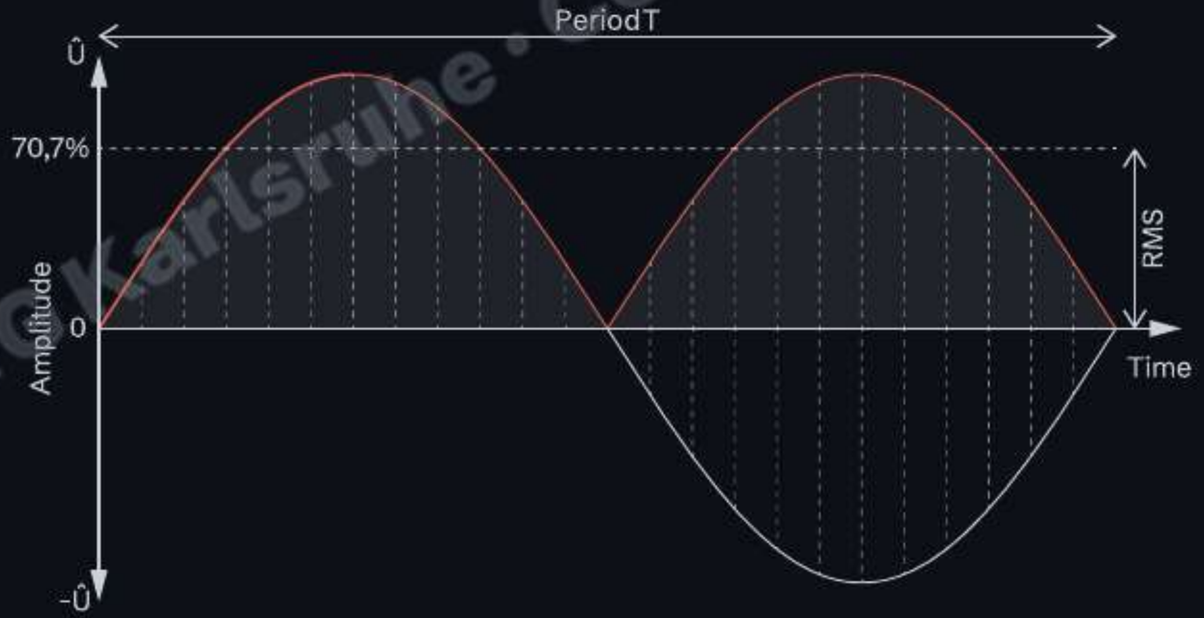
Waveform	RMS
Sine	$0.707 \times A_{\text{peak}}$
Square	A_{peak}

→ *Electrical and acoustic power are proportional to the square of the RMS value.*

Calculating RMS

Square root of the mean of the squares:

- Square all values of the signal: $a^2(t)$ (makes all values positive)
- Compute the area under the squared curve: $\int_0^T a^2(t)\,dt$
- Divide by the period T: $\frac{1}{T}\int_0^T a^2(t)\,dt$
- Take the square root: $\sqrt{}$



$$A_{\mathrm{RMS}} = \sqrt{\frac{1}{T}\int_0^T a^2(t)\,dt}$$

RMS and power

Electrical and acoustic power are proportional to the square of RMS values:

- Amplifier power ratings use RMS voltage and current
- Sound intensity is proportional to RMS sound pressure squared

→ *RMS enables meaningful comparison of signal strength and power delivery across different waveforms.*

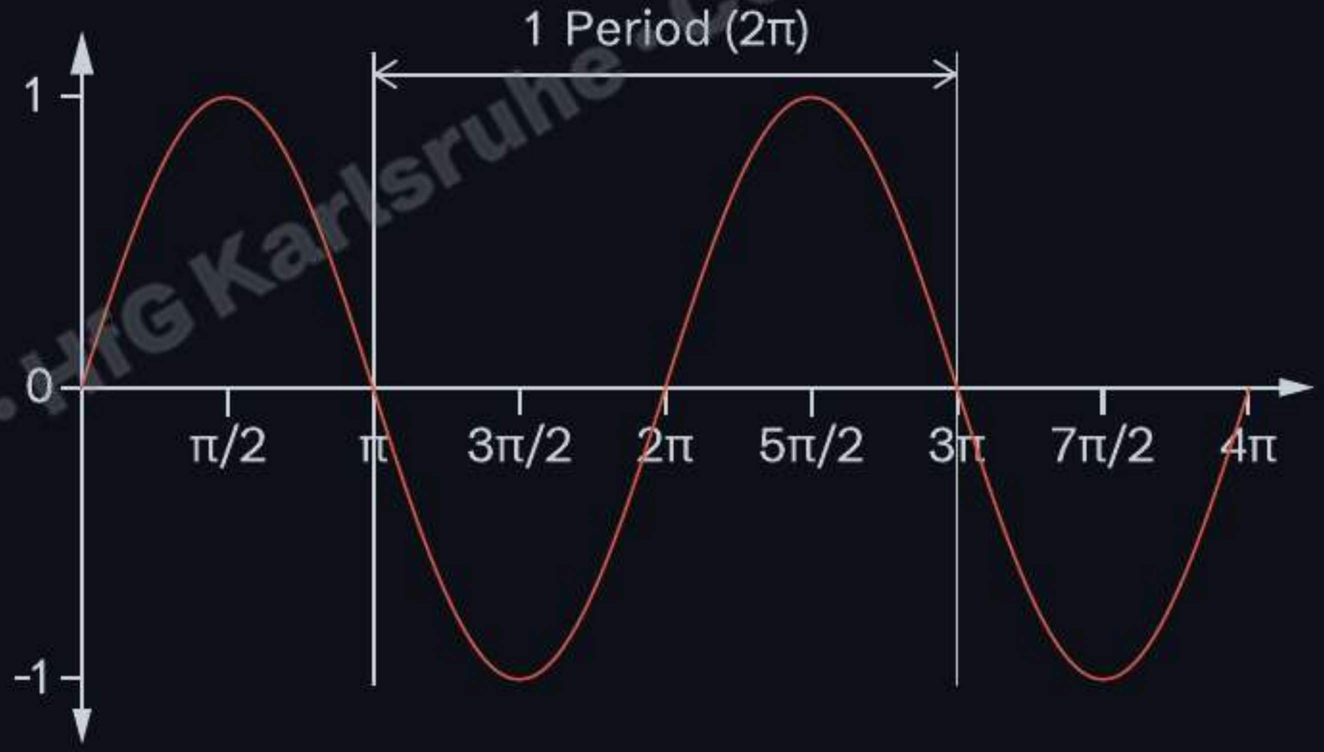
Period T

Time required for a wave to complete one wave cycle (2π)

$$T = \frac{2\pi}{\omega} = \frac{1}{f}$$

Each multiple of a period is also a period, but we usually refer to the smallest positive one as the period.

view in graphing calculator



Period

Example: $f = 440\,\mathrm{Hz}$

$$T = \frac{1}{f} = \frac{1}{440}\,\mathrm{s} = 2.27\,\mathrm{ms}$$

Frequency

Variations or alterations between 0,05 ms (20 kHz) and 50 ms (20 Hz) are perceived as sound.

- Ultrasound: higher than 20 000 Hz.
- Infrasound: lower than 20 Hz.

→ *Hearing range for humans is 20 Hz to 20 000 Hz.*

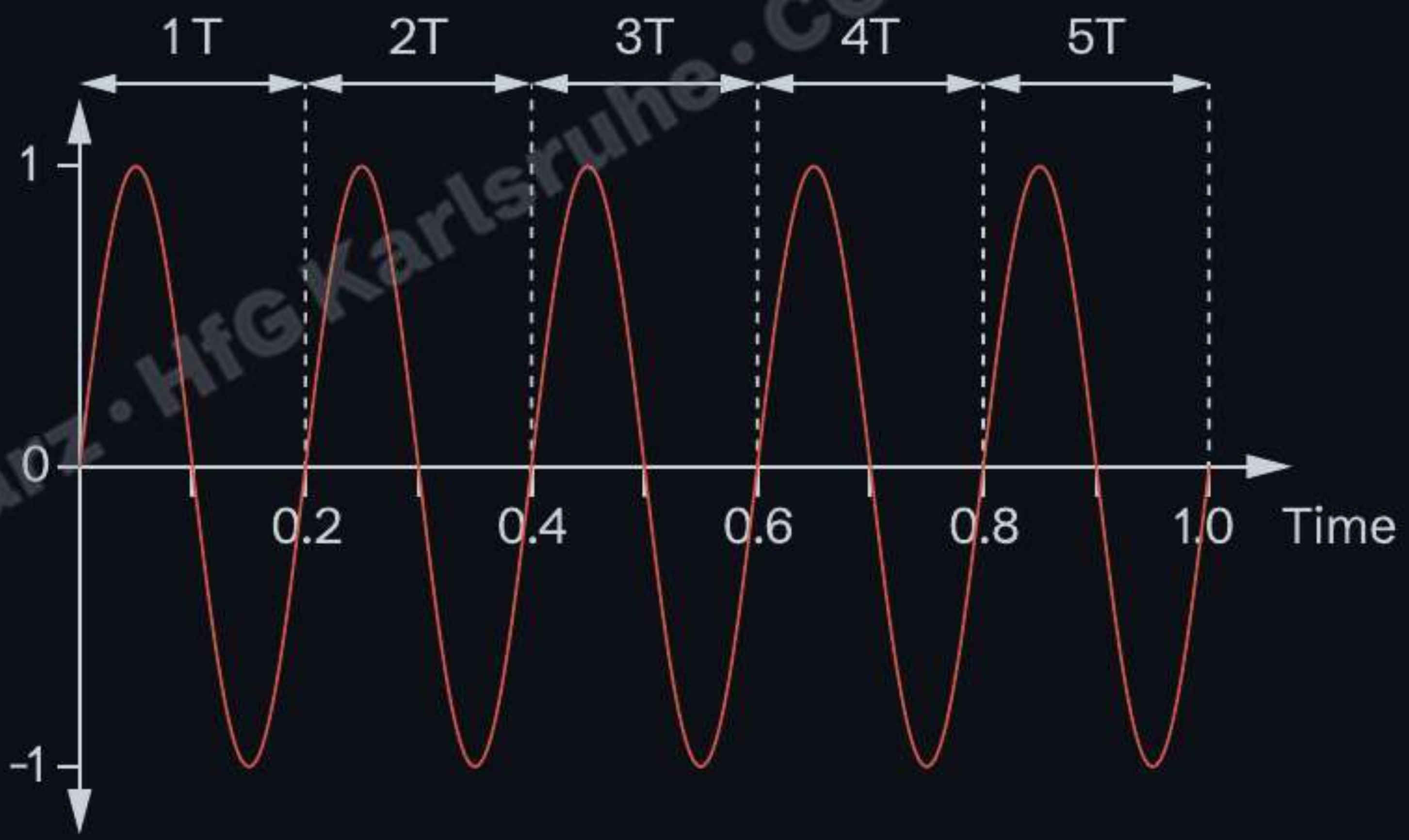
Frequency f

Number of wave cycles per second, expressed in Hertz [Hz]

$$f = \frac{1}{T} = \frac{\omega}{2\pi}$$

f: frequency
T: period
ω: angular frequency

$$\omega = 2\pi f$$



Angular frequency (ω)

Whereas Hertz [Hz] counts cycles per second, radians per second [rad/s] measure the angle swept per second by a rotating pointer.

- **Hz**: cycles per second (cps)
- **rad/s**: angle swept per second (rotating motion)

Hertz and radian can be expressed as reciprocal seconds:

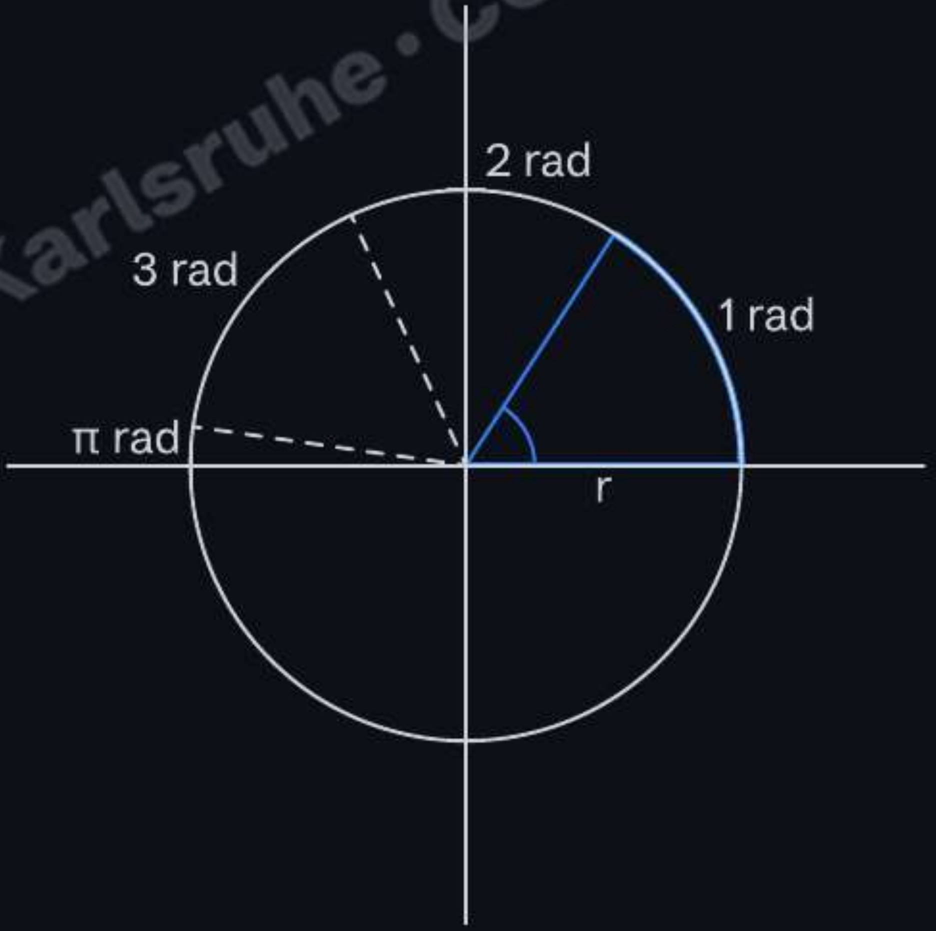
$$[Hz] = [\omega] = s^{-1}$$

Radian

Radian is the angle subtended at the center of a circle by an arc equal in length to the radius.

- 1 full circle = 2π radians
- Radians are dimensionless (ratio of lengths)
- Used in angular measures: ω in rad/s, φ in rad

view in graphing calculator



Example: Relating rad/s to Hz

One radian per second:

$$f_{(Hz)} = \frac{\omega_{(rad/s)}}{2\pi} = 0.1591549433\,Hz$$

Frequency and period

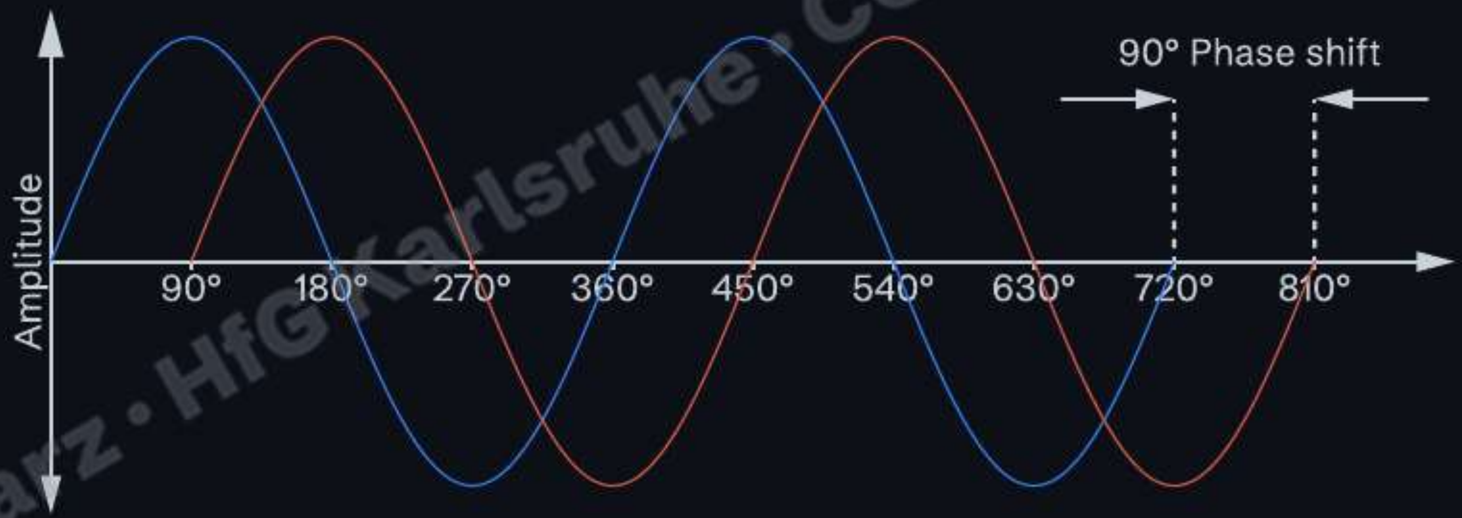
The period is the reciprocal of the frequency and vice versa.

$$T = \frac{1}{f} \qquad f = \frac{1}{T}$$

Phase

Position of a sine wave in time.

- defined for two sine waves
- not for music signals or noise



Cosine and sine have a mutual phase difference of 90°

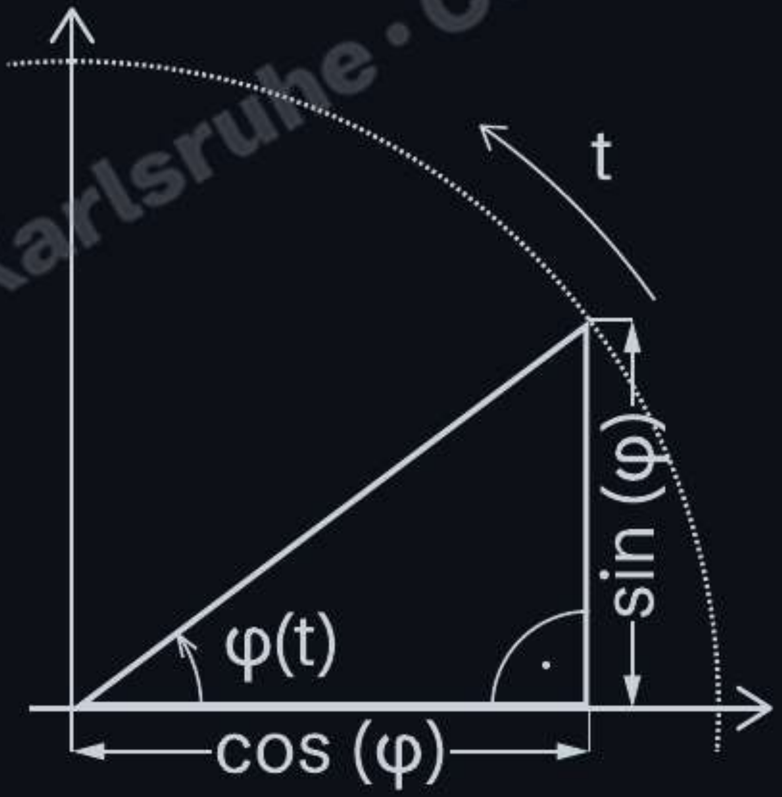
$\varphi(t) = 90^{\circ} = \frac{\pi}{2}$

Phase angle φ

The phase indicates the angular position in the cycle of a periodic process as a function of time.

$$\varphi(t) = 2\pi f t = \omega t$$

view in graphing calculator



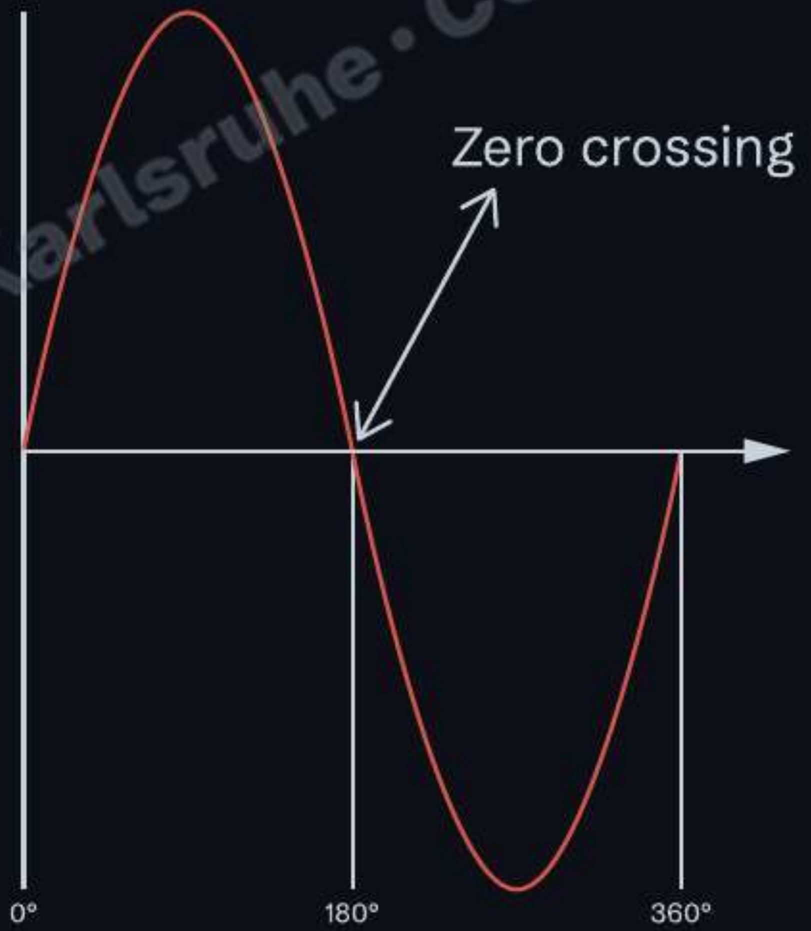
One complete cycle is 2π in radians or 360 in degrees:
0° starting point (zero position), 90° highest point, 270° lowest point

Zero crossing

Zero crossing is the point where the signal's amplitude is zero and it changes sign:

- Occurs twice per cycle in simple waveforms (e.g., sine, sawtooth, triangle, square)

→ *In speech processing, the zero-crossing rate helps distinguish between voiced and unvoiced speech sounds.*



The wave equation

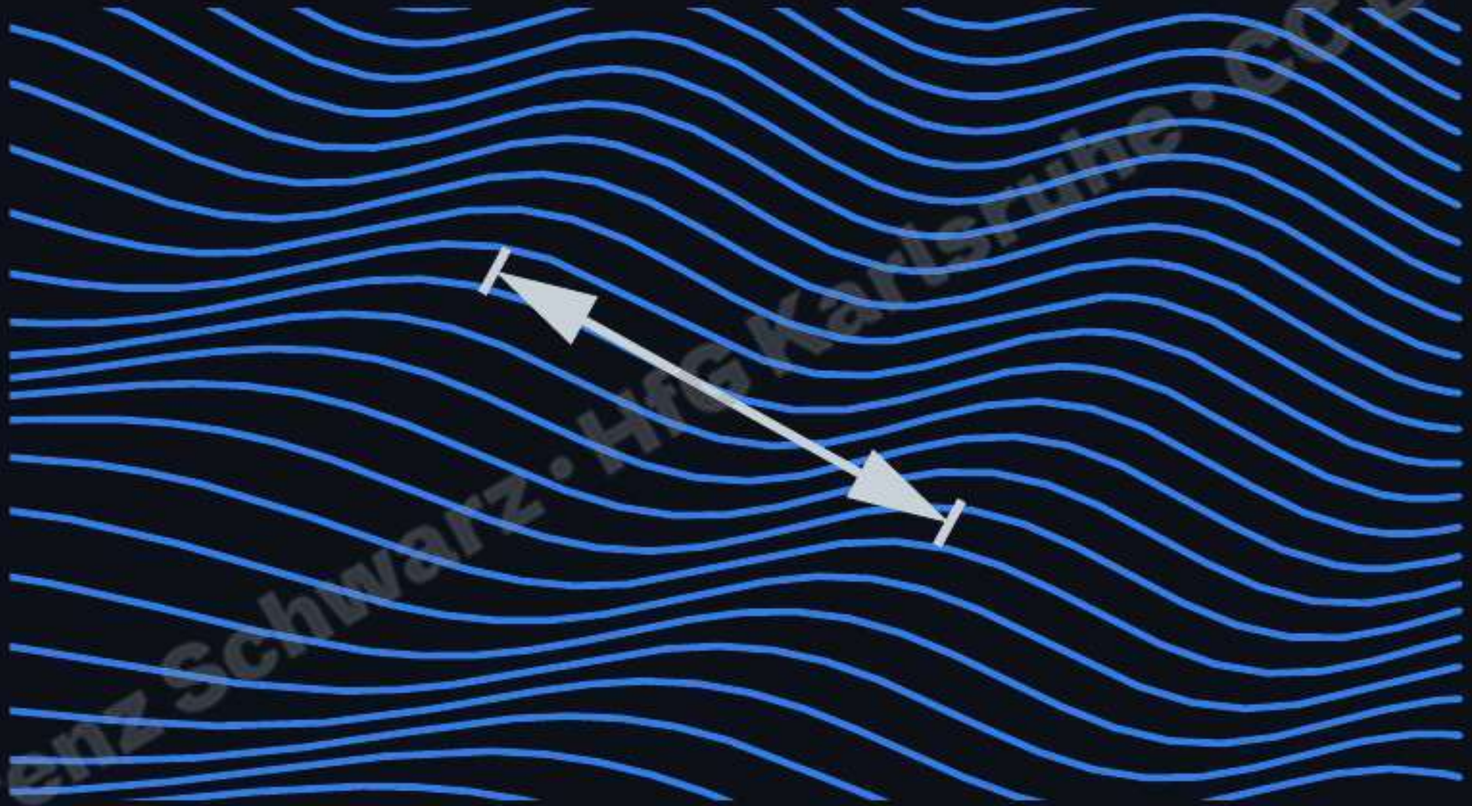
All wave properties are interconnected through a single relationship:

$$c = f\lambda = \frac{\lambda}{T}$$

- c is the propagation speed of the wave (e.g. 343 m/s in air)
- Amplitude determines energy, independent of the other properties
- Frequency and period are reciprocals: $f = \frac{1}{T}$
- Wavelength depends on both frequency and medium: $\lambda = \frac{c}{f}$
- Phase describes position within a cycle

→ *Changing one property (except amplitude) necessarily affects others.*

Wavelength λ



Wavelength

Spatial period:

- distance over which the wave's shape repeats (related to frequency)
- parallel to the direction of propagation

$$\lambda = \frac{c}{f}$$

λ Wavelength
f Frequency
c Speed of sound

Calculating wavelength from frequency

Example:

What is the wavelength of a 440 Hz tone in air, where sound speed is 343 m/s?

$$\lambda = \frac{c}{f} [m]$$

$$\lambda = \frac{343}{440} = 0.78m$$

Wavelength λ

Wavelength and frequency

Frequency and wavelength are inversely proportional to each other:

$$\text{low frequency} \longleftrightarrow \text{long wavelength}$$
$$\text{high frequency} \longleftrightarrow \text{short wavelength}$$

Wavelengths of various sound frequencies

Frequency (Hz)	Wavelength in Air (m)
31.5	11
63	5.5
125	2.7
250	1.4
500	0.7
1k	0.344
2k	0.172
4k	0.086
8k	0.043
16k	0.021

Speed of sound

The speed of sound is the distance a sound wave travels per unit of time through a medium.

Speed of sound:

$$c \approx 343\,\frac{\mathrm{m}}{\mathrm{s}} \text{ at 20°C in air}$$

$$c \approx 1481\,\frac{\mathrm{m}}{\mathrm{s}} \text{ in fresh water}$$

Factors affecting the speed of sound

- **Elasticity:** The more elastic (less compressible) the medium, the faster sound travels.
- **Density:** Higher density generally slows sound in gases but may increase speed in solids or liquids if accompanied by high elasticity.
- **Temperature:** In gases, higher temperatures increase the speed of sound by reducing the medium's density and increasing molecular energy.

Speed of sound in gases

$$c = \sqrt{\kappa \frac{p}{\rho}} = \sqrt{\kappa \frac{RT}{M}} = \sqrt{\kappa \frac{k_B T}{m}}$$

κ heat capacity ratio
ρ density
p pressure
R molar gas constant
$M = m \cdot N_A$ molar mass
T thermodynamic temperature in kelvin
k_B Boltzmann constant
m molecular mass

Example: Time-distance relationship of sound

- **34 cm/ms**
  - sound travels about 34 cm per millisecond
- **3 ms/m**
  - Sound takes roughly 3 ms to travel 1 meter

(Assuming speed of sound = 343 m/s)

Applying the speed of sound formula

Example:
Determining the distance of a lightning bolt (and thunderstorm cell):

- Every **3 seconds** of delay ≈ **1 kilometre** distance from the lightning bolt.

Question:
How long does sound need to travel **2 m**?

$$t = \frac{d}{c} = \frac{2\,\mathrm{m}}{343\,\mathrm{m/s}} \approx 5.8\,\mathrm{ms}$$

Speed of sound in different media (at 20°)

Media	Meters/Second
Air	344
Helium	981
Water, fresh	1480
Seawater	1500
Ice (-4°C)	3250
Acrylic Glass	2670
Beech wood	3300
Concrete	5850-5920
Mild Steel	5050
Aluminium	6250-6350

Wave propagation

Sound speed in an elastic medium depends on temperature.

- Lower temperature → lower speed
- 1°C change ≈ 60 cm/s change in sound speed

Practical use: delay lines in PA systems

When multiple speaker systems cover a large area, sound from distant speakers must be delayed to match the time it takes for sound from closer speakers to arrive.

Speed of sound and related terms

Generally, sound travels faster in denser and less compressible media.

- **Subsonic:** Motion or speed less than the speed of sound in a given medium.
- **Infrasonic:** Sound waves with frequencies lower than ~20 Hz (below the range of human hearing).

Doppler effect

The change in frequency or pitch of sound waves perceived by an observer due to the relative motion between the sound source and the observer.

- The pitch is higher than the stationary pitch as the source approaches.
- The pitch decreases as the source passes the observer.
- The pitch becomes lower than the stationary pitch as the source moves away.

Used in rotary (Leslie) speakers and film sound design plug-ins.

▶ Doppler effect applied to a moving sound source (plugin simulation)

Doppler effect formula

$$f' = \frac{c \pm v_r}{c \mp v_s} \cdot f_0$$

- f' = observed frequency
- f_0 = emitted (source) frequency
- c = speed of sound in the medium
- v_r = velocity of the receiver
- v_s = velocity of the source

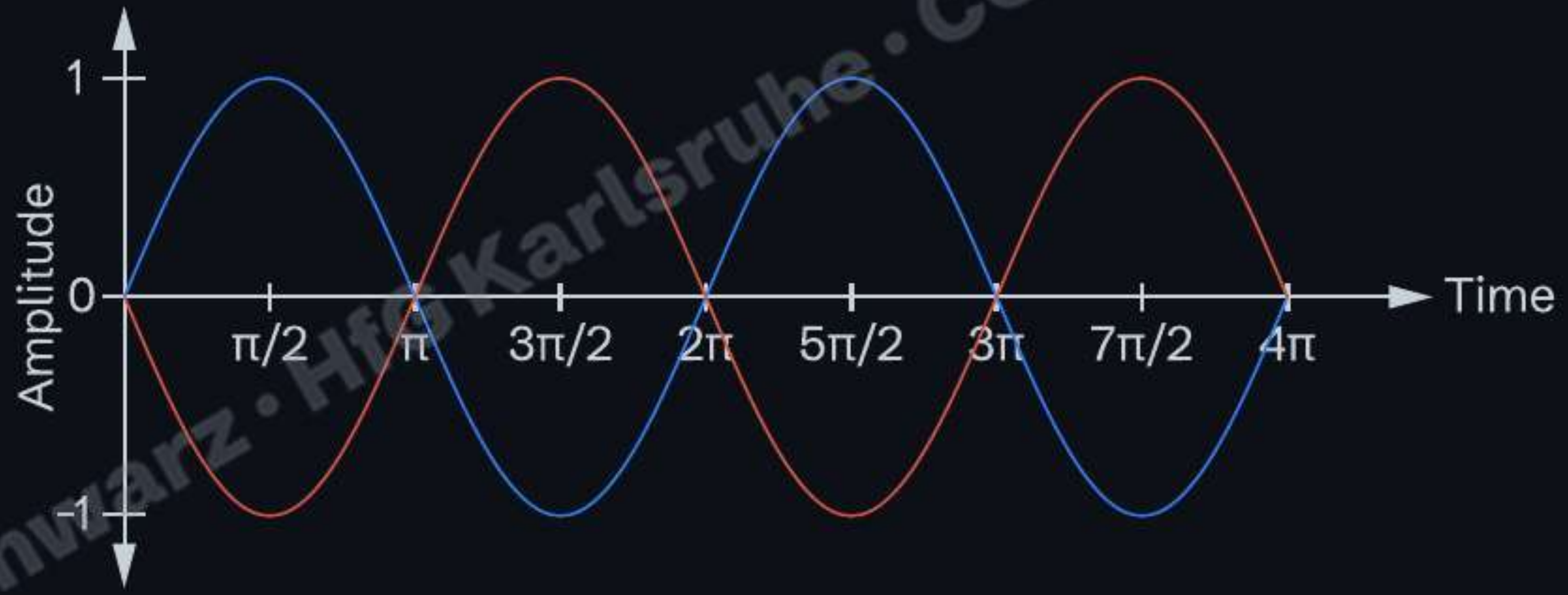
For a stationary receiver ($v_r = 0$):

$$f' = \frac{c}{c \mp v_s} \cdot f_0$$

Doppler effect and sound barrier

Polarity inversion

- opposite amplitude
- inverted signal
- no time shift



The red graph shows an inverted version of the blue graph (same shape, opposite sign)

Polarity inversion

Applications:

- Differential signalling for transmitting analog audio.
- "Phase" button on mixing desks to avoid phase cancellation

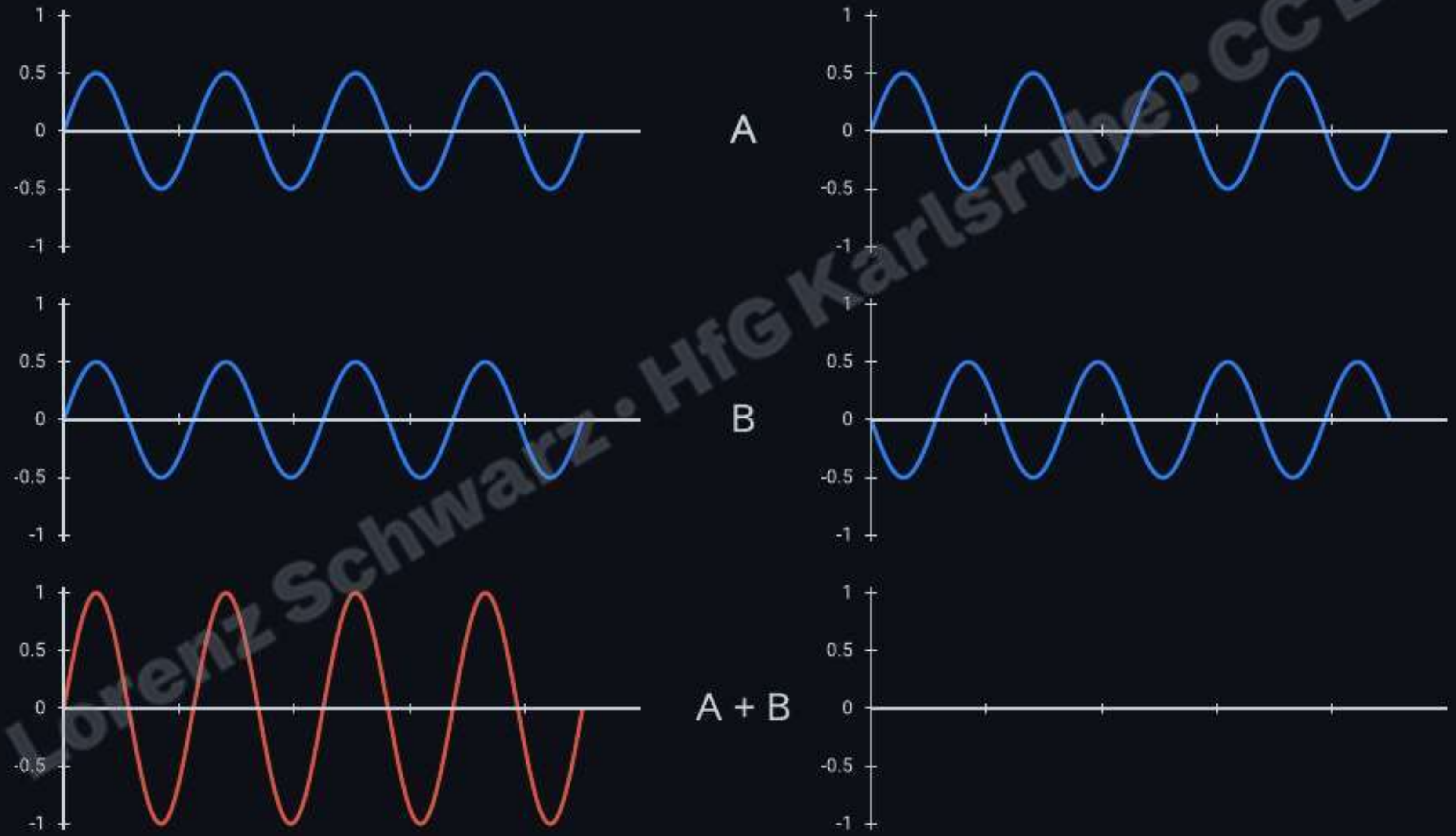
Superposition and interference

Interference, a consequence of superposition, describes the interaction between sound waves. The resultant amplitude is the sum of the individual amplitudes:

- amplification (constructive, even multiple of π)
- attenuation (destructive)
- cancellation (destructive, odd multiple of π)

view in graphing calculator

Constructive and destructive interference



Applications:

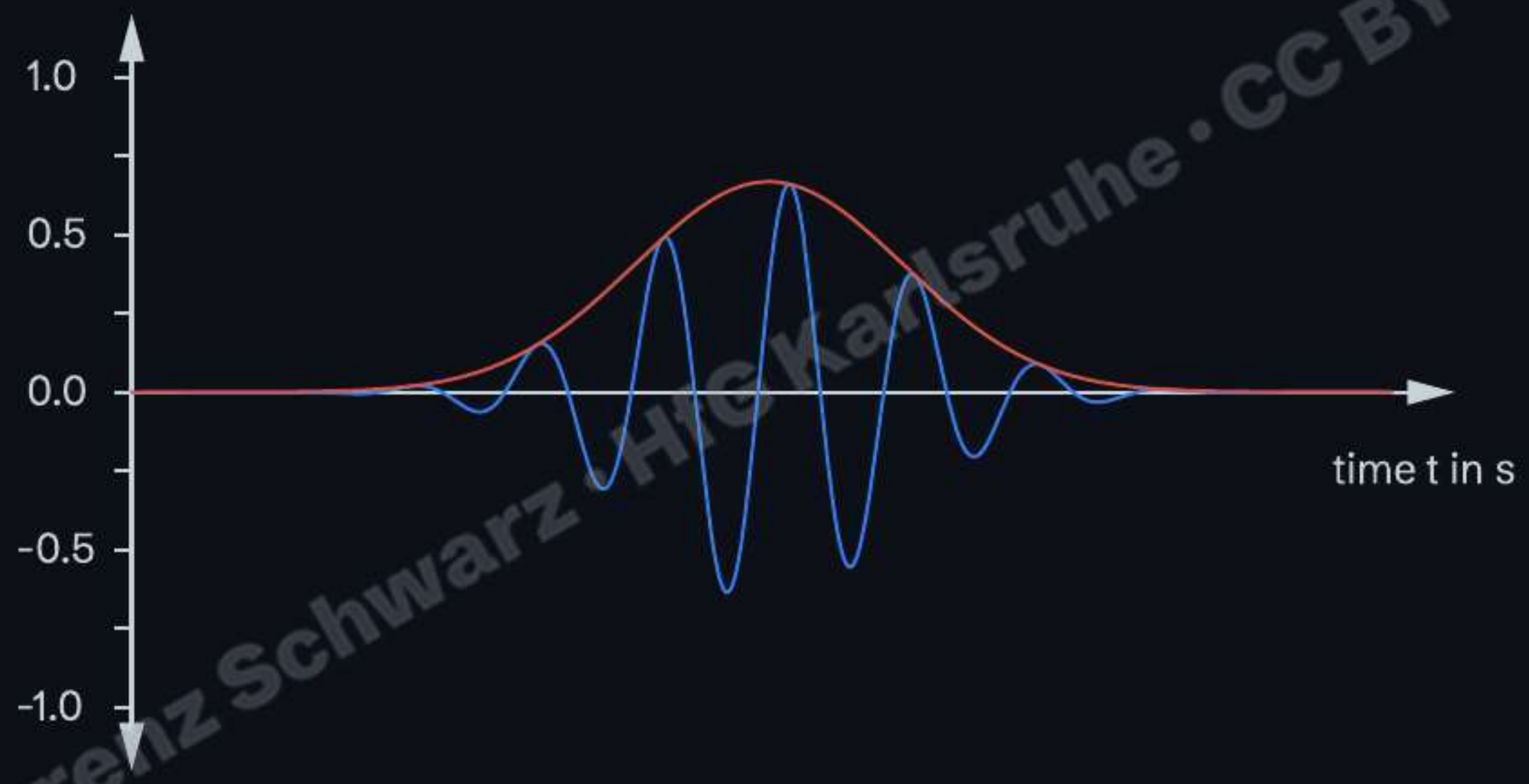
- Chorus (multiple copies of the same signal, slightly delayed and out of tune)
- Phaser (copied signal runs through an all-pass filter and is then mixed with its original)
- Active noise control

▶ Phasing effect applied to white noise

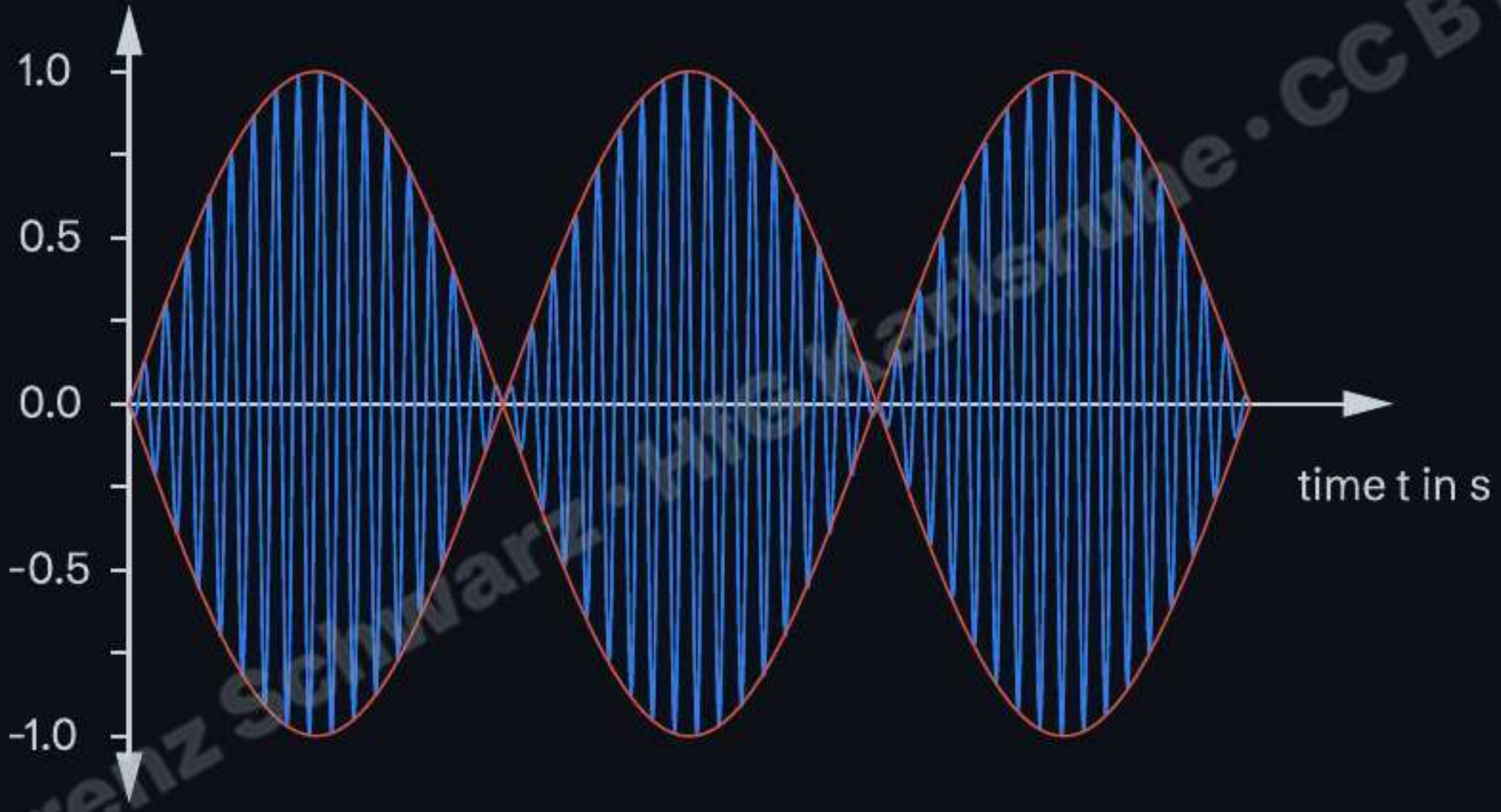
Wave properties and sound design

The following slides on envelope and amplitude modulation connect amplitude concepts to time-varying behavior.

Envelope over time



Upper and lower envelope



Listening examples: Amplitude modulation

Varying the amplitude of a 400 Hz sound with a lower-frequency modulation signal:

1. Slow rates (< 4 Hz) → Pulsation. ▶
2. Moderate rates (4 - 30 Hz) → Tremolo. ▶
3. Faster rates (30 - 70 Hz) → Roughness. ▶
4. Very fast rates (> 70 Hz) → Spectral coloration. ▶

WAVEFORMS

Complex vibratory systems

While a spring-mass system produces a single sinusoidal vibration, real vibrating systems (strings, air columns, membranes, plates) produce many simultaneous vibration modes.

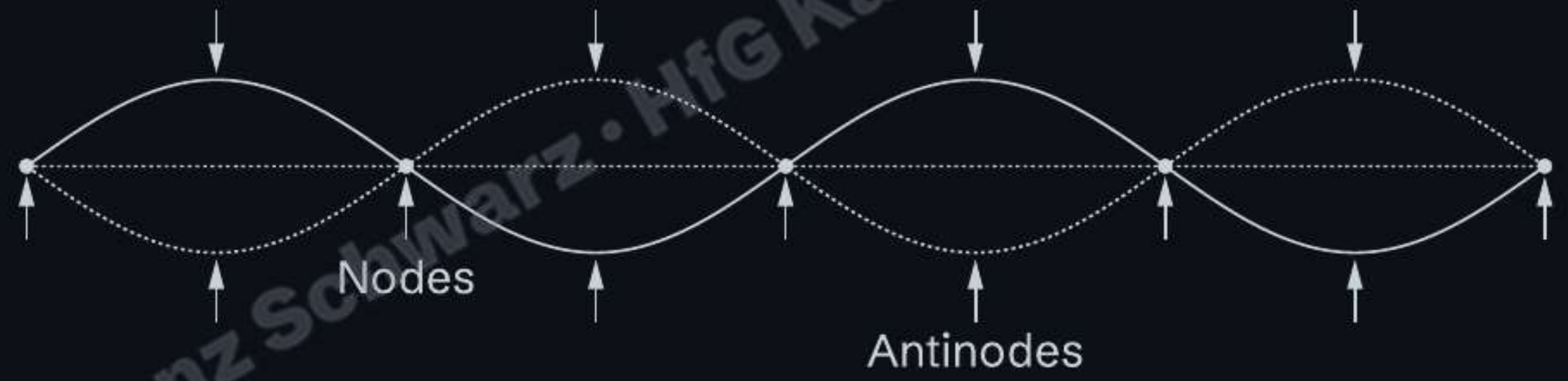
Modes and spectra in real sounds

Pure sine waves are rare in the real world:

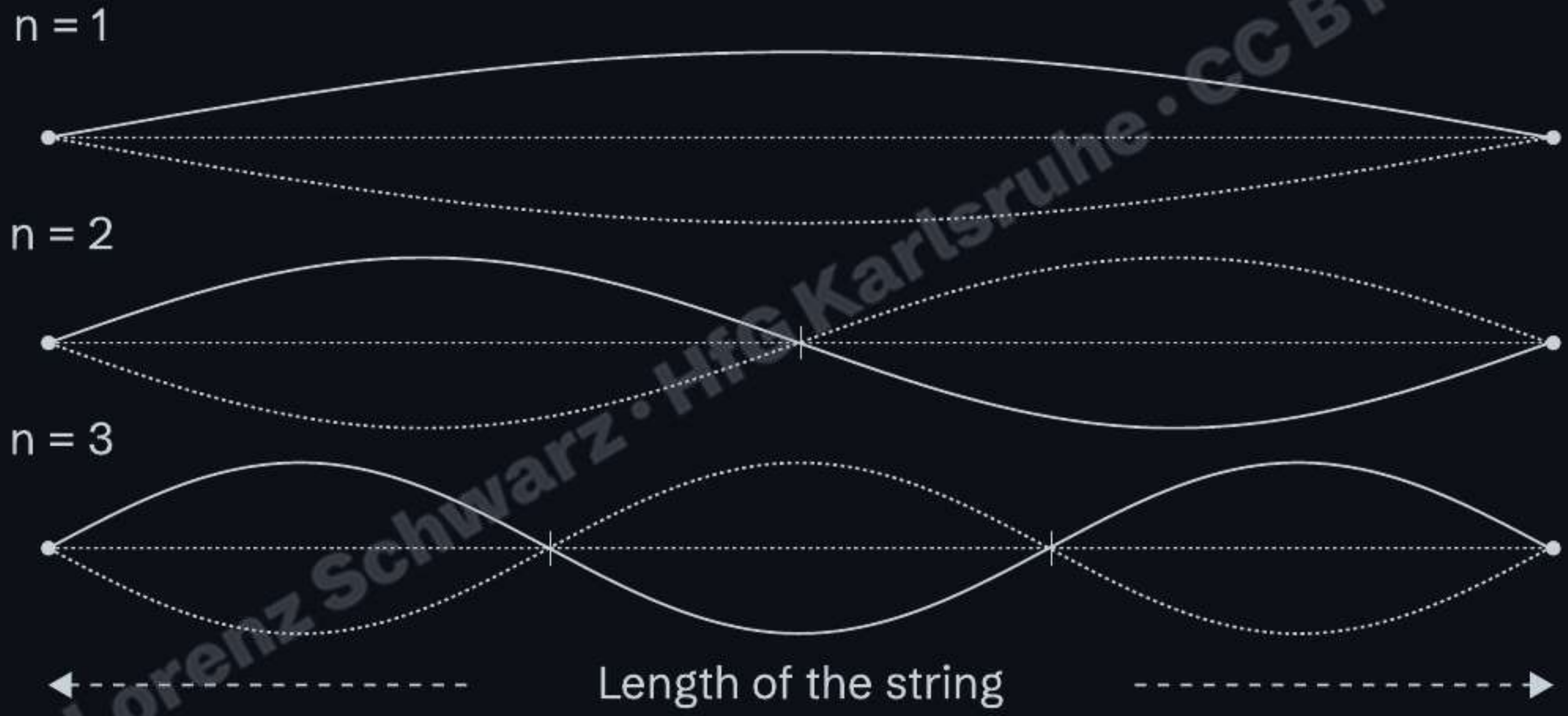
- Real sound sources excite many modes, producing complex spectra.
- Each normal mode is a sinusoid.
- Superposition of modes gives the harmonic series and timbre.

Standing waves on a vibrating string

On a vibrating string, waves travel both ways, interfere, and form standing waves with nodes (no motion) and antinodes (max motion):



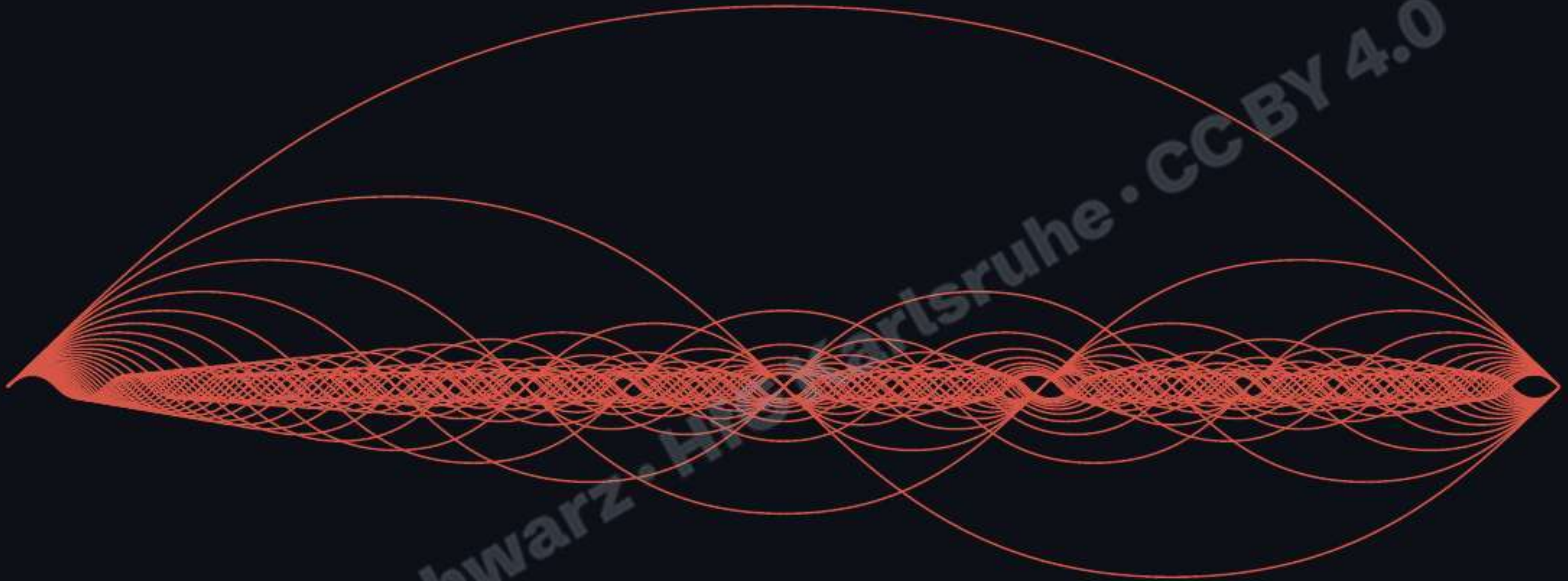
Nodes appear at rational fractions of string length.



Normal modes on a string

Normal modes are standing-wave patterns that fit the string's boundary conditions.

→ *Their frequencies are integer multiples of the fundamental.*



Standing waves in a string: it vibrates as a whole (fundamental) and in integer fractions of its length (harmonics).

Mode frequencies

Allowed frequencies are integer multiples of the fundamental f_1:

$$L = n\,\frac{\lambda_n}{2} \;\Rightarrow\; \lambda_n = \frac{2L}{n}, \qquad f_n = \frac{n\,v}{2L}\;(n = 1, 2, 3, \ldots)$$

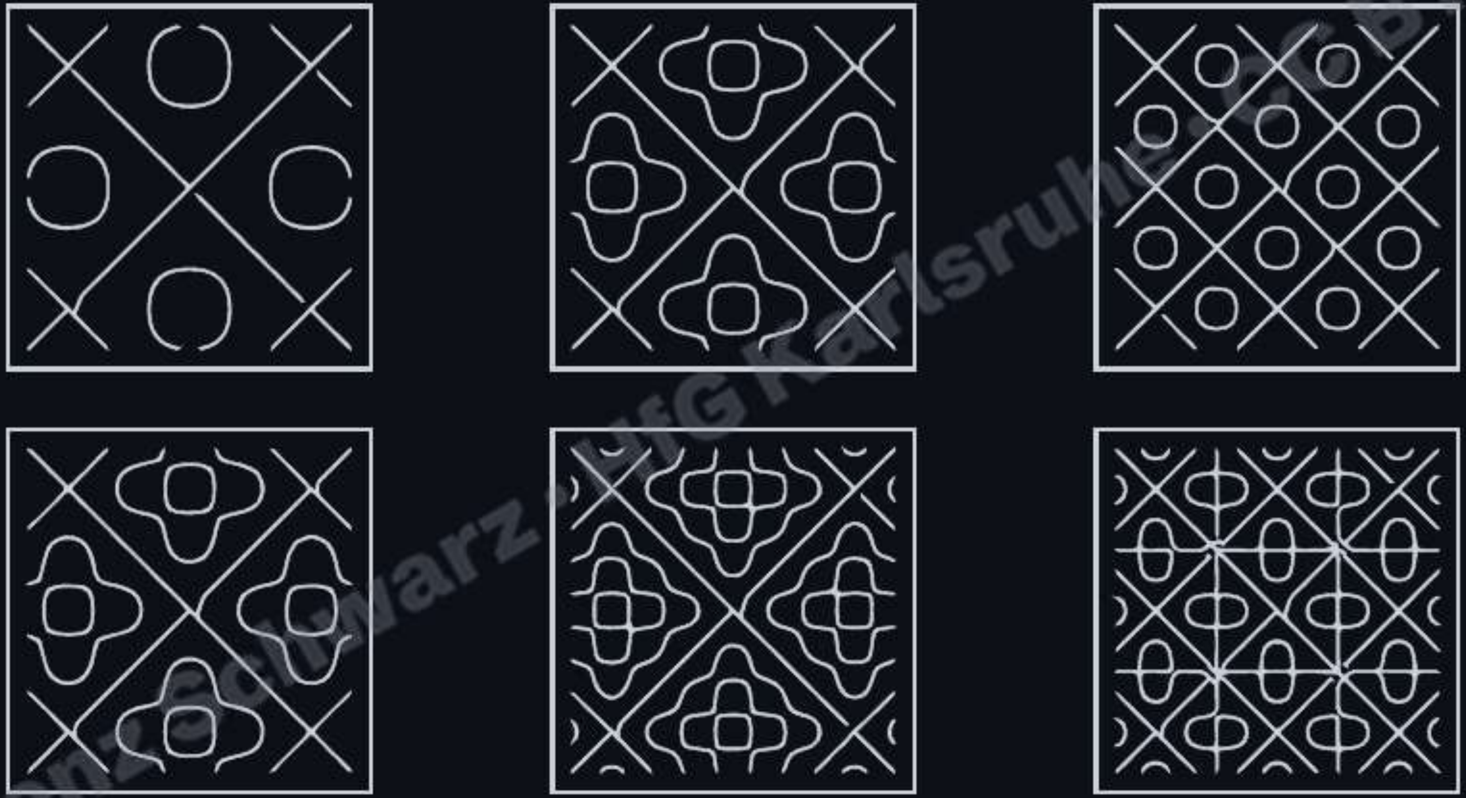
- n = mode (harmonic) number
- L = string length
- λ_n = wavelength of the n-th mode
- f_n = frequency of the n-th harmonic
- v = wave speed on the string

Inharmonic partials in real sounds

Plates, gongs, bells, drumhead membranes and other real world sound sources have **inharmonic partials.**

→ *Inharmonic Partials that are not integer multiples of the fundamental frequency.*

Chladni figures reveal allowed vibrational modes.



[View on Wolfram]

Periodic functions

A periodic function in audio describes a waveform that repeats its shape at regular time intervals. Understanding these fundamental shapes and their spectral properties is essential for sound synthesis.

1. Sine wave
2. Sawtooth wave
3. Triangle wave
4. Square wave
   - Pulse wave

Basic shapes of periodic waveforms

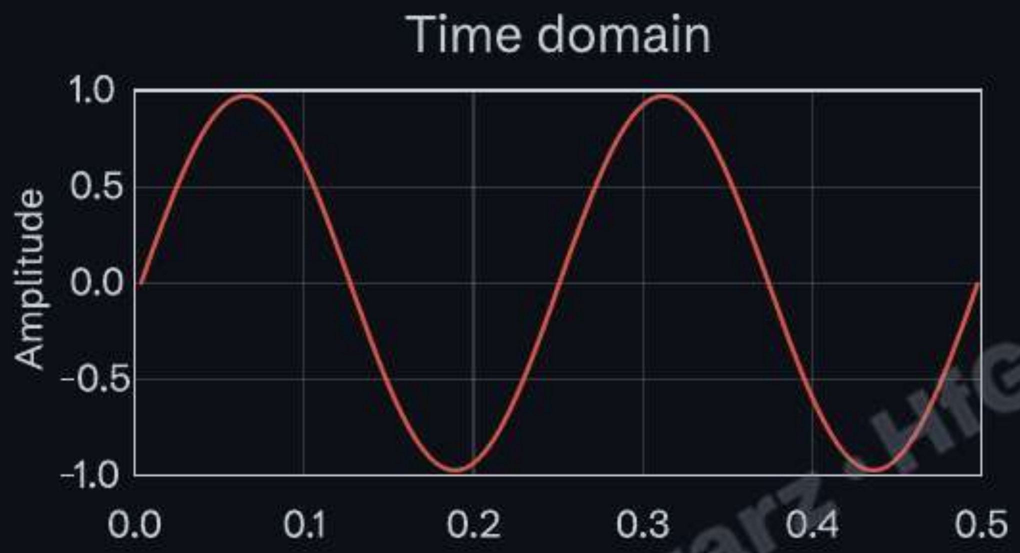
Sine wave

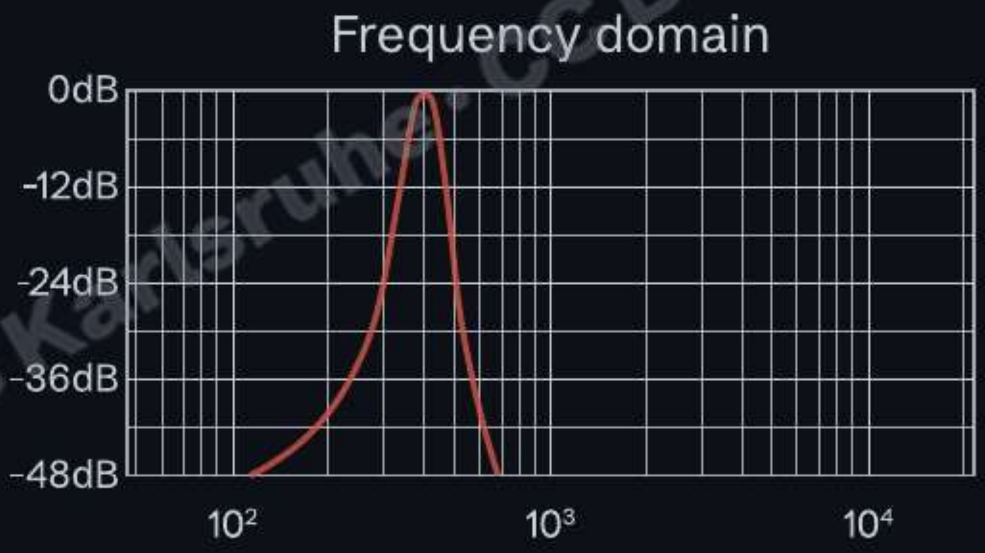
Symmetrical and curved rise and fall with no abrupt changes:

- has no overtones (single frequency, fundamental only)
- serves as a building block of periodic signals (additive synthesis)
- rarely exists alone in nature
- resembles a pipe sound, like a flute or an organ
- used often as a test tone to assess signal integrity

▶ Sine 400 Hz

Sine wave





$$x(t) = A\sin(2\pi f_0 t + \varphi)$$

[View sine wave on Desmos]

Sawtooth wave

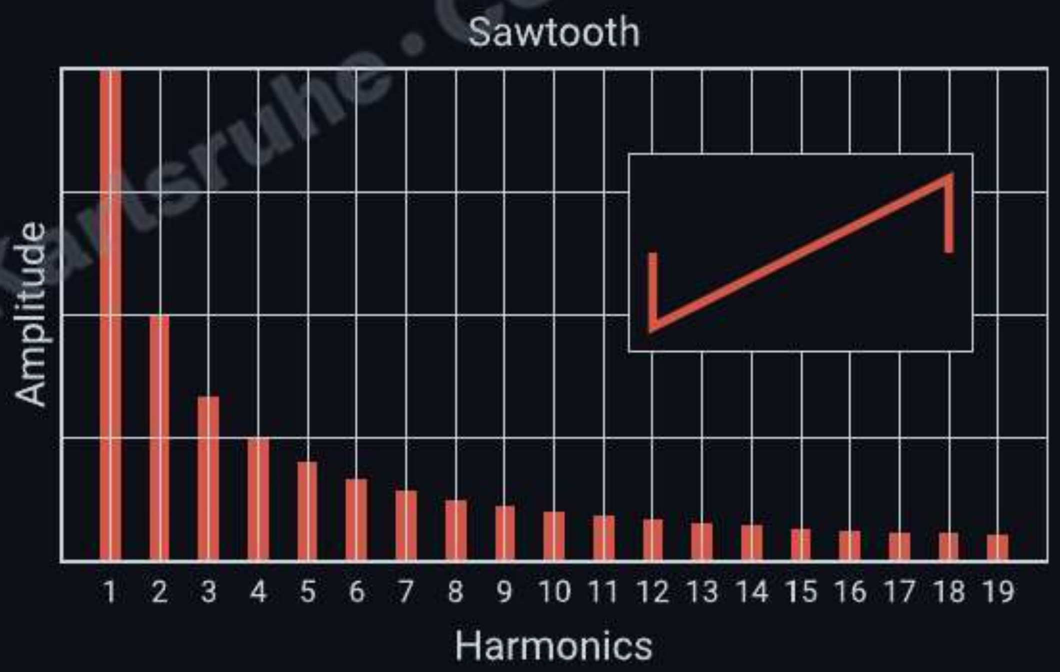
A sawtooth is characterized by a linear rise followed by an abrupt drop:

- also called ramp ('up' or 'down')
- ramp down: shifting the phase of the even harmonics by 180°
- rich and full, great for powerful synth bass and lead sounds

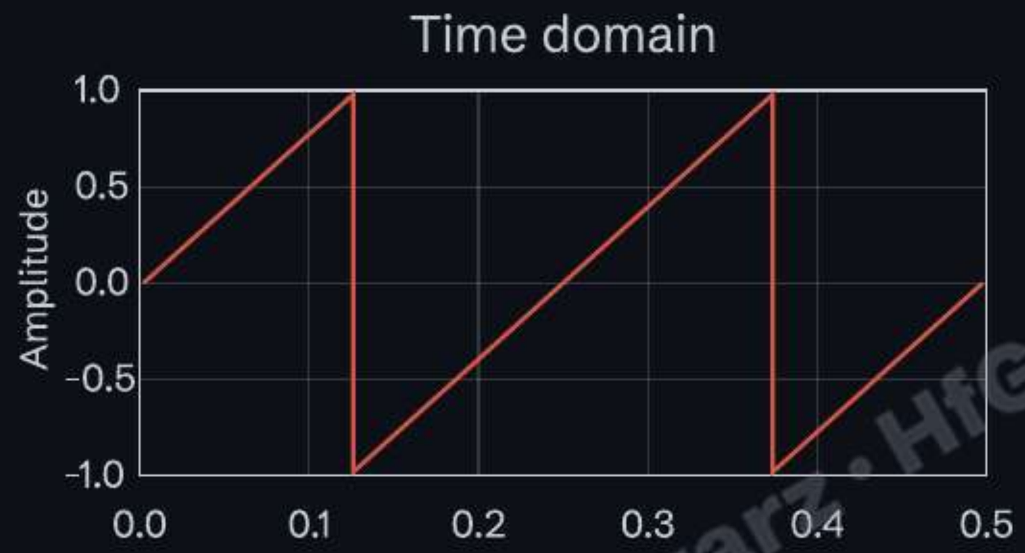
▶ Sawtooth 400 Hz

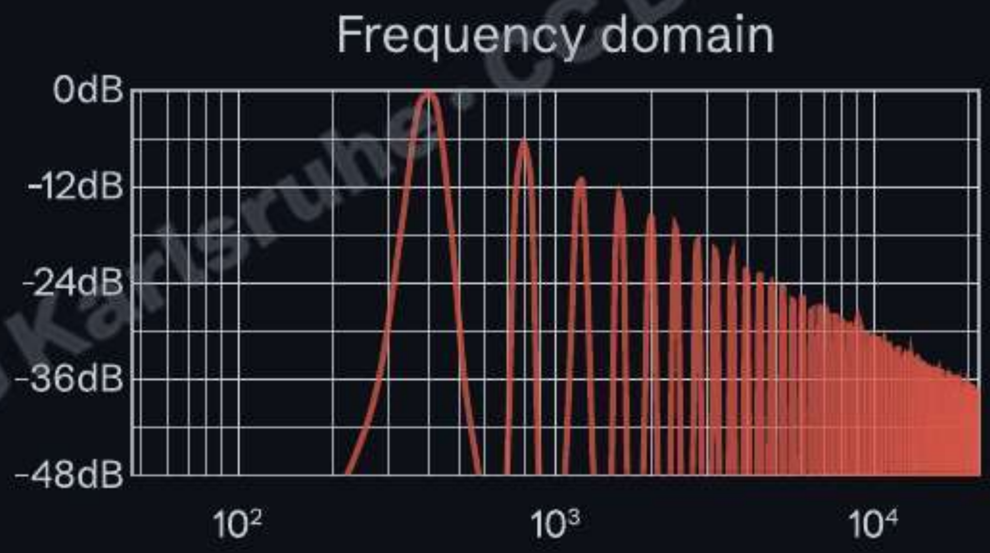
Harmonic spectrum of a sawtooth

- contains both even and odd harmonics
- relative amplitudes of harmonics are $\frac{1}{n}$



Sawtooth wave





The formula shows the waveform as a sum of sine waves [(view on Desmos)]

$$x(t) = \frac{2A}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \sin(2\pi n f_0 t)$$

Triangle wave

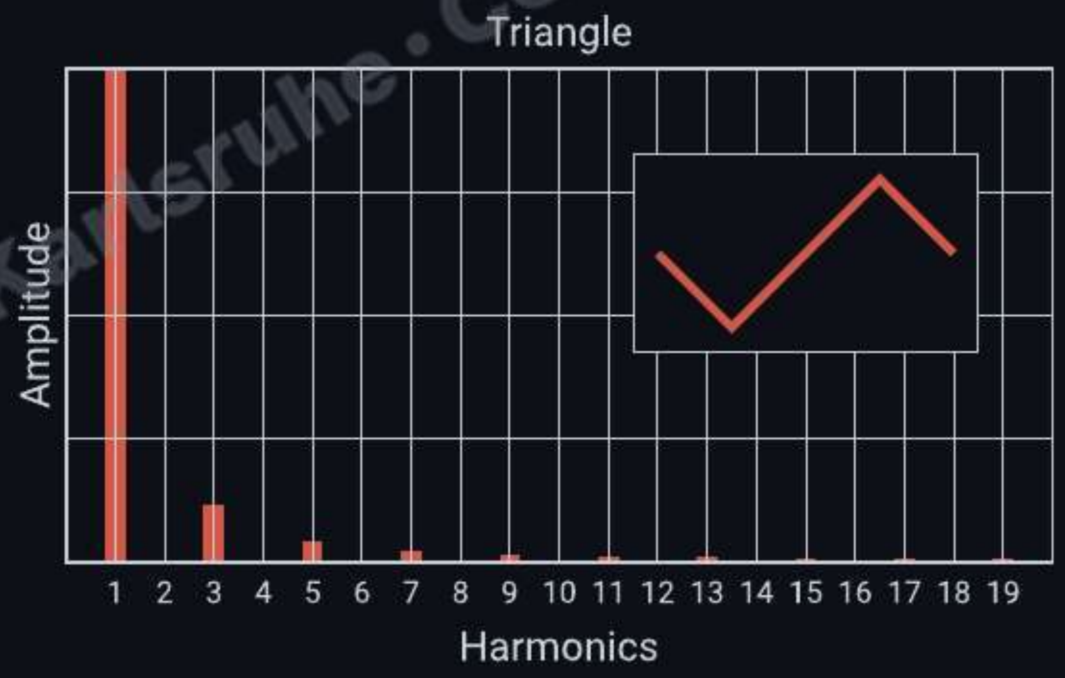
Continuous, linear rise and fall between its maximum and minimum values, forming a symmetric triangle:

- closer to a sine wave

▶ Triangle 400 Hz

Harmonic spectrum of a triangle wave

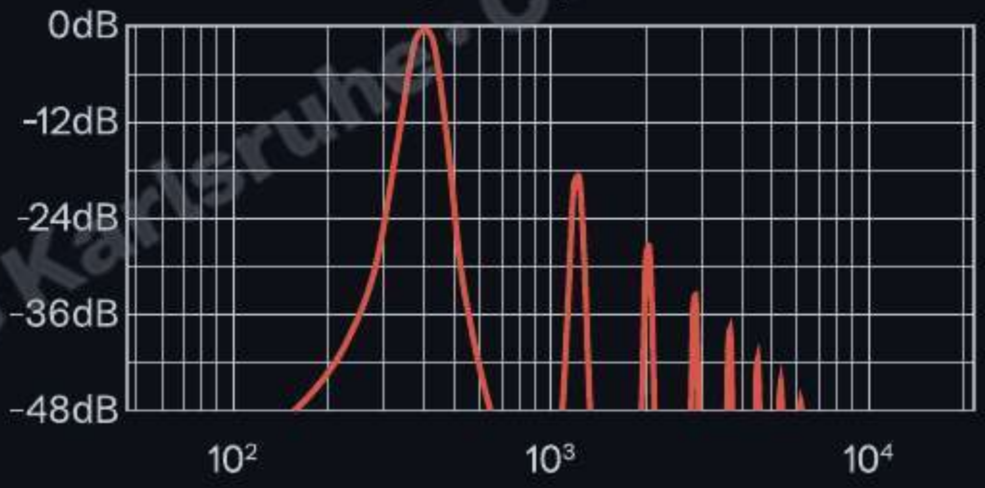
- contains only odd harmonics (1,3,5,7...)
- relative amplitudes decay as $\frac{1}{(2n-1)^2}$
- every other harmonic is 180 degrees out of phase



Triangle wave

Time domain

Frequency domain



The formula shows the waveform as a sum of sine waves (view on Desmos)

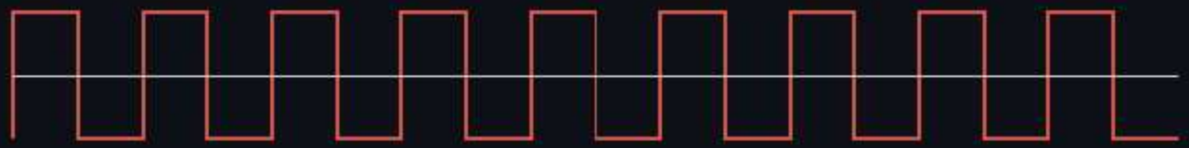
$$x(t) = \frac{8A}{\pi^2} \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{(2n-1)^2} \sin(2\pi(2n-1)\, f_0\, t)$$

Square wave

The signal spends equal time at the maximum (high) and minimum (low) levels, making it a symmetrical waveform with a 50% duty cycle (T_{ON} = T_{OFF}):

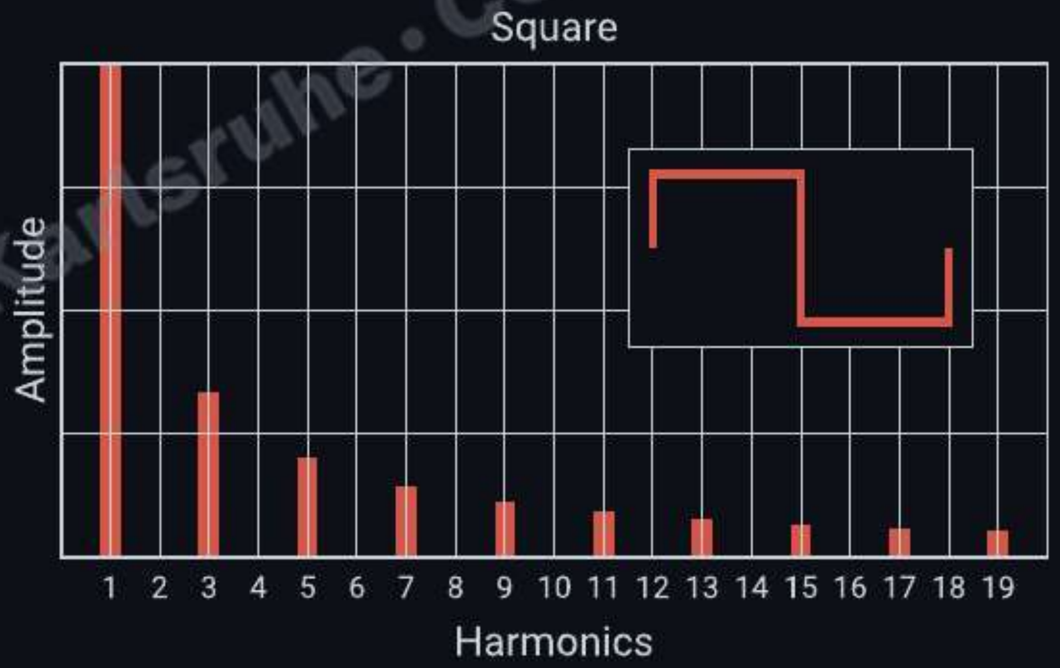
- Square waves are often described as sounding "hollow" or "nasal". This means that they are good for creating wind instruments, like a clarinet.

▶ Square wave 400 Hz

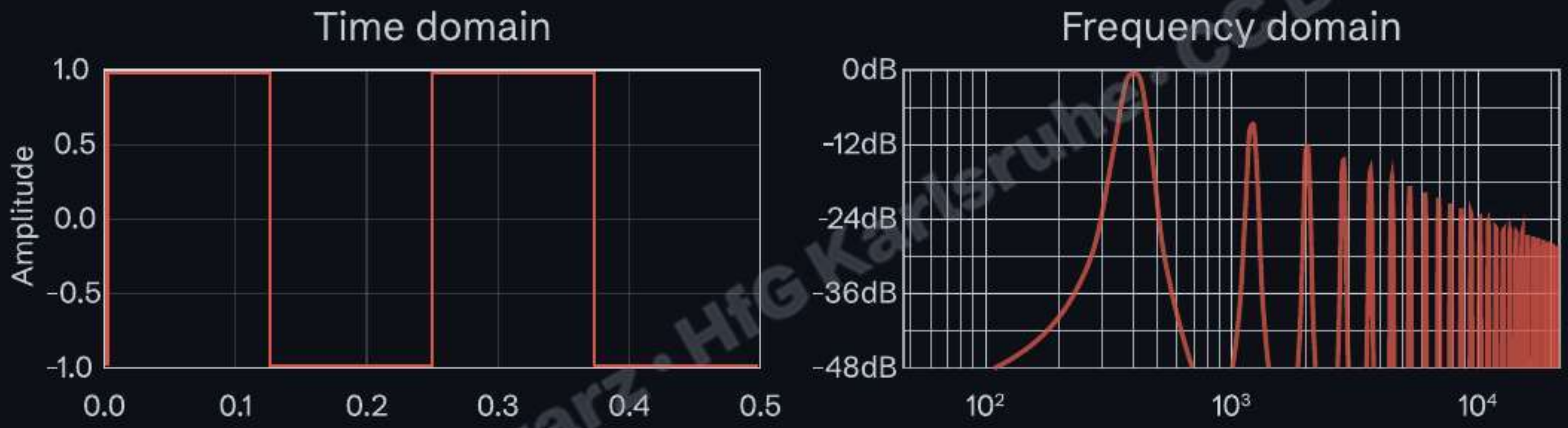


Harmonic spectrum of a square wave

- contains only odd harmonics
- relative amplitudes of harmonics are $\frac{1}{(2n-1)}$
- duty cycle of a square wave is always 50%
- all harmonics in phase



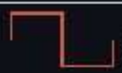
Square wave



The formula shows the waveform as a sum of sine waves [(view on Desmos)]

$$x(t) = \frac{4A}{\pi} \sum_{n=1}^{\infty} \frac{1}{2n-1} \sin(2\pi(2n-1)f_0 t)$$

Harmonic content of periodic waveforms

Waveform	Harmonics	Amplitude
 Sine	Fundamental only	-
 Sawtooth	Odd and even	-6 dB/octave ($\propto 1/n$)
 Triangle	Odd only	-12 dB/octave ($\propto 1/n^2$)
 Square	Odd only	-6 dB/octave ($\propto 1/n$)

Pulse wave

A pulse wave is a non-sinusoidal periodic signal characterized by abrupt alternation between two amplitude levels: a maximum (T_{ON}) and a minimum (T_{OFF}):

- Durations of the high and low states differ.

→ *Asymmetrical form of a square wave.*

Pulse width and duty cycle *D*

The duty cycle (D) is the percentage of a waveform's period (T_{ON}) during which the signal is in the "high" or "on" state (value of 1 for a square wave), calculated as the ratio of the on time to the total period (T_{ON} + T_{OFF}).

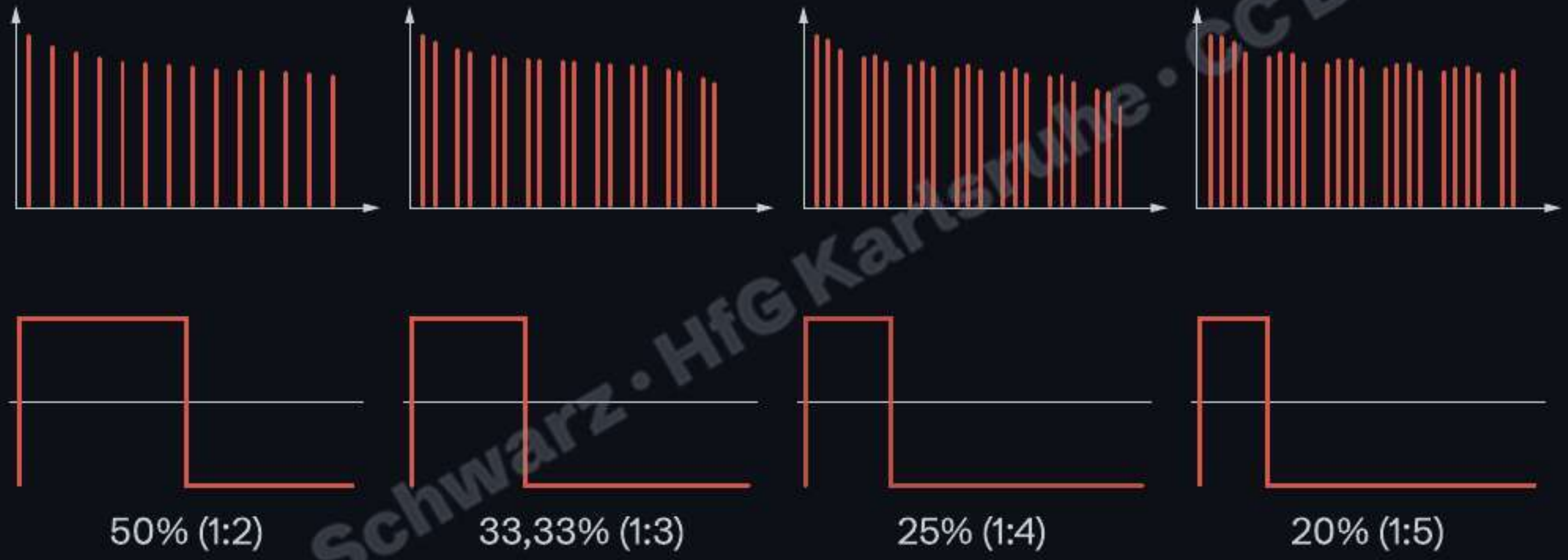
- Ratio of the pulse width to the total period

$$D = \frac{T_{ON}}{(T_{ON} + T_{OFF})} \times 100\%$$

Duty cycle

Pulse waves with different duty cycles.

Harmonic spectra of pulse waves with various duty cycles.



→ *The duty cycle determines the harmonic spectrum of the pulse wave.*

▶ Play examples

Pulse width and timbre

Changing the duty cycle alters the harmonic structure and perceived timbre of a pulse wave.

- Narrowing the duty cycle from 50 % produces a thinner, more nasal sound
- Very narrow pulses create a characteristic reed-like quality
- Commonly used for string- and brass-like synth sounds

→ *Modulating the duty cycle over time (PWM) creates dynamic, evolving timbres*

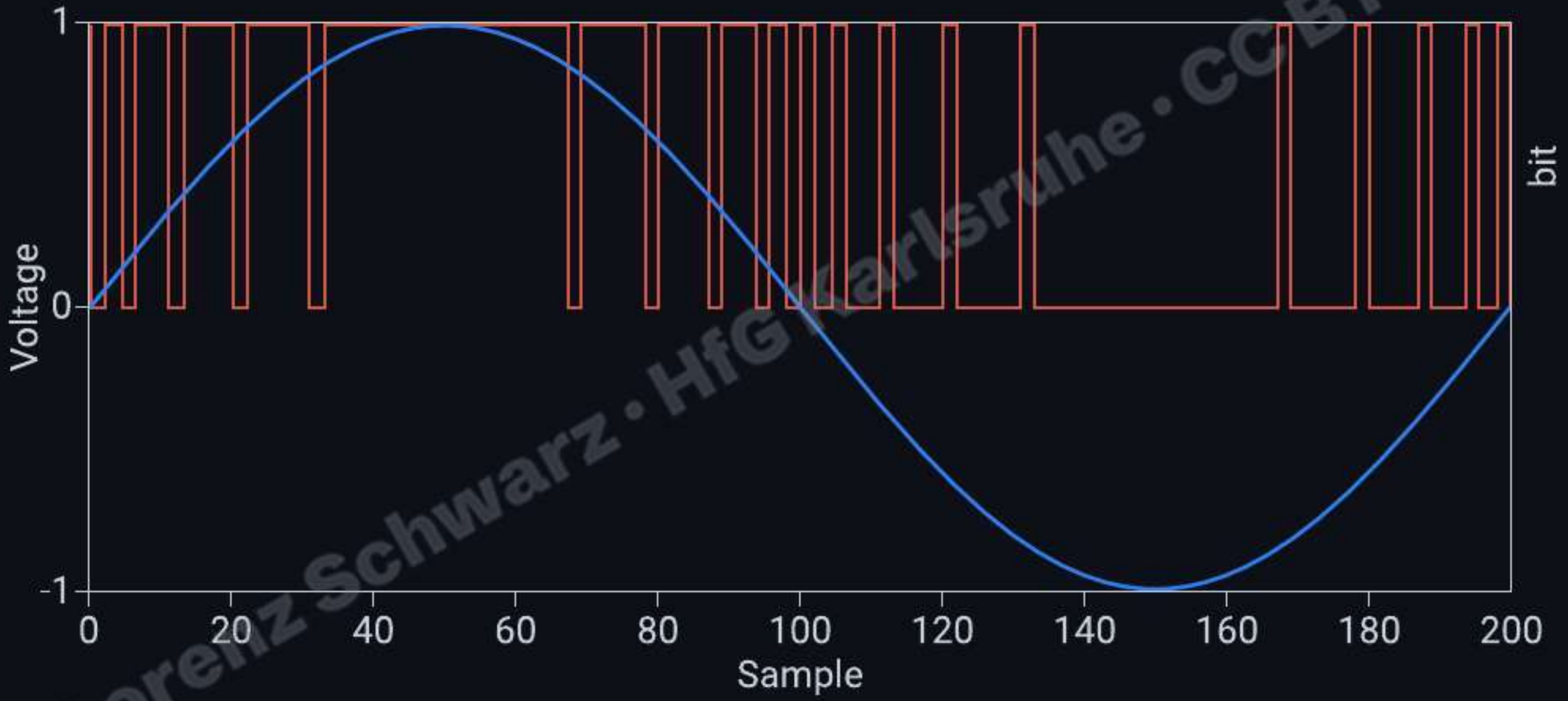
Pulse Width Modulation (PWM)

PWM (Pulse Width Modulation) is a type of signal modulation that converts an analog signal into a binary-coded signal by varying the duty cycle of a pulse wave in direct proportion to the amplitude of the analog signal.

Applications of PWM:

- Class-D amplifiers, light dimmers/LED brightness control
- Variable-speed control for computer fans, and servo motor
- Sound synthesis

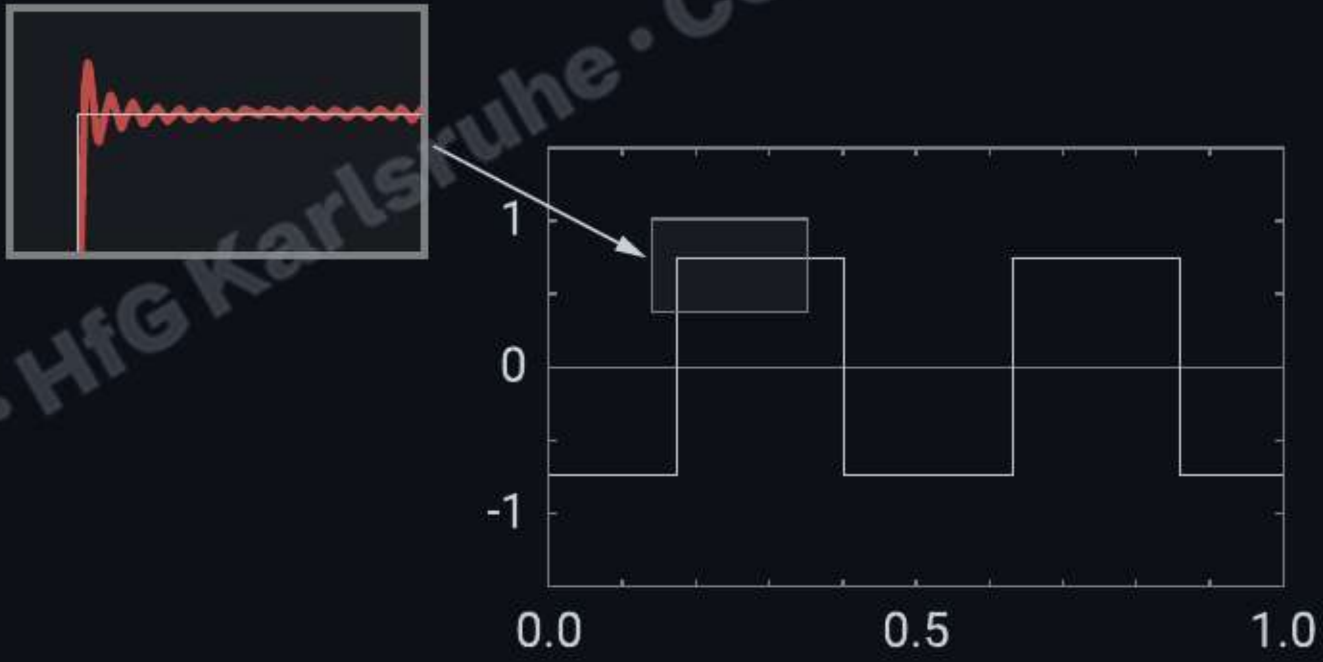
Amplitude values are encoded into pulses.



Gibbs phenomenon

Approximating a discontinuous function (such as a square wave or a sawtooth wave) by a finite sum of continuous sine waves causes:

- Oscillations at the jump discontinuities occur.
- The overshoot does not vanish, even as more terms are added.

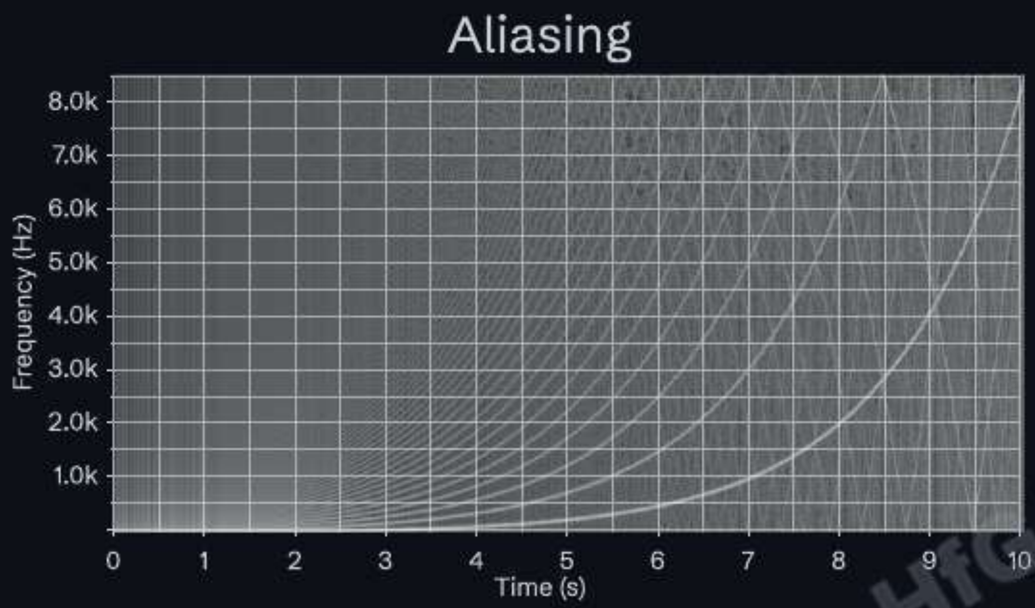


Digital waveforms

In digital systems, frequency components above the Nyquist frequency (half the sampling rate) are mirrored back into the audible range, creating new, non-harmonic frequencies.

This effect is called aliasing.

→ *Steep low-pass filtering before sampling minimizes aliasing (band-limited synthesis)*



- Square wave sweep without band-limiting
- Square wave sweep with band-limiting

Stochastic signals

In contrast to periodic signals, stochastic signals (noise) are random and non-repeating, and are described primarily by their spectral distribution rather than their waveform shape.

1. White noise
2. Pink noise
3. Brownian/Red noise
4. Blue/Azure noise
5. Violet/Purple noise

Types of power-law colored noise

The term colored noise refers to signals whose power distribution across frequencies is roughly similar to the corresponding spectra of visible light.

→ *An imprecise analogy inspired by filtering white light.*

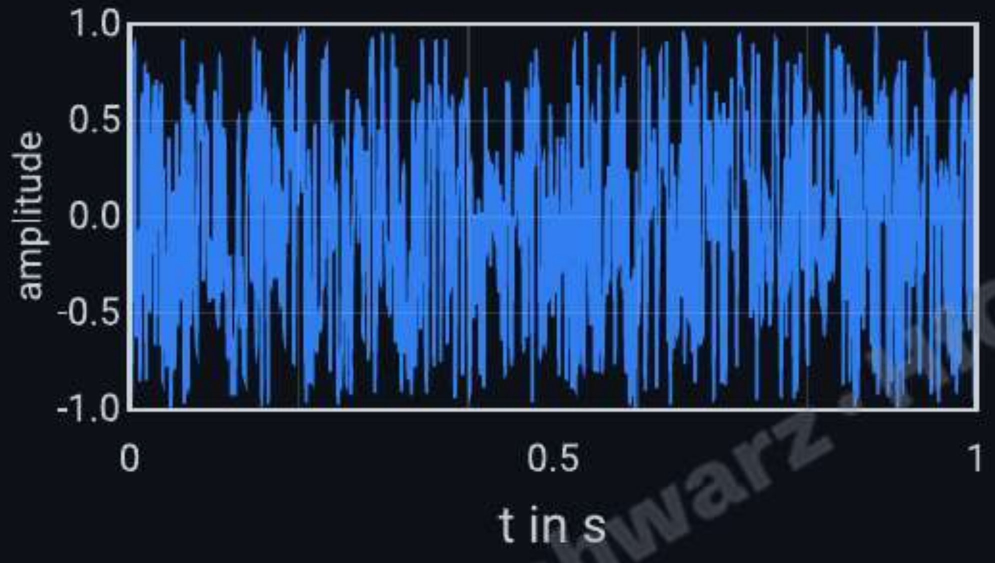
White noise

White noise, analogous to white light which contains all spectral components, has equal energy distributed across all frequencies with a constant power spectral density.

Applications:

- Audio synthesis
- Testing and calibrating audio equipment

▶ White noise


1.0
0.5
0.0
-0.5
-1.0
amplitude
0
0.5
1
t in s

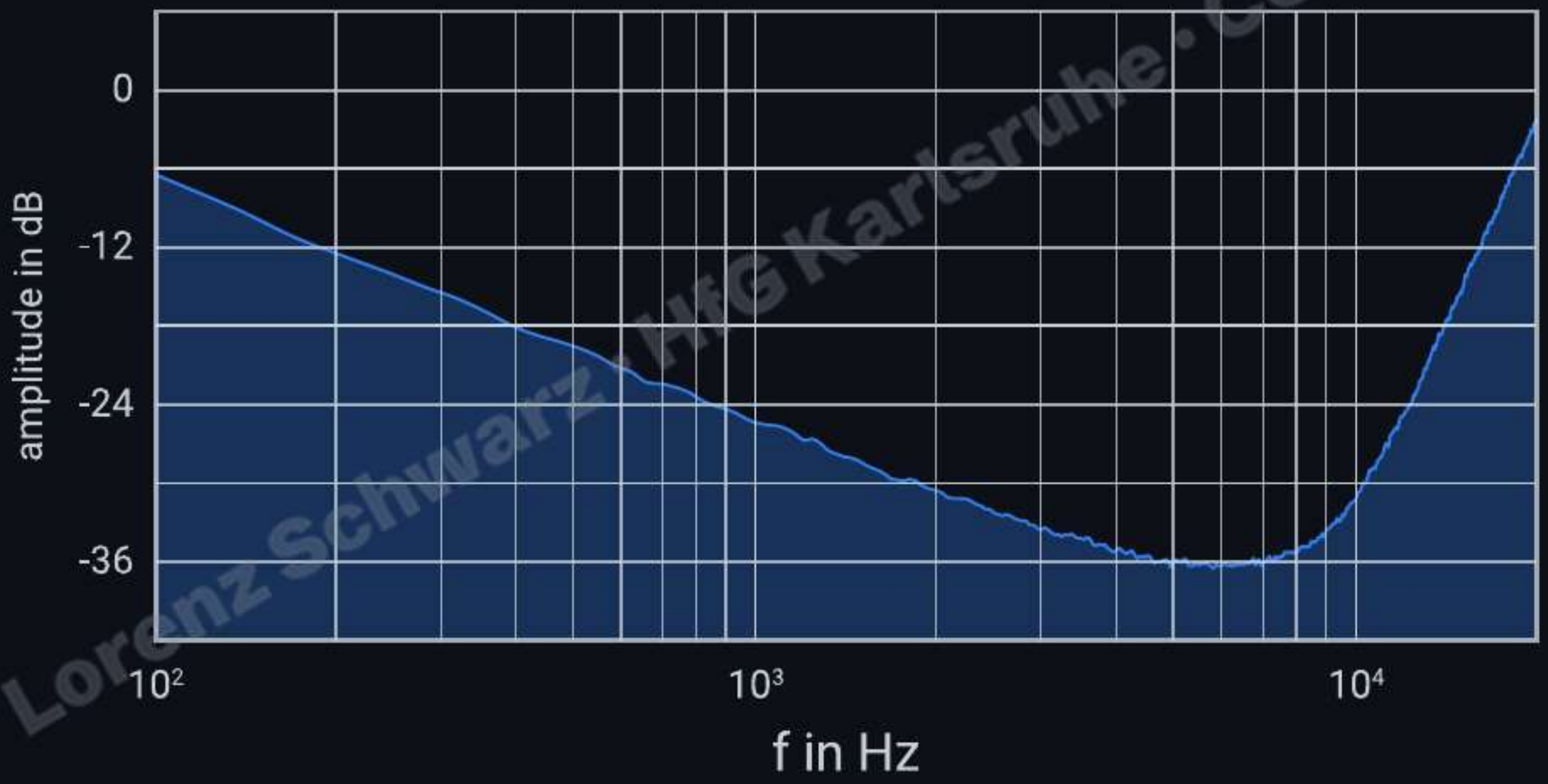

0
-12
-24
-36
amplitude
10^2
10^3
10^4
f in Hz

Grey noise

While white noise has physically equal energy at all frequencies, grey noise has perceptually equal loudness.

→ *Inverse of the equal-loudness curve (A-weighting) compensates for the human ear's varying sensitivity across the frequency spectrum.*

▶ Grey noise



Pink noise

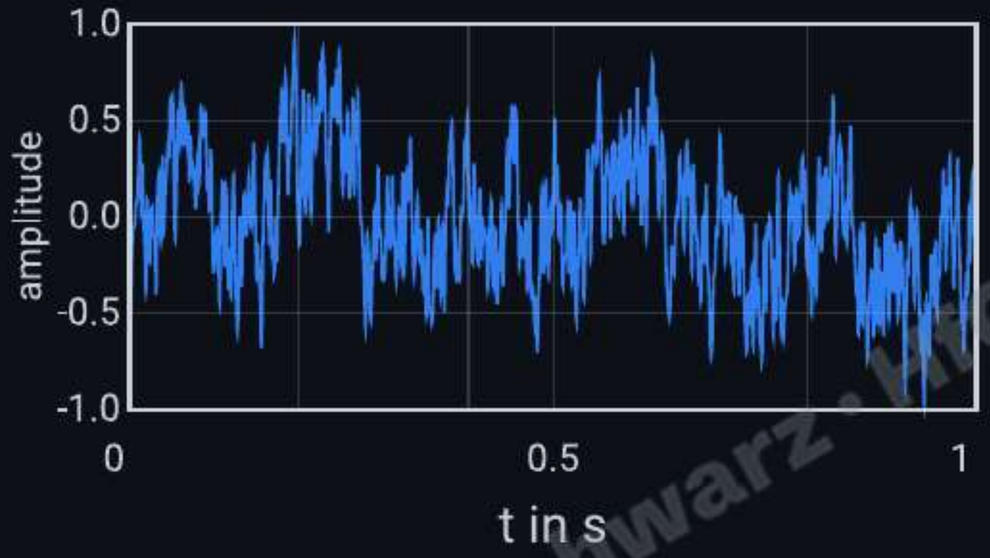
Pink noise has equal power per octave, meaning its power decreases as frequency increases.

- Power spectrum is inversely proportional to frequency ($\sim 1/f$)
- Found in natural and biological systems
- Sounds like a steady waterfall or rainfall
- Also called flicker noise in electronics

Application:

- Commonly used for acoustic measurements and sound system calibration.

▶ Pink noise


1.0
0.5
0.0
-0.5
-1.0
amplitude
0
0.5
1
t in s

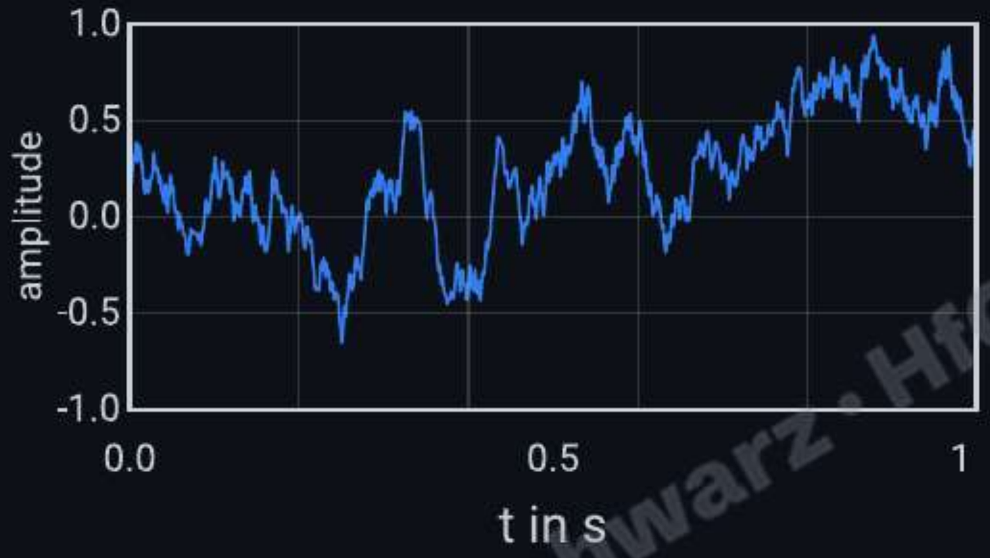

0
-12
-24
-36
amplitude
10^2
10^3
10^4
f in Hz

Brownian noise ($\sim 1/f^2$ noise)

Brownian noise, also known as Brown noise or Red noise, has a power density that decreases by 6 dB per octave (or 20 dB per decade), emphasizing lower frequencies.

- It approximates the random patterns of Brownian motion.
- Named after Robert Brown, who discovered Brownian motion in 1827.

▶ Brownian noise



Blue/Azure noise

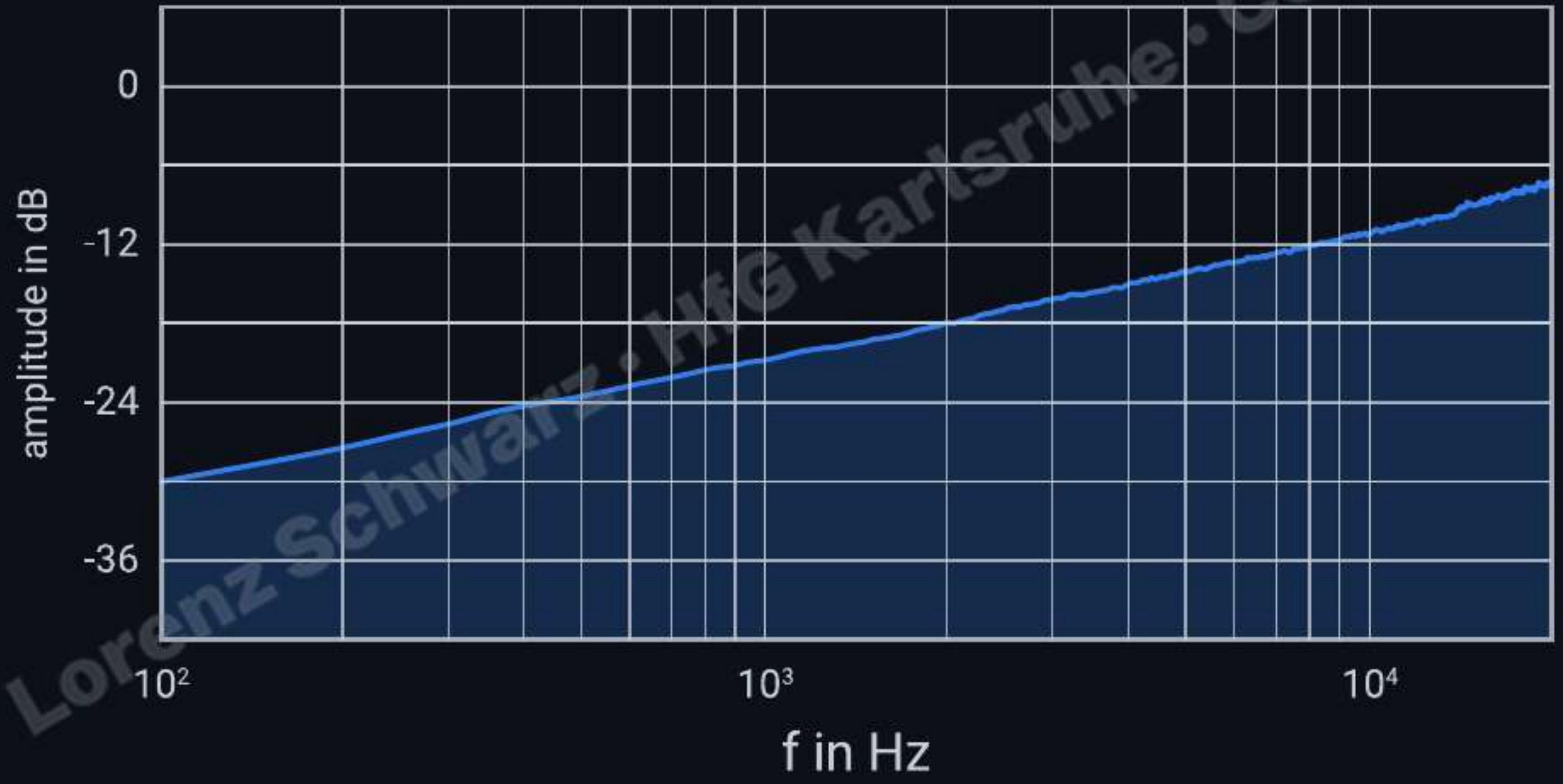
Blue noise's power density increases by 3 dB per octave as the frequency increases.

- Inverse of pink noise.
- Proportional to frequency.

Application:

- Dithering

▶ Blue noise

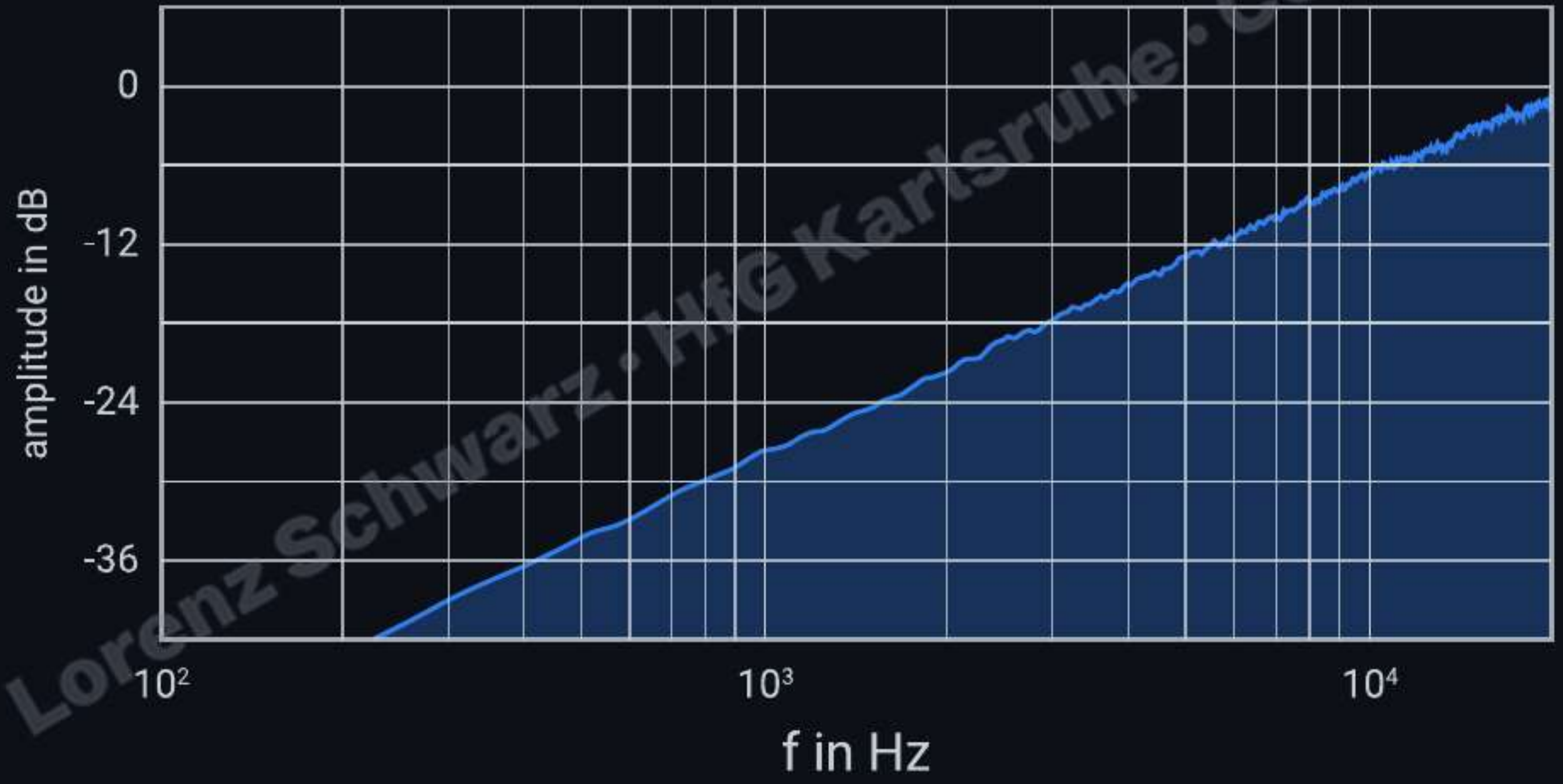


Violet/Purple noise

Power density increases by 6 dB/octave with frequency.

- Differentiated white noise
- Opposite of brown noise

▶ Purple noise



Spectral properties of noise types

Type of Noise	Spectral Density	Change per Octave (dB)	$1/f^\alpha$
White	$S(f) \propto 1$	0 dB	$\alpha = 0$
Pink	$S(f) \propto 1/f$	-3 dB	$\alpha = 1$
Brownian / Red	$S(f) \propto 1/f^2$	-6 dB	$\alpha = 2$
Blue / Azure	$S(f) \propto f$	+3 dB	$\alpha = -1$
Violet / Purple	$S(f) \propto f^2$	+6 dB	$\alpha = -2$

Additive synthesis

Build complex sounds by adding sine waves together

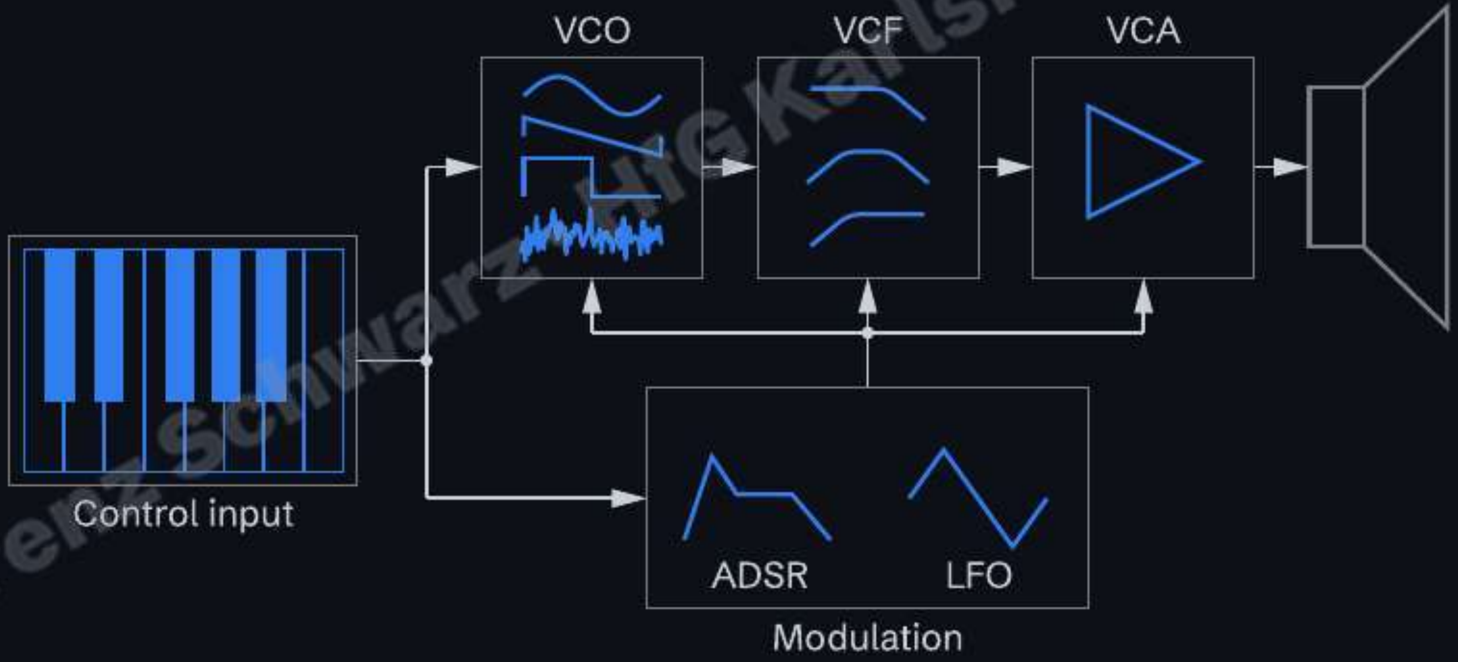
$$x(t) = \sum_{k=1}^{K} A_k(t) \sin(2\pi k f_0 t + \varphi_k)$$

- Requires many oscillators (one per partial)
- Controlling timbre changes over time is complex
- Computationally expensive

→ *Additive synthesis: conceptually simple, practically expensive*

Subtractive sound synthesis

Subtractive synthesis starts with rich, periodic waveforms (like sawtooth or square) and removes frequencies using filters.



Signals for measurement and analysis

Beyond sound synthesis, certain signals are designed specifically for measuring and analyzing acoustic systems, such as room reverberation and loudspeaker response.

Dirac delta function (δ distribution)

A mathematical function with infinite amplitude at a single point and infinitely small duration.

Application:

- Impulse response measurement

Impulse and unit sample

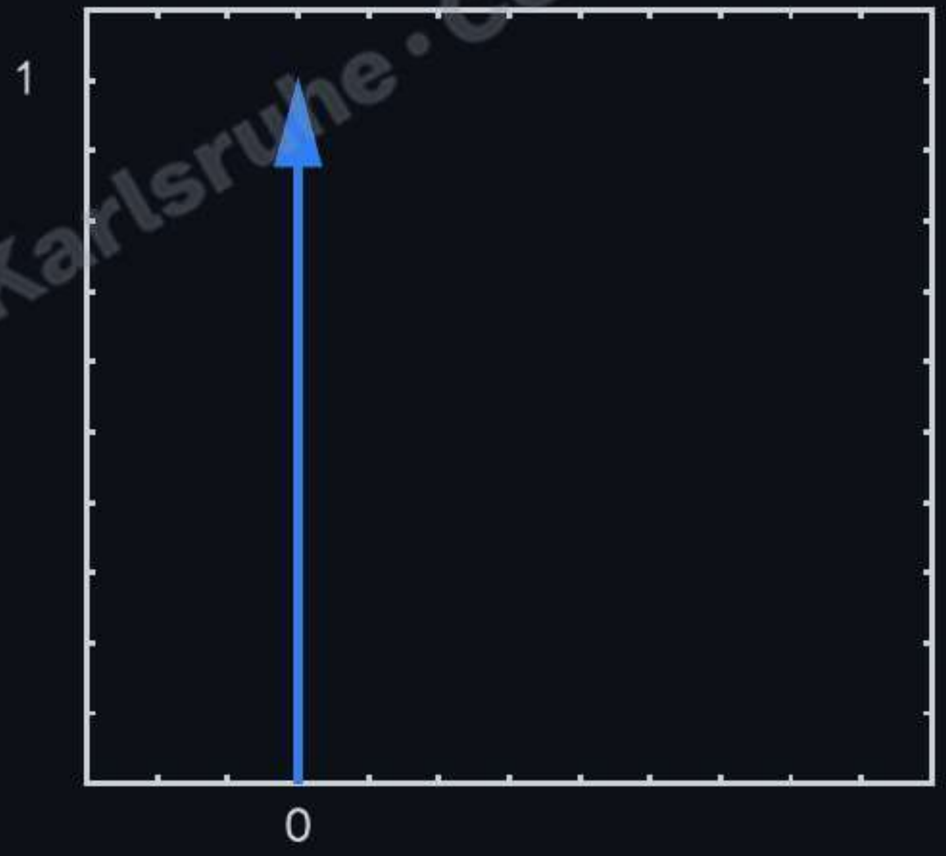
The discrete unit sample is the digital equivalent of the Dirac delta:

$$\delta(n) \mathrel{\widehat{=}} \begin{cases} 1, & n = 0 \\ 0, & n \neq 0. \end{cases}$$

Real-world approximations:

- Balloon pop, gunshot
- Electromagnetic interference

▶ Dirac Impulse



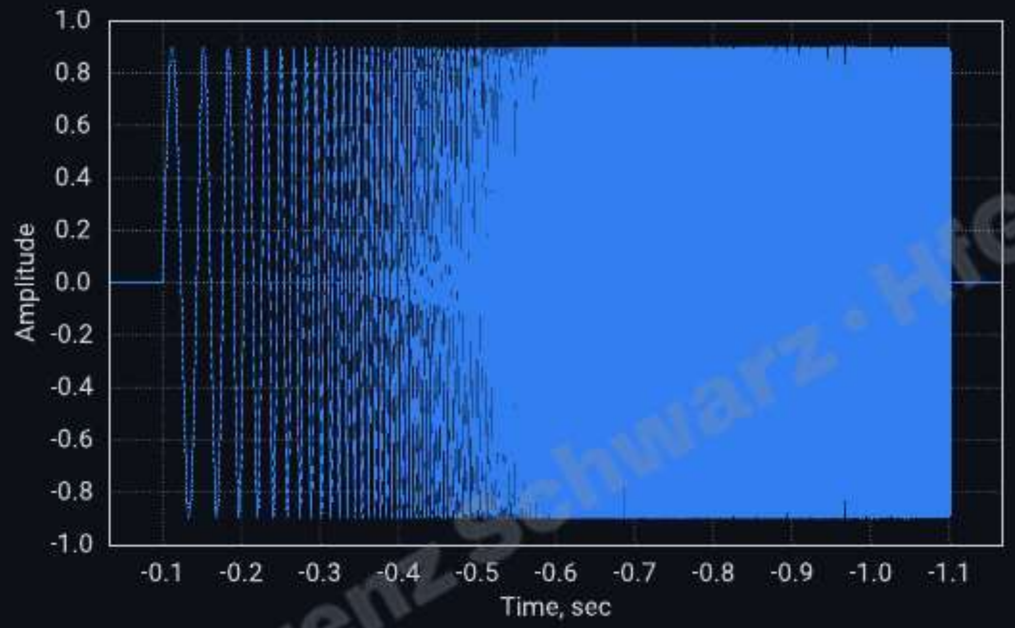
Sine sweep (chirp)

A sine sweep uses a sinusoid with an increasing frequency to excite an acoustic system, enabling the calculation of its impulse response.

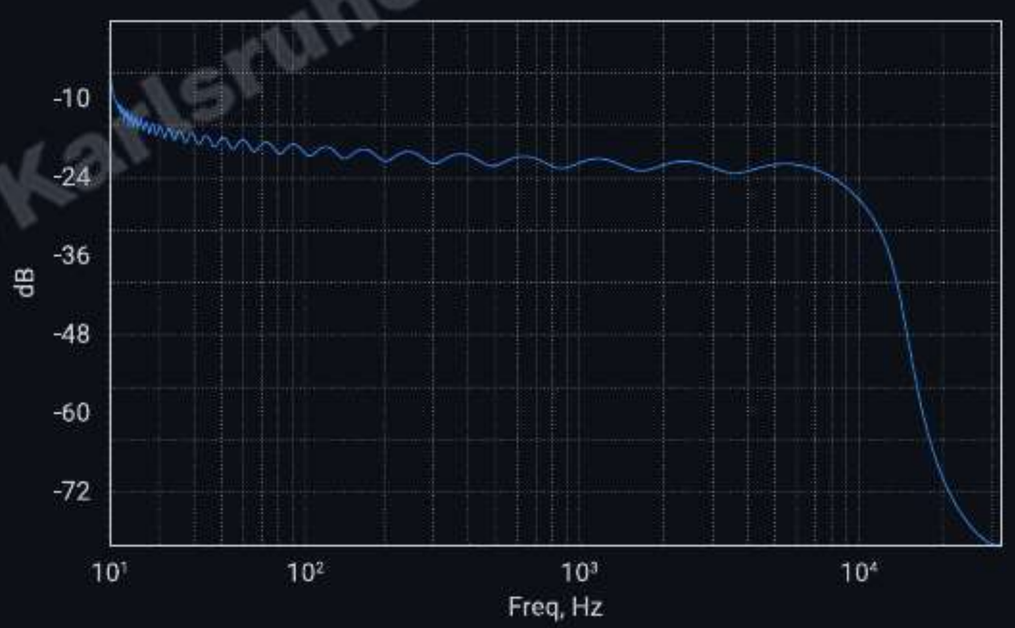
Application:

- Impulse response measurement

▶ Sine sweep



1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
Amplitude
-0.1 -0.2 -0.3 -0.4 -0.5 -0.6 -0.7 -0.8 -0.9 -1.0 -1.1
Time, sec



-10
-24
-36
-48
-60
-72
dB
10^1 10^2 10^3 10^4
Freq, Hz

SPECTRAL ANALYSIS

Fourier transform and frequency domain

Sound in different domains

An audio signal can be described from different perspectives, depending on which aspect of sound is being analyzed.

- Time-based descriptions reveal changes over time
- Frequency-based descriptions reveal spectral content

→ *Both views describe the same signal, but reveal different information.*

Time domain vs. frequency domain

Audio signals are commonly represented in two domains:

- Time domain:
  - Amplitude as a function of time
- Frequency domain:
  - Distribution of energy across frequencies (magnitude and phase)

(Analog signals are continuous; digital signals are discrete.)

Time domain

The time domain shows how a signal's amplitude changes over time.

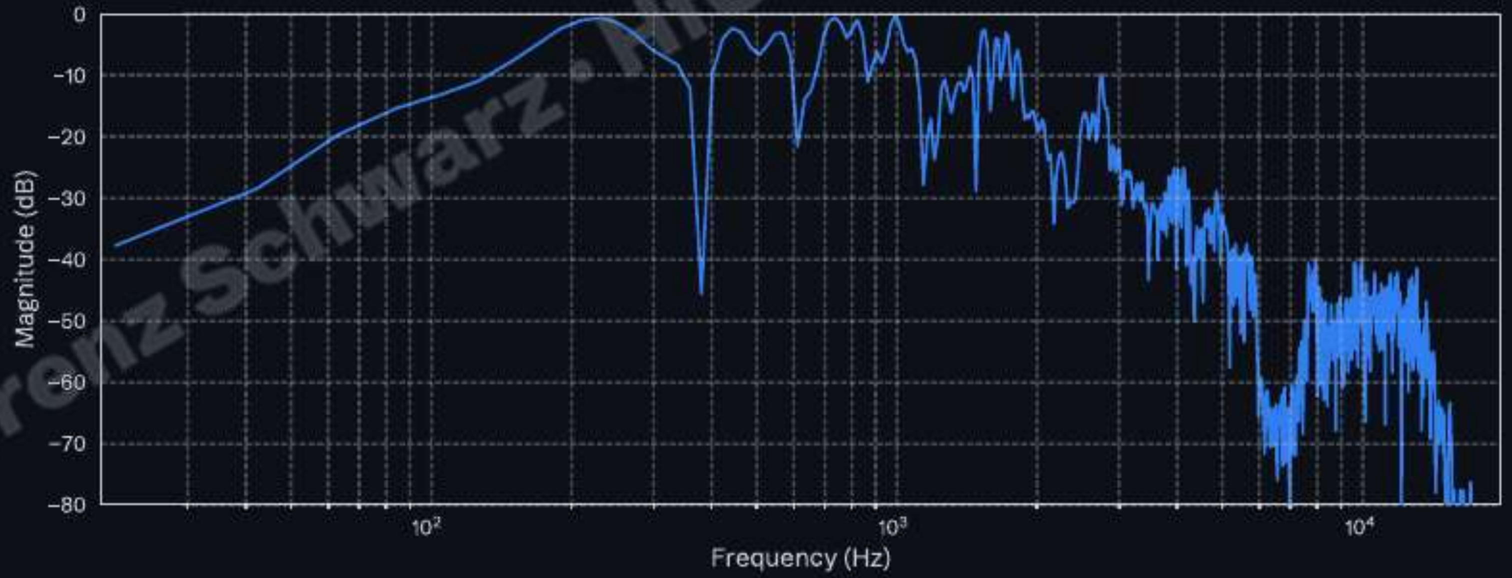
- Analog signals: continuous in time and amplitude
- Digital signals: discrete samples in time and amplitude

→ *The time domain reveals transients, timing, and amplitude changes.*

Frequency domain

The frequency domain describes a signal in terms of its frequency components rather than time.

- shows how much energy each frequency contributes



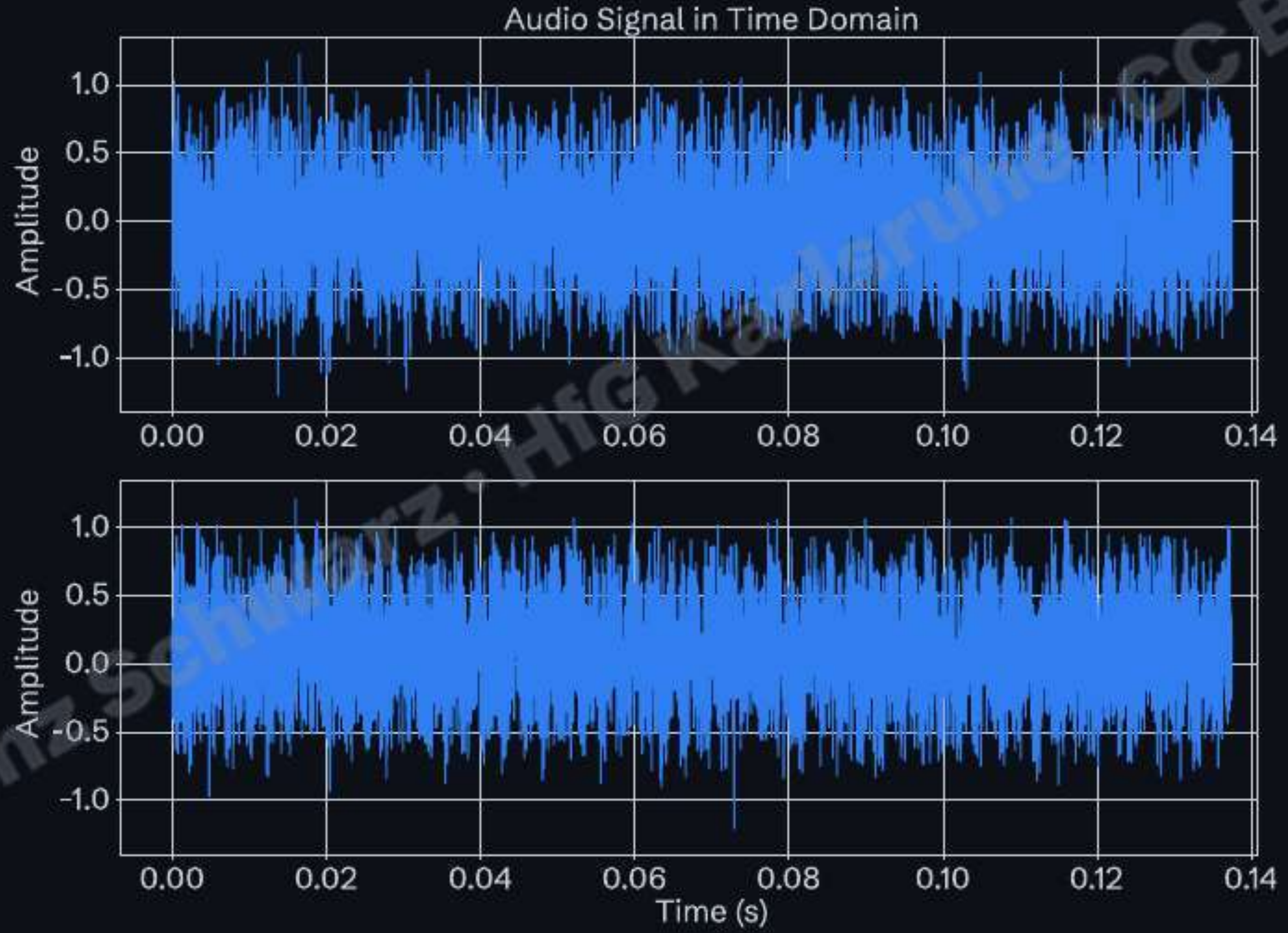
Spectrum

The spectrum shows a moment of a signal's frequency content.

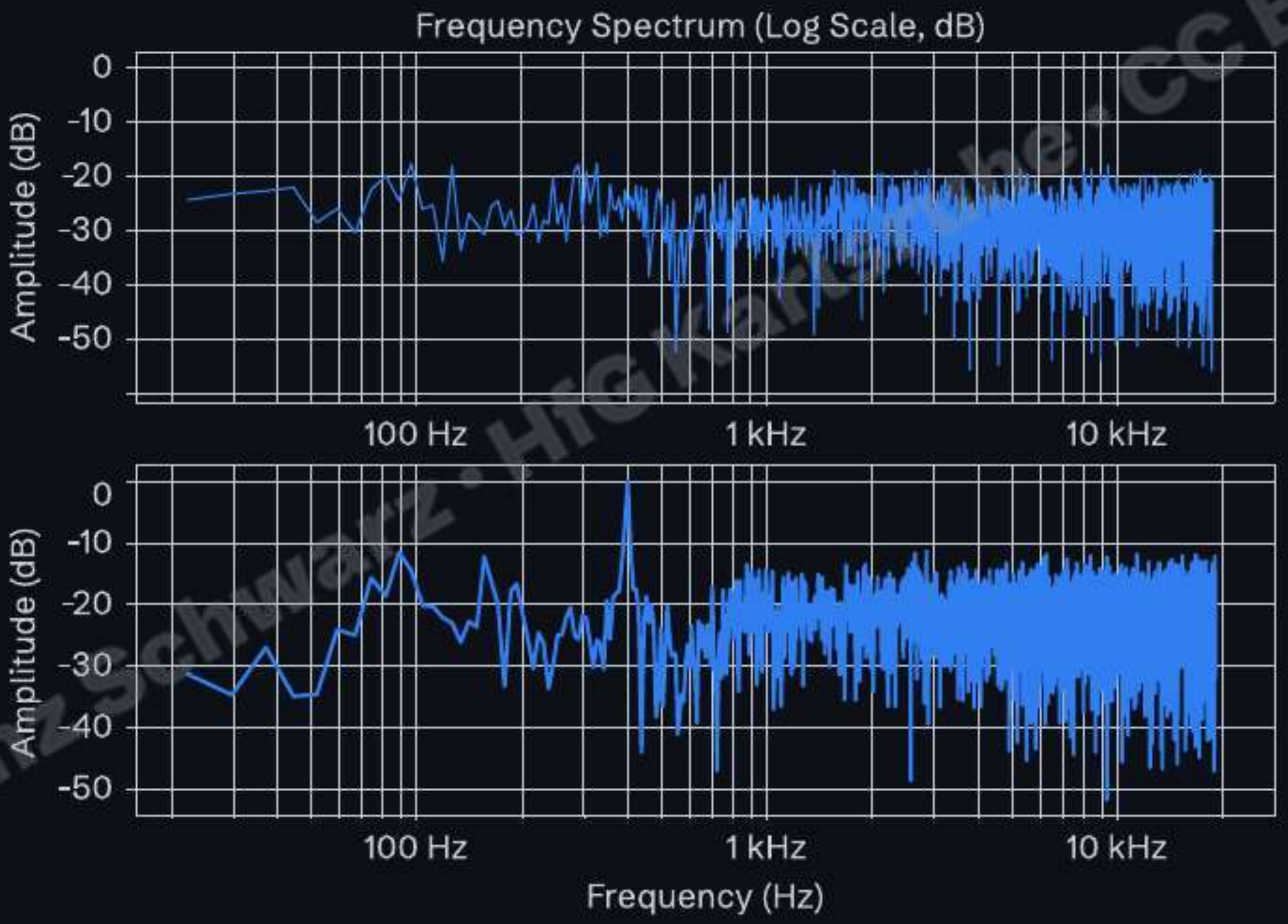
- Amplitude (or magnitude) as a function of frequency
- Optionally includes phase information

→ *The spectrum is the primary tool for analyzing timbre and harmonic structure.*

Which signal contains a 400 Hz sine?



The second spectrum shows a spike at 400Hz



Timbre and spectrum

Timbre (perceptual):
The sonic quality that distinguishes instruments playing the same pitch

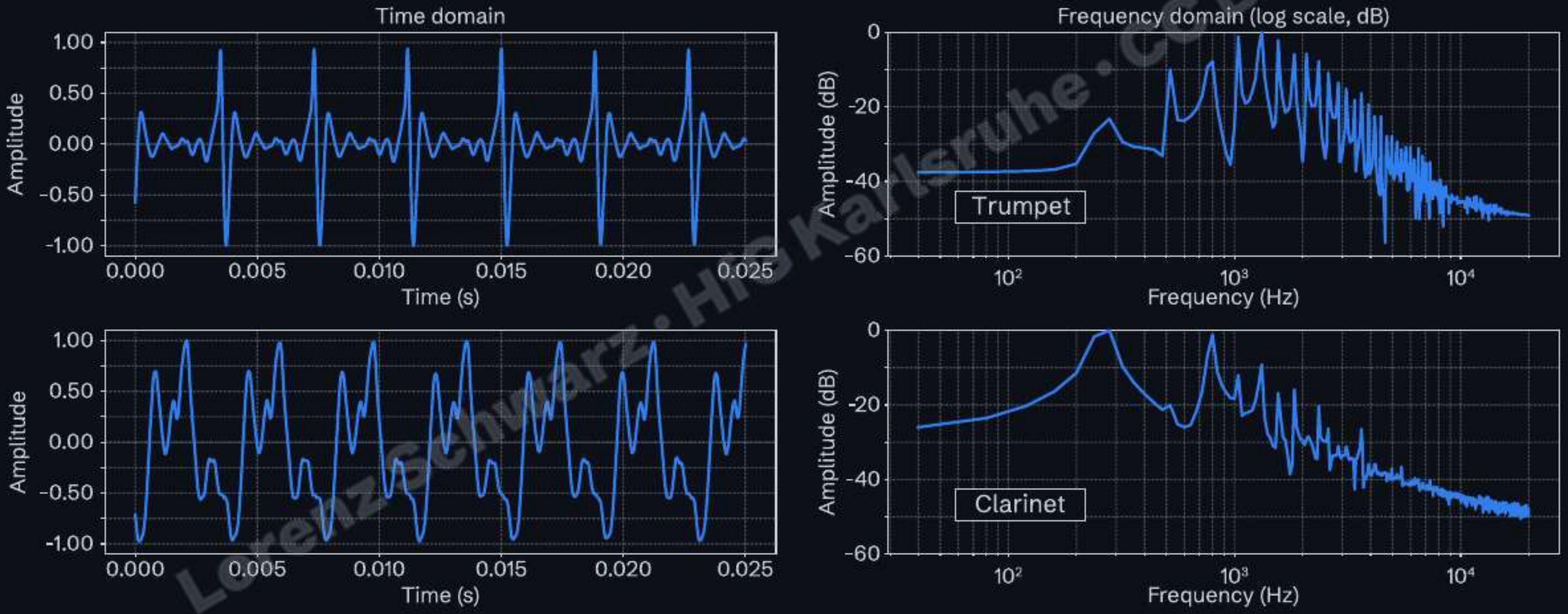
Spectrum (technical):
The distribution of frequency components and their amplitudes

Listen to the same pitch (C4 ≈ 261.6 Hz):

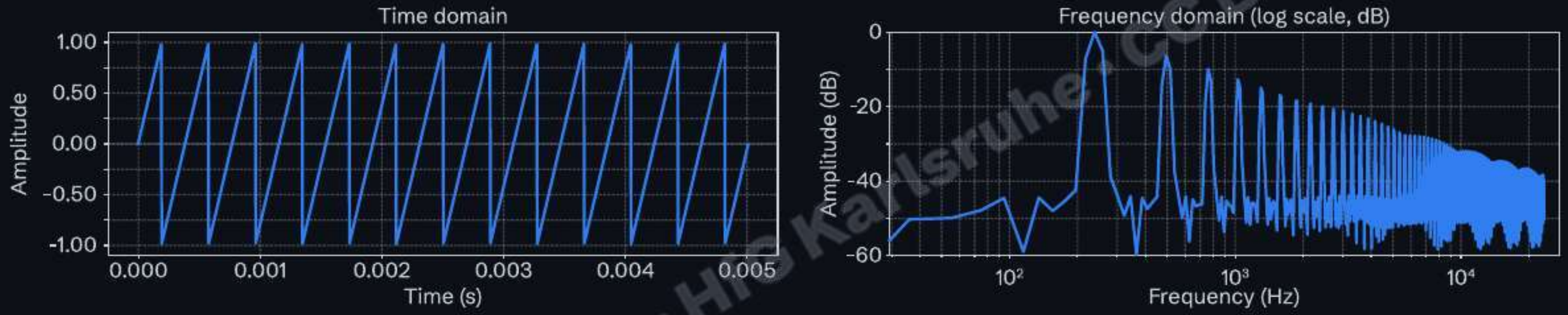
Trumpet: ▶ Play **Clarinet:** ▶ Play

→ *Same pitch, different timbre*

Comparing clarinet and trumpet at 260 Hz



Understanding spectra with a sawtooth wave

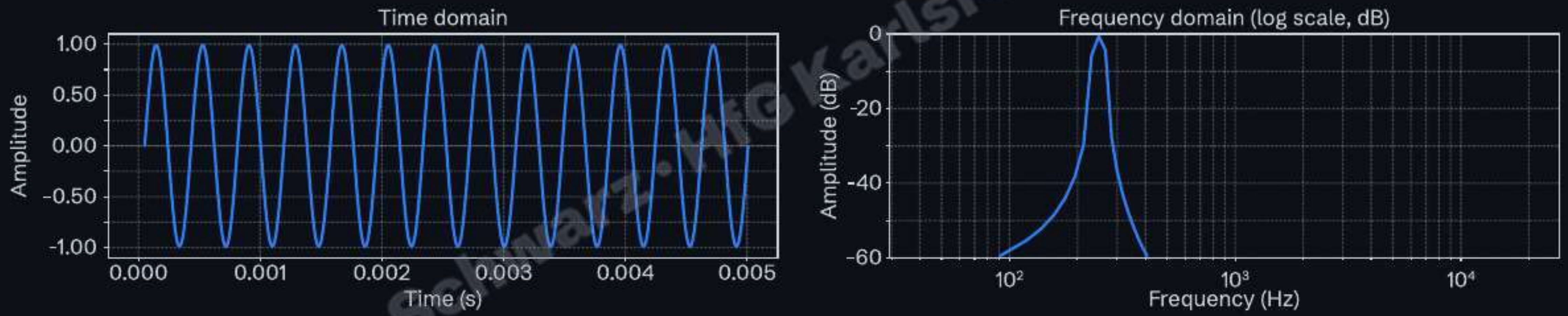


- Each peak in the spectrum represents one sine wave (partial or harmonics)
- Harmonic series 260, 520, 780, 1040… Hz

▶ Play sawtooth wave C4 ≈ 260 Hz

Pure tone (sine wave)

A sine wave is a single frequency component, the fundamental building block of all sounds

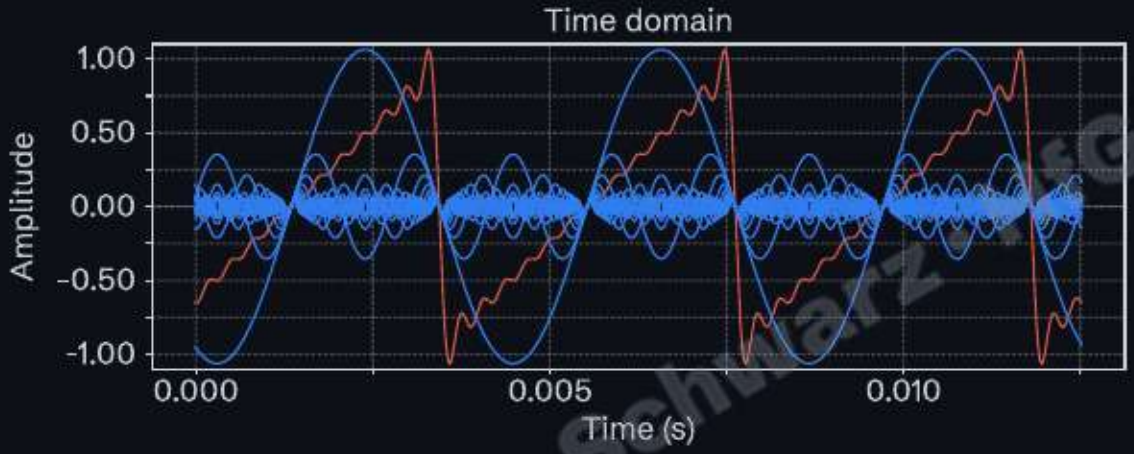


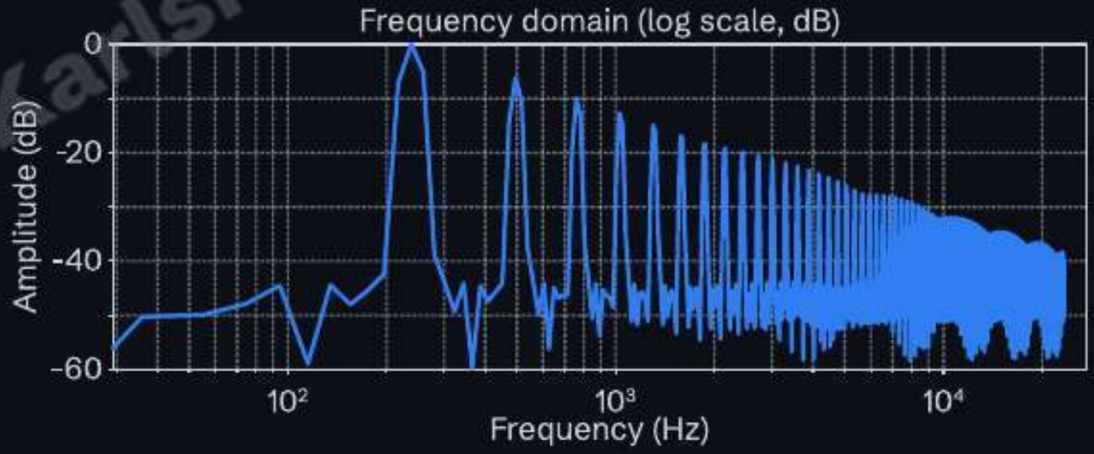
$$x(t) = A\sin(2\pi f_0 t + \varphi)$$

View sine wave on Desmos

Complex tones (example: sawtooth)

Musical instrument sounds and basic waveforms (except sine) contain many sine waves (click for graphing calculator)





$$x(t) = \frac{2A}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \sin(2\pi n f_0 t)$$

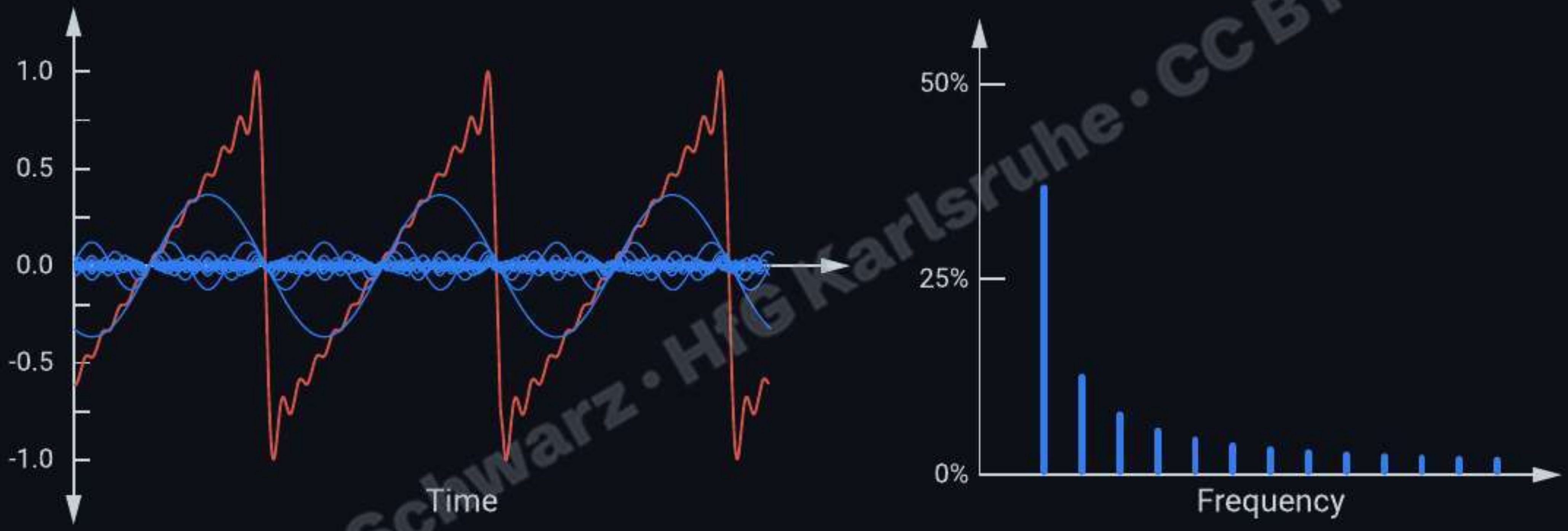
Fourier transform

Decomposes a signal from the time domain (waveform) into the frequency domain (spectrum):

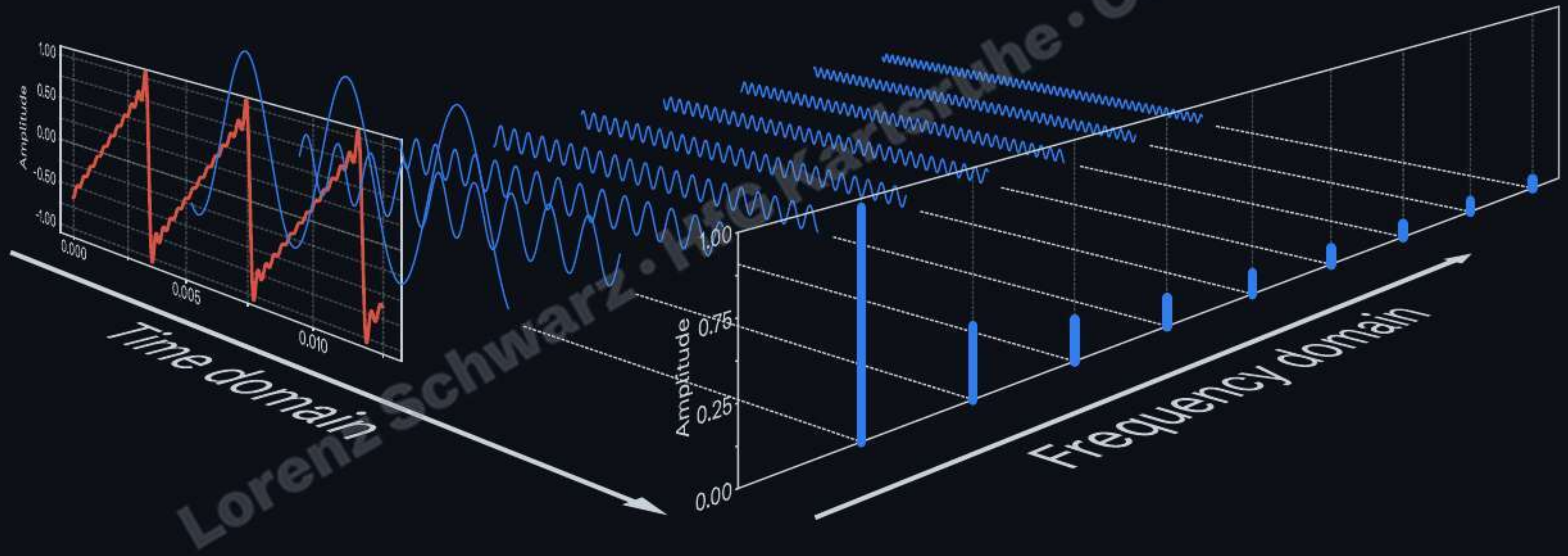
- Reveals the individual sine wave components and their amplitudes (and phases)
- Shows which frequencies are present and how strong they are

→ *Any complex sound can be represented as a sum of sine waves.*

Fourier series of a saw tooth wave (approximation)



Fourier transform of a sawtooth wave



Time-domain signals and spectral analysis

A time-domain signal $x(t)$ represents amplitude values over time, either continuous or discrete $x(n)$. The spectrum $X(\omega)$ is a weighting function that describes how harmonic components are combined to reconstruct the time-domain signal as a sum.

- Input: time-domain signal $x(t)$
- Output: frequency-domain spectrum $X(\omega)$

→ *The spectrum represents the amplitude and phase of each frequency component.*

Fourier Transform

Fourier transform (analysis formula):

- Break the signal into its frequency components

$$X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} \mathrm{d}t, \quad \omega \in (-\infty, \infty)$$

Inverse transform (synthesis formula):

- Rebuild the signal from its frequency components.

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{j\omega t} \mathrm{d}\omega$$

Euler's formula and complex numbers

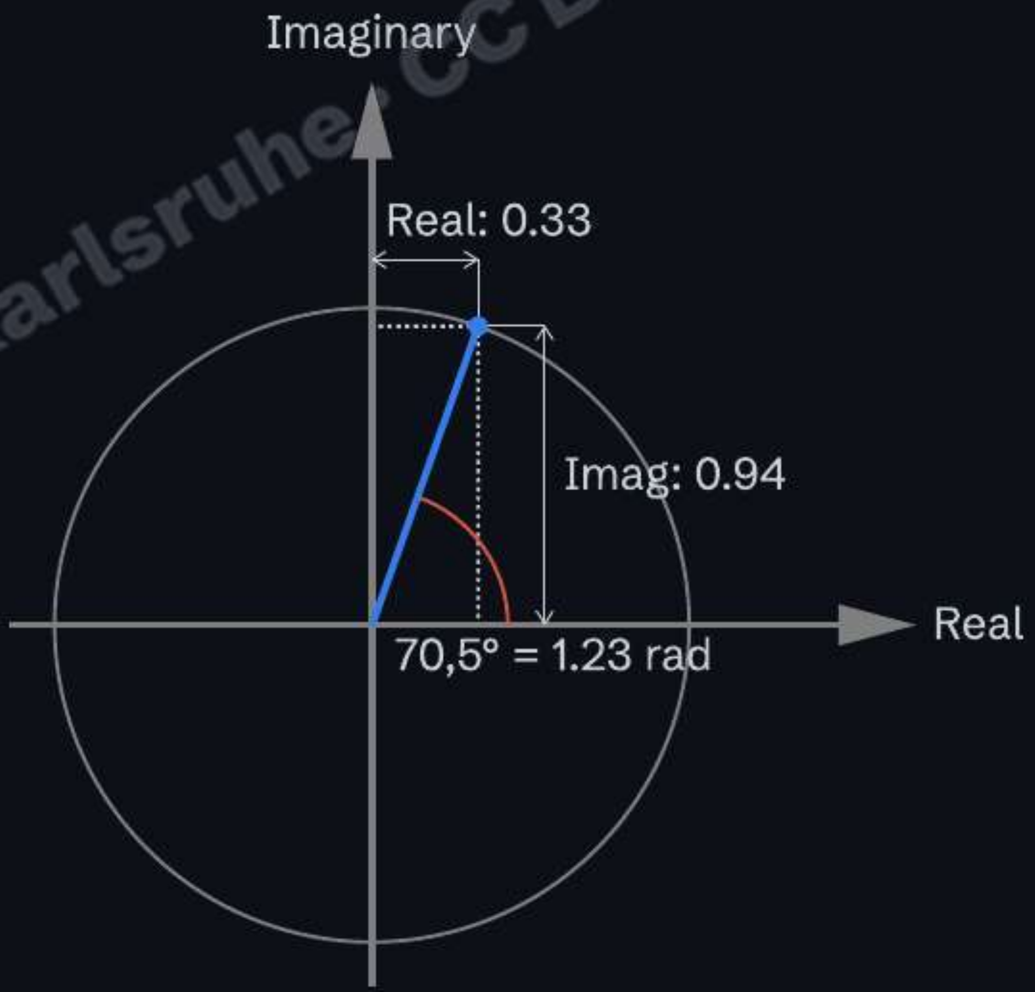
A complex exponential combines cosine and sine into a single expression representing sinusoidal components:

$$e^{j\omega t} = \cos(\omega t) + j\sin(\omega t)$$

where $j = \sqrt{-1}$ is the imaginary unit.

- **Real part:** $\cos(\omega t)$ — cosine component
- **Imaginary part:** $j\sin(\omega t)$ — sine component

→ *This representation is fundamental to the Fourier transform, allowing efficient encoding of both amplitude and phase.*



Discrete Fourier transform (DFT)

For digital audio, the DFT is used, which operates on discrete, finite-duration signals by replacing the continuous integral with a finite sum. The Fast Fourier Transform (FFT) is an efficient algorithm for computing the DFT.

$$X(\omega_k) = \sum_{n=0}^{N-1} x(t_n) e^{-j\omega_k t_n}, \quad k = 0, 1, 2, \ldots, N-1$$

inverse DFT:

$$x(t_n) = \frac{1}{N} \sum_{k=0}^{N-1} X(\omega_k) e^{j\omega_k t_n}, \qquad n = 0, 1, 2, \ldots, N-1$$

Discrete Fourier transform (DFT)

- No calculus needed — uses finite sums, avoids infinities.
- Assumes finite, sampled signals: Digital processing uses sampled signals.

→ *DFT is simpler and more computationally relevant than FT.*

Quantities of the DFT formula:

$$
\begin{aligned}
\sum_{n=0}^{N-1} &= f(0) + f(1) + \ldots + f(N-1) \\
x(t_n) &= \text{input signal amplitude at time } t_n \text{ (sec)} \\
t_n &= nT = nth \text{ sampling instant (sec), n an integer} \geq 0 \\
T &= \text{sampling interval (sec)} \\
X(\omega_k) &= \text{spectrum of x at frequency } \omega_k \\
\omega_k &= k\Omega = kth \text{ frequency sample (rad/s)} \\
\Omega &= \frac{2\pi}{NT} = \text{radian-frequency sampling interval (rad/s)} \\
f_s &= 1/T = \text{sampling rate (samples/second, or Hertz (Hz))} \\
N &= \text{number of time samples} = \text{number of frequency samples (integer)}
\end{aligned}
$$

Fast Fourier Transform (FFT)

The FFT is an efficient algorithm for computing the DFT, reducing computational complexity from N^2 to $N \log N$ operations.

This allows efficient computation for:

- Real-time spectrum analysis
- Frequency-domain filtering and equalization
- Convolution-based processing (reverberation, time-stretching)

→ *The FFT is an algorithmic optimization of the DFT computation.*

Frequency bins

The FFT produces discrete frequency values called bins, each representing a specific frequency component. Each bin contains amplitude and phase information for its frequency component.

Frequency of bin k:

$$f_k = \frac{k}{N} f_s$$

where $k = 0, 1, 2, \ldots, N-1$ is the bin index, f_s is the sampling rate, and N is the FFT size.

For real signals: Number of bins = $\frac{N}{2} + 1$ (due to symmetry)

Frequency resolution

Frequency resolution (bin spacing) determines how finely the spectrum is divided:

$$\Delta f = \frac{f_s}{N}$$

where f_s is the sampling rate and N is the FFT size.

- Larger FFT size → smaller Δf → better frequency resolution
- Frequencies separated by less than Δf cannot be distinguished

Example

Sampling rate: 44.1 kHz
FFT 1024-sample

$$\Delta f = \frac{44100}{1024} \approx 43 \text{ Hz per bin}$$

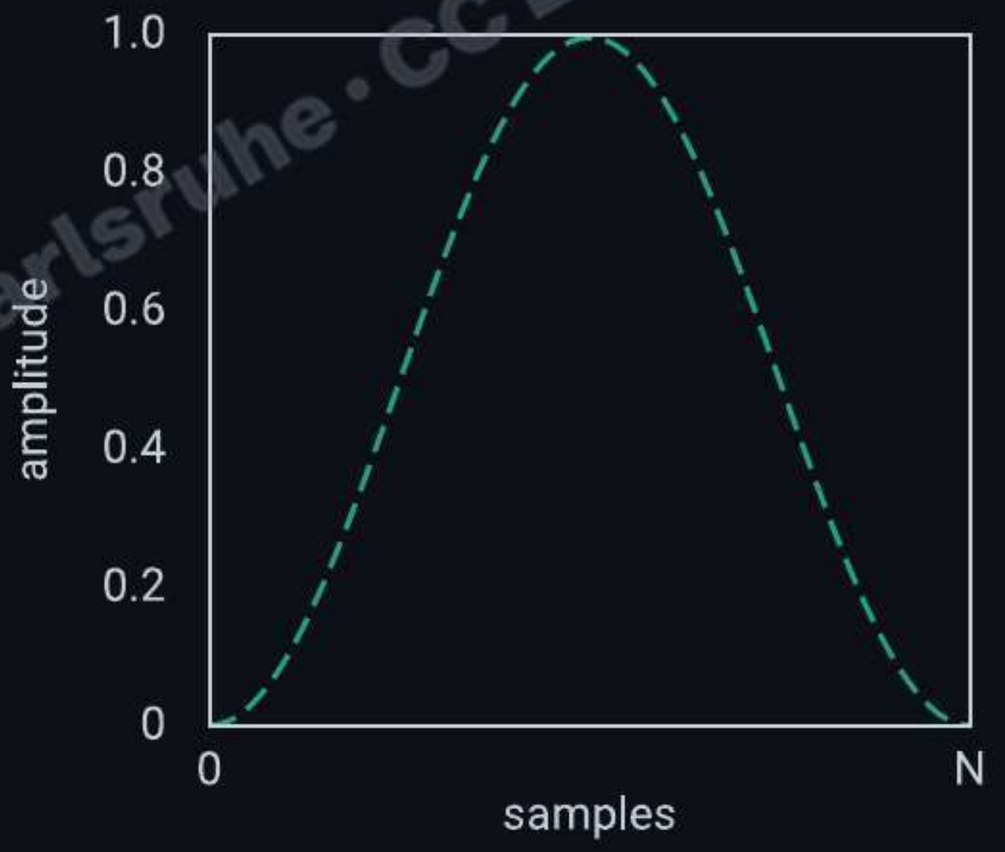
This means frequencies within a frequency band 43 Hz fall in the same bin and cannot be distinguished.

Window function

Tapering function that smoothly reduces signal amplitude to zero at analysis window boundaries, minimizing discontinuities and spectral leakage.

- Applied when signals contain non-integer periods within the FFT window
- Typically symmetrical, bell-shaped functions
- Common types: Hann, Hamming, Blackman-Harris

→ *Trade-off: Reduced leakage vs. reduced frequency resolution.*



Hann window

Spectral leakage and windowing

Spectral leakage occurs when the analysis window doesn't contain an exact integer number of wave cycles.

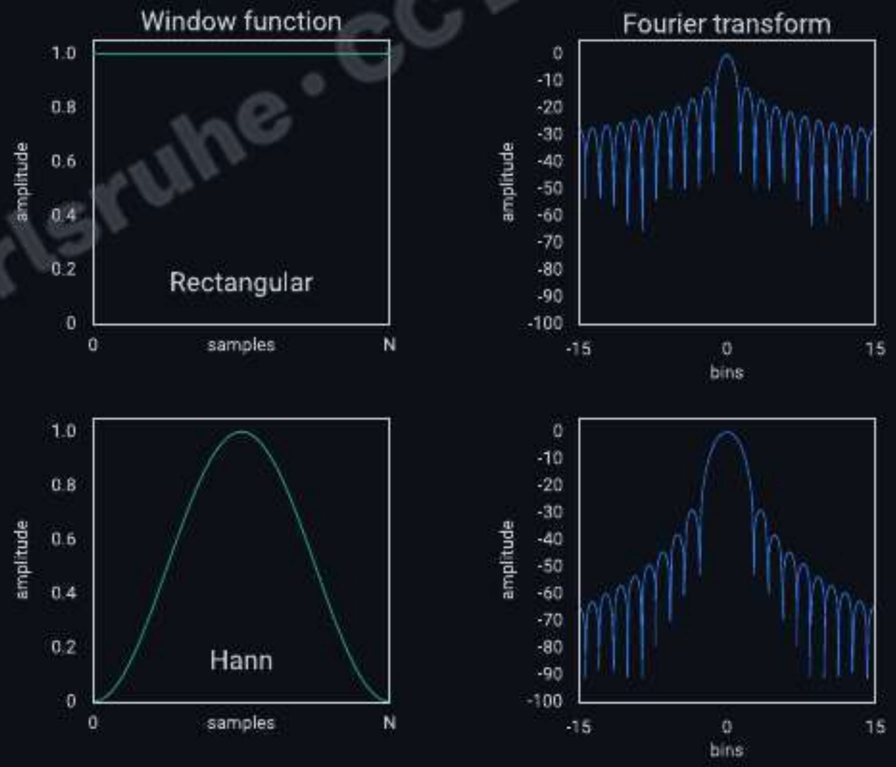
- The signal appears discontinuous at window edges
- This discontinuity creates artificial frequency components
- Energy 'leaks' from the true frequency into neighboring bins

→ *A window function tapers the signal smoothly to zero at the edges.*

Window characteristics in frequency domain

Each window function has a characteristic frequency response with a main lobe and side lobes:

- **Main Lobe:**
  - Central peak determining frequency resolution. Width measured between first zeros (null points).
- **Side Lobes:**
  - Secondary peaks flanking the main lobe. Height (in dB) indicates leakage suppression quality.
- **Trade-off:**
  - Lower side lobes require wider main lobes, reducing frequency resolution.



Rectangular window: narrow main lobe, high side lobes (-13 dB)

Selecting a window function

Window	Main lobe width	Side lobe level	Use case
Rectangular (no window)	Narrowest (2 bins)	Highest (-13 dB)	Maximum frequency resolution, integer number of periods
Hann	Medium (4 bins)	-31 dB	General purpose, good balance of resolution and leakage
Hamming	Medium (4 bins)	-42 dB	Better side lobe suppression, 8-bit systems, telephony
Blackman-Harris	Widest (6 bins)	-92 dB (4-term)	High dynamic range, very low leakage critical applications

Trade-off: Better side lobe suppression = wider main lobe = reduced frequency resolution.

FFT size (window size)

Number of samples per FFT computation. Determines the time-frequency resolution trade-off:

- Larger size: Better frequency resolution, worse time resolution
- Smaller size: Better time resolution, worse frequency resolution

Common sizes: 256, 512, 1024, 2048, 4096 (powers of 2)

FFT Size	Frequency Resolution	Time Resolution
Small (256)	Poor (coarse bins)	Good (fast response)
Large (4096)	Good (fine bins)	Poor (slow response)

Applications of the Fourier transform

Theoretical approaches:

- Organs: Additive synthesis for sound creation.
- Tone Wheels: Used in the Telharmonium by Thaddeus Cahill (1898).

Spectral audio signal processing:

- Additive synthesis
- Digital filter design
- Vocoder: Manipulation of speech and audio signals.

Applications of spectral analysis

- Analysis: spectrum analyzers in DAWs
- Processing: convolution reverb, spectral effects, noise reduction, time stretching
- Synthesis: additive synthesis, spectral resynthesis

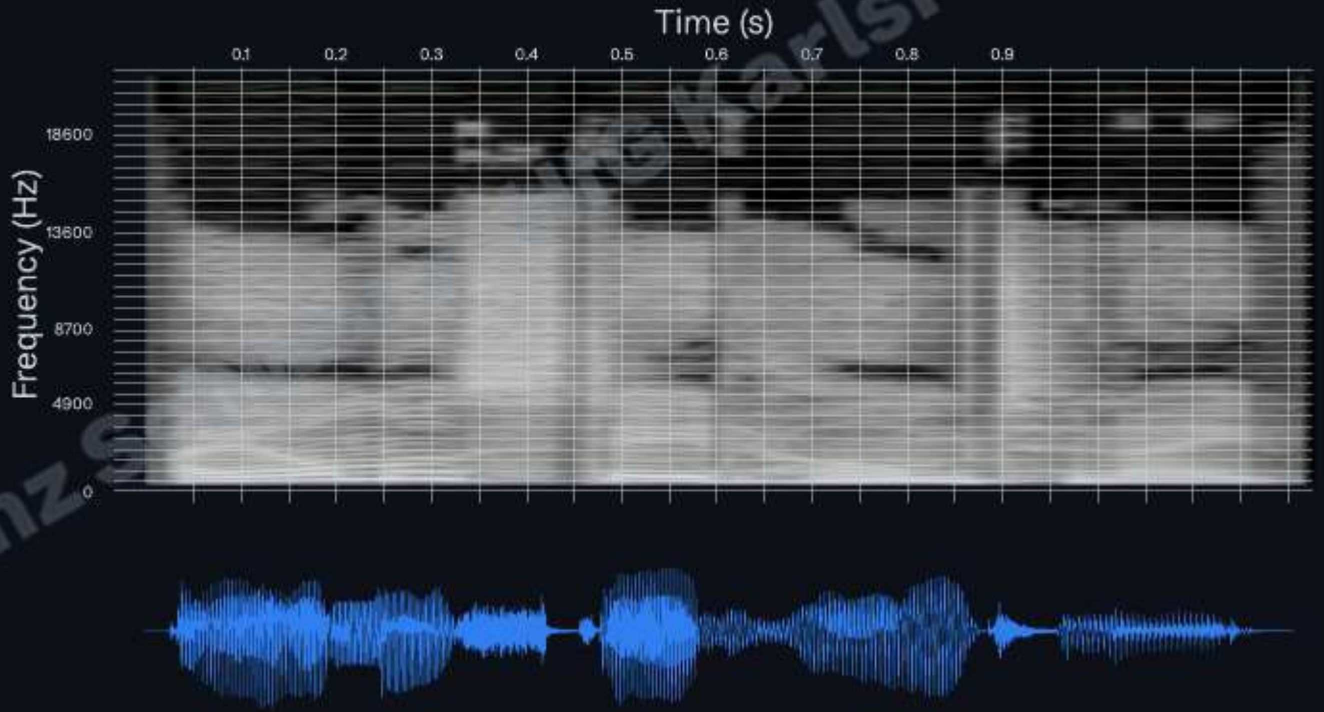
Spectrogram, sonogram

A time-varying visual representation of a signal's frequency content:

- **X-axis:** Time
- **Y-axis:** Frequency
- **Color/brightness:** Amplitude

Spectrogram

Spectrograms reveal temporal evolution of spectral content. Helpful for analyzing speech, music, and environmental sounds.



FILTERS

Shaping the spectrum of audio signals

Examples of a filter

- **Vocal tract:** Dynamic filtering.
- **Human hearing (e.g., pinna, head, torso):** Acts as a direction-selective filter.
- **Musical instruments (e.g., Helmholtz resonator):** Filtering through resonances
- **Rooms:** Altering the sound waves through boundary behaviors.
- **Loudspeakers and microphones:** Shape sound through their frequency response.
- ...

Filter

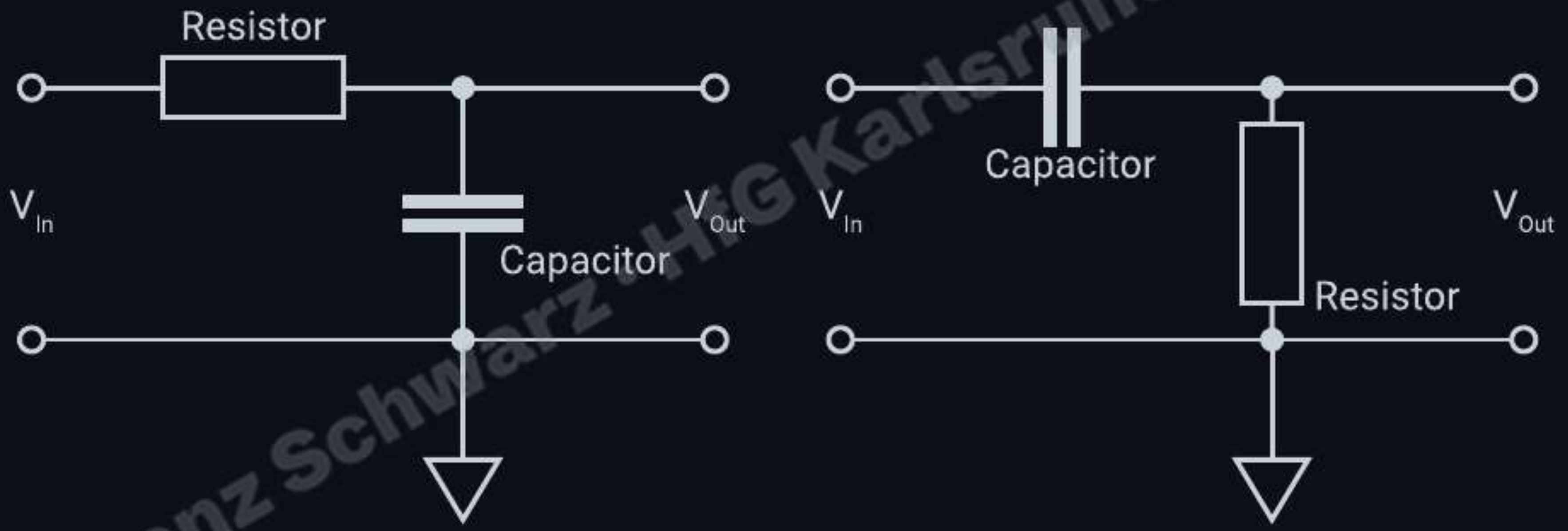
A filter selectively emphasizes or attenuates certain frequencies in a signal:

- commonly a device, electronic circuit or software algorithm
- used to shape the frequency spectrum of an audio signal

→ *Phase response is often perceptually less critical, but may be relevant in some contexts.*

Electronic filter

First order lowpass filter (left) and high pass filter (right)

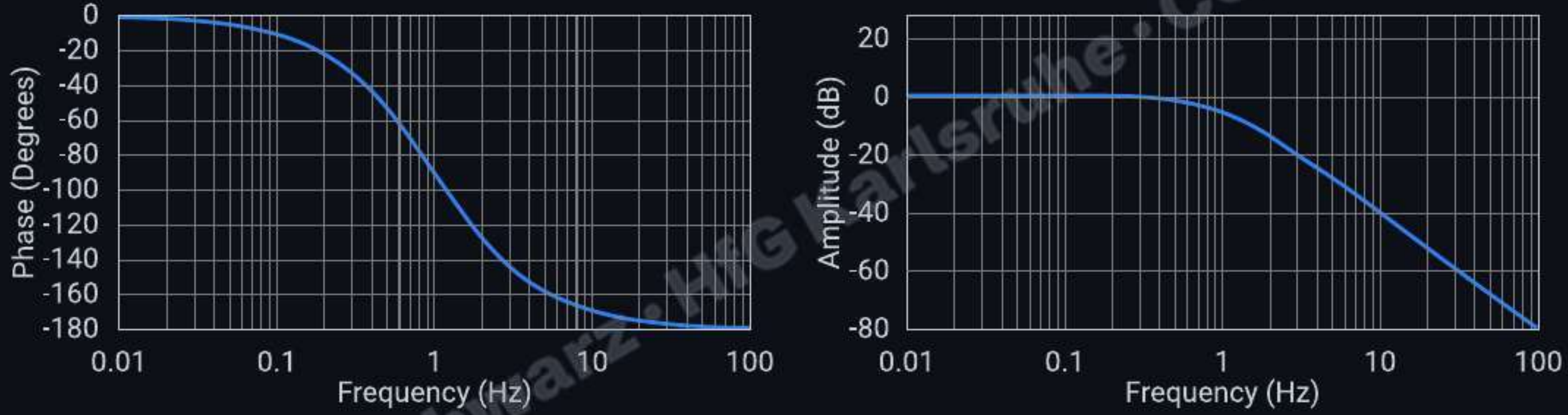


Visualization of a filters behavior

Transfer function plots graphically represent how a filter modifies input signals across frequencies.

- **Frequency (x-axis):** Plotted on a logarithmic scale (in Hz).
- **Magnitude (y-axis):** Plotted on a logarithmic scale (in dB).
- **Phase angle (y-axis):** Plotted on a linear scale (in degrees or radians).

Phase (left) and magnitude response (right) of a low-pass filter



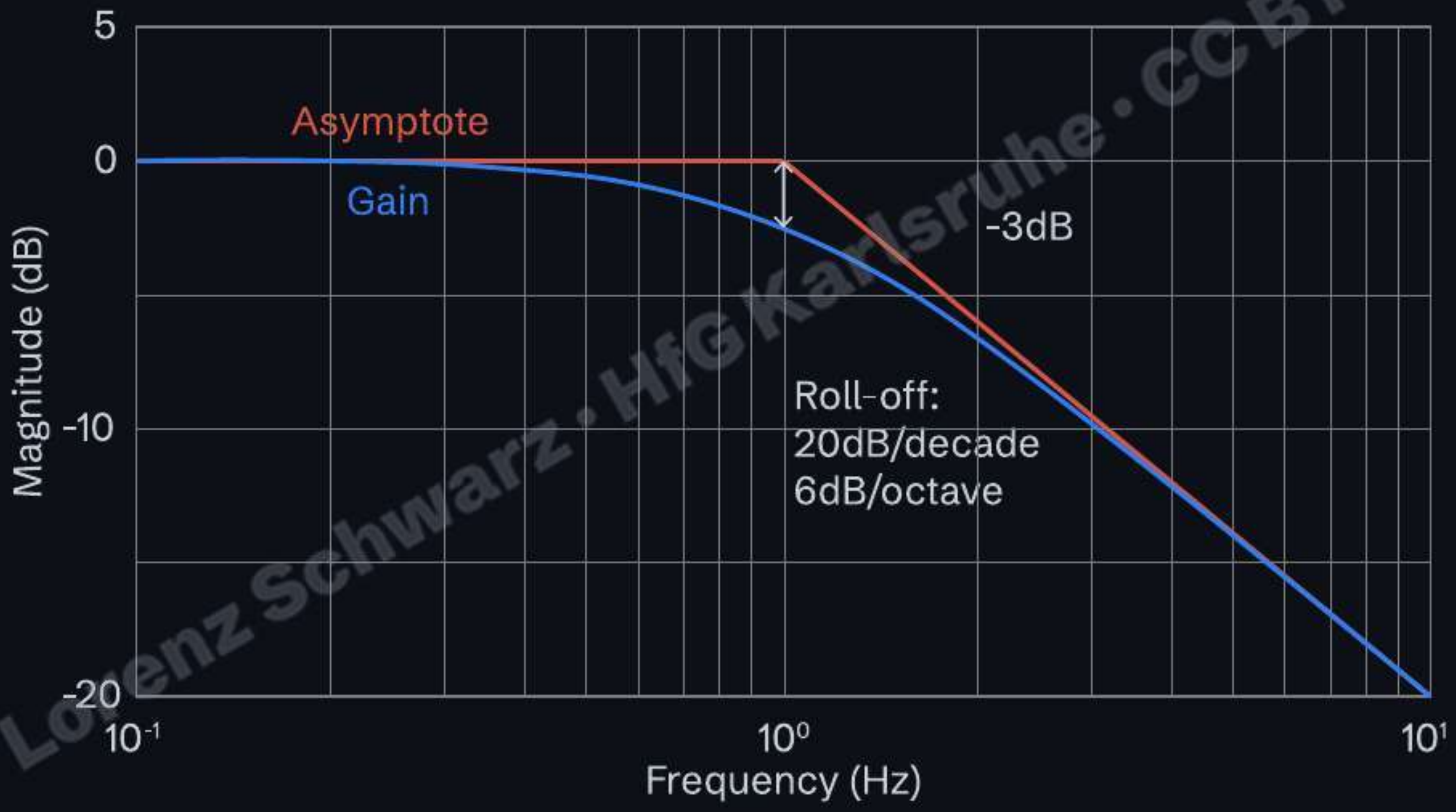
Magnitude response of a filter

The Bode magnitude response plot shows the gain (ratio of output amplitude to input amplitude) of a filter as a function of frequency, typically displayed on a logarithmic scale.

Gain versus frequency:

- **Vertical axis (magnitude):** Logarithmic scale, typically in decibels (dB).
- **Horizontal axis (frequency):** Logarithmic scale in Hertz (Hz).

Bode magnitude plot

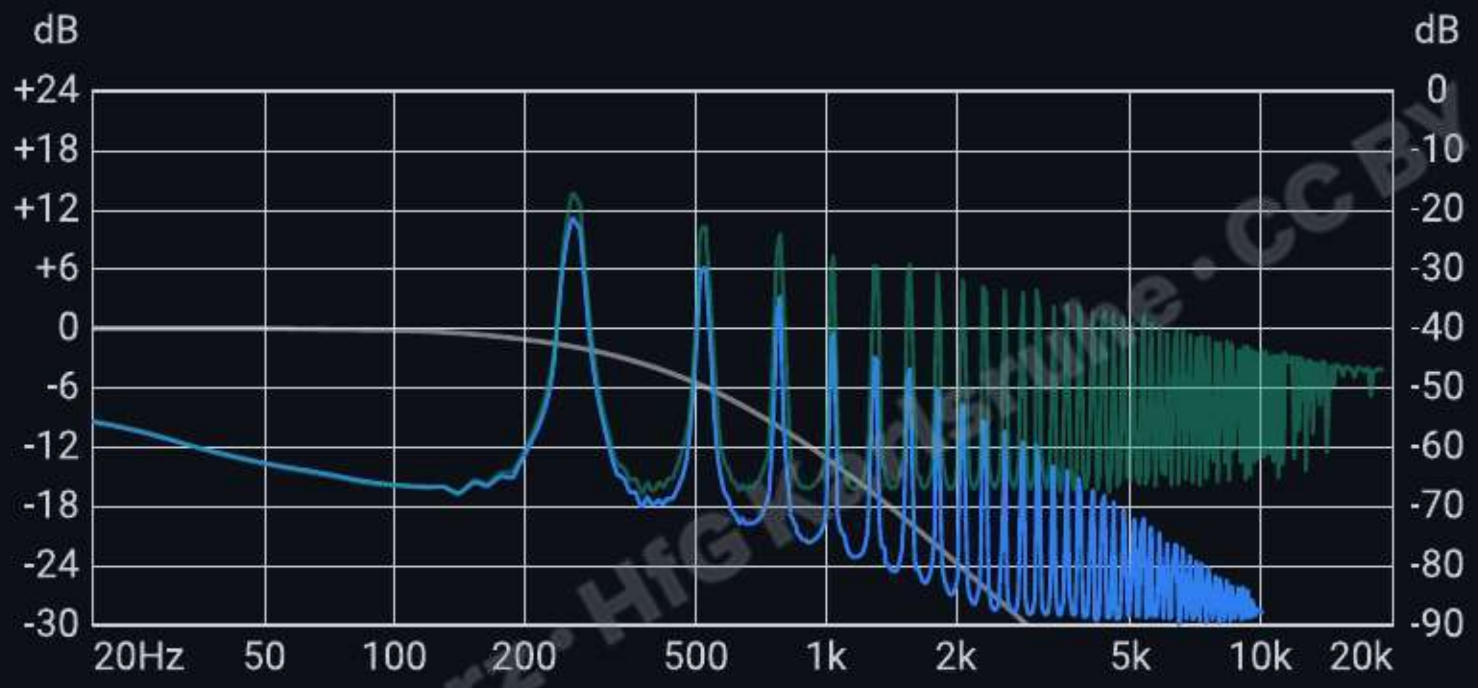


Magnitude of gains

$\left\lvert \frac{V_{out}}{V_{in}} \right\rvert$	10^{-4}	0.001	0.01	0.1	1	10	100	1000	10^{4}
Gain	-80dB	-60dB	-40dB	-20dB	0dB	20dB	40dB	60dB	80dB

Attenuation — Unity — Amplification

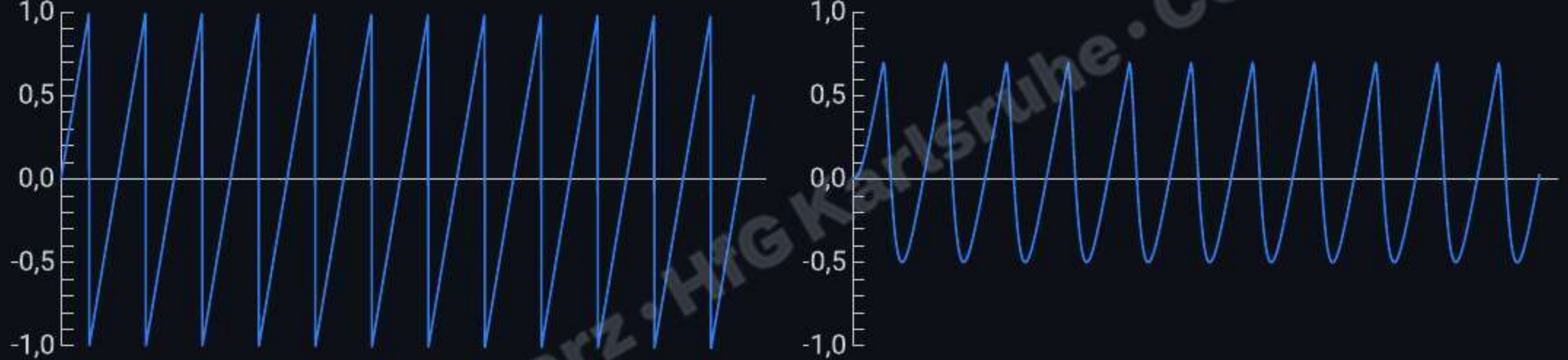
$$gain\ in\ dB = 20 \log_{10} \left\lvert \frac{V_{out}}{V_{in}} \right\rvert$$



The output spectrum (blue) is the input spectrum (green) multiplied by the filter's magnitude response (grey curve).

▶ Sawtooth 260 Hz unfiltered - filtered

Filtered waveform in the time domain

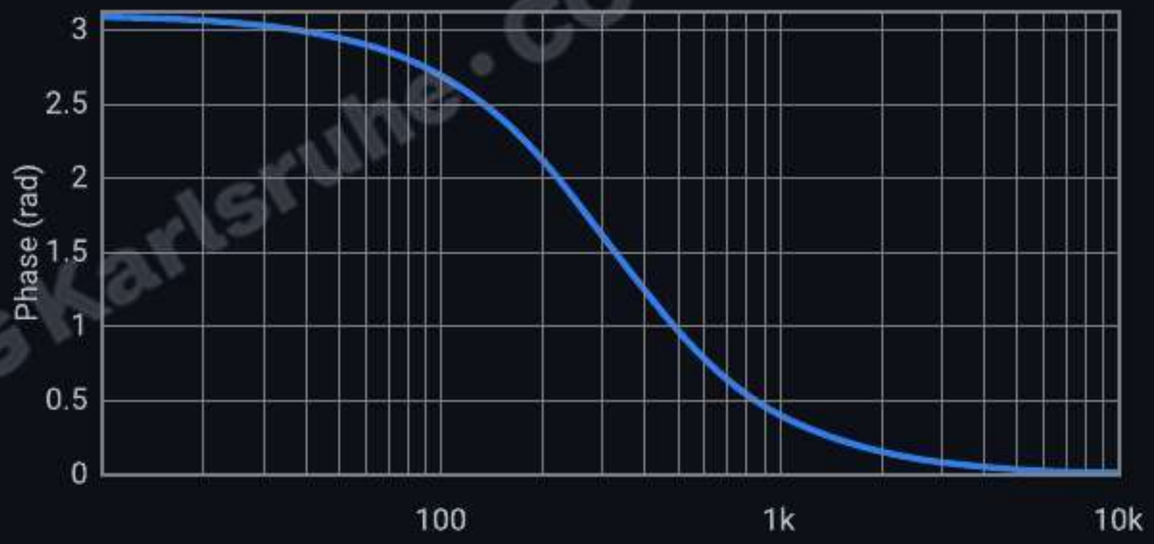


Input Output

Phase response of a filter

A Bode phase plot is a graph that shows the phase relationship between a sinusoidal input signal and the output signal of a filter, as a function of frequency.

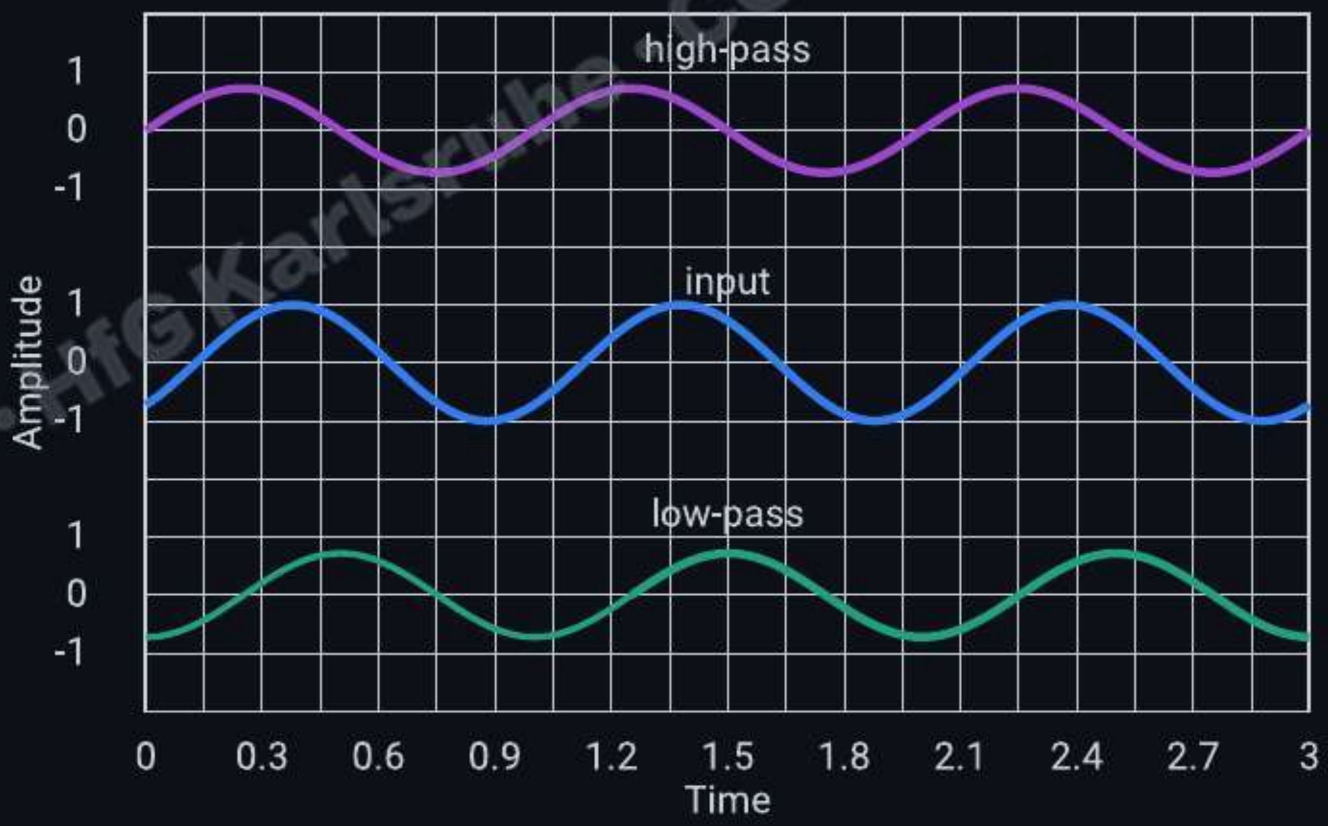
- **Vertical axis (phase):** Linear scale, typically in degrees or radians, representing the phase shift introduced by the filter.
- **Horizontal axis (frequency):** Logarithmic scale in Hertz (Hz).



Phase and time-domain filters

The following filters don't primarily remove or boost frequencies, but make use of a filters property to manipulate phase relationships of a signal to create interference patterns.

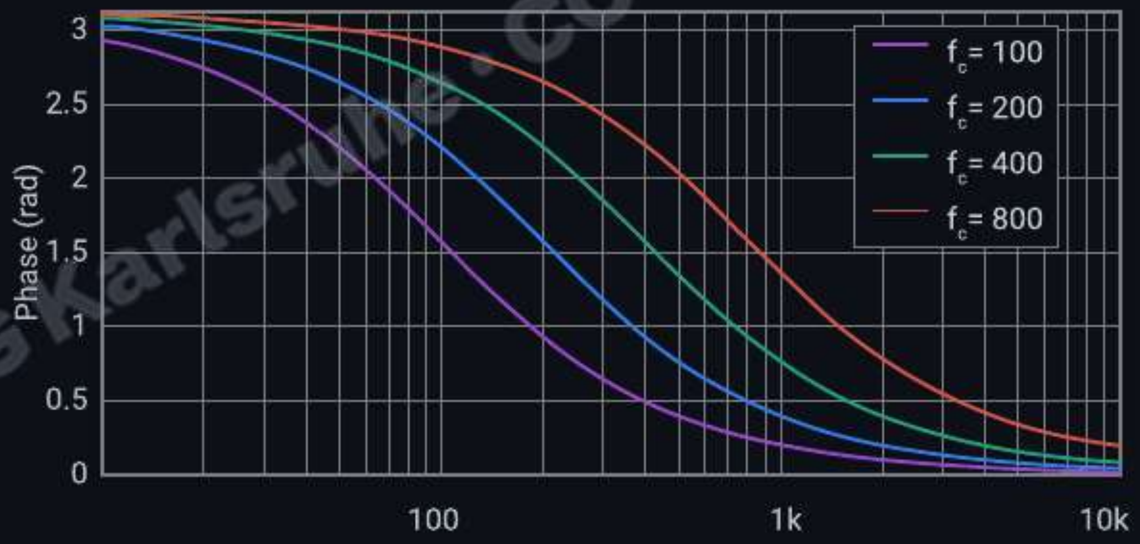
left: Input and outputs of a single-pole high-pass and low-pass filter



All-pass filter

An all-pass filter changes the phase relationship between frequencies by introducing a frequency-dependent phase shift, while allowing all frequency components to pass with equal amplitude (unity gain).

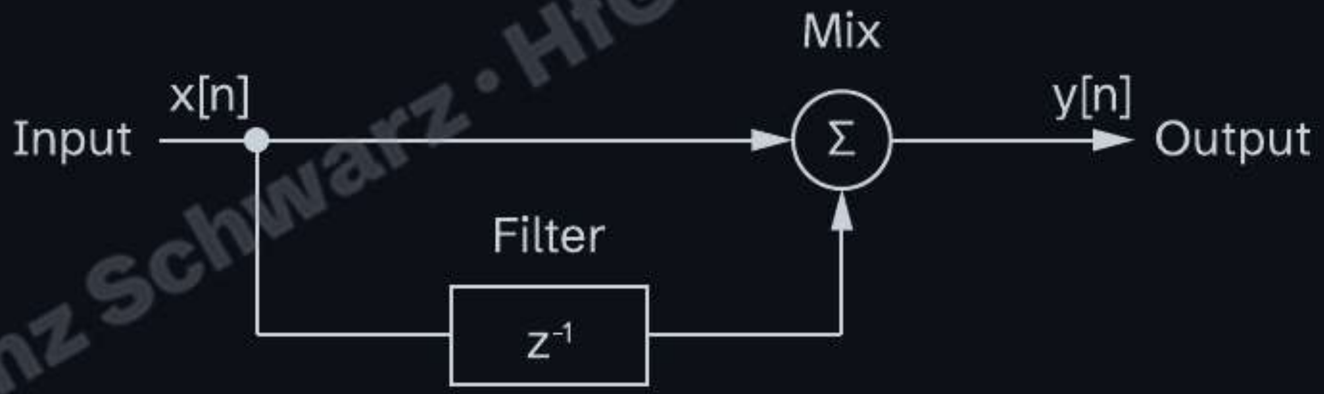
Applications: Phaser audio effect



Comb filtering in sound (phaser)

A comb filter mixes a signal with a delayed copy of itself:

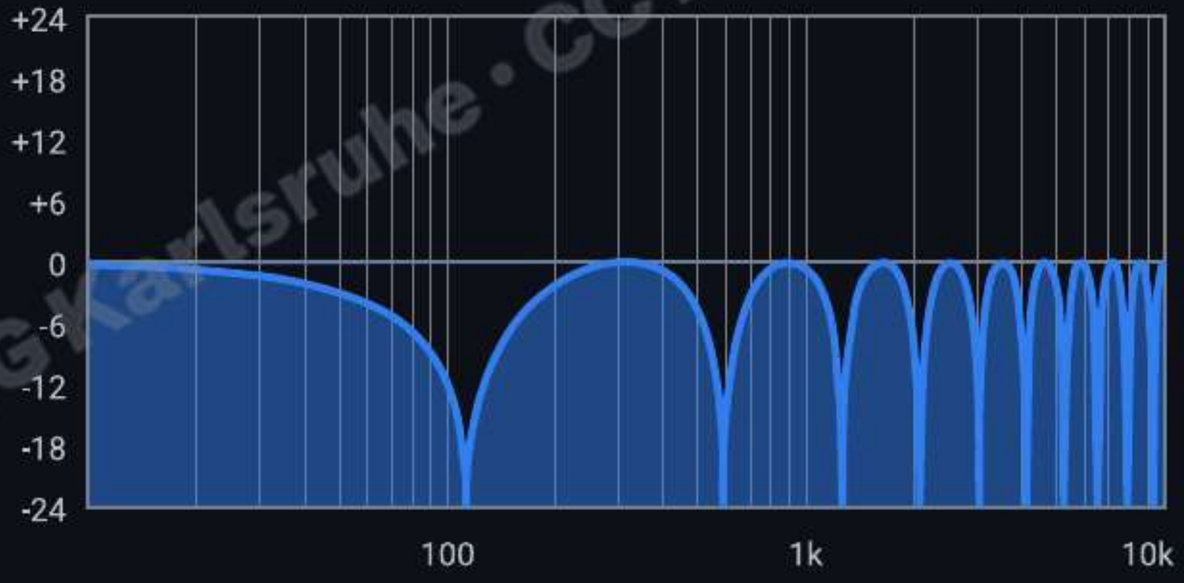
- causes constructive and destructive interference
- produces characteristic spectral notches (resemble the teeth of a comb)



▶ White noise comb filtered

Natural occurrences of comb filtering

- Early reflections from hard surfaces (floor or wall reflections)
- Multiple microphones capturing the same source at different distances

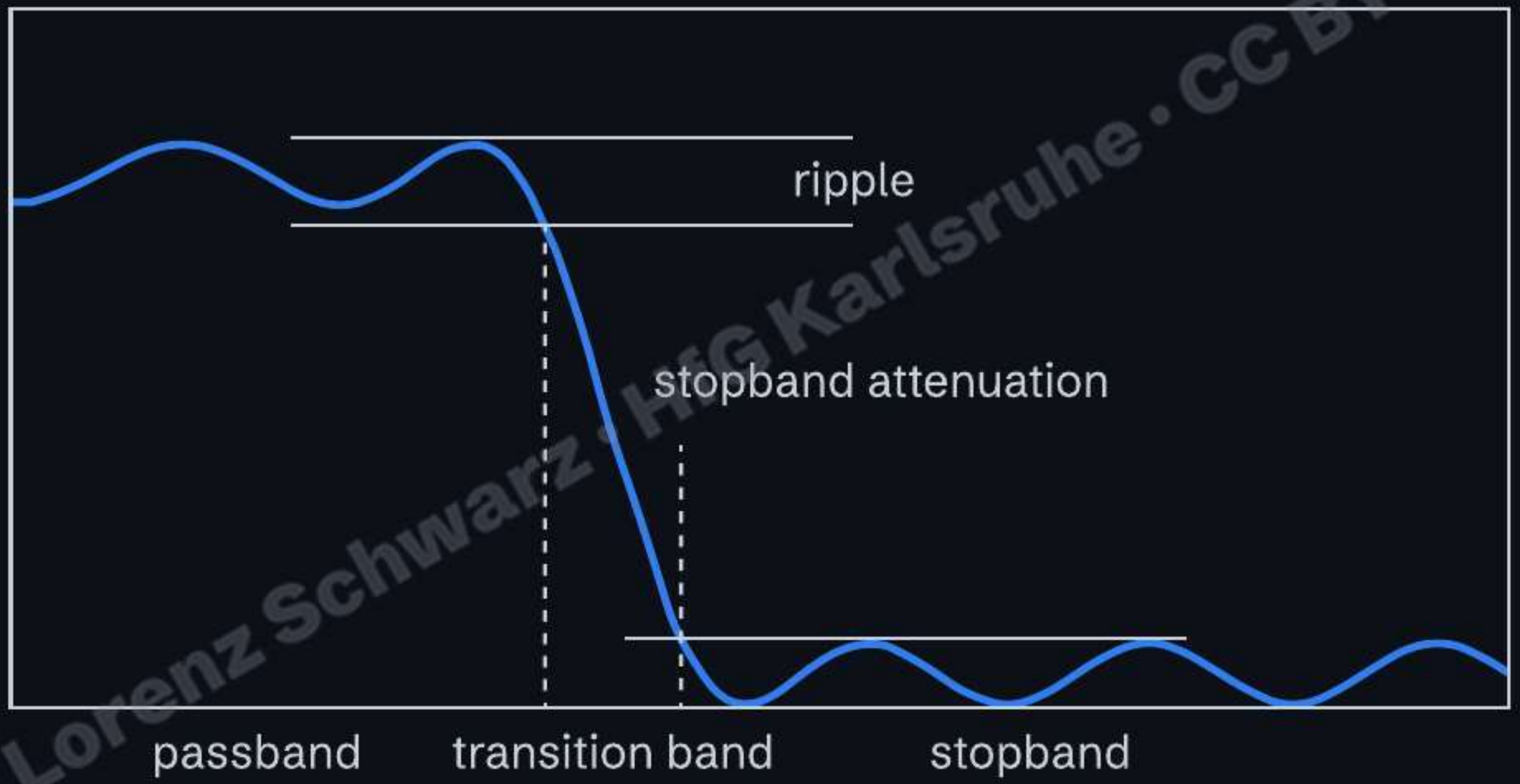


→ *Comb filtering can be an unwanted acoustic artifact or a deliberate effect.*

Passband, stopband, transition band, ripple

- **Passband:** The range of frequencies that pass through the filter with no attenuation.
- **Stopband:** The range of frequencies that are significantly attenuated or blocked by the filter.
- **Transition band:** The region between the passband and stopband where attenuation gradually increases.
- **Ripple:** Deviation from flatness in a filter's magnitude response, showing the ratio of max to min gain in the passband or stopband.

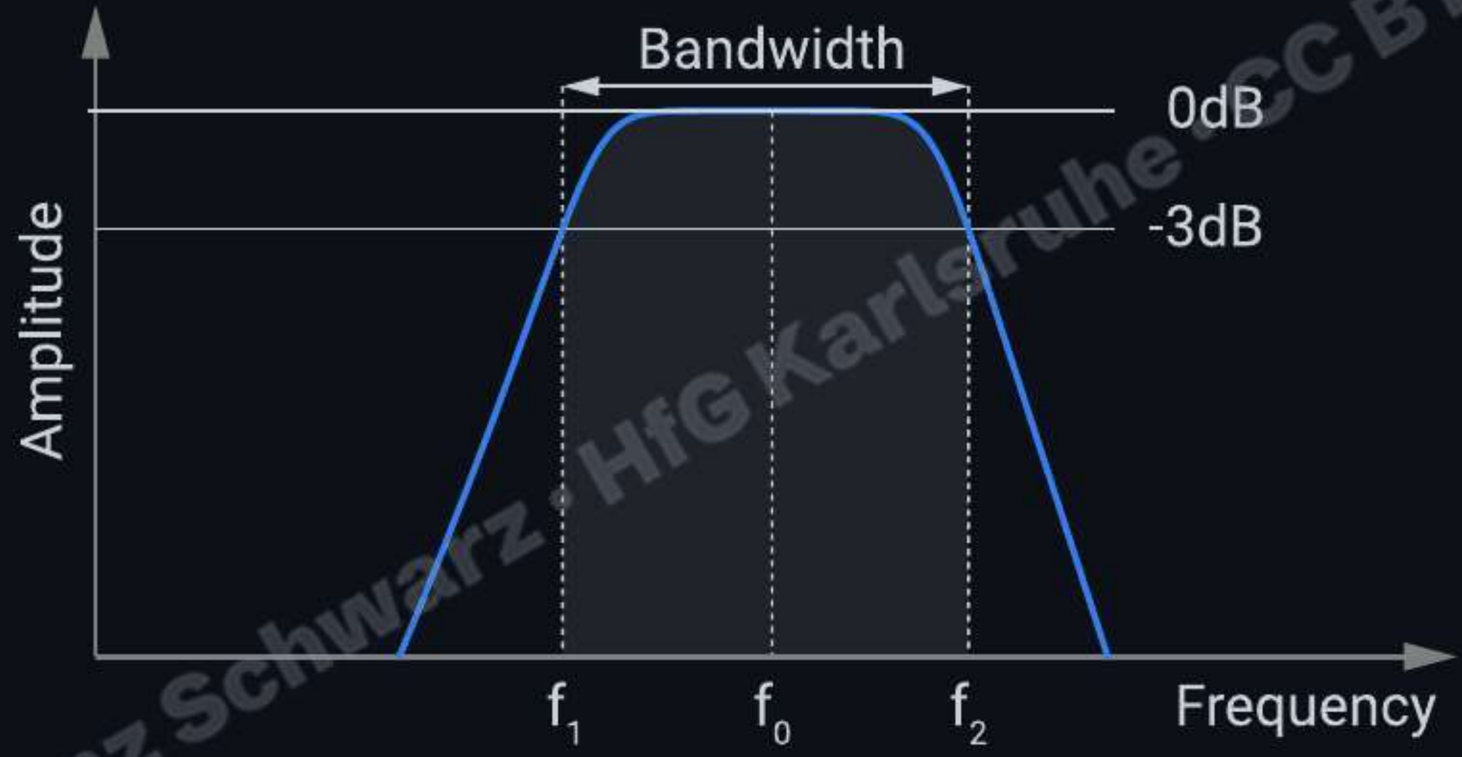
Description of the magnitude response



Bandwidth

The range of frequencies passed by a band-pass filter or attenuated by a notch filter, defined as the difference between the upper and lower cutoff frequencies. It represents the section of the frequency spectrum that the filter affects most significantly.

Bandwidth, lower and upper corner frequency



Cutoff or corner frequency

The cutoff frequency (also known as the half-power point) is the frequency at which the output voltage level decreases by 3 dB compared to the input voltage level (0 dB). Beyond this point, the output voltage progressively decreases relative to the input voltage.

The -3 dB level corresponds to the factor:

$$\sqrt{\frac{1}{2}} = \frac{1}{\sqrt{2}} = 0.7071$$

This means the output voltage is about 0.7071 times the input voltage.

Center frequency f_0

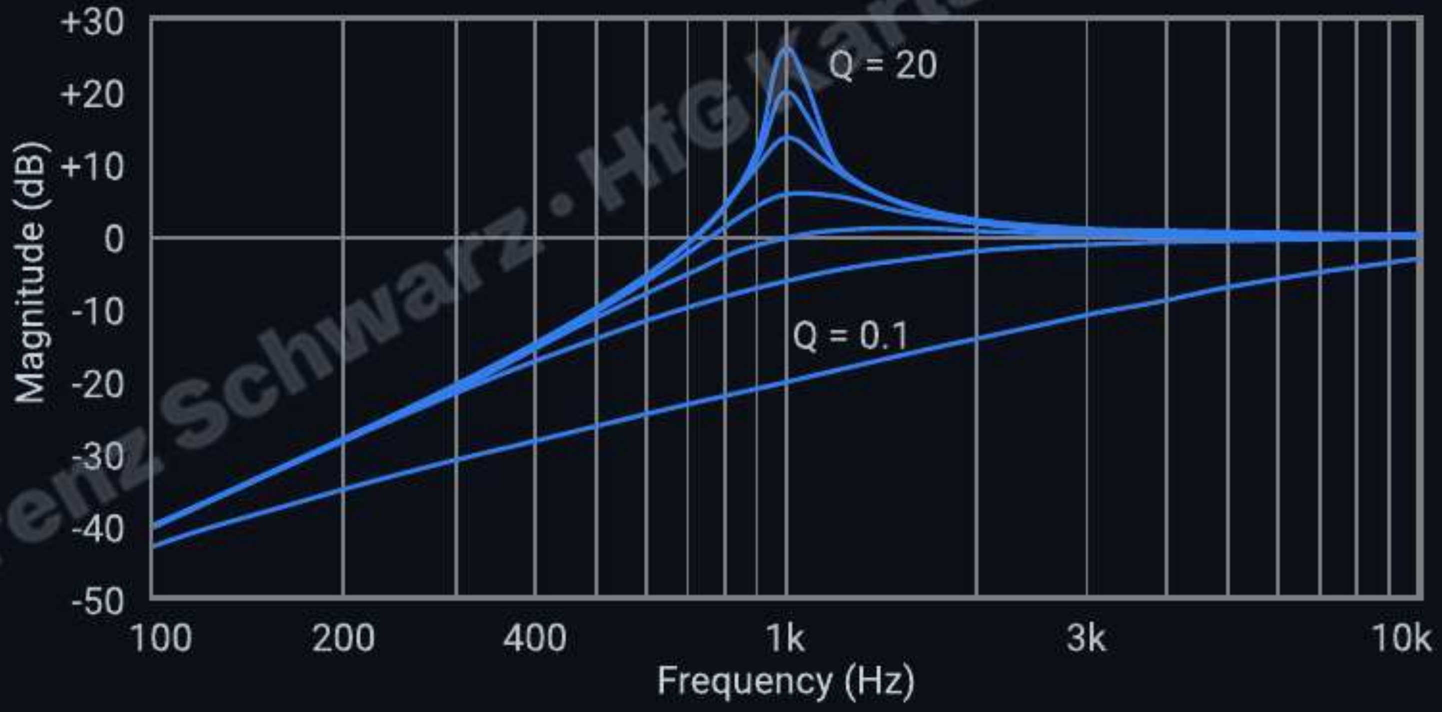
The midpoint center frequency of a band pass filter is the geometric mean between the lower cuttoff frequency and the upper cutoff frequency.

Geometric mean:

$$f_{0geo} = \sqrt{f_1 \cdot f_2}$$

Quality factor (Q-factor)

Parameter that describes the selectivity of a filter, defined as the ratio of the center frequency to the bandwidth of the filter



Filter resonance

Filter resonance refers to the phenomenon where a filter exhibits a peak in amplitude at a specific frequency, often near the cut-off frequency.

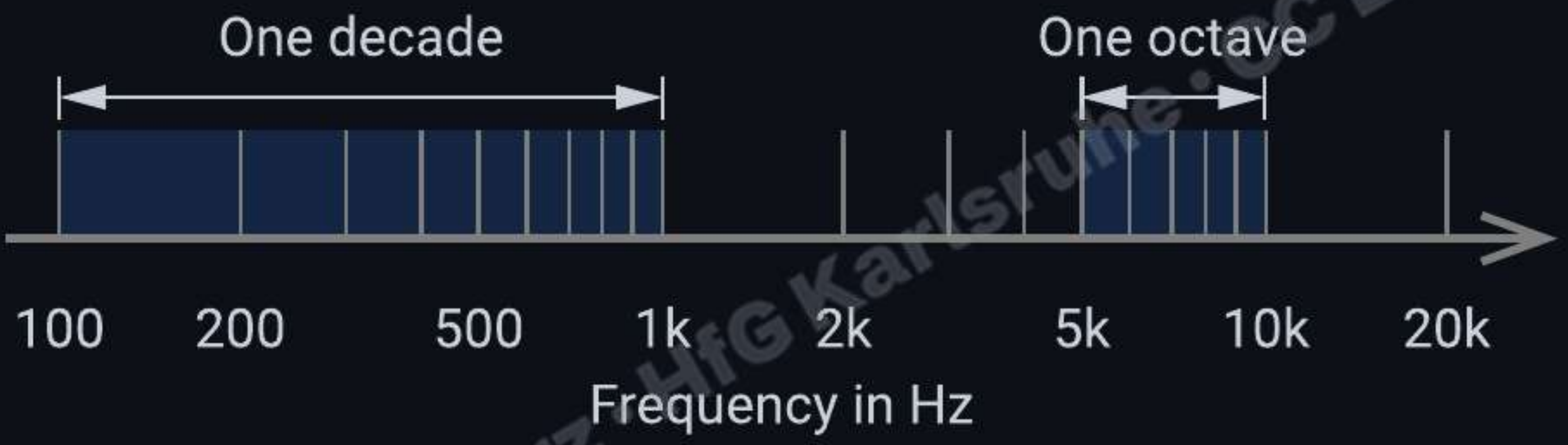
- A sufficiently high Q-factor can lead to self-oscillation even without an input signal.

▶ Low pass sweep

▶ Low pass sweep with resonance

▶ Low pass self-oscillating

Frequency ratio units



- **Octave:** Doubling or halving of frequency
- **Decade:** Tenfold increase or decrease in frequency (factor-of-ten)

Relationship of filter key parameters

Center frequency f_0

$$f_0 = \sqrt{f_1 \cdot f_2}$$

Bandwidth B

$$B = f_2 - f_1$$

Quality factor Q

$$Q = \frac{f_0}{B}$$

Example: calculating bandwidth

Relation between center frequency (f_0), bandwidth (B), and Q-factor (Q)

Example: $f_0 = 800Hz$, $Q = 10$

$$B = \frac{f_0}{Q} = \frac{800}{10} = 80Hz$$

- high Q $\longleftrightarrow$ narrow bandwidth
- low Q $\longleftrightarrow$ wide bandwidth

Relation between Q-factor and bandwidth

large Q = narrow bandwidth $\longleftrightarrow$ small Q = broad bandwidth.

BW in octaves	Q
2.0	0.667
1.0	1.414
2/3	2.145
1/2	2.871
1/3	4.318
1/6	8.651
1/10	14.424
1/30	43.280

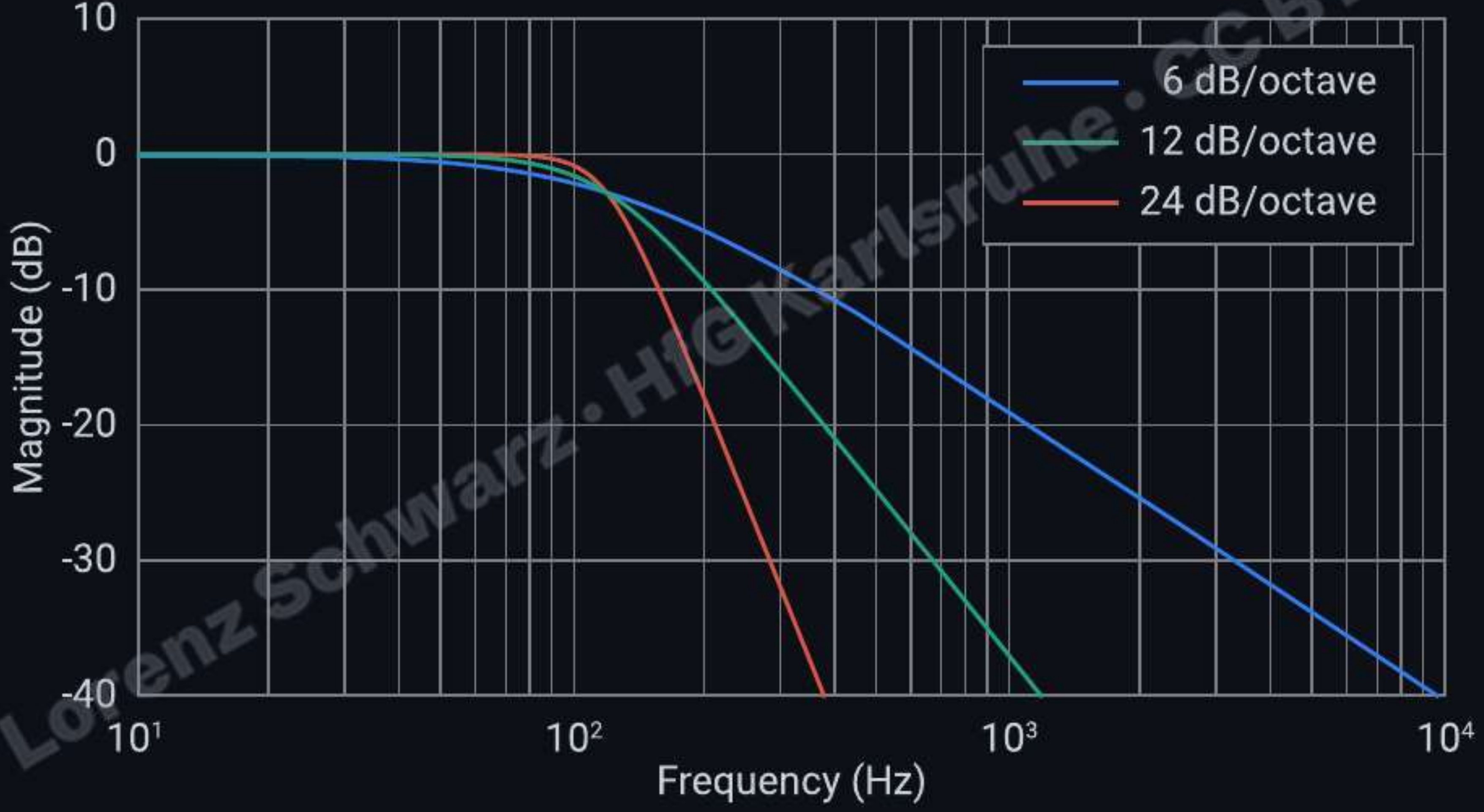
Roll-off (filter slope)

Roll-off describes the steepness of the transition from the passband to the stopband in a filter's transfer function graph.

- Indicates how rapidly the filter attenuates the signal beyond the cutoff frequency.
- Measured in decibels per octave (dB/octave) or decibels per decade (dB/decade).

▶ White noise corner frequency 500 Hz

Roll-off



10
0
-10
-20
-30
-40
Magnitude (dB)
6 dB/octave
12 dB/octave
24 dB/octave
10^1
10^2
10^3
10^4
Frequency (Hz)

Filter order

The filter order refers to the highest power of the variable in the polynomial of the filter's transfer function, which is the algebraic representation of the filter's behavior. The order determines the steepness of the filter's attenuation beyond the cutoff frequency.

For example 3rd order Butterworth polynomial:

$$B(s) = (s^2 + s + 1)(s + 1)$$

→ *Higher-order filters produce steeper slopes.*

Filter order and corresponding roll-off rates

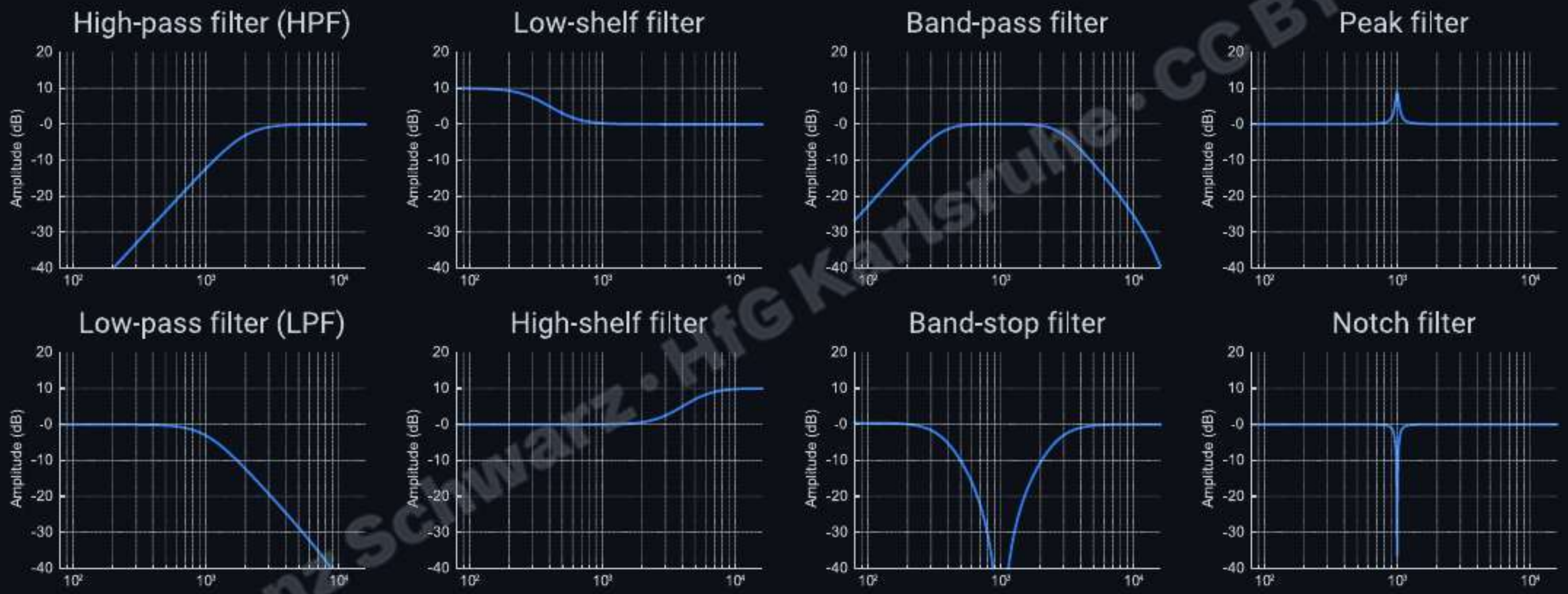
Each increase in filter order results in a roll-off rate increase of 6 dB per octave.

Filter order	dB/Octave	dB/decade
first order	6	20
second order	12	40
third order	18	60
fourth order	24	80
fifth order	30	100
sixth order	36	120
seventh order	42	140
eighth order	48	160

Filter types

- Low-pass filter
- High-pass filter
- Shelving filters
- Band-pass filter
- Band-stop filter
- Notch filter
- Peak filter

Overview of common filter types



Spectrum-limiting filters

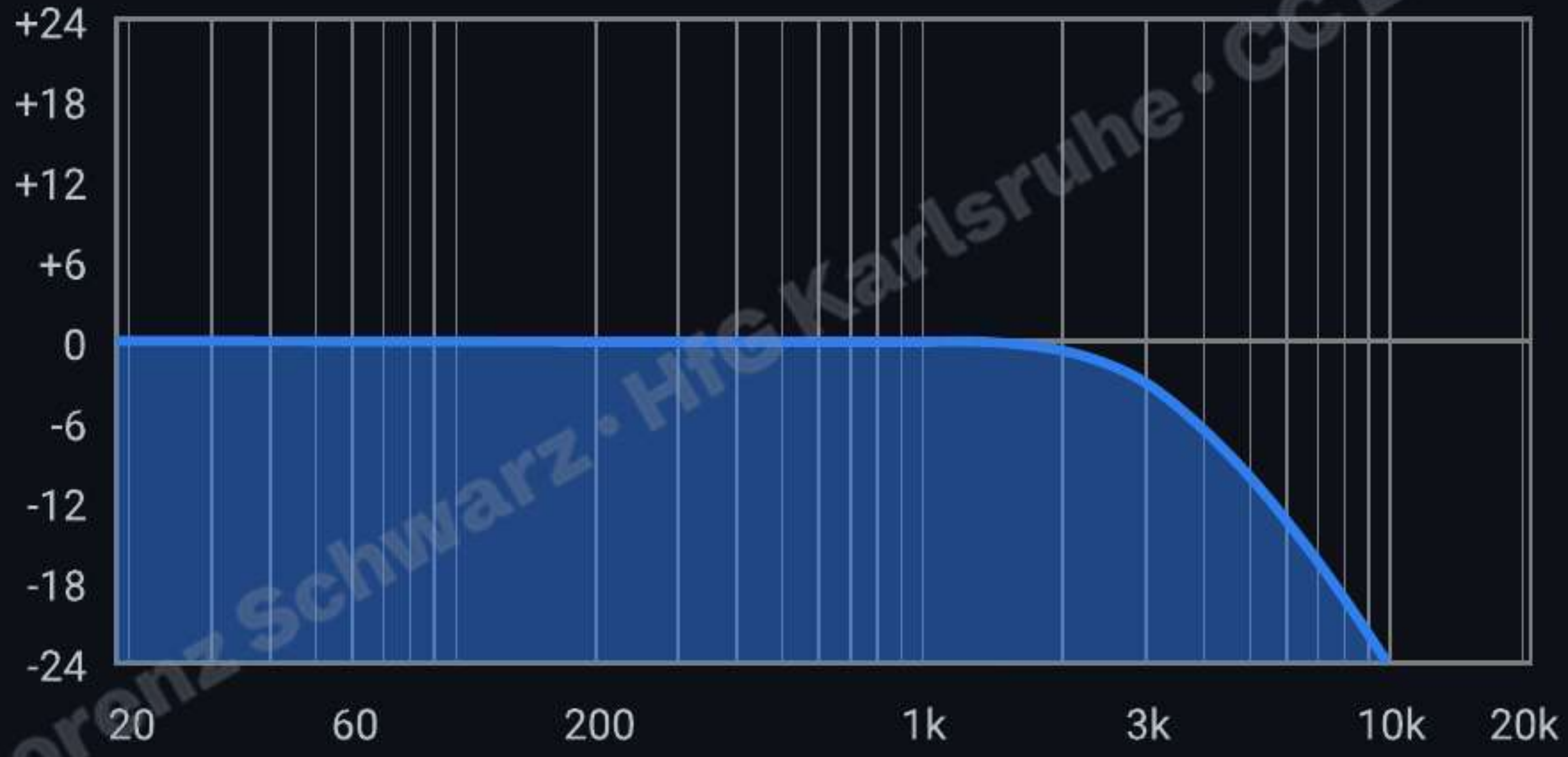
Low-pass and high-pass are the most fundamental filter types. They define a boundary frequency and progressively attenuate everything on one side of it.

Low-pass filter (LPF) - high cut

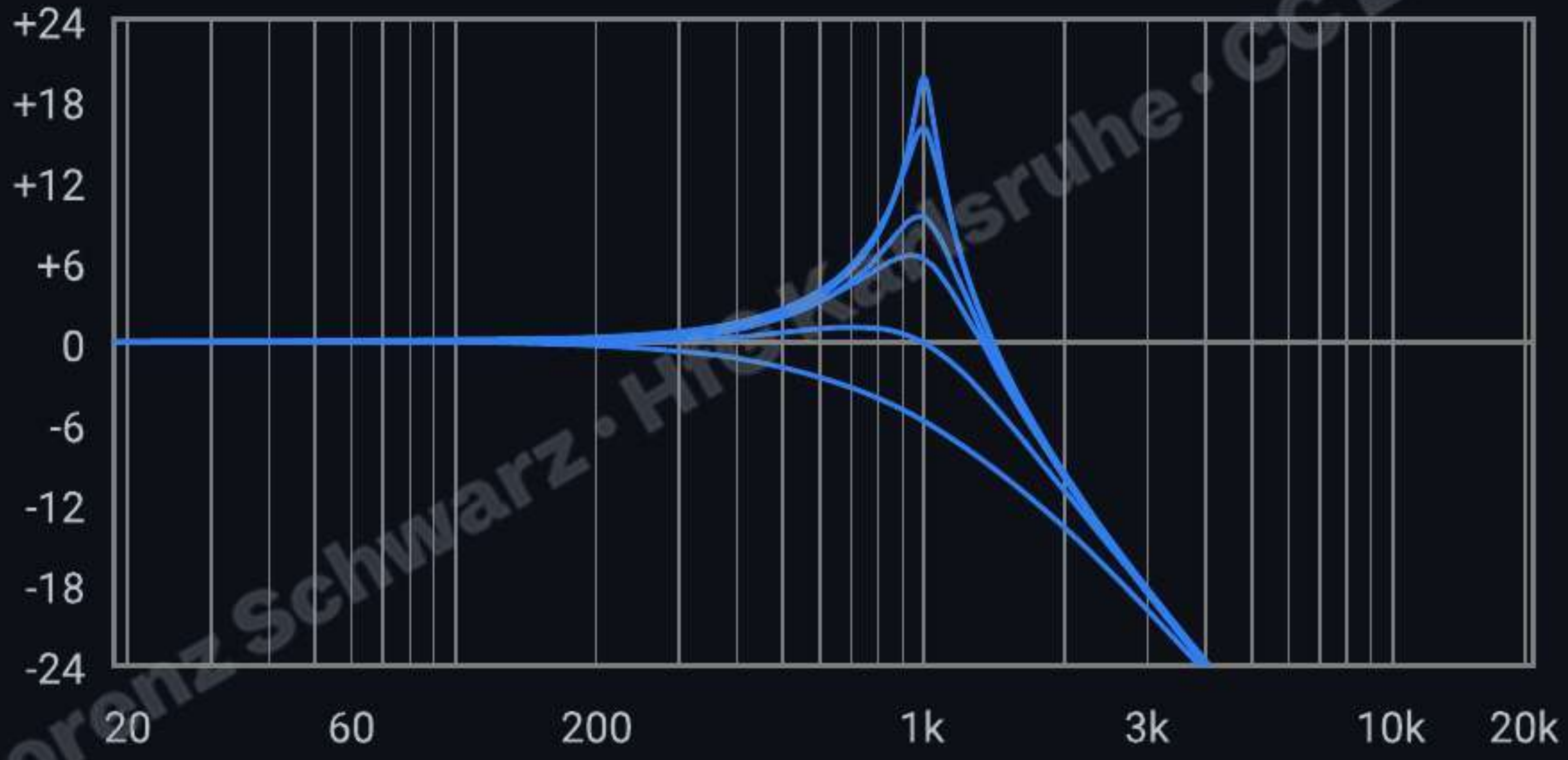
A low-pass filter is the opposite of a high-pass filter:

- It allows low-frequency components to pass while attenuating frequencies above the cutoff frequency.
- Commonly used to remove high-frequency noise.

Low-pass filter



Low-pass with different Q factors



Applications of low-pass filtering

Low-pass filters attenuate high-frequency content and are widely used in audio processing and sound design.

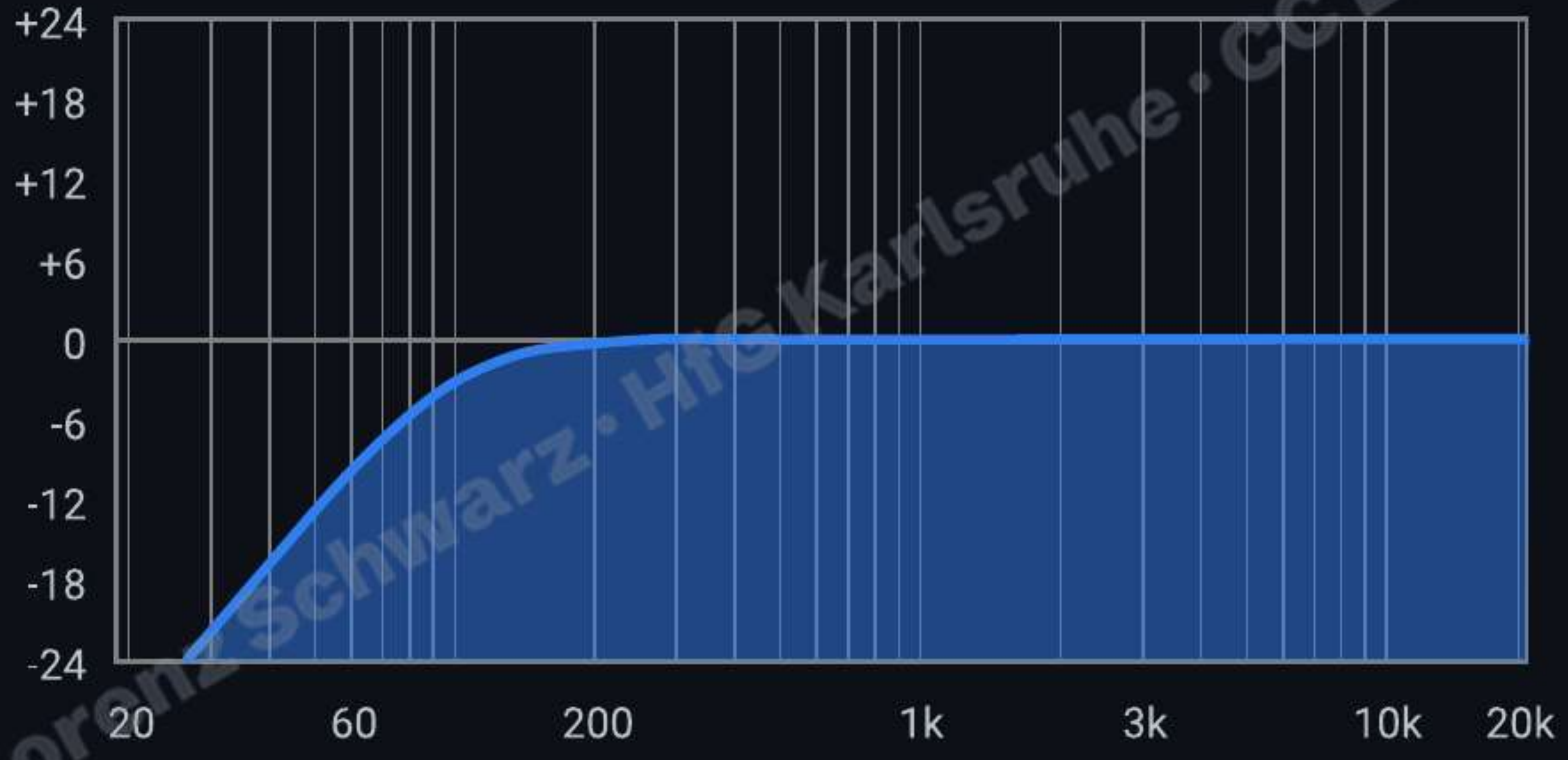
- Removing high-frequency hiss or noise from recordings
- Creating a muffled sound to simulate distance or obstruction
- Subtractive synthesis by shaping timbre from harmonically rich waveforms
- Anti-aliasing before digital sampling

High-pass filter (HPF) - low cut

A high-pass filter is the opposite of a low-pass filter:

- It allows high-frequency components to pass while attenuating frequencies below the cutoff frequency.
- Commonly used to remove low-frequency noise or rumble from audio signals.

High-pass filter



Applications of high-pass filtering

High-pass filters attenuate low-frequency content and are commonly used to improve clarity and technical efficiency in audio systems.

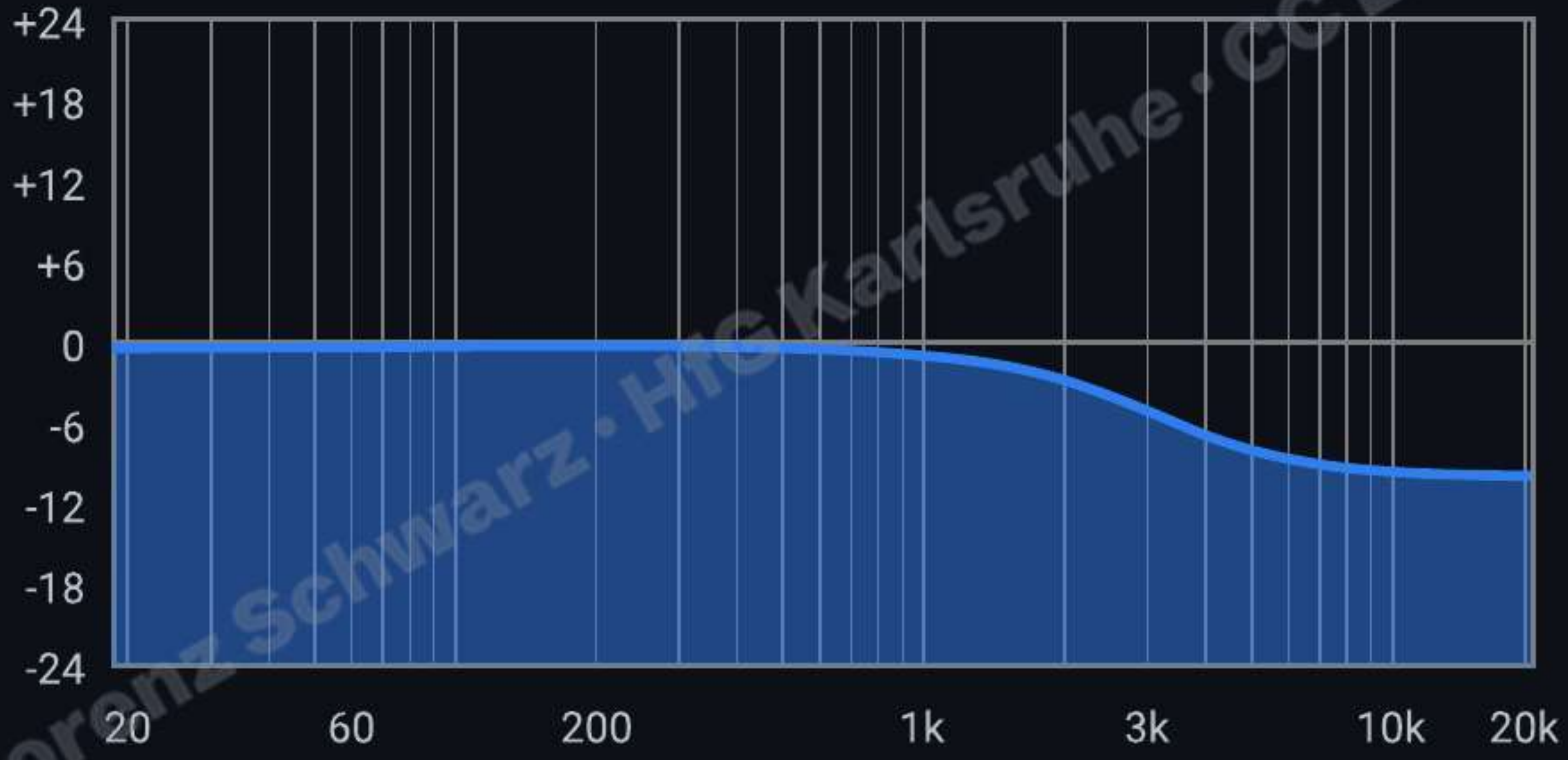
- Removing rumble or handling noise from recordings
- Filtering instruments that do not require bass content
- Preventing low frequencies from wasting amplifier headroom

Shelving filters

Unlike low-pass and high-pass filters, shelving filters boost or cut frequencies to a fixed level and then plateau rather than continuing to attenuate.

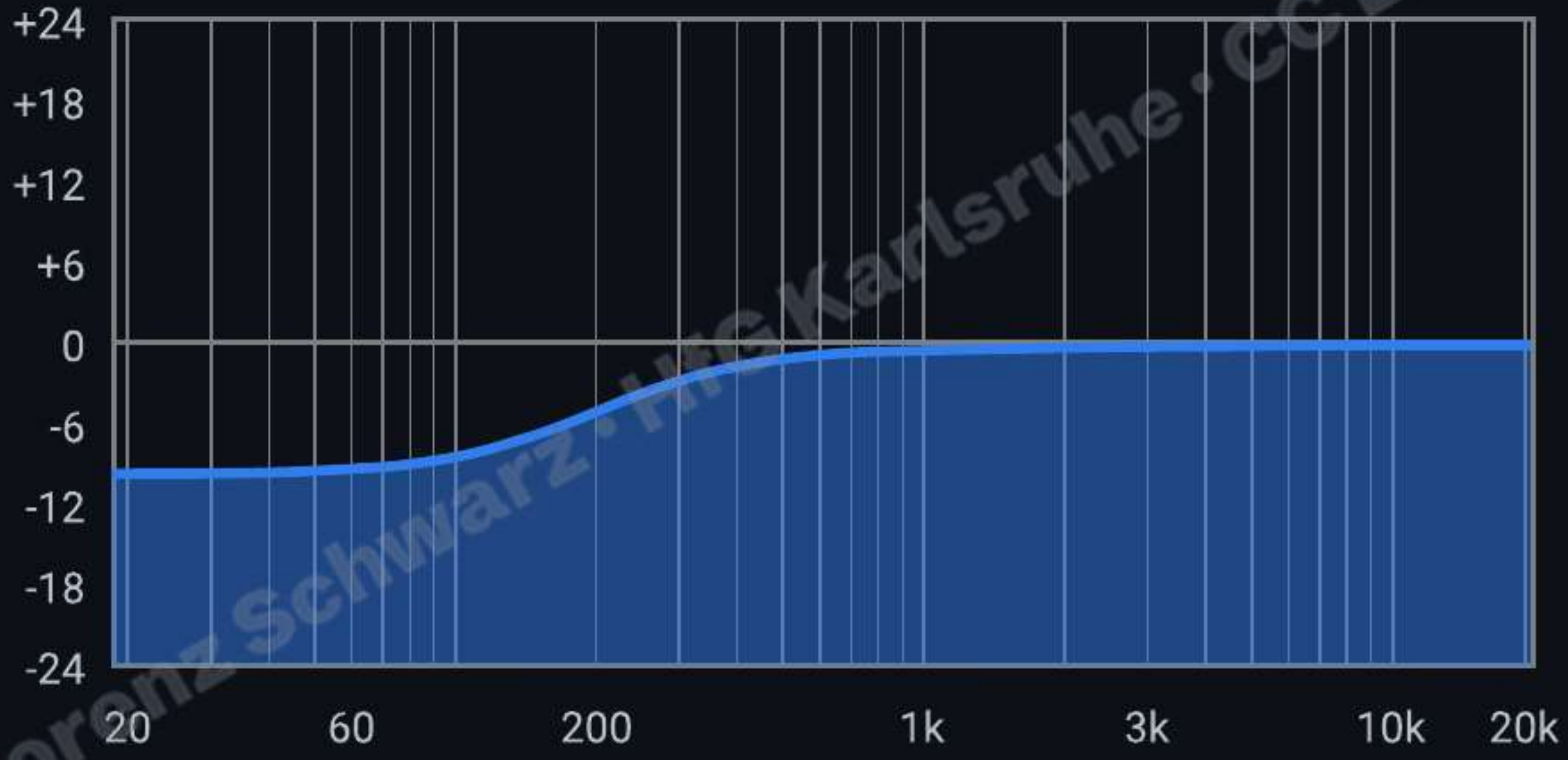
- Low-shelf filters affect frequencies below a cutoff frequency
- High-shelf filters affect frequencies above a cutoff frequency
- Commonly used for general tone control in audio systems and mixing desks

High shelf (cut)



High shelf (boost)

Low shelf (cut)



Low shelf (boost)

Creative use of shelving EQ

Shelving equalizers are commonly used for broad, musical tonal shaping rather than precise corrective filtering.

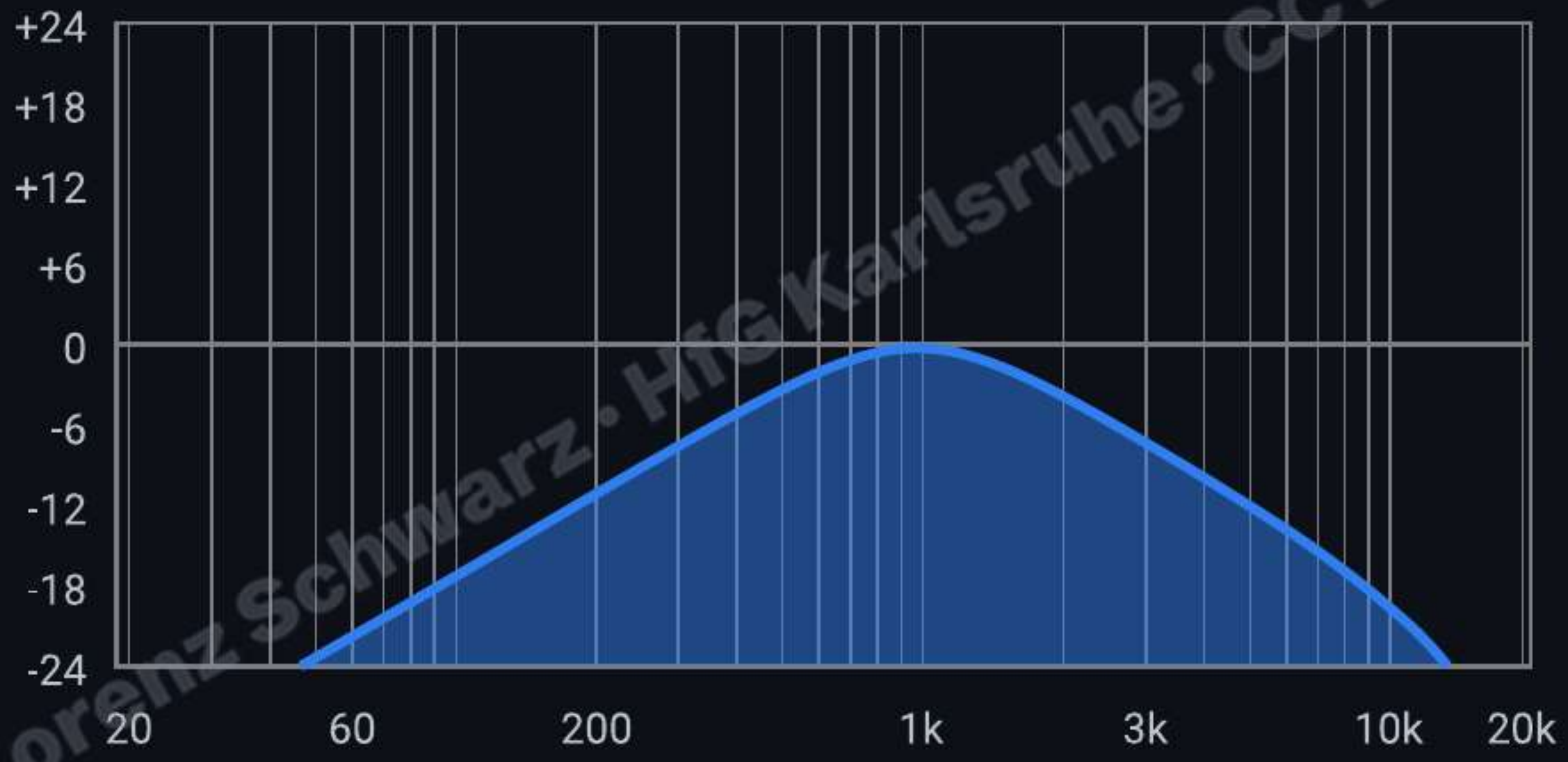
- **High-shelf boost:** adds “air” or presence (≈ 10–12 kHz and above)
- **High-shelf cut:** tames harshness; can act as a gentle de-essing alternative
- **Low-shelf boost:** adds warmth and weight (use carefully to avoid muddiness)
- **Low-shelf cut:** reduces boominess and proximity effect

Band-pass filters

Band-pass filters target a specific frequency range rather than everything above or below a single cutoff, making them useful for isolating or emphasizing selected spectral content.

- Defined by a lower and an upper cutoff frequency
- Frequencies outside this band are attenuated
- Commonly used to isolate or remove specific frequency components

Band pass filter



Applications of band-pass filtering

Band-pass filters isolate a defined frequency range.

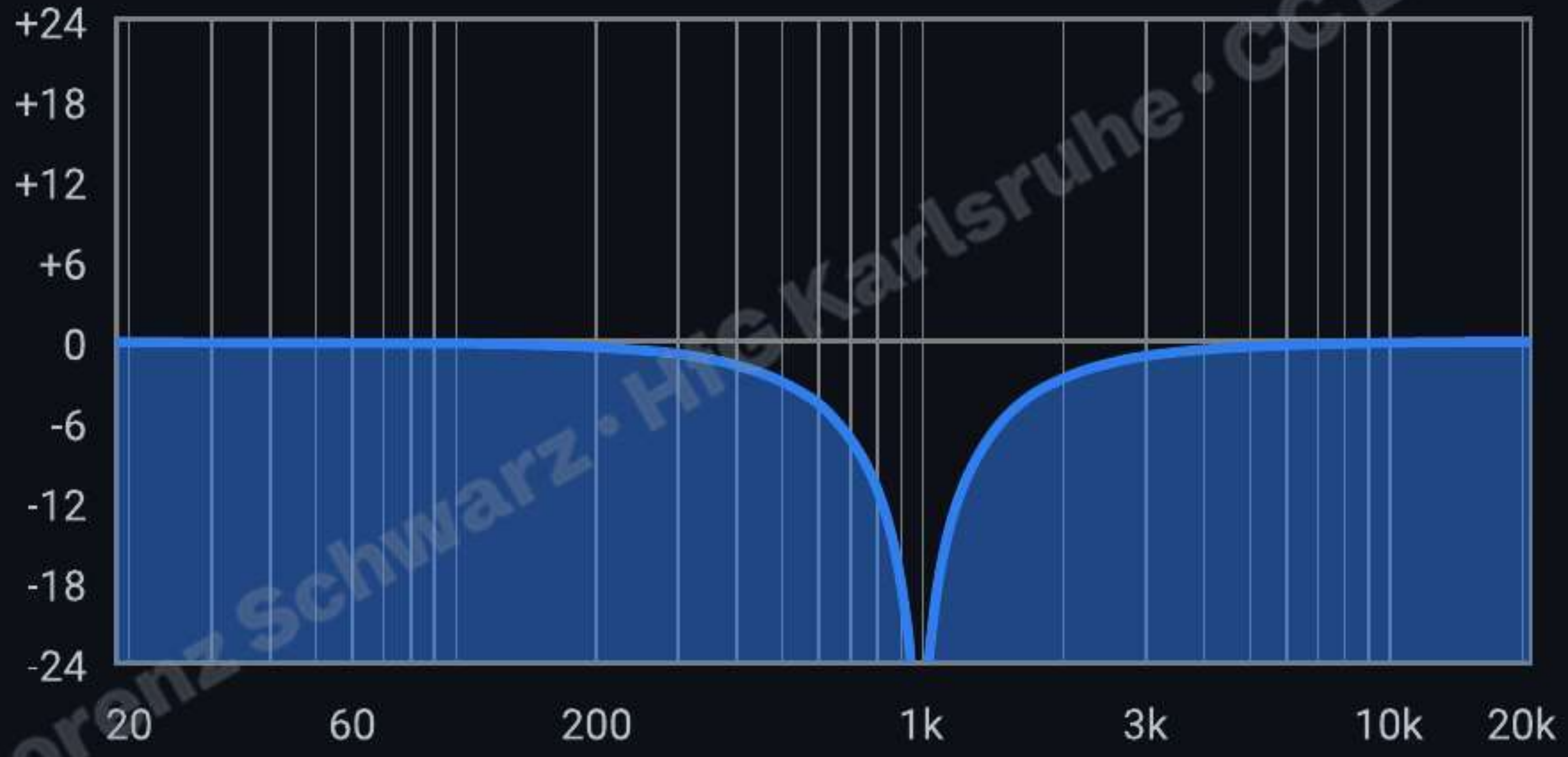
- Creating “telephone” or “radio” voice effects (narrow band-pass ≈ 300 Hz–3 kHz)
- Formant filtering in vocoders
- Wah-wah effects using a swept band-pass filter

Band-stop

A band-stop filter is the opposite of a band-pass filter:

- Frequencies within a specified range (the filter band) around the center frequency are suppressed, while frequencies outside this range are allowed to pass.

Band-stop filter

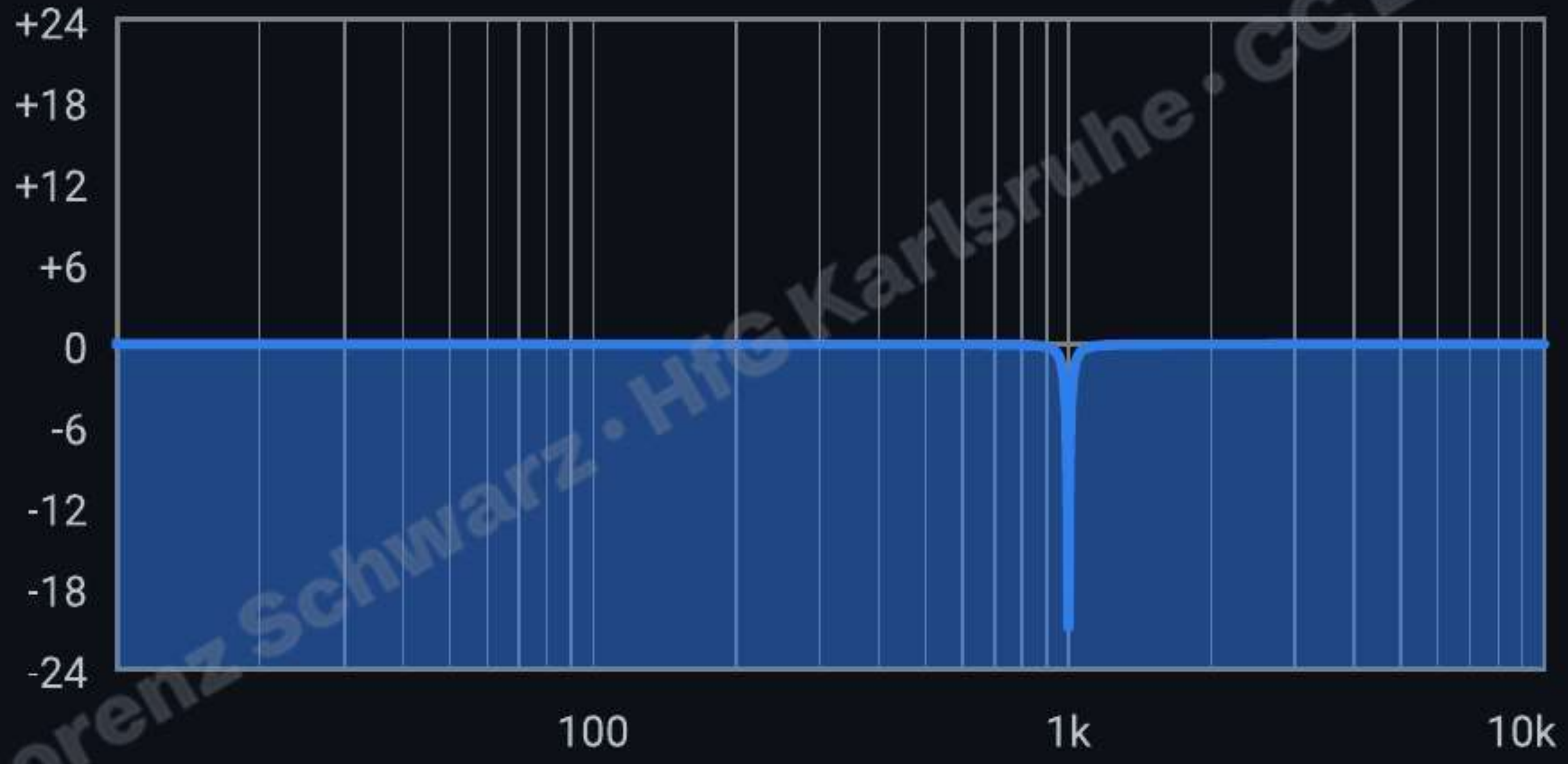


Notch filter

A notch filter is an extremely narrow banded type of band-stop filter, designed to attenuate a very specific frequency or a small range of frequencies while leaving other frequencies unaffected.

- Commonly used to eliminate unwanted sounds such as power line hum (e.g., 50 Hz or 60 Hz) or unwanted resonances (e.g., feedback).

Notch filter



Notch filters

Notch filters remove a very narrow frequency band while leaving the rest of the spectrum largely unaffected.

- Eliminating mains hum (50 Hz in Europe, 60 Hz in North America) and its harmonics
- Feedback suppression in live sound by targeting problematic resonant frequencies
- Removing room modes in live sound mixing or on location recordings

→ *Notch filters enable precise removal of unwanted frequencies without altering overall timbre.*

Peak filter

A peak filter boosts or attenuates frequencies within a specified range around a center frequency, forming a "bell-shaped" response.

- Commonly used in multi-band EQs to target specific frequency bands for precise adjustments without affecting surrounding frequencies.

Peak filter

Peak (bell) filters

Peak (bell) filters are the primary building blocks of parametric equalizers, allowing localized gain adjustments around a center frequency.

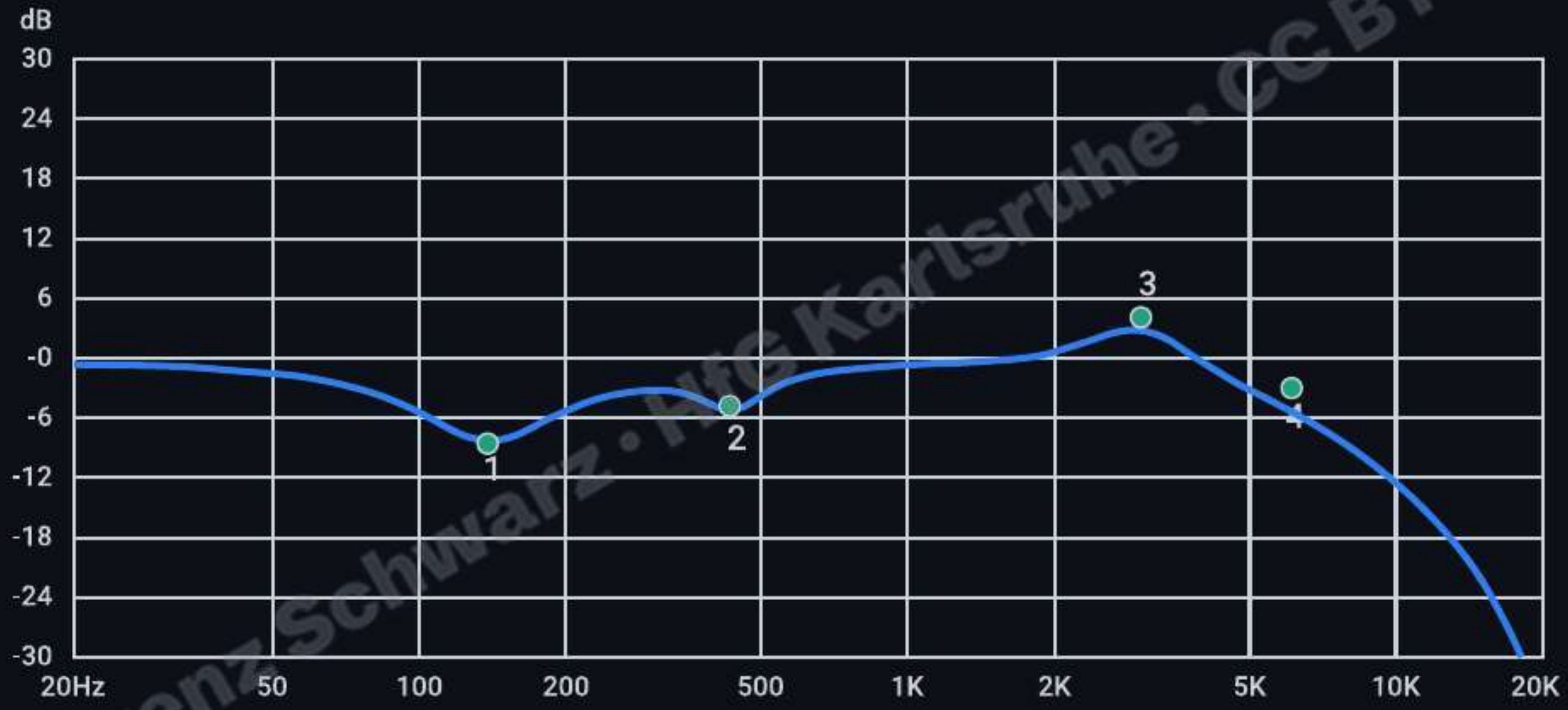
- Used for corrective (problem-focused) and tonal shaping
- Narrow bandwidth (high Q): precise attenuation of resonances
- Wide bandwidth (low Q): broad spectral shaping

Parametric EQ

Parametric equalizers are more versatile than graphic equalizers:

- Multiple bands can be adjusted independently.
- Variable Parameter such as center frequency, gain (boost or cut), bandwidth, and Q-factor of each band can be controlled.

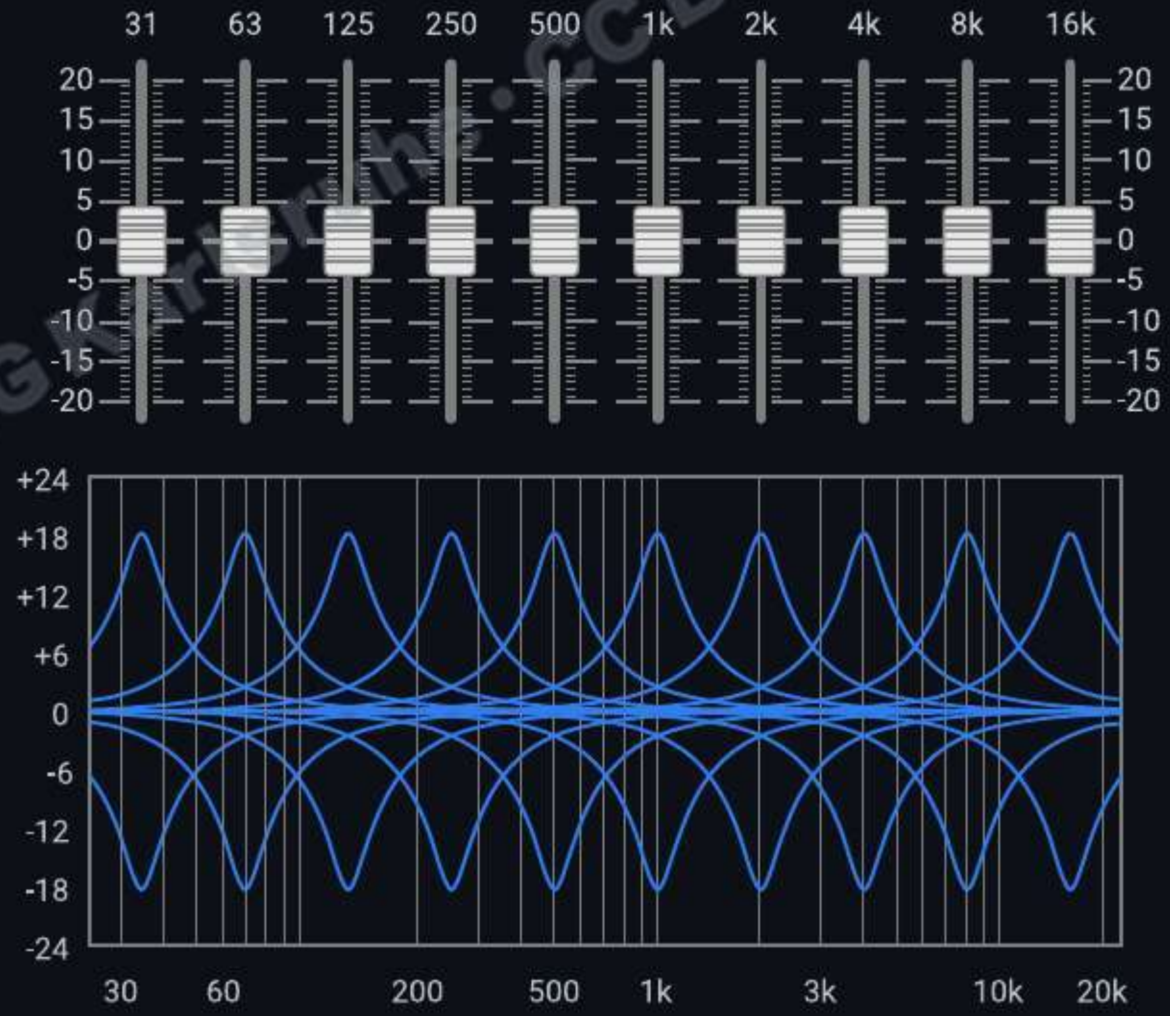
4-band parametric EQ (low pass)



Graphic EQ

A graphic equalizer allows amplification or attenuation of predetermined frequency bands using adjustable faders.

- Position of the faders visually represent the frequency curve of the filter
- The width of each frequency band (Q) remains constant.



Parametric vs. graphic EQ

Aspect	Parametric EQ	Graphic EQ
Precision	Exact control of center frequency and bandwidth	Fixed frequency bands, less precise
Operation	More complex, multiple parameters per band	Intuitive, one control per band
Typical use	Studio mixing and mastering	Live sound and room correction

Common filter types

Filter Type	Design Characteristic	Typical Application
Butterworth	Maximally flat passband, no ripple	General-purpose filtering, natural response
Chebyshev	Steeper roll-off, ripple in passband or stopband	Narrow transition band required
Bessel	Linear phase response (linear group delay), gentle roll-off	Preserving transients, pulse shaping
Elliptic	Steepest roll-off, ripple in both bands	Maximum frequency selectivity needed

FIR and IIR filters (digital filters)

The impulse response (IR) describes a filter's output when excited by a single-sample impulse.

Finite Impulse Response (FIR)

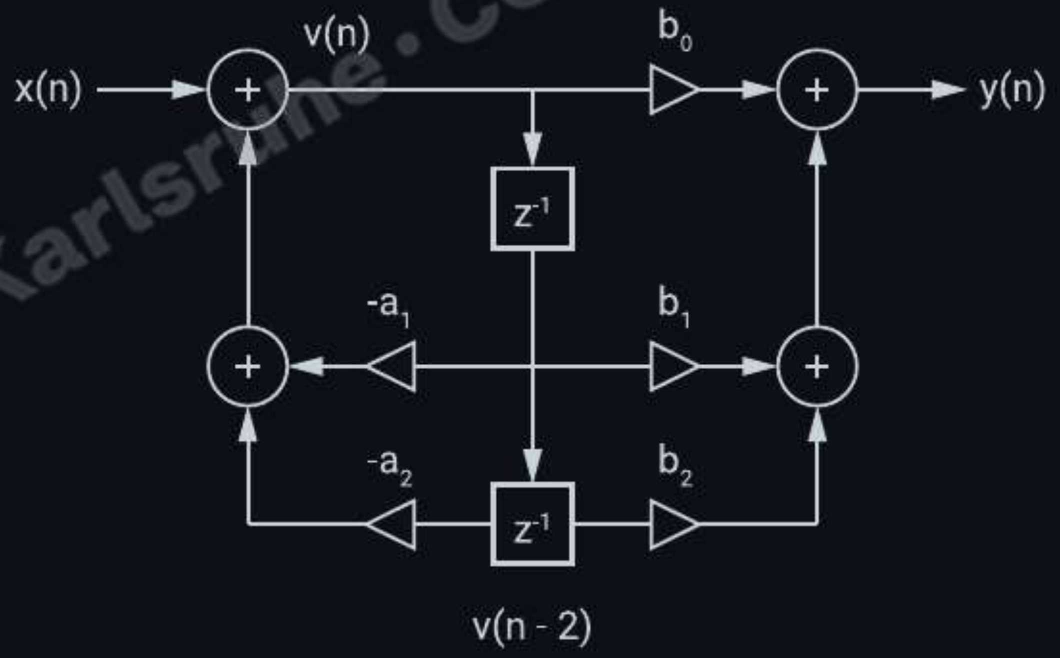
- IR decays to zero in finite time
- Linear phase response
- Higher computational cost and latency
- No direct analog counterpart
- Always stable
- Common in mastering

Infinite Impulse Response (IIR)

- IR is theoretically infinite
- Nonlinear phase response
- Computationally efficient, low latency
- Models traditional analog filters
- Stability depends on filter design
- Common in real-time processing

Biquad filter

Biquad filters are the building blocks of most digital EQs and filters. A single biquad can implement LP, HP, BP, notch, peak, or shelf filters by changing its coefficients. Higher-order filters are built by cascading multiple biquads.



Second-order DF-II structure

ANATOMY OF HEARING

Human hearing

The process by which variations in ambient atmospheric pressure (sound waves) are:

1. Captured by the ear
2. Converted into neural signals by the auditory system
3. Interpreted by the brain

→ *Unlike other senses, the ears cannot be 'closed,' making them continuously receptive to sound stimuli.*

Hearing as remote tactile perception

Hearing can be described as a 'tactile remote sense' because it involves detecting variations in air pressure through specialized mechanoreceptors in the inner ear.

- Vision: electromagnetic radiation
- Smell: chemical detection
- Hearing: mechanical pressure waves

Auditory signal processing stages

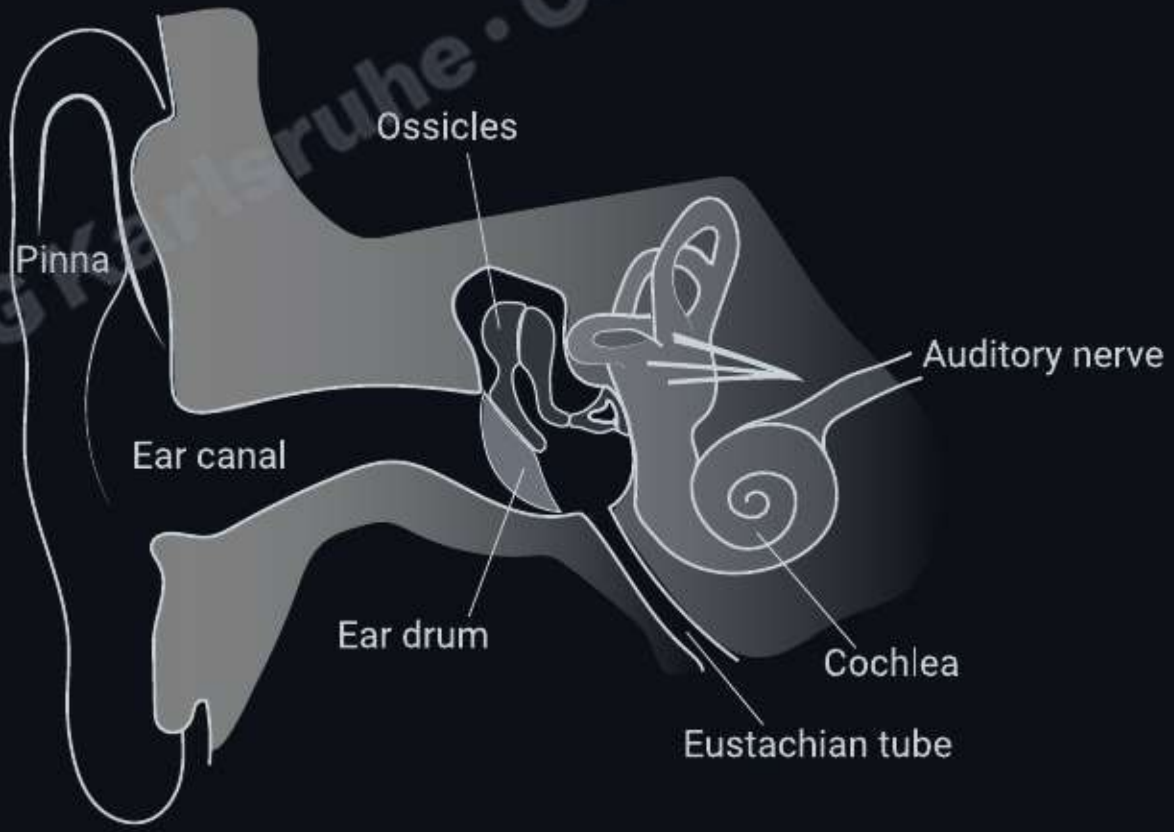
Sound waves → Outer ear (collection) → Middle ear (impedance matching) → Inner ear (transduction) → Auditory nerve → Brain (perception)

Three main stages:

1. **Physical:** Sound field characteristics
2. **Mechanical/neural:** Encoded into neural signals by the auditory system
3. **Perceptual:** Processed by the central nervous system and integrated with other sensory information

Anatomy of the ear

- **Outer ear:** Pinna and ear canal; collects and funnels sound
- **Middle ear:** Eardrum and ossicles; impedance matching
- **Inner ear:** Cochlea (hearing) and vestibular system (balance); transduction
- **Auditory nerve:** Transmits neural signals to brain

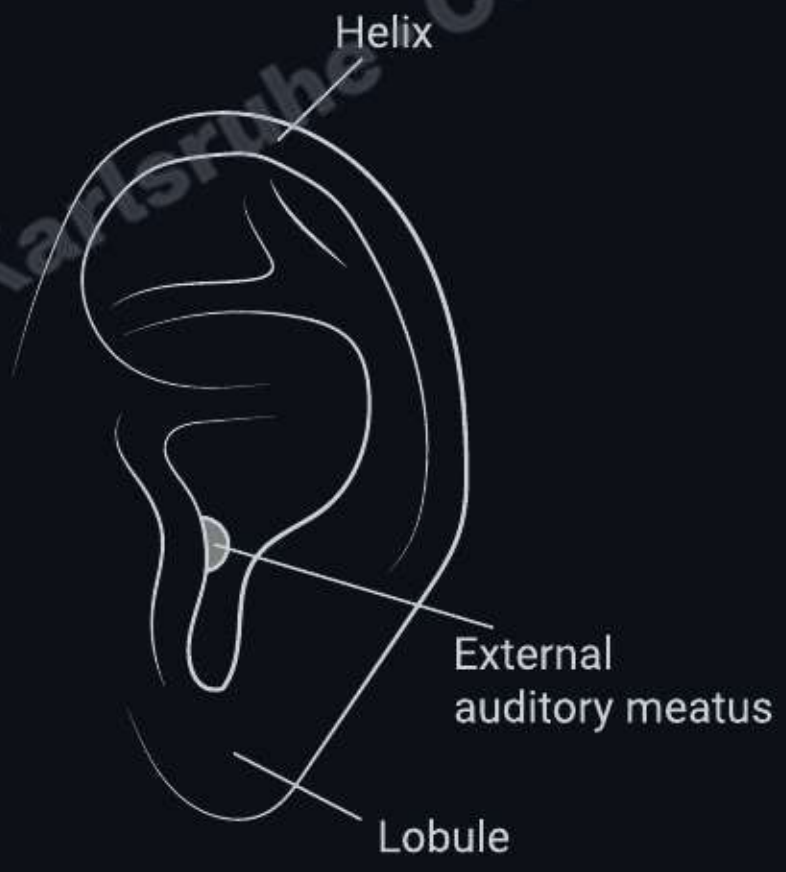


Outer ear

The auricle (pinna) is the visible, irregularly shaped part of the outer ear that encloses the ear canal.

- Directional filtering for sound localization (HRTF)
- Amplifies 2–4 kHz by ~10–15 dB (speech range)
- Protects the eardrum

→ *The unique shape of each person's pinna creates personalized spatial audio cues.*



Middle ear

The eardrum (tympanic membrane) vibrates in response to sound pressure. Three small bones (ossicles) transmit vibrations:

1. Malleus (hammer)
2. Incus (anvil)
3. Stapes (stirrup)

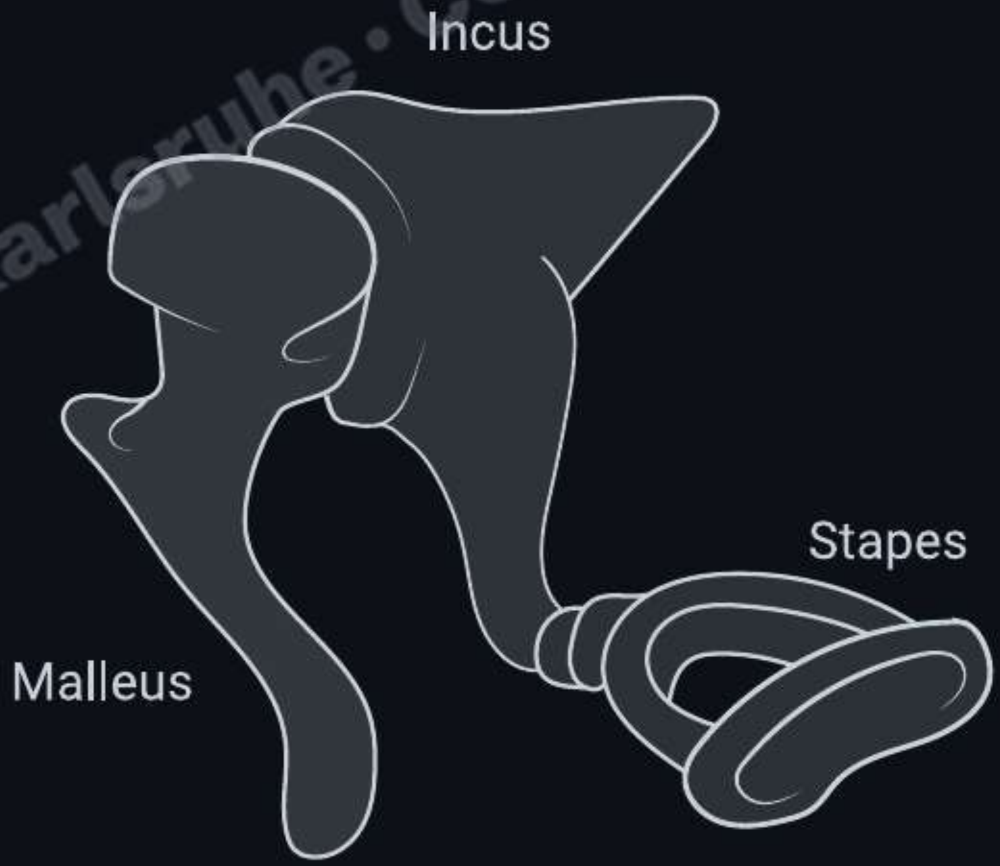
→ *The middle ear enables impedance matching between air and fluid-filled inner ear.*

Auditory ossicles

The ossicles mechanically amplify sound vibrations:

- **Area ratio:** Eardrum (~55 mm²) is ~17× larger than oval window (~3.2 mm²)
- **Lever action:** Ossicles act as a lever system

The stapes footplate connects to the oval window, transmitting vibrations into the fluid-filled cochlea.

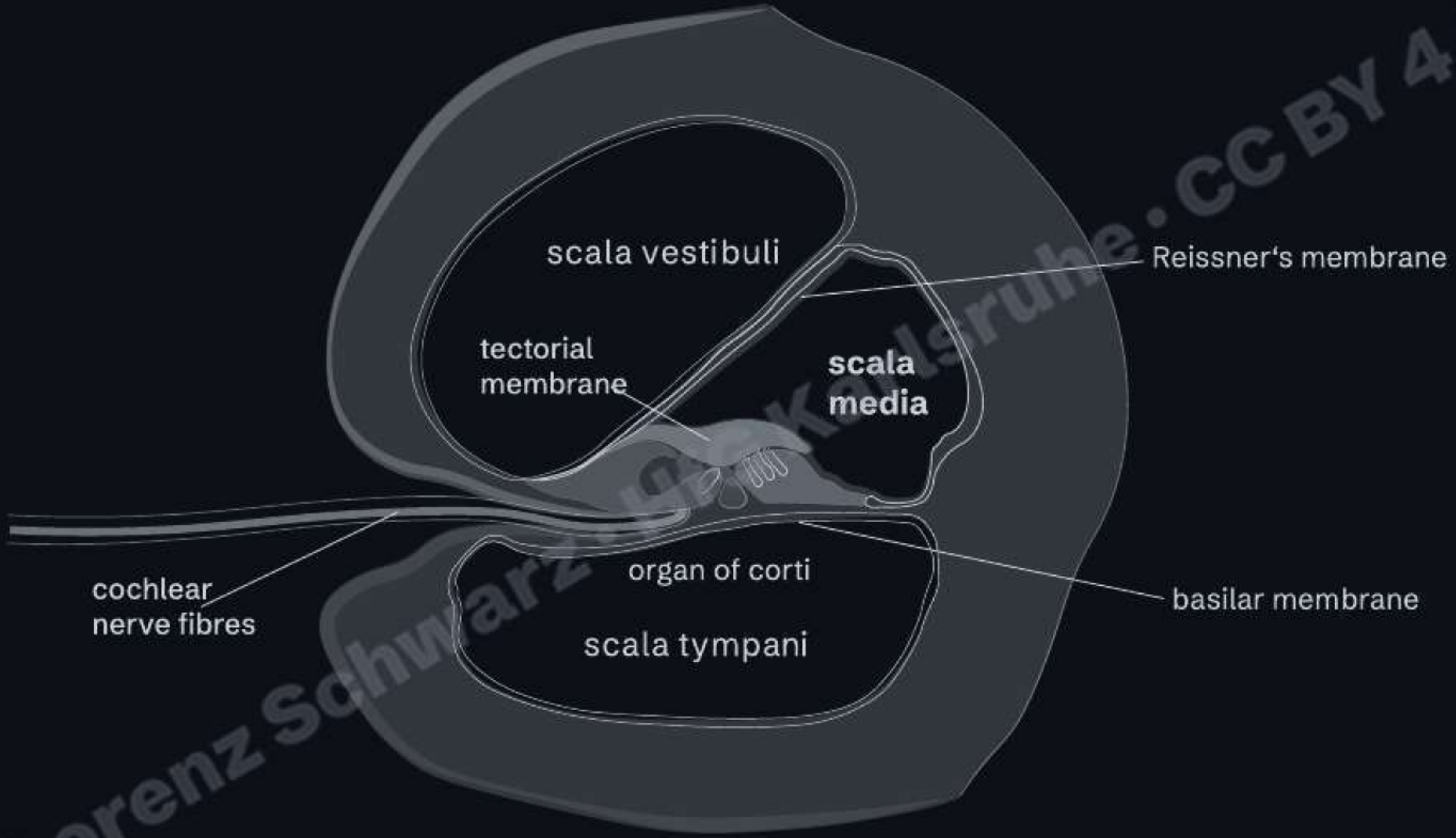


Inner ear and the cochlea

The cochlea is a spiral-shaped, fluid-filled structure, consisting of three compartments:

- Scala vestibuli (perilymph)
- Scala media (endolymph)
- Scala tympani (perilymph)

The oval window receives vibrations from stapes, while the round window allows pressure release for fluid displacement.



Cross section of cochlea

Traveling wave and frequency analysis

Sound creates a traveling wave along the basilar membrane. Different frequencies cause the wave to peak at different locations:

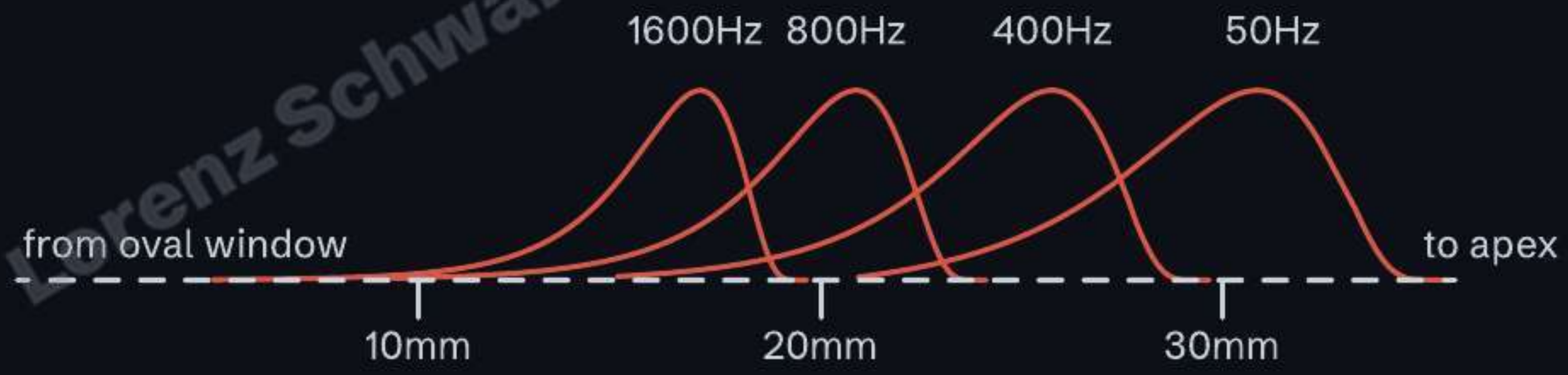
- **High frequencies (20 kHz):** Peak near the base (oval window)
- **Low frequencies (20 Hz):** Peak near the apex (tip of cochlea)

→ *The cochlea performs mechanical frequency analysis through tonotopic organization.*

Frequency tuning of the basilar membrane

The basilar membrane exhibits a frequency-specific gradient along its length, with each position responding to a specific frequency in an approximately logarithmic mapping.

- **Base:** Narrow and stiff → sensitive to high frequencies
- **Apex:** Wide and flexible → resonates at low frequencies



Organ of Corti

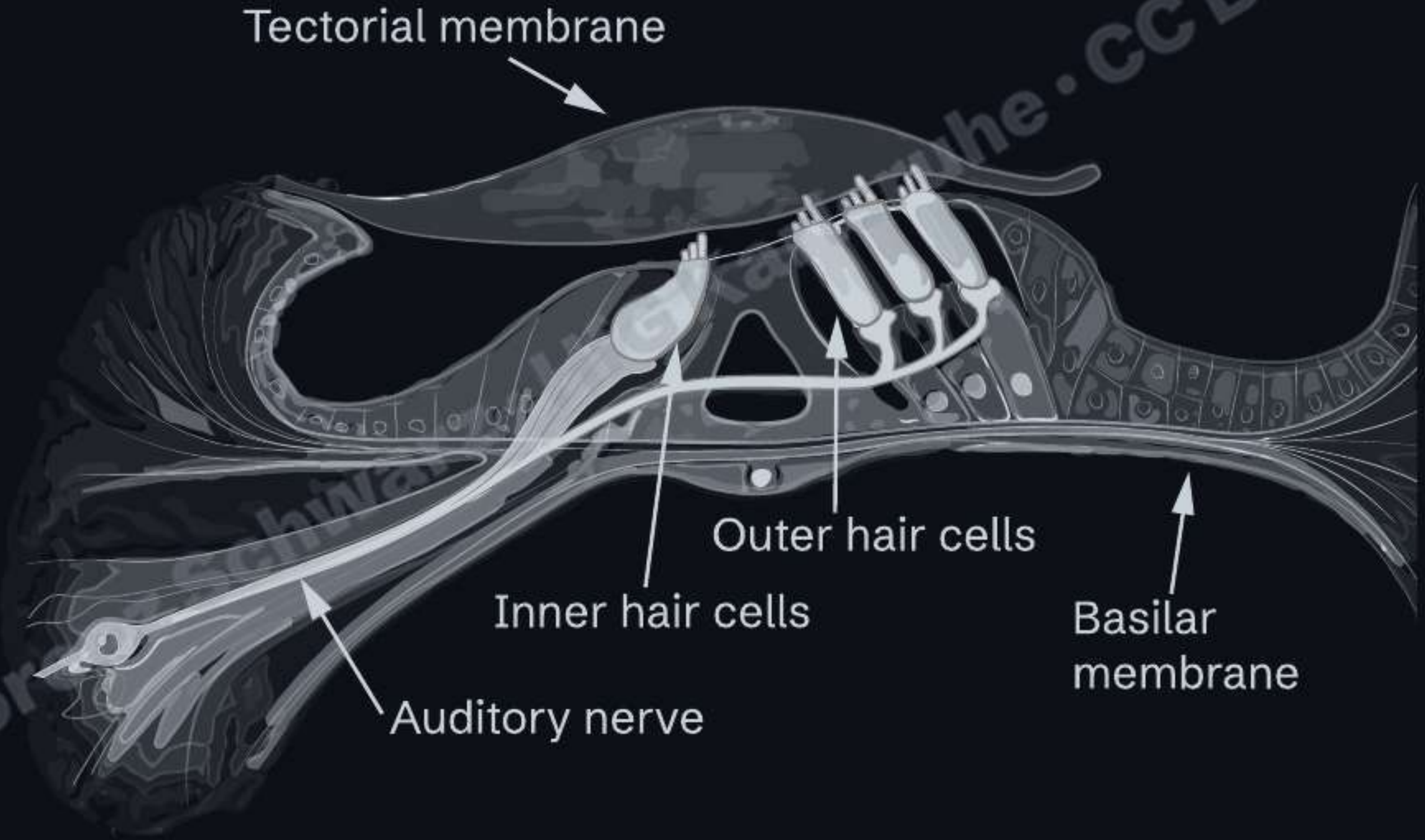
The organ of Corti sits on the basilar membrane and contains the sensory hair cells.

Four rows of hair cells:

- **Inner hair cells (1 row):** ~3,500 cells; transduce vibrations into neural signals
- **Outer hair cells (3 rows):** ~12,000 cells; amplify vibrations (cochlear amplifier)

Tectorial membrane: Overlying structure that bends hair cell stereocilia during basilar membrane motion

Organ of Corti



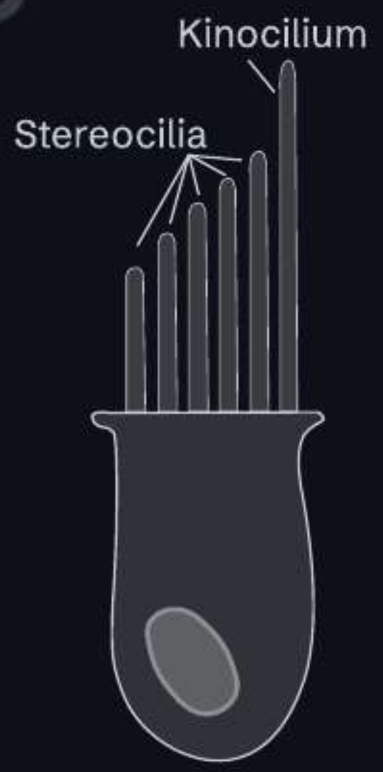
Hair cell transduction

Mechanical to neural conversion:

1. Basilar membrane movement bends hair cell stereocilia
2. Bending opens ion channels
3. Ion flow triggers neurotransmitter release
4. Auditory nerve fibers fire action potentials

Encoding:

- Firing rate encodes sound intensity
- Firing timing encodes frequency information



Hair cell

Cochlear nerve

~30,000 auditory nerve fibers transmit signals from the cochlea to the brain.

- Each inner hair cell connects to 10–30 nerve fibers
- Different fibers have different thresholds and dynamic ranges
- Collectively encode intensity range of ~120 dB

Tonotopic organization preserved: Nerve fibers maintain frequency-specific organization through brainstem to auditory cortex.

Frequency response of human hearing

Human hearing is most sensitive to mid-frequencies (2–5 kHz):

- **Ear canal resonance:** Amplifies 2–4 kHz
- **Evolutionary advantage:** Speech intelligibility
- **Sensitivity varies with level:** Low frequencies are less audible at low SPLs

→ *A-weighting for noise measurements approximates this frequency-dependent sensitivity.*

PSYCHOACOUSTICS

Auditory signal processing

Human hearing is not merely the translation of mechanical processes into neural action potentials; it involves complex physiological and psychoacoustic signal processing in the inner ear and brain.

Psychoacoustics

Psychoacoustics examines the relationship between the physical properties of sound waves and the subjective perception they evoke in the listener.

Academic fields:

- Perceptual psychology
- Neuroscience
- Physics and acoustics
- Computer science

Listening tests and analysis

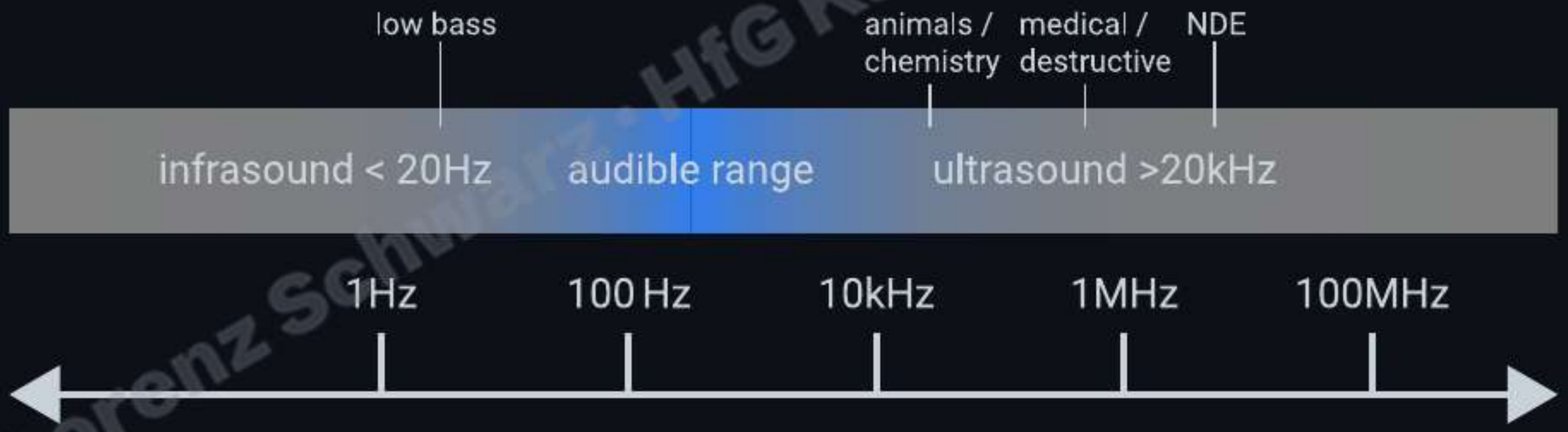
Psychoacoustic methods involve listening tests and statistical analysis of a large number of subjective judgments.

Applications of psychoacoustics

- **Music and sound design:** Shaping aesthetic and compositional approaches by optimizing sound for human perception.
- **Audio compression:** Reducing file sizes by leveraging perceptual limitations (e.g., MP3, AAC).
- **Auditory disorders:** Improving hearing aid design based on psychoacoustic principles.
- **Workplace & medical environments:** Designing alarms and machine-status signals for improved perception in noisy conditions (e.g., ICU, factories).

Human frequency range

- Humans perceive roughly **20 Hz - 20 kHz**
- **Below 20 Hz:** perceived as vibration (infrasound).
- **Above 20 kHz:** ultrasonic, inaudible but used in technology.

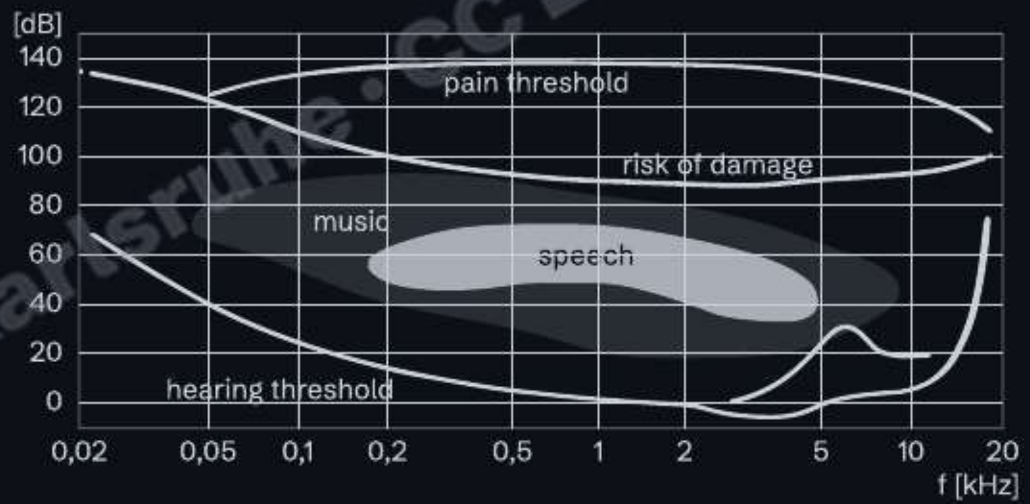


Human dynamic range

The ear detects an enormous dynamic range from the faintest sound to the threshold of pain:

- **Lower limit (threshold of hearing):** ~0.00002 Pa (20 µPa) at mid frequencies (~1 kHz), corresponding to 0 dB SPL.
- **Upper limit (threshold of pain):** Peaks of up to ~200 Pa, corresponding to 140 dB SPL.

$$DR = 20 \log \frac{200}{0.00002} = 140 \text{ dB}$$



Subjectivity of pitch and loudness

Frequency

→ Pitch
physical vs. perceived height of tone

Amplitude

→ Loudness
physical vs. perceived strength of sound

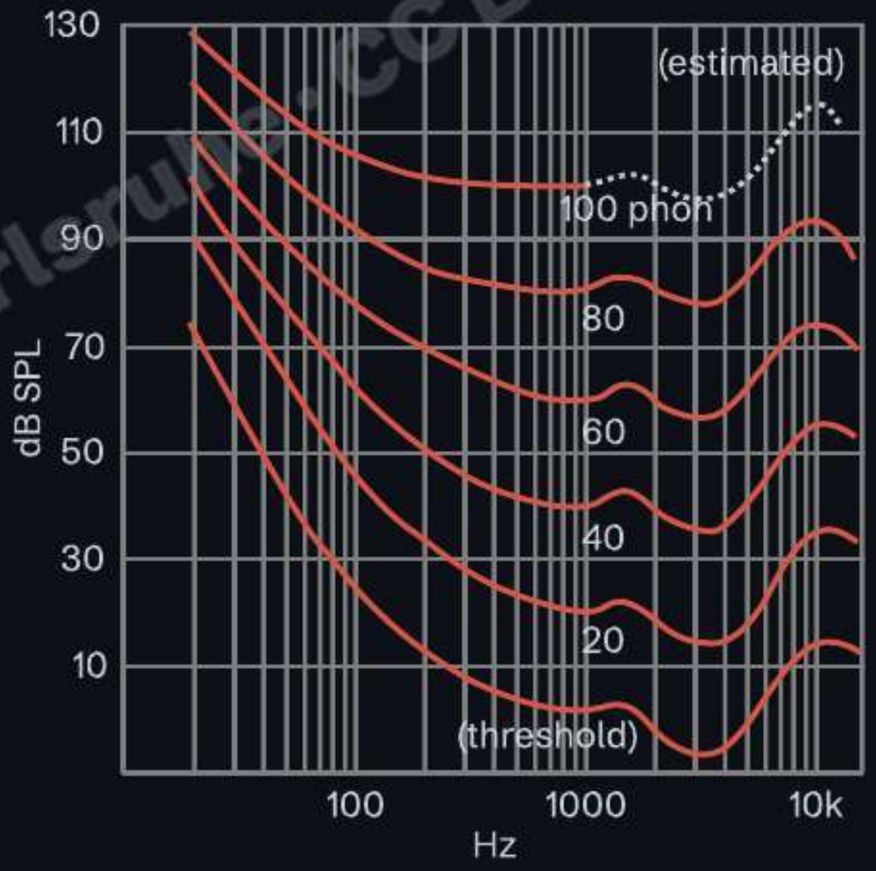
LOUDNESS · INTENSITY

Loudness

Human hearing is nonlinear, and sensitivity changes with frequency and sound level. We are most sensitive to the range of speech perception.

Two tones with identical physical amplitude can be perceived as having different loudness:

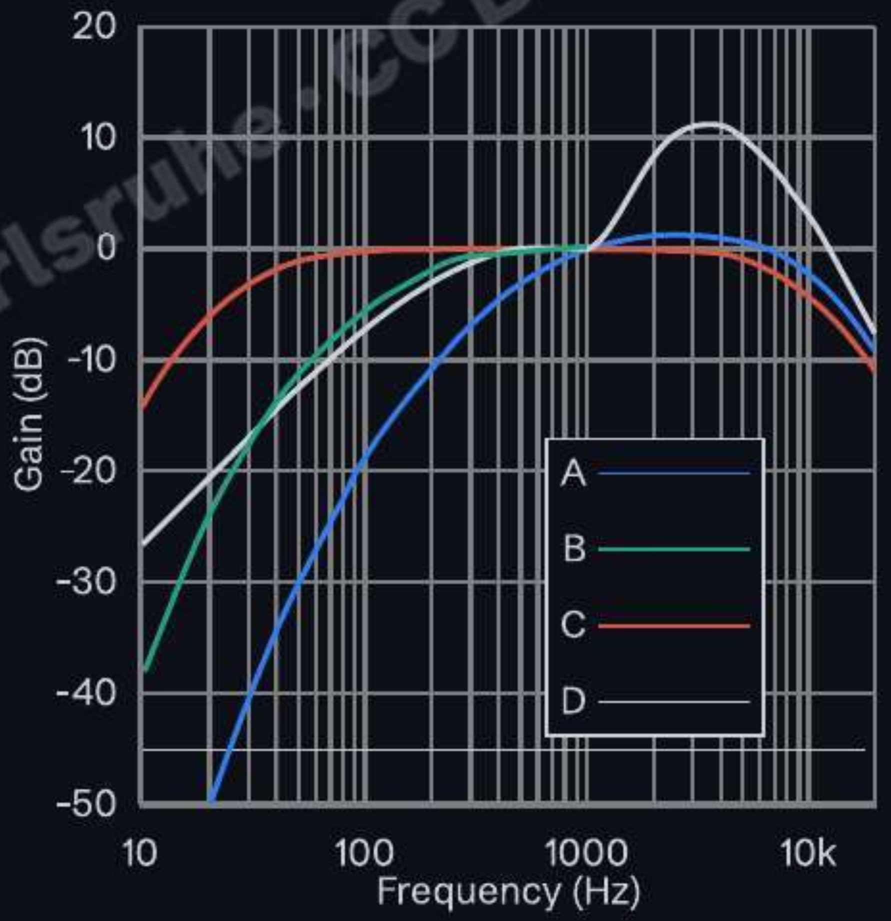
- ▶ 300 Hz sine wave
- ▶ 3000 Hz sine wave



Frequency weighting curves

Frequency weighting curves apply frequency-dependent filters to sound level measurements to approximate human hearing sensitivity at different SPLs.

A-weighting (dBA) approximates the inverse of the 40-phon equal-loudness contour and is the standard for audio engineering, environmental noise, and equipment specifications.



Quantifying perceived loudness

- **Phon:** The unit phon is derived from equal-loudness contours (isophones) and corresponds to the sound pressure level (SPL) in dB of a 1 kHz sine tone that is perceived as equally loud as a sound of any other frequency.
- **Sone:** The sone is a unit of perceived loudness where 1 sone corresponds to the loudness of a 1 kHz tone at 40 dB SPL. The loudness in sones doubles for every 10 dB increase in SPL, reflecting the logarithmic nature of loudness perception.

Phon	40	50	60	70	80	90	100	110	120	130	140
Sone	1	2	4	8	16	32	64	128	256	512	1024

Hearing protection

Understanding psychoacoustics also has practical consequences for health and safe listening practices

Safe exposure levels:

- 85 dB SPL: 8 hours maximum
- 88 dB SPL: 4 hours maximum
- 91 dB SPL: 2 hours maximum
- 100 dB SPL: 15 minutes maximum

→ *Noise-induced hearing loss is permanent. Hair cells do not regenerate!*

FREQUENCY PERCEPTION

Frequency vs. pitch

Frequency: Physical property of sound wave (Hz)

Pitch: Subjective perception, ordering sounds low to high

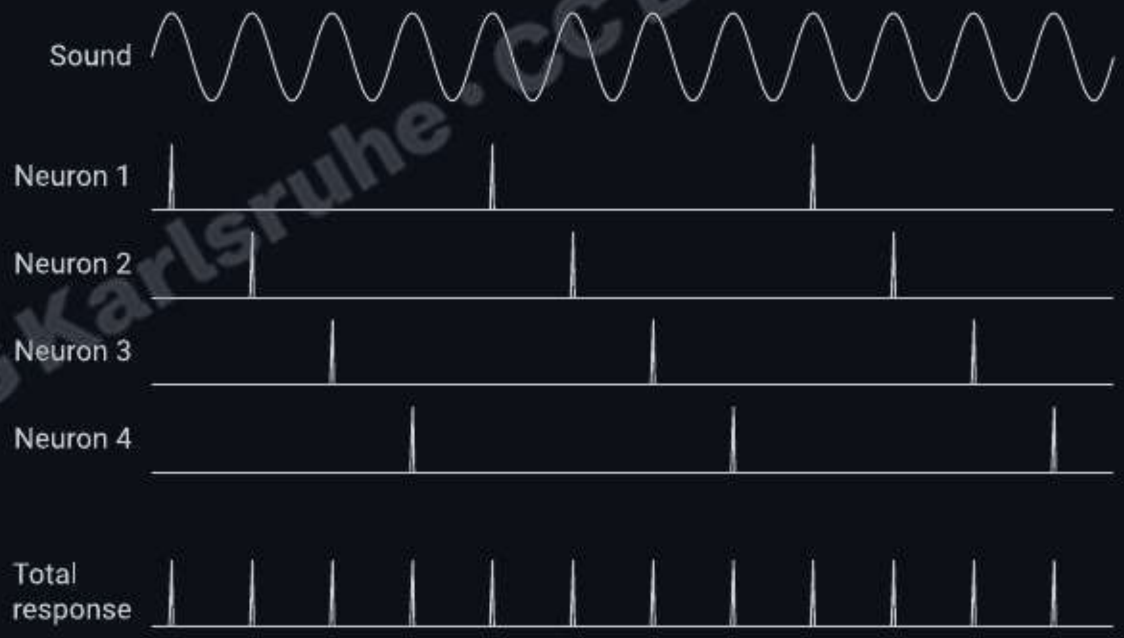
Many pitch-related illusions arise from the interaction of place and temporal coding:

- Missing fundamental
- Binaural beats
- Combination tones

Place and temporal theories

- **Place theory:**
  - Pitch is determined by the location of maximum excitation on the basilar membrane.
  - Dominant for frequencies above 5000 Hz.
- **Temporal theory:**
  - Pitch perception is based on neural firing patterns that synchronize with the sound wave's period (phase-locking).
  - Effective for frequencies below 1000 Hz.

→ *Neurons fire at a lower rate than the maximum audible frequency.*



Missing fundamental

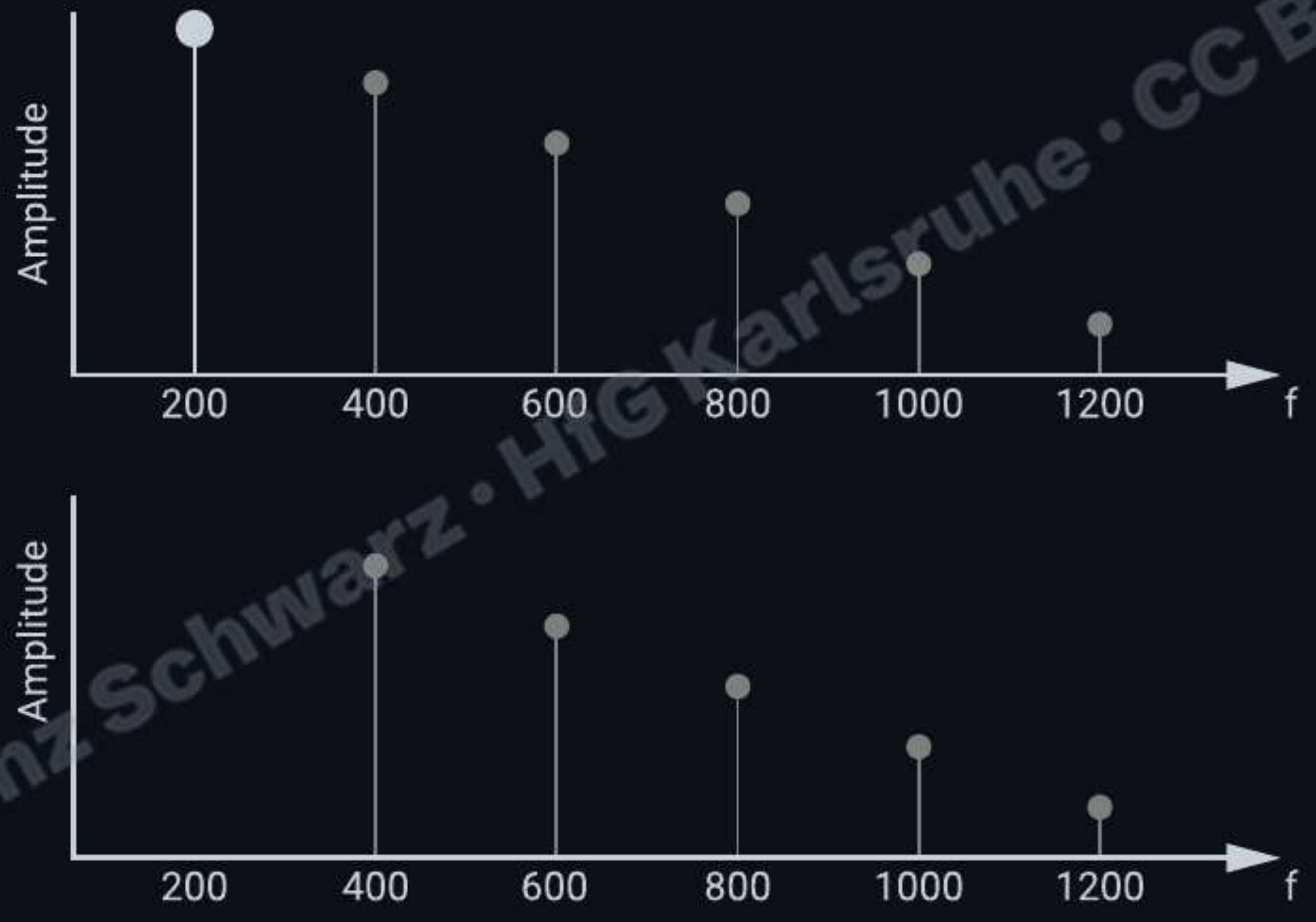
The "missing fundamental" of a complex tone is a psychoacoustic phenomenon where the brain perceives a fundamental frequency even when it is absent from the actual sound signal:

- repetition patterns of higher harmonics (periodicity)
- **Application:** pseudo low frequency psycho-acoustic sensation (MaxxBass)

▶ Melody with fundamental and harmonics

▶ Same melody without the first three harmonics

Missing fundamental



Frequency filtering in the auditory system

The auditory system functions as a series of bandpass filters, modeling the frequency-selective response of the basilar membrane.

Equivalent Rectangular Bandwidth (ERB) approximates auditory filter bandwidth:

$$ERB(f) = 0.108f + 24.7$$

where f is the center frequency in Hz.

→ *ERB increases with frequency—auditory filters become broader at higher frequencies.*

Just noticeable difference (JND)

The just noticeable difference (JND) is the smallest change in a sound property that can be reliably detected by human listeners.

The JND for loudness is approximately 1 dB SPL.

The JND for pitch varies with frequency:

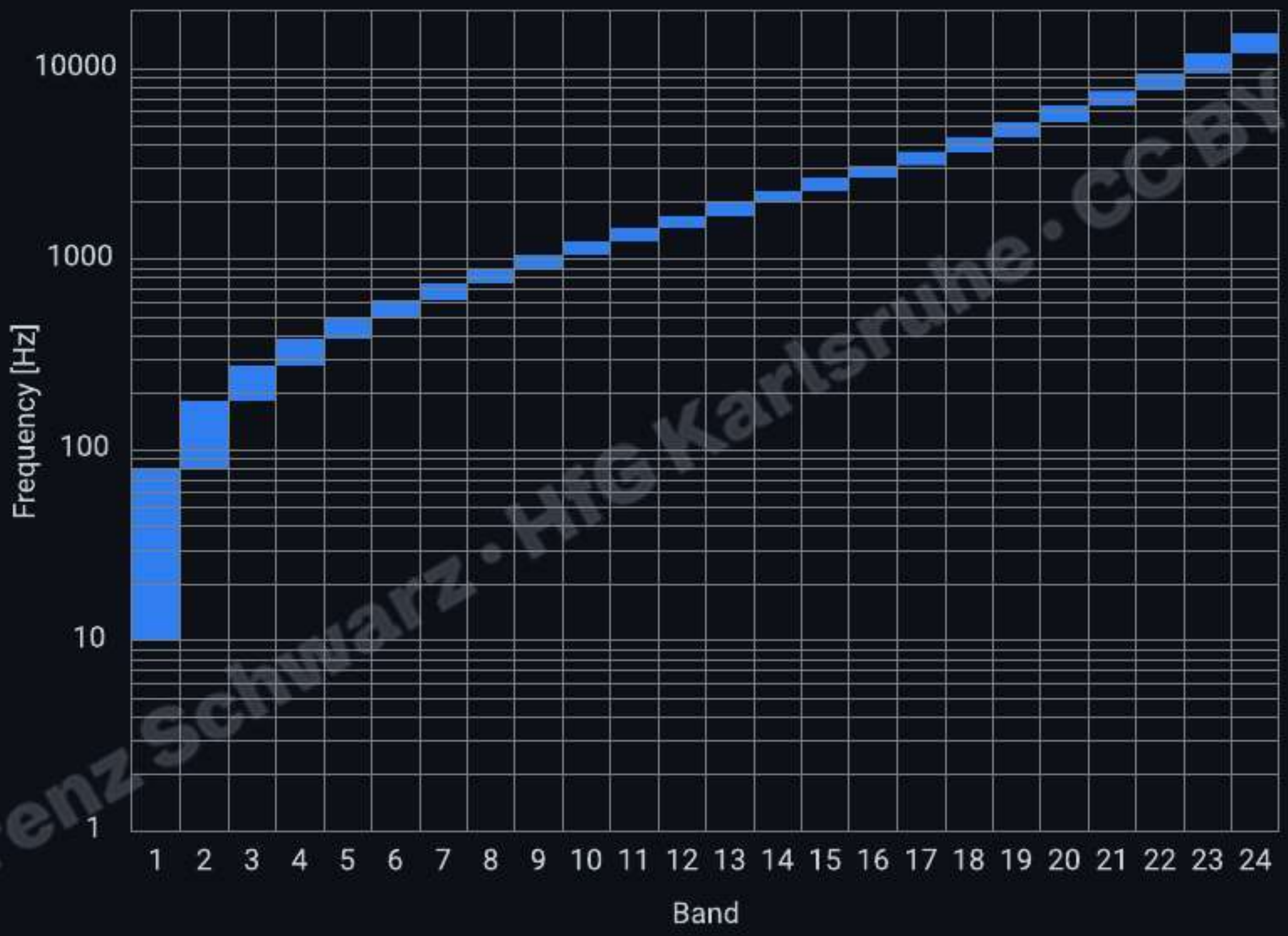
- Below 500 Hz: ~3 Hz for sine waves
- Above 1000 Hz: ~0.6% of the frequency (approximately 10 cents)

→ *The human auditory system can perceive approximately 1400 distinct pitch steps across the hearing range.*

Bark scale

The Bark scale, named after Heinrich Barkhausen, is a psychoacoustic scale that reflects the spatial frequency mapping of the **basilar membrane.**

- Models auditory frequency resolution
- Based on critical bands
- Used in perceptual audio processing


Frequency [Hz]
10000
1000
100
10
1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Band

Mel scale

The Mel scale models **perceived pitch.**

- **Below 500 Hz:** Nearly proportional to the linear frequency scale
- **Above 500 Hz:** Pitch intervals are perceived as **smaller** than their physical spacing

MASKING · CRITICAL BANDS

Auditory masking

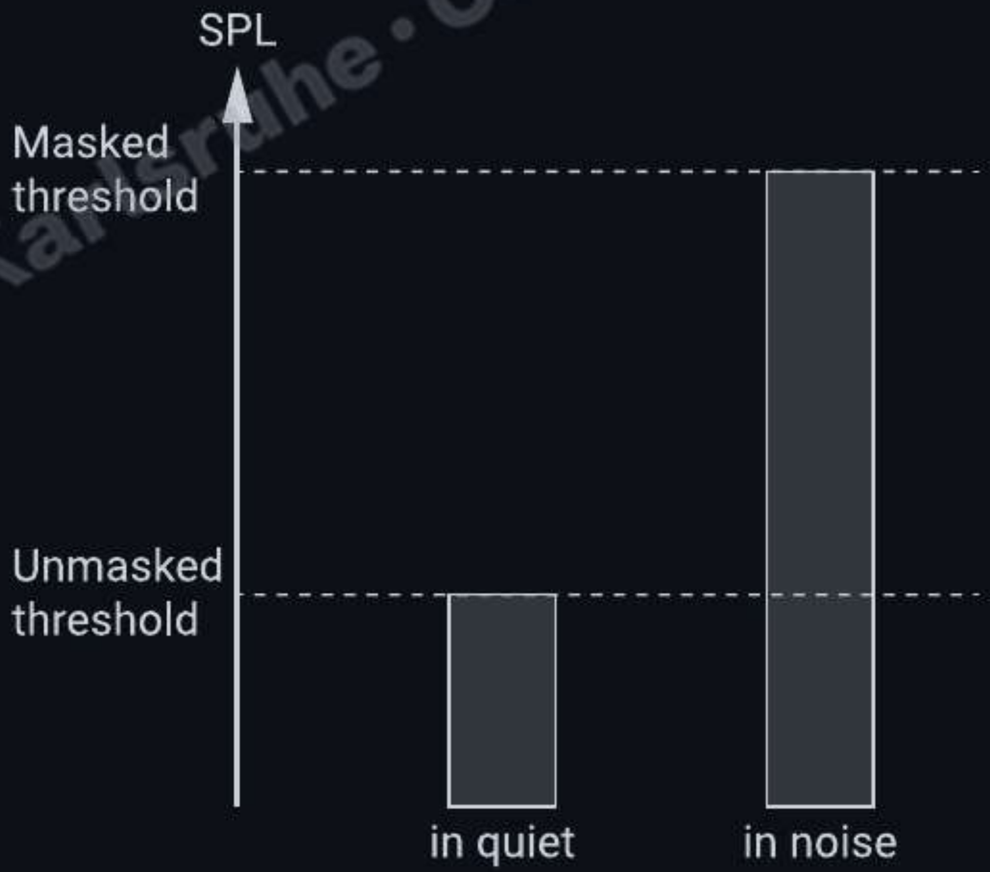
Auditory masking is a psychoacoustic phenomenon where the perception of one sound is affected by the presence of another, making the target sound less audible, depending on factors like frequency, intensity, and temporal proximity.

- temporal masking
- spectral masking

Masking threshold

The degree of masking is defined as the difference between the masked threshold and the unmasked threshold.

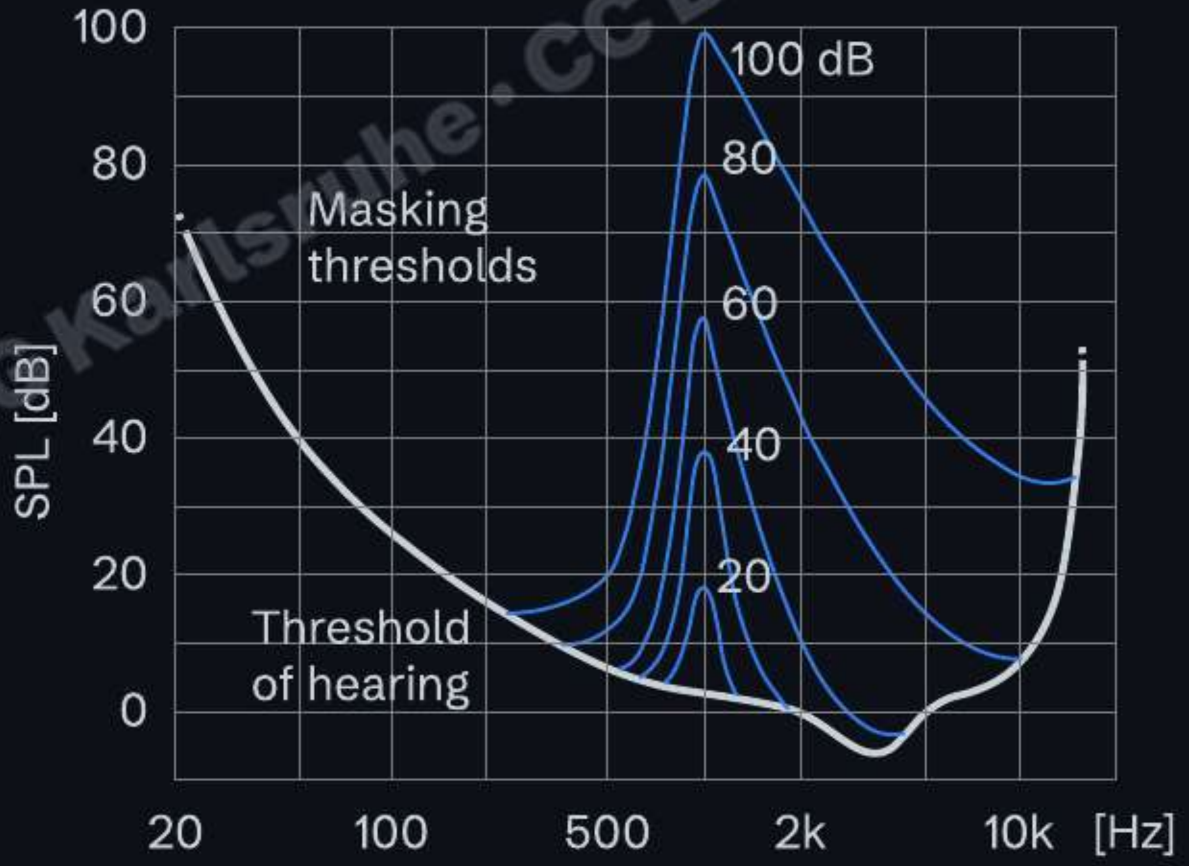
▶ Broadband noise masking sine tone



Spectral masking (or frequency masking)

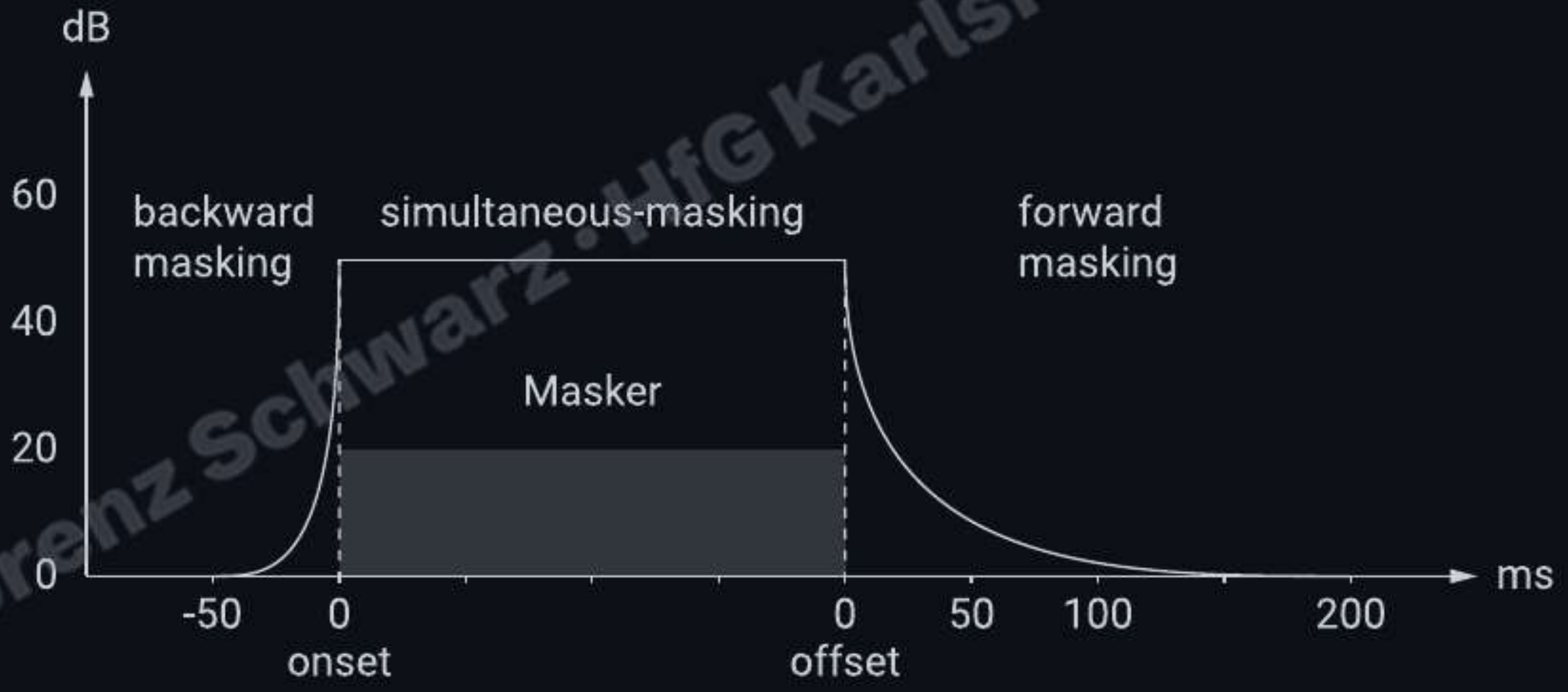
A strong sound in one frequency band masks weaker sounds in nearby frequencies. (1kHz example).

1 kHz masker increases hearing thresholds



Temporal masking

A loud sound masks a softer one that occurs before (pre-masking) or after (post-masking) it within a short time window.



Masking & Critical Bands

Sounds close in frequency share the same critical band, making one mask the other.

In mixing:

- Bass and kick can mask each other.
- EQ or panning helps separate them.

Critical Bands

Critical bands are ranges where tones influence each other during perception.

This interaction explains effects such as:

- **Auditory masking** – one sound hides another
- **Roughness** – harshness when tones are too close in frequency
- **Pitch shift & loudness change** – perception of pitch and intensity influenced by neighboring frequencies.

→ *All effects result from limited frequency resolution of the auditory system (critical bands).*

Audio Compression and Masking

Perceptual audio codecs (MP3, AAC, Opus) use psychoacoustic masking to reduce file size without perceptible quality loss:

- Analyzes the audio spectrum
- Identifies masking thresholds (which frequencies mask others)
- Removes or reduces data below the masking threshold
- Encodes only the perceptible information

→ *Psychoacoustic models allow a reduction from ~50 MB to ~5 MB without audible loss.*

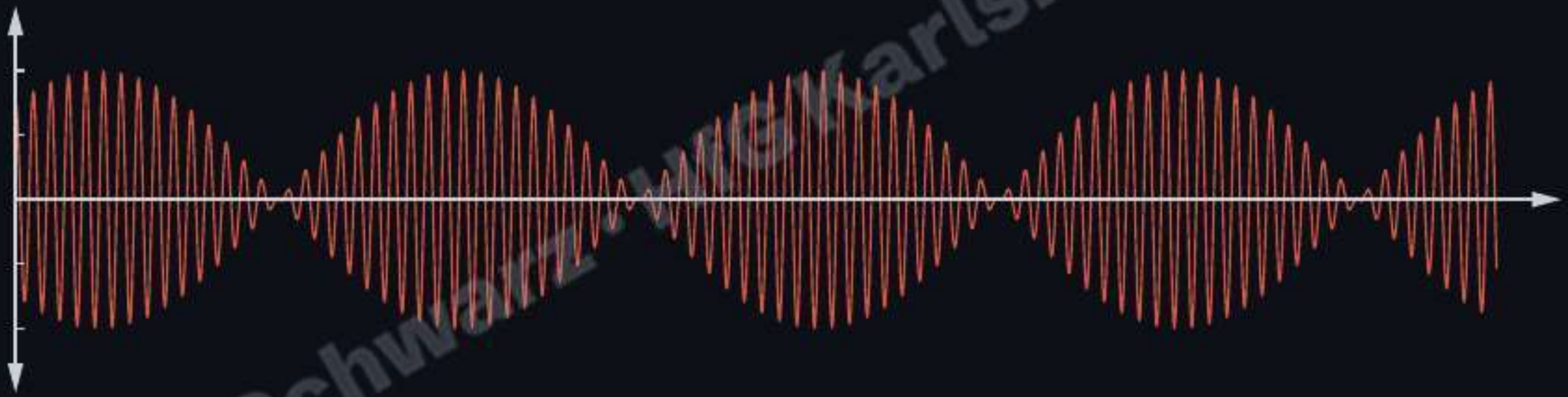
Temporal perception

- **Resolution:** The human auditory system can detect very short changes in sound, such as interruptions as brief as 2–3 ms.
- **Integration:** Sound energy can be integrated over 200–300 ms to improve sound detection.

INTERFERENCE PHENOMENA

Beat

Interference of sound waves with slightly different frequencies.



Beat

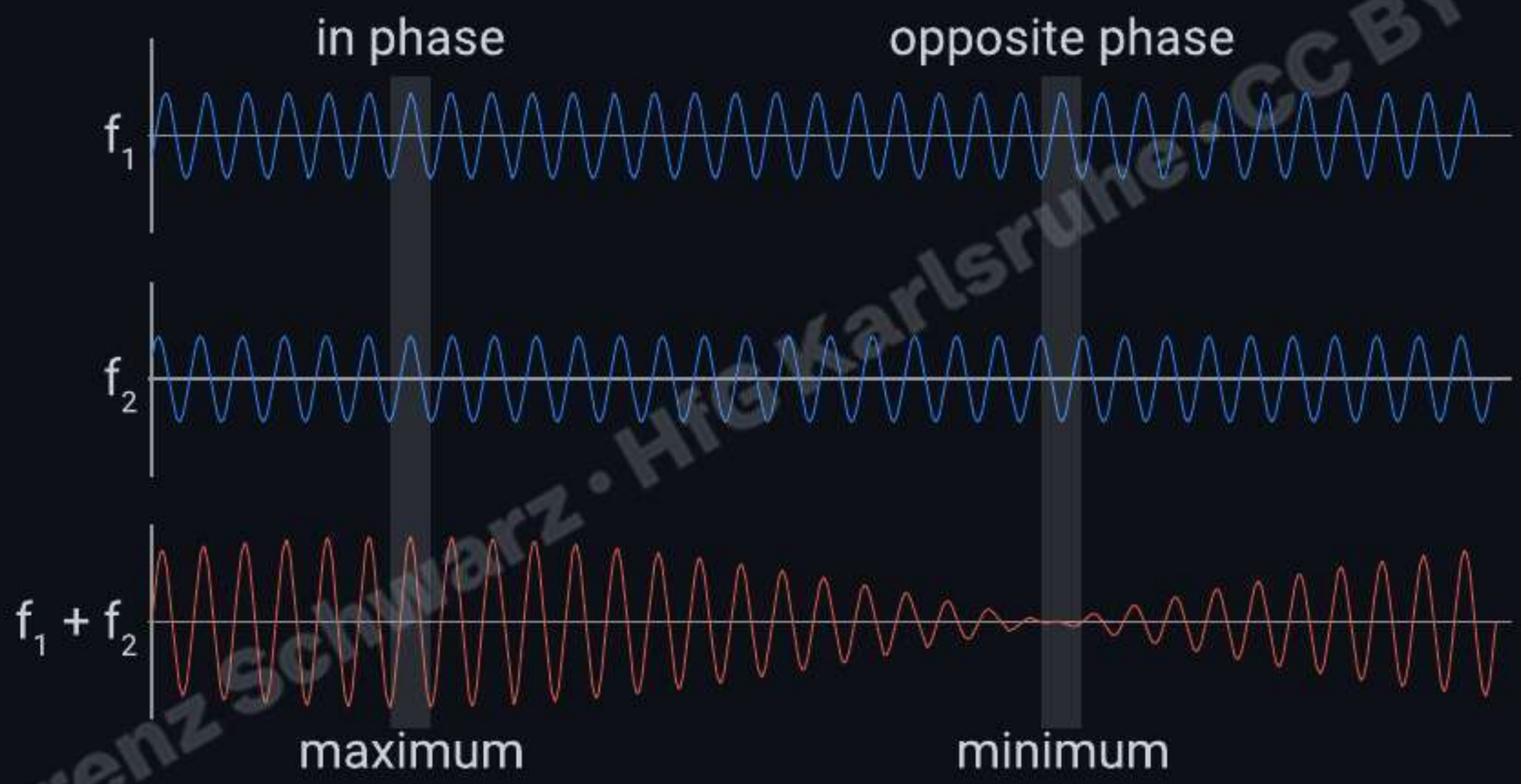
Two sound waves with slightly different frequencies interfere, causing a cyclical change in volume known as beating. This beat frequency equals the absolute difference between the two frequencies.

$$f_{beat} = |f_1 - f_2|$$

Used for tuning instruments.

- ▶ 400 Hz sine tone
- ▶ 404 Hz sine tone
- ▶ Beating

Interference of sound waves with slightly different frequencies.



[view in graphing calculator]

Binaural beats

Auditory illusion from two slightly different frequencies presented dichotically (one to each ear).

- Perceived beat is generated in the brainstem (not physically present.)
- Requires low-frequency carriers (<1500 Hz).

▶ Binaural beats 200 Hz and 202 Hz (listen with headphones)

Roughness

Roughness occurs when two tones of similar amplitude fall within the same critical bandwidth.

→ *Limited frequency resolution of the basilar membrane.*

▶ Two tones

▶ Beating

▶ Roughness

Combination tone

When two tones sound simultaneously, an additional tone may be perceived, which corresponds to the sum or difference of the fundamental frequencies of the two tones:

- Difference tone: $f_2 - f_1$ (most audible)
- Sum tone: $f_2 + f_1$ (less audible, usually masked)

→ *Similar to electronic ring modulation*

▶ Combination tone 1500 Hz and 1000 Hz

TIMBRE · FORMANTS

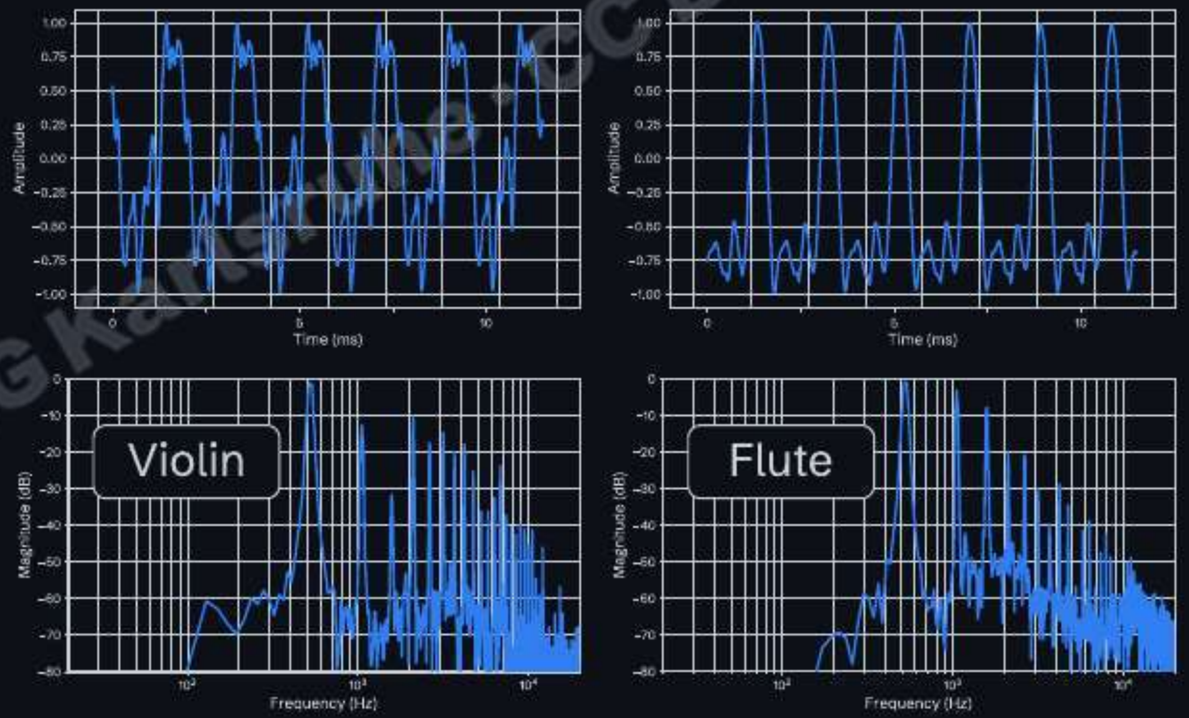
Timbre

The sonic characteristics of a sound, determined by the combination of its fundamental tone, overtones, noise, and the amplitude signature of its frequency components.

→ *This explains why the same musical pitch (note) sounds different when played on different instruments.*

▶ Violin C5 (left)

▶ Flute C5 (right)



Formants

A formant is a local maximum of energy around a specific frequency in the sound spectrum, caused by resonances.

Formants contribute to the characteristic timbre of instruments and, in phonetics, are critical for distinguishing vowels and voiced sounds.

→ *Formant regions are largely independent of the pitch of the fundamental*

Examples: Vowels /ä/, /i/, /o/ have distinct formant patterns regardless of pitch

▶ Vowels /ä/, /i/, /o/

AUDITORY SCENE ANALYSIS

Auditory scene analysis (ASA)

ASA is how the auditory system groups sounds into distinct, meaningful streams, making sense of complex environments.

Key concepts:

- Segmentation
- Integration
- Segregation

Cocktail party effect

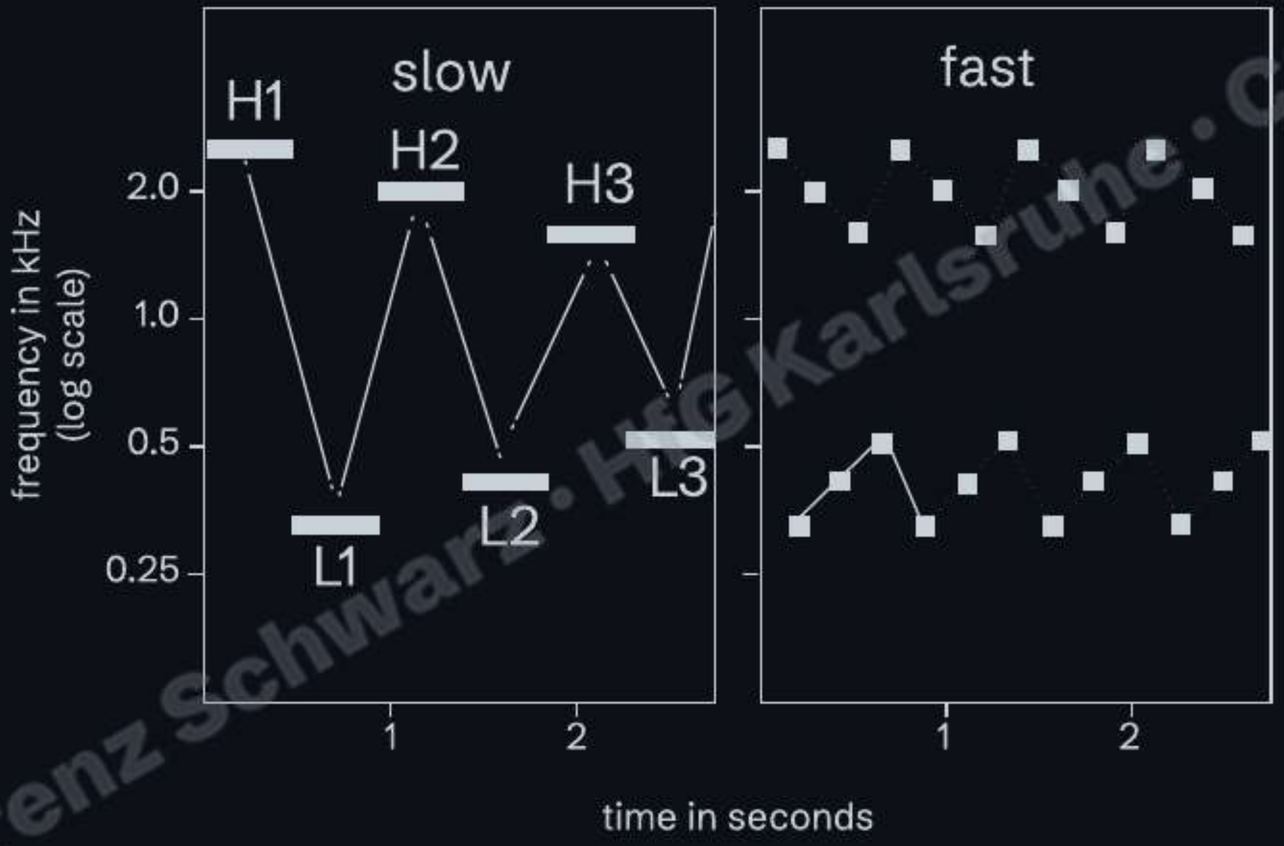
The ability to focus on a single conversation in a noisy environment while filtering out competing sounds.

- Binaural effect (requires both ears)
- Based on spatial separation and interaural differences (ITD, ILD)
- Neural cross-correlation extracts the target source

▶ Cocktail party, mono

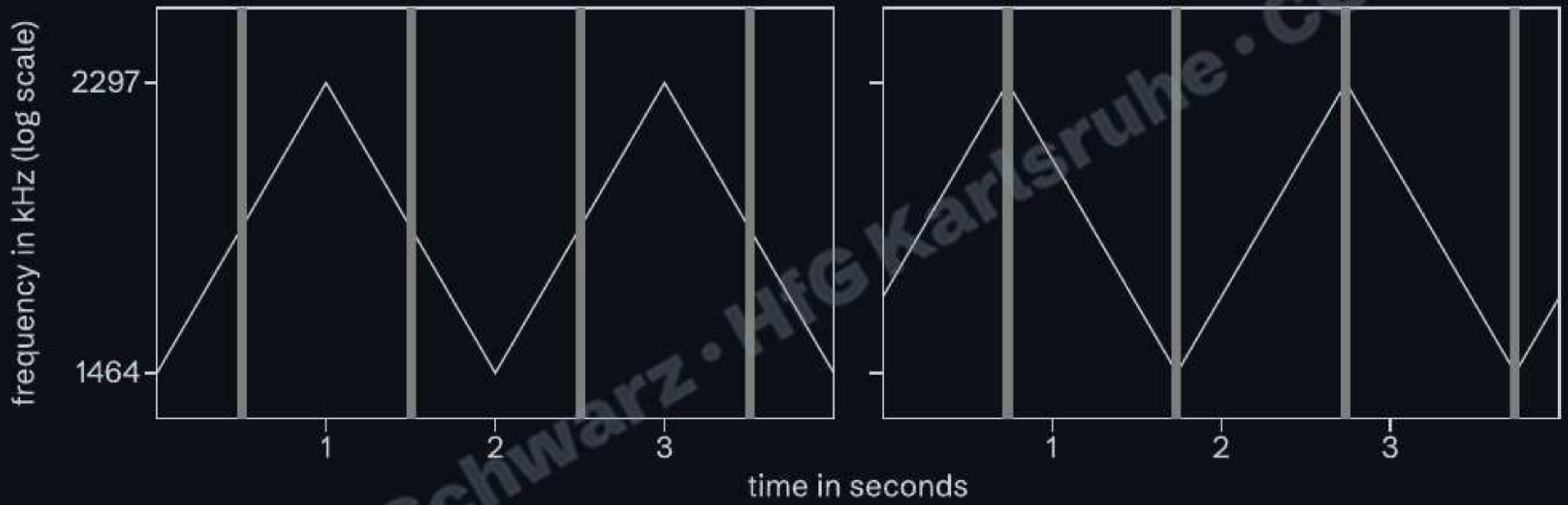
▶ Cocktail party with localization cues

Stream segregation in a cycle of six tones



▶ Stream segregation in a cycle of six tones

Perceptual continuation of a gliding tone through a noise burst



- ▶ Perceptual continuation of a gliding tone through a noise burst

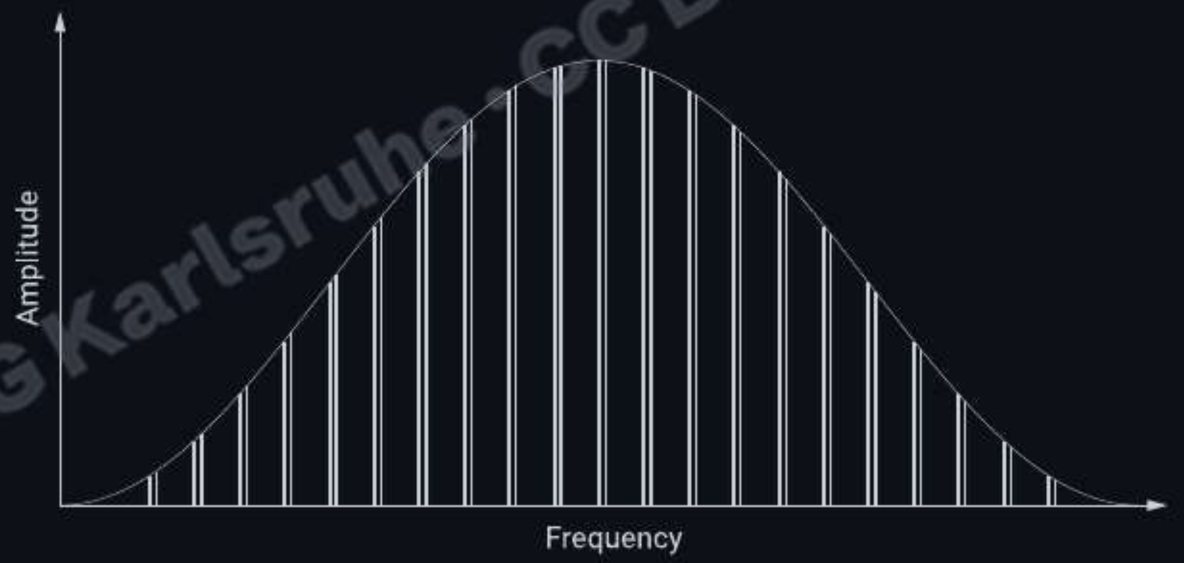
ILLUSIONS

Shepard-Risset glissando

The Shepard-Risset glissando is an auditory illusion where a continuously ascending or descending pitch appears to rise or fall endlessly, even though the physical stimulus is cyclically structured

▶ Risset glissando

▶ Excerpt Georg Friedrich Haas: In Vain



Summary

Psychoacoustics connects physics with perception:

- Loudness is nonlinear (equal-loudness contours, phon, sone)
- Pitch perception combines place and temporal coding
- Masking (spectral and temporal) enables audio compression
- Critical bands explain roughness, masking, and pitch shifts
- Auditory scene analysis separates concurrent sound sources
- Timbre distinguishes instruments despite identical pitch

SOUND LOCALIZATION

Spatial Hearing in Audio Practice

Spatial hearing informs how we capture, shape, and evaluate sound scenes:

- Microphone technique and stereo array selection (localization cues, stereo width)
- Panning, depth, and spatial processing (level/time differences, reverberation cues)
- Immersive and binaural production workflows
- Assessment of monitoring and room acoustics (imaging, translation)

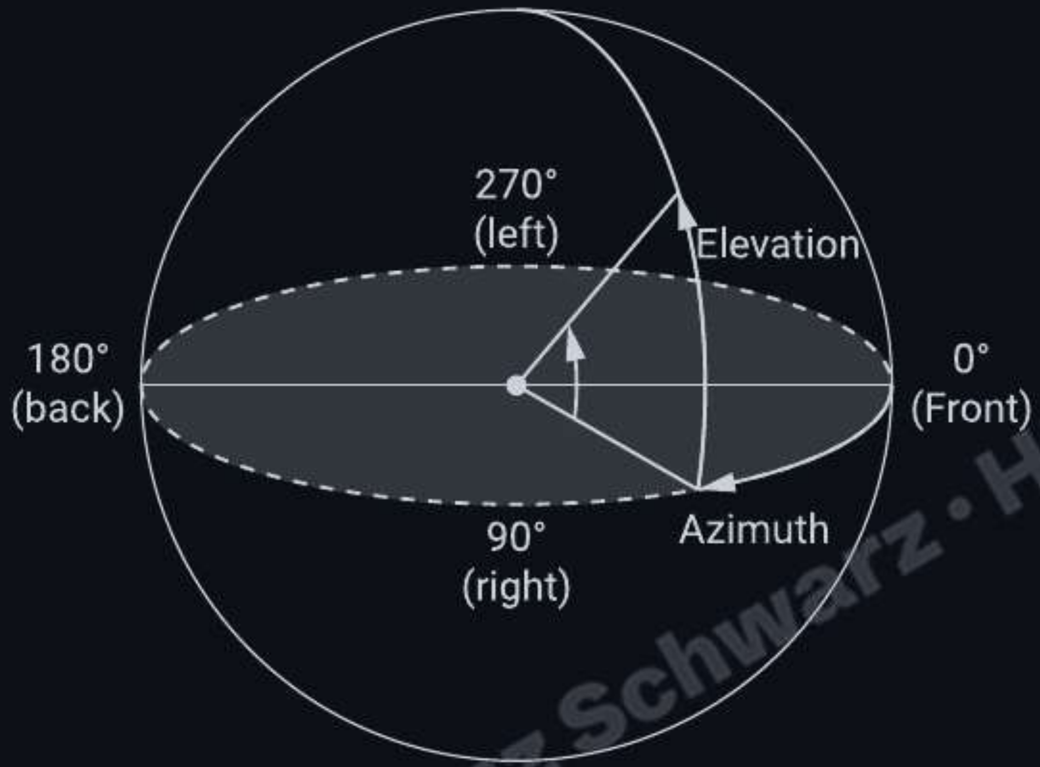
BINAURAL HEARING

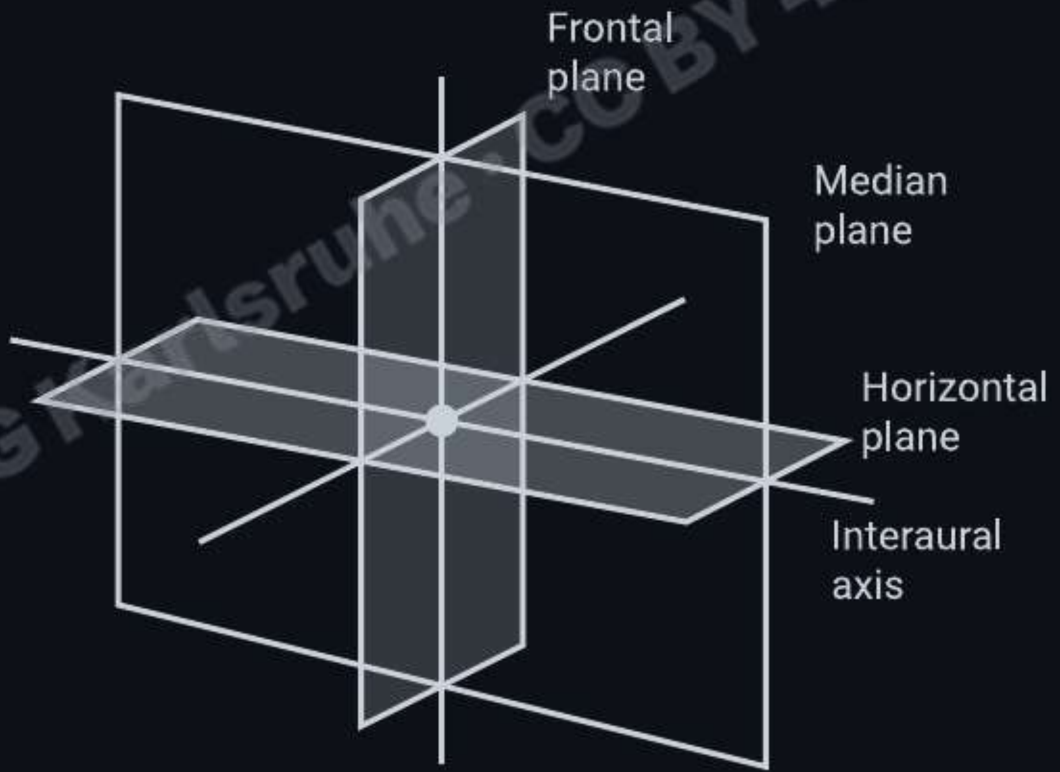
Spatial references in binaural hearing

- Azimuth (horizontal angle)
- Elevation (vertical angle)
- Distance

Anatomical planes:

- **Frontal plane:** Divides the body into front and back halves.
- **Horizontal plane (transverse plane):** Divides the body into upper and lower halves.
- **Median Plane:** Divides the body into left and right halves.





Horizontal coordinate system (left) and anatomical planes (right)

Lateralization and localization

- Lateralization: Perceived left-right position of a sound presented over headphones, typically experienced inside the head.

▶ Lateralization (listen via headphones)

- Localization: Perception of a sound source at a specific position in external space, as with loudspeakers or real sound sources.

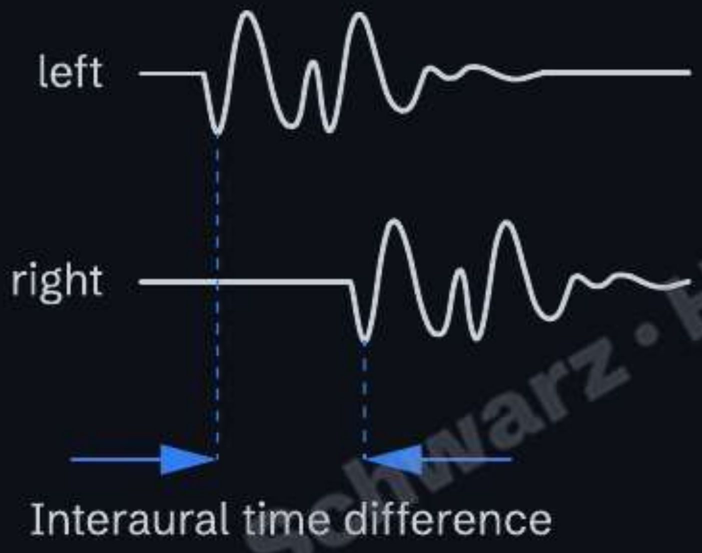
▶ Localization (listen via headphones)

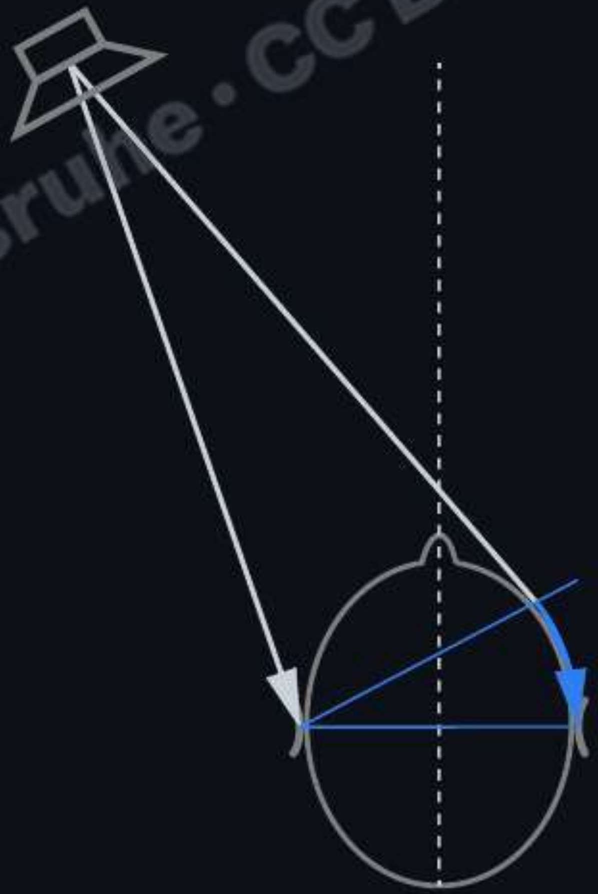
Binaural hearing and source localization

The first theory of lateral sound localization was proposed by Lord Rayleigh as "duplex theory." Later extended with Blauert's spectral cues.

- **Interaural Time Difference (ITD):**
  - Difference in arrival time between ears. Effective for low frequencies (<1500 Hz) because long wavelengths maintain phase coherence between ears.
- **Interaural Level Difference (ILD):**
  - Difference in sound intensity between ears. Effective for high frequencies (>1500 Hz) due to head shadow effect (short wavelengths are blocked by the head).
- **Spectral Cues:**
  - Additional information from how the head and pinnae filter sound (HRTF).

Interaural time difference (ITD)





Interaural time difference (ITD)

ITD helps localize low-frequency sounds.

For an ear distance of $\lambda \approx 22$ cm, the natural interaural time delay is approximately $\approx 660\mu s$

$$f = \frac{c}{\lambda} = \frac{343\frac{m}{s}}{0.22m} \approx 1500Hz$$

- Phase delays at low frequencies (if wavelength is greater than half the distance between the ears $< 800Hz$)

▶ ITD noise bursts (listen via headphones)

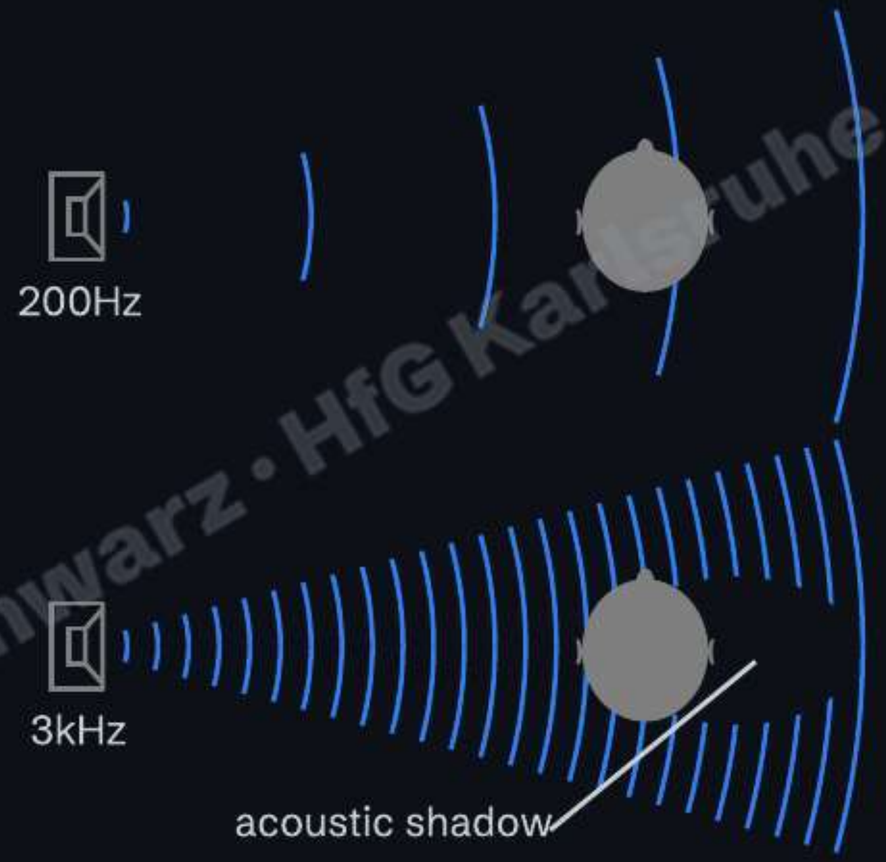
Interaural level difference (ILD)

Interaural Level Difference (ILD), also known as Interaural Intensity Difference (IID), plays a key role in the localization of high-frequency sounds.

- Low frequencies bend around the head with minimal attenuation
- High frequencies are significantly attenuated due to the head shadow effect
- Level differences increase above ~1600 Hz

▶ ILD noise bursts (listen via headphones)

Interaural level difference (ILD)



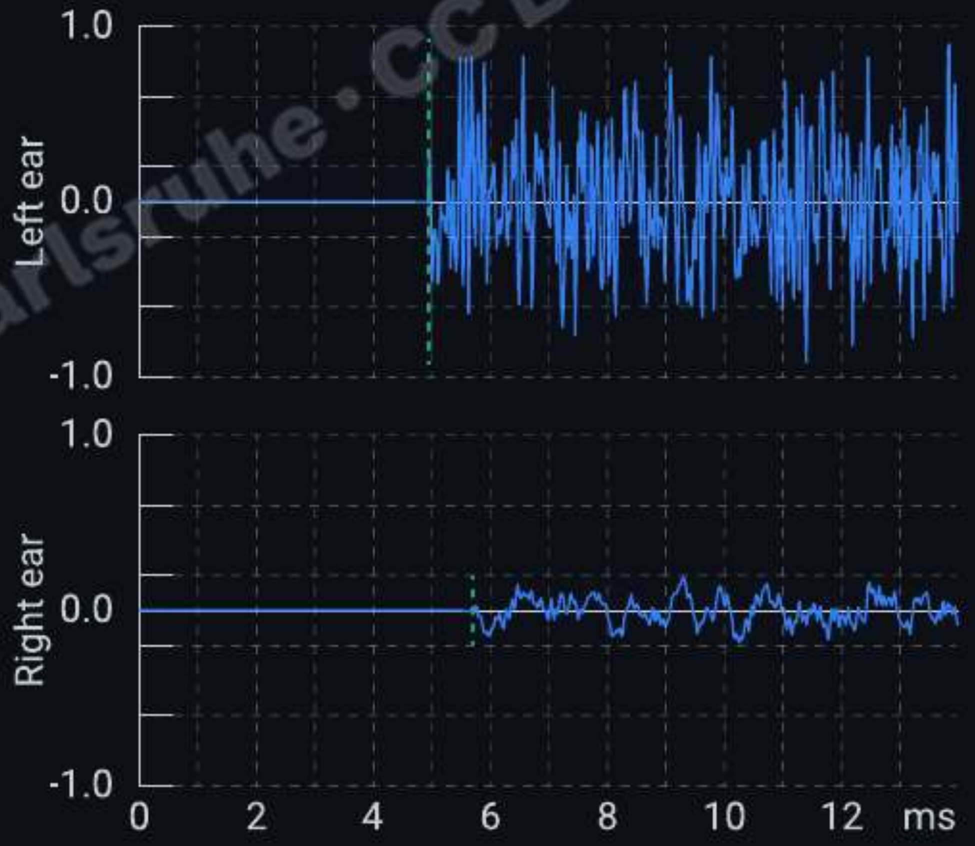
ITD and ILD

Localization accuracy:

- 1° for sources in front
- 15° for sources to the sides

Frequency ranges:

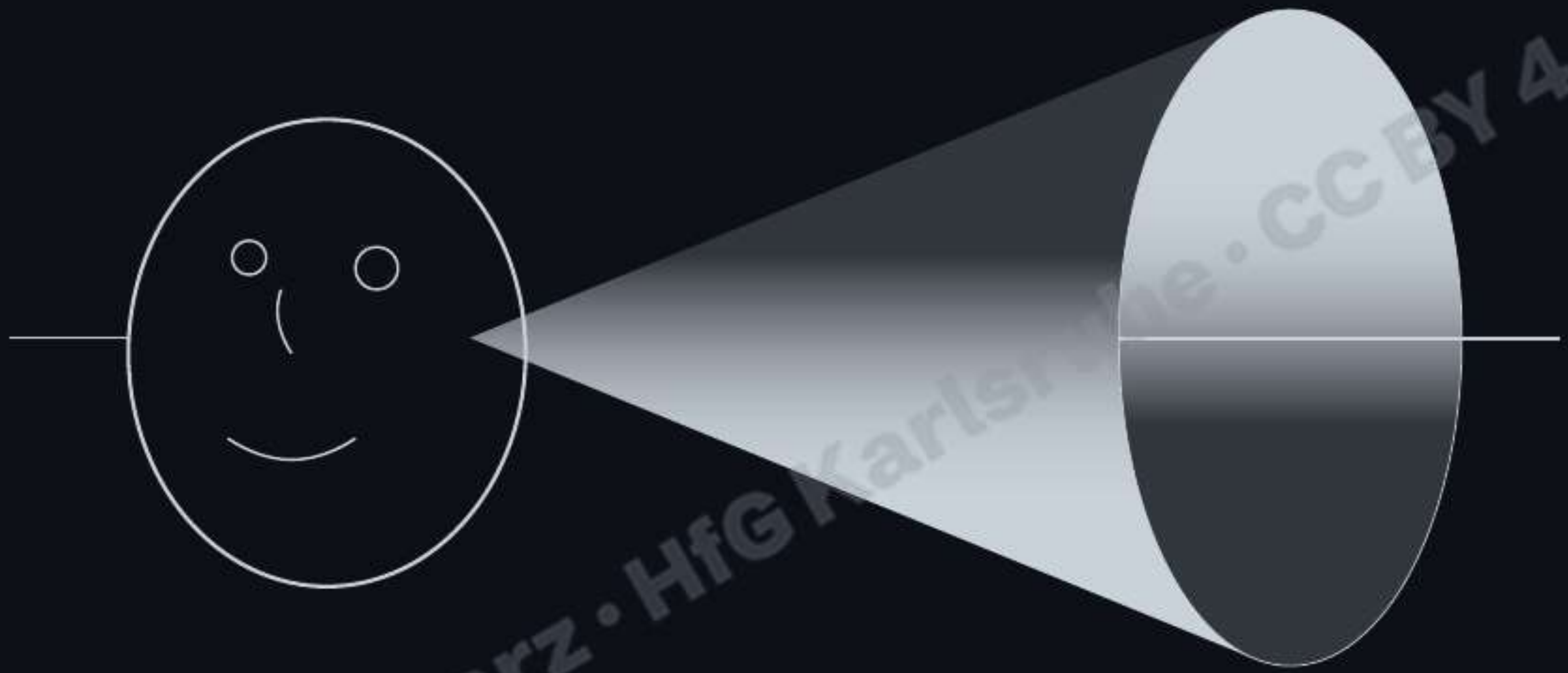
- ITD dominant for $f < 800$ Hz
- Both contribute between 1000-1500 Hz
- ILD dominant for $f > 1600$ Hz
- No localization below 80 Hz



White noise at -90° azimuth: maximum ITD of 0.63 ms

Cone of confusion

- **Front-back ambiguity:** Cannot discriminate between sounds originating from the front or rear.
- **No elevation cues:** ITD and ILD do not provide information about vertical localization (elevation).



A region in space where sound sources produce identical ITDs and ILDs, making localization ambiguous.

Resolving ambiguity

- **Tilting the head:** Introduces new timing and level information to help resolve the location of the sound source.
- **Spectral cues:** The filtering effects of the pinnae and torso shape the sound spectrum, helping to distinguish elevation and front-back positioning.

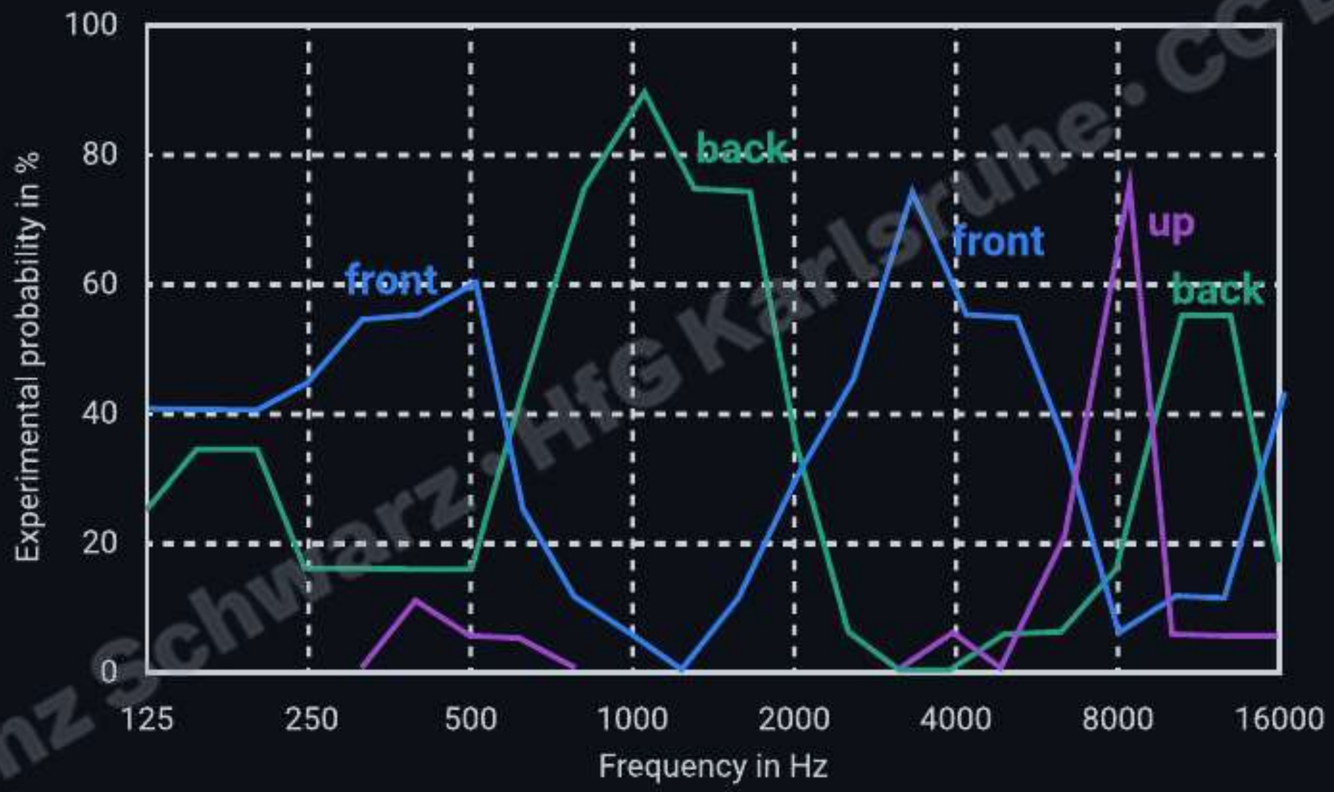
SPECTRAL CUES

Spectral cues

The outer ear, auricle (pinna), and ear canal act as a resonator system, shaping sound based on its direction of incidence.

- Frequency-dependent filtering provides spatial information, aiding in sound localization.
- Encodes vertical and front-back localization cues (median plane), primarily through frequency bands affected by the pinna.

Directional bands (after J. Blauert)



▶ Pink noise back and front ear level and above (listen via headphones)

Head related transfer function (HRTF)

HRTF describes how spatial audio cues are encoded in the sound reaching the ears, allowing for sound source localization.

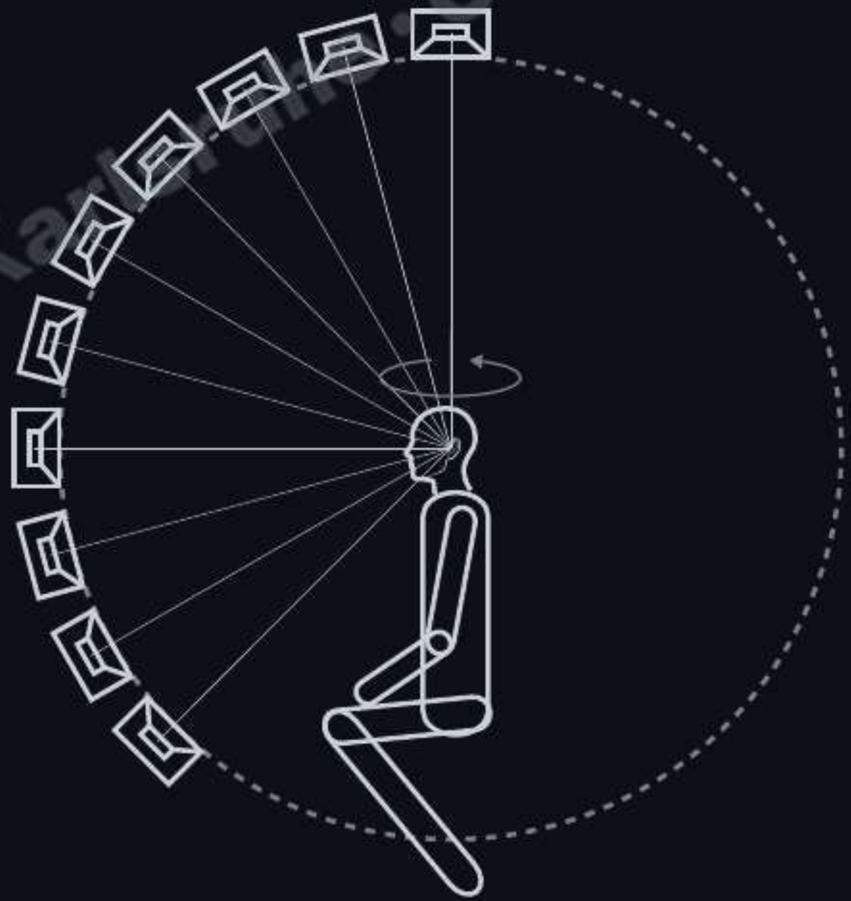
The torso, head, and pinna act as direction-dependent filters, introducing frequency-specific alterations to the sound.

→ *This effect can be mathematically represented as a transfer function.*

▶ Clicks rotating 360° up - down (listen via headphones)

HRTF measurement

HRTFs are measured at small angular increments in an anechoic chamber, with interpolation used to estimate unmeasured positions.



Applications of HRTF

- 3D audio rendering
- VR & AR
- Binaural recording

→ *The auditory system adapts to a modified head-related transfer function over time.*

Franssen effect

The Franssen effect is a psychoacoustic phenomenon in which the perceived location of a tone remains tied to its initial onset position, even if the actual sound source subsequently moves or changes.

The auditory system relies heavily on initial transients or onsets to localize the source of a sound.

▶ Play Franssen effect

Franssen effect

Auditory distance perception

The auditory system has limited ability to determine distance, relying on:

- **Initial Time Delay Gap (ITDG):** Time between direct sound and first reflection
- **Direct-to-reverberant ratio:** Closer sources have more direct sound
- **Reverberation density:** More diffuse reflections indicate greater distance
- **High-frequency absorption:** Distant sounds lose high-frequency content
- **Loudness:** Closer sources perceived as louder
- **Motion parallax:** Closer sources shift position faster for moving listeners

▶ Distance close - far

STEREOPHONIC REPRODUCTION

Stereophonic Sound

When two channels are played through separate speakers, listeners perceive a soundstage extending between those speakers.

- **Summing localization:** Brain combines signals from both speakers to create phantom sound sources
- **Phantom center:** Virtual sound source perceived between the speakers

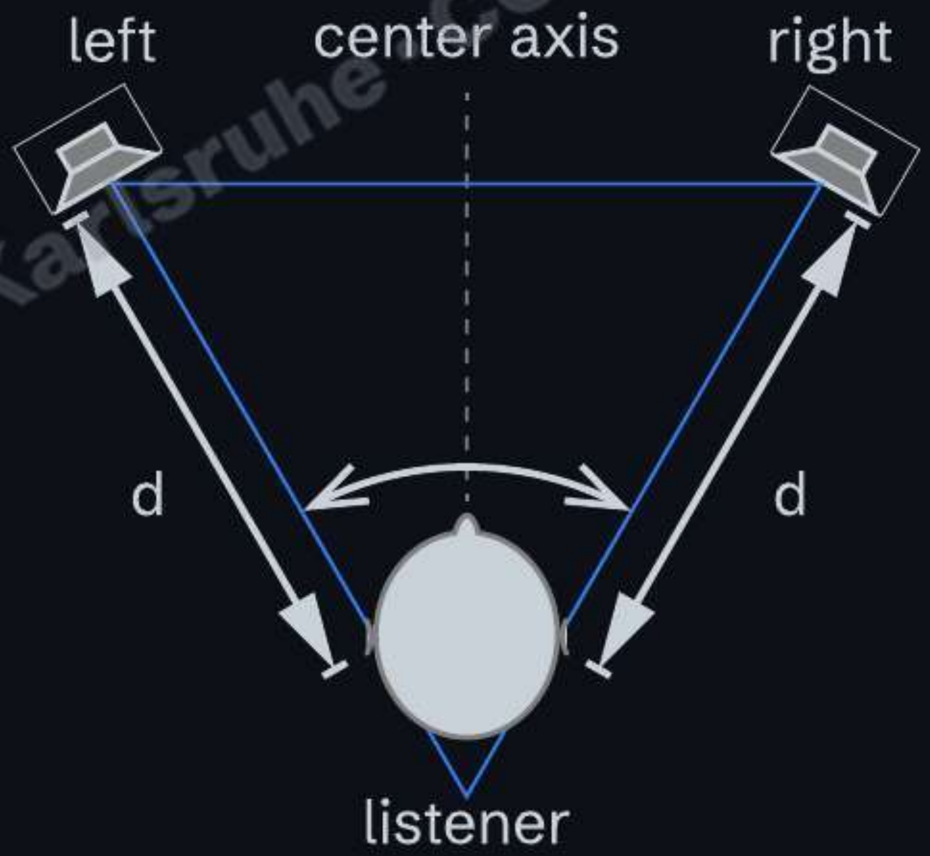
Creates an illusion of multi-directional spatial perspective.

Artificial stereo and phantom center

The recommended placement forms an equilateral triangle: each speaker and the listener at equal distances:

- Sweet spot or reference listening position

▶ Panning from left to right



Dual-mono signal

Mono source material played back through two stereo channels with identical signals on both the left and right channels.

- Identical waveform in both L and R channels
- Creates phantom center image when played through speakers
- Perceived as center image in headphones (lateralization)
- Common in broadcast, mono recordings, and centered mix elements

Summing localization

Two speakers create a phantom sound source between them by manipulating the same binaural cues the brain uses for natural localization:

- **Amplitude panning** (ILD)
- **Time delay** (ITD)

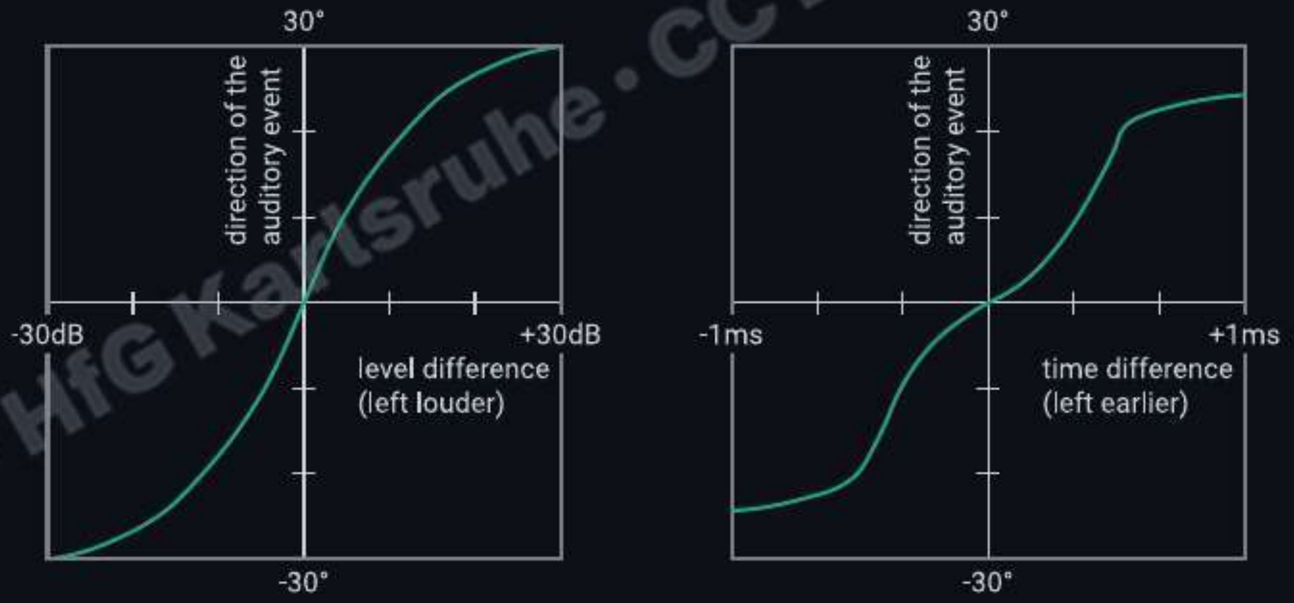


Diagram: perceived source location (Wendt, 1963)

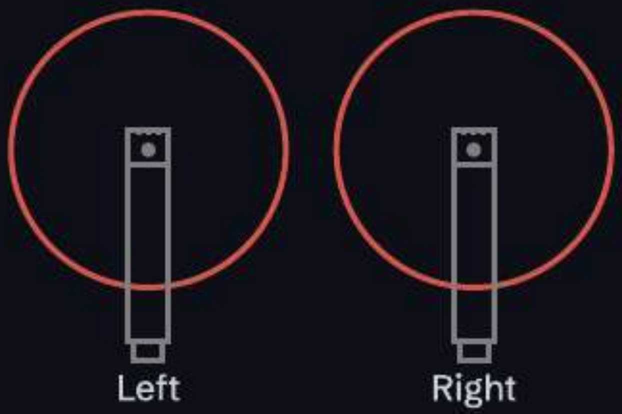
Stereo recording

ITD and ILD are used in common stereo recording practices:

Stereo recording techniques:

- Spaced pair (A-B): Wide stereo image
- Coincident pair (X-Y): Accurate localization

XY: level differences



A-B: time-of-arrival stereo

Summary: sound localization

- ITD (time differences) for low frequencies
- ILD (level differences) for high frequencies

Ambiguity (Cone of confusion) resolved through head movement and spectral cues (HRTF) for elevation and front-back.

Applications:

- Stereo reproduction relies on phantom imaging
- HRTF measurements enable binaural 3D audio

ROOM ACOUSTICS

Room acoustics

Room acoustics shapes every sound we hear and create. Understanding how spaces interact with sound enables:

- Microphone placement decisions
- Informed choice of performance venues
- Creative use of acoustic environments

Historical development of acoustics

Antiquity:

- Pythagoras & Hippasus: Mathematical basis of sound (monochord)
- Archytas: Physical vibrations as source of sound
- Vitruvius: Acoustic design in amphitheaters (echea/resonators)

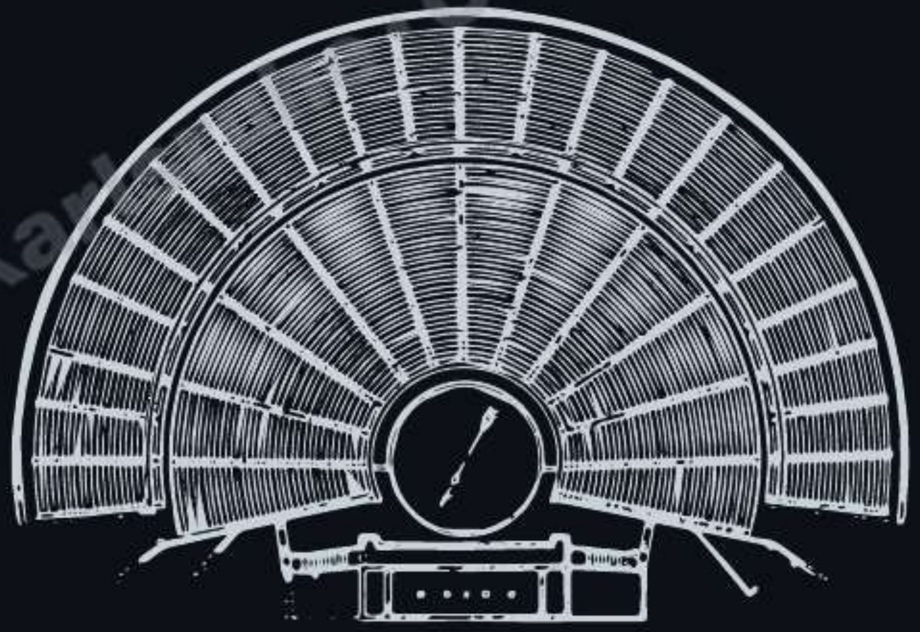
Modern era:

- Wallace Clement Sabine (1868–1919): Quantitative architectural acoustics
- Developed reverberation time measurement (RT_{60})
- First scientifically designed concert hall (Boston Symphony Hall)

Amphitheater

Amphitheaters exhibit remarkable acoustics, achieved through their architectural design, and are characterized by high speech intelligibility across the audience area:

- Amplification of relevant frequency bands
- Balanced reverberation time
- Uniform sound distribution
- Low background noise levels



Layout of the ancient theatre of Epidaurus

FIELDS · PROPAGATION · BOUNDARIES

Acoustic Fields

Understanding sound propagation requires considering:

- The **acoustic environment** (open or enclosed space)
- The **listening position** relative to the sound source

These factors determine how sound energy behaves and is perceived.

Free field vs. diffuse field

Free field (open space):
Sound propagates without reflections. Only direct sound is present; intensity decreases by inverse-square law.

Diffuse field (closed space):
Sound reflects off boundaries, creating a mix of direct and reflected sound.

▶ Free field (direct sound)

▶ Diffuse field (reflections)

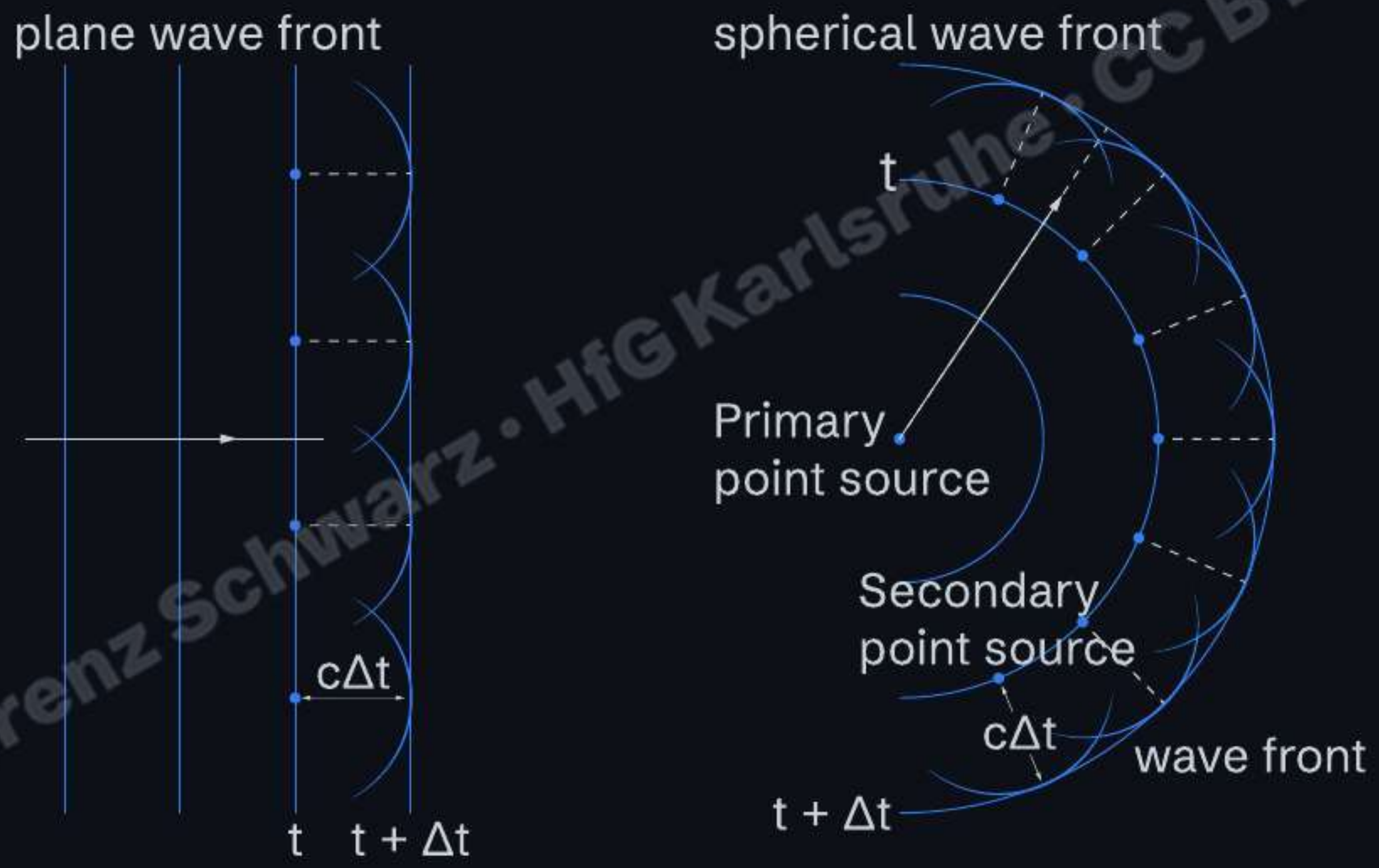
Wave propagation and Huygens' principle

Huygens' principle states that every point on a wavefront acts as a source of secondary spherical wavelets. The superposition of these wavelets forms the new wavefront as the wave propagates.

This principle explains:

- **Diffraction:** bending and spreading of waves around obstacles or through openings
- **Refraction:** change in wave direction when passing between media

Wavefront construction from secondary wavelets

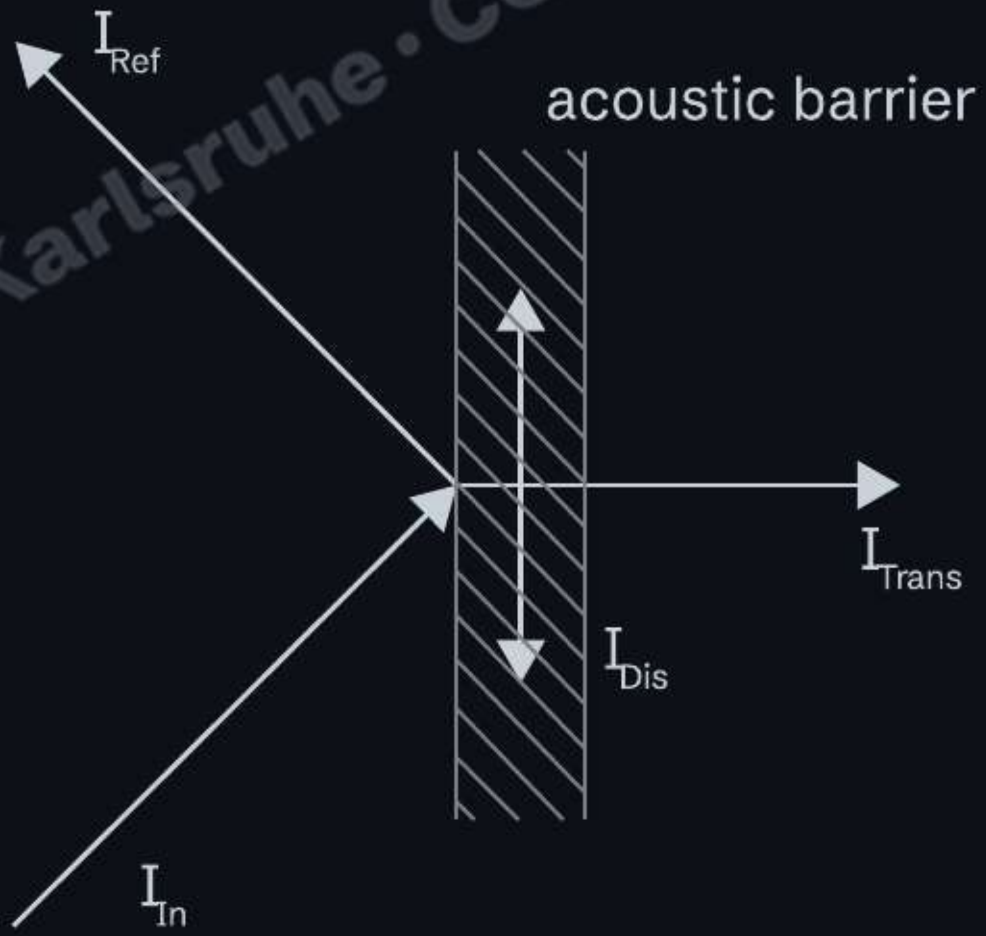


REFLECTION · DIFFUSION · ABSORPTION

Wave behaviour at acoustic barriers

Possible paths of sound waves when encountering an acoustic barrier:

- I_{In} Incoming sound wave.
- I_{Ref} Reflected sound wave.
- I_{Dis} Absorbed sound wave, dissipated as heat within the barrier.
- I_{Trans} Transmitted sound wave passing through the barrier.



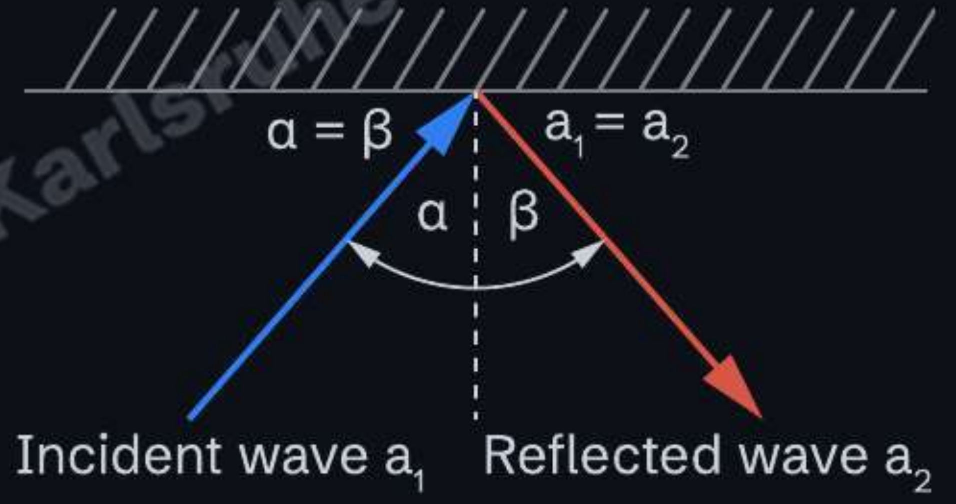
Acoustic barriers and boundaries

Surface shapes and materials have a profound effect on the behavior of sound waves:

- Reflection
- Diffusion
- Absorption
- Diffraction
- Refraction

Reflection

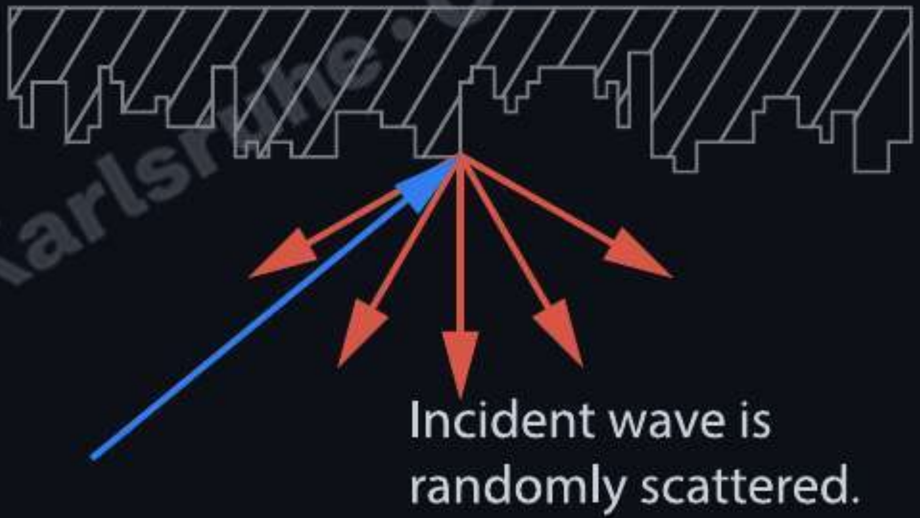
Sound waves interact with surfaces through reflection, where the angle of incidence equals the angle of reflection.



Diffusion

Irregular surfaces diffuse sound waves by scattering them in multiple directions.

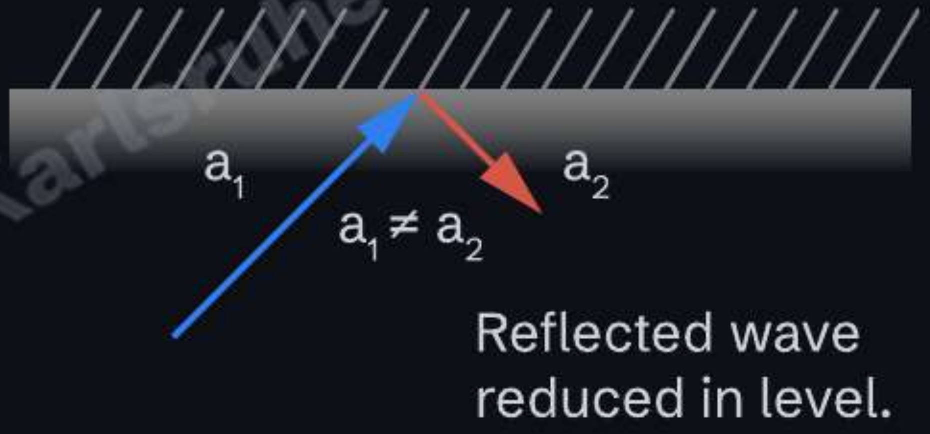
- Reduces echoes
- Enhances acoustic quality
- Creates a more balanced sound environment



Absorption

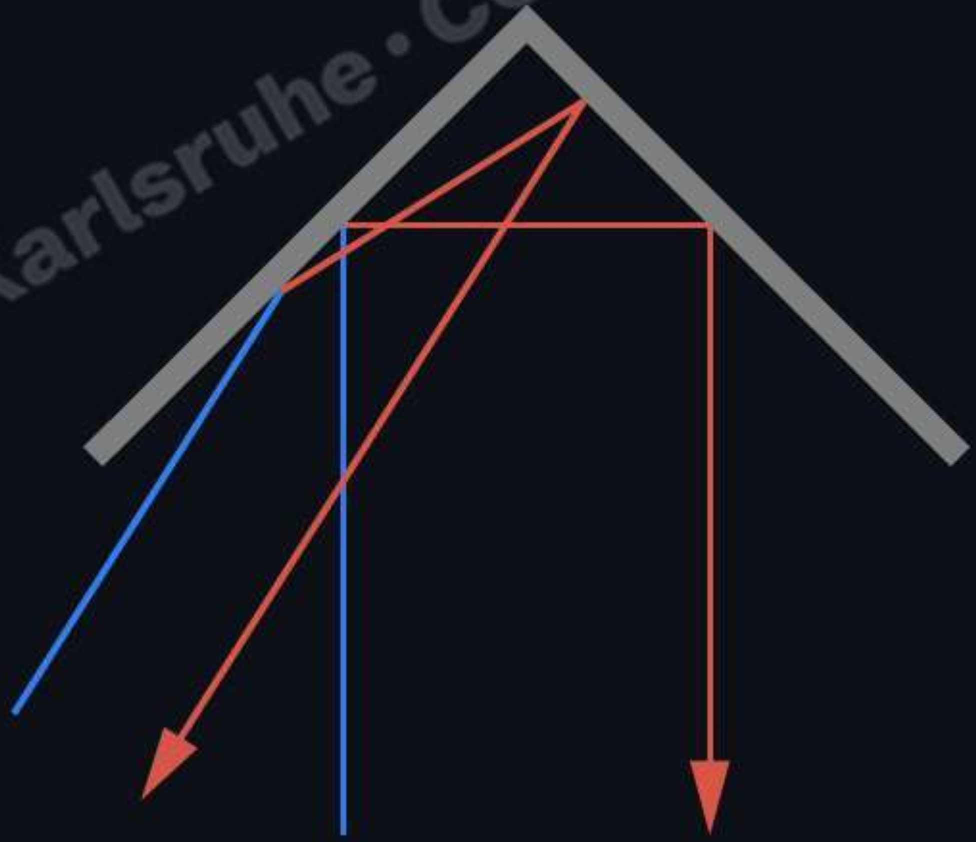
Porous or soft materials absorb sound energy, converting it into heat.

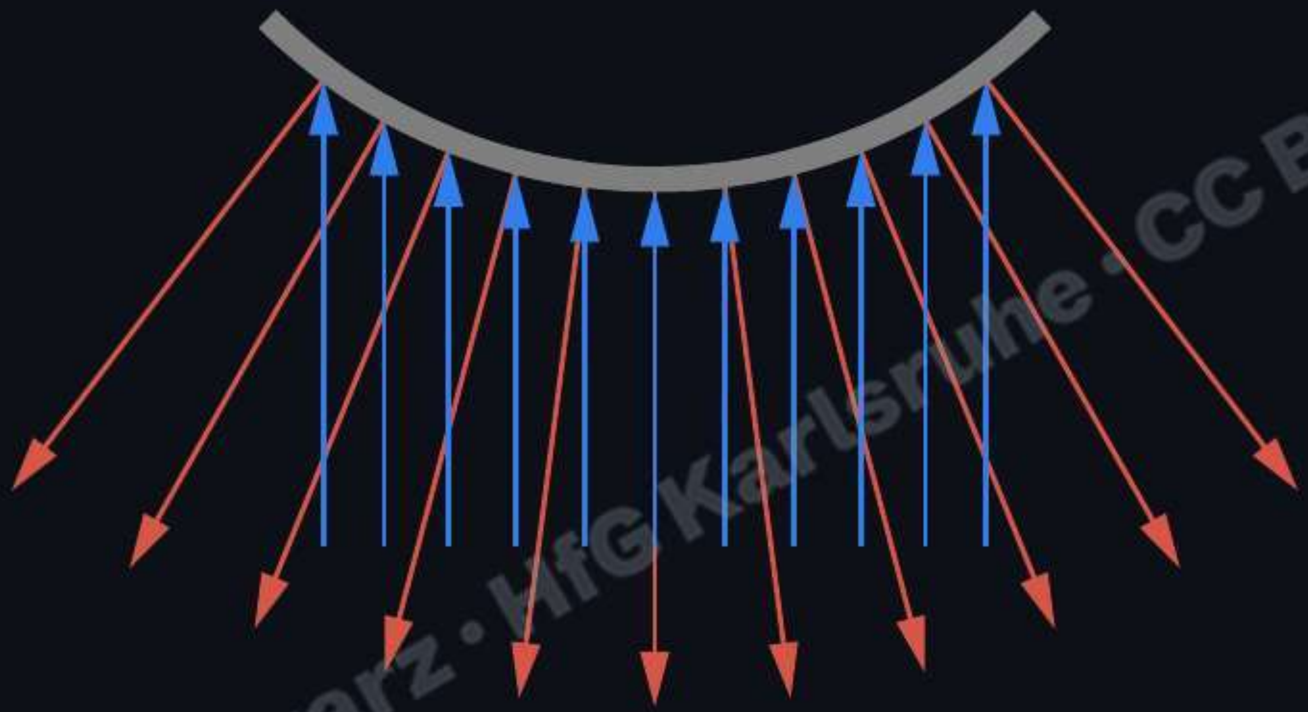
- Reduces reverberation and noise levels



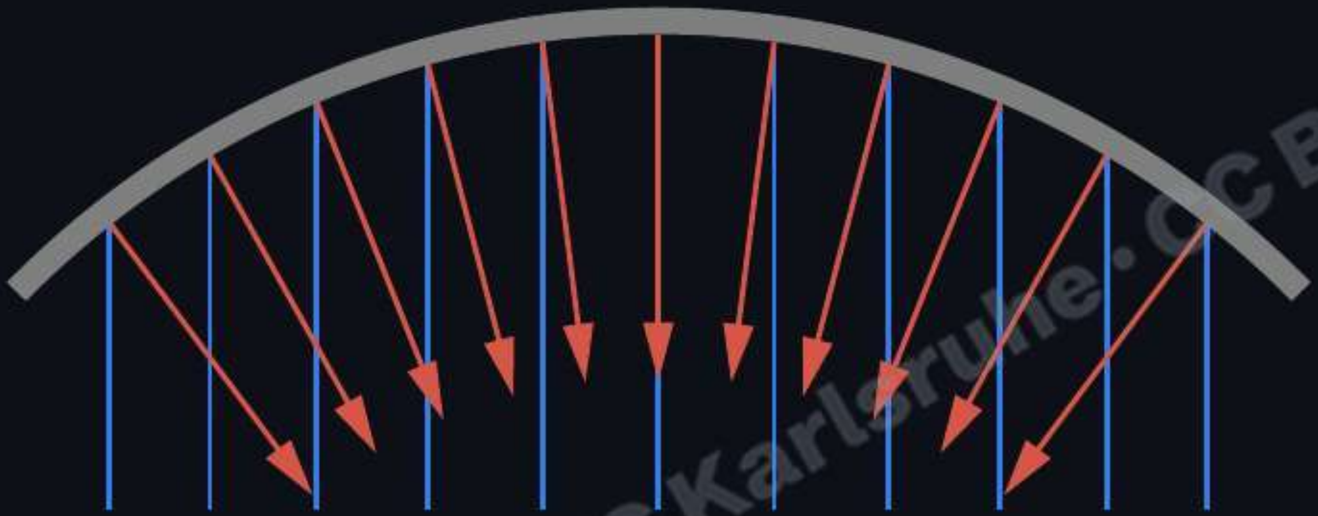
Room corners

- Corners return sound to its source.
- Directivity control at low frequencies.
- Problematic reflections at high frequencies.





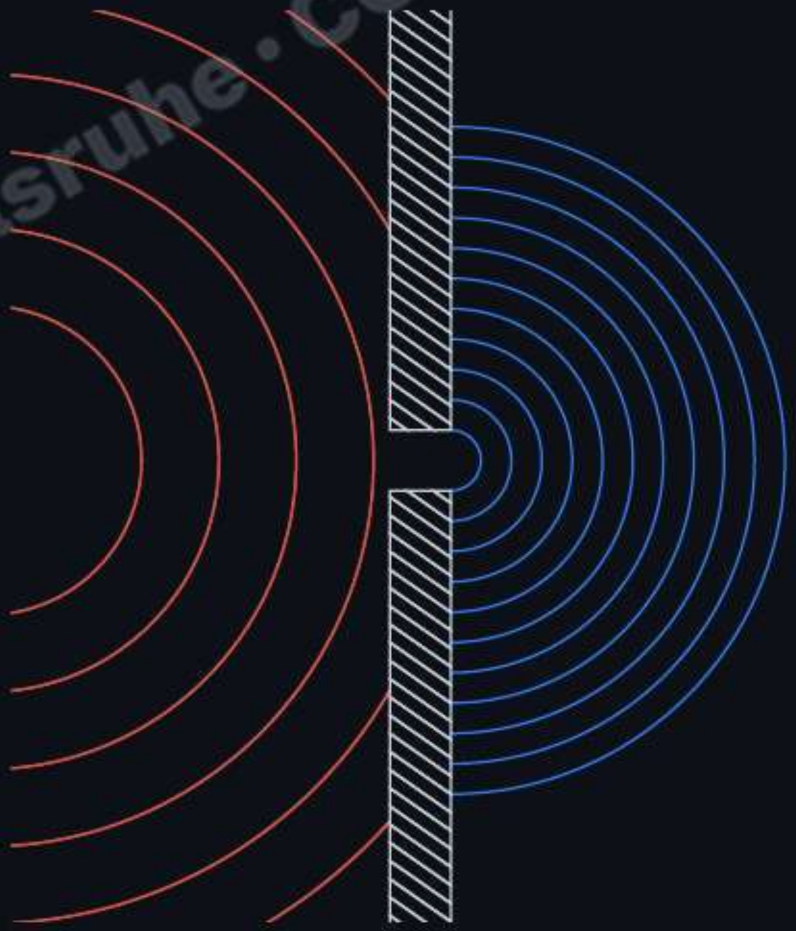
Convex surfaces scatter sound.



Concave surfaces focus sound.

Diffraction

Sound waves bend around obstacles and spread through small openings, allowing them to travel beyond direct lines of sight.

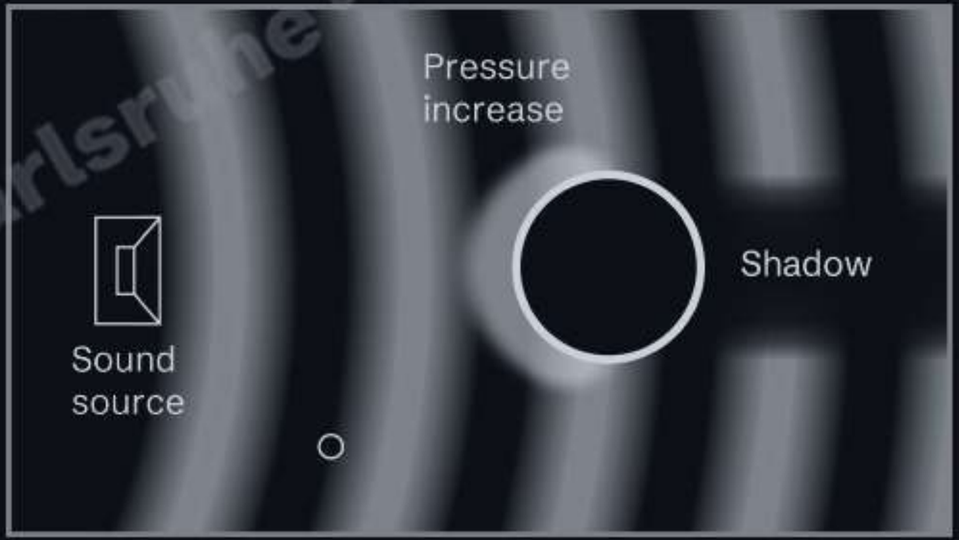


Acoustic shadow

Acoustic shadow and diffraction depend on wavelength and obstacle size.

$$d = 5 \cdot \lambda = 5 \cdot \frac{c}{f}$$

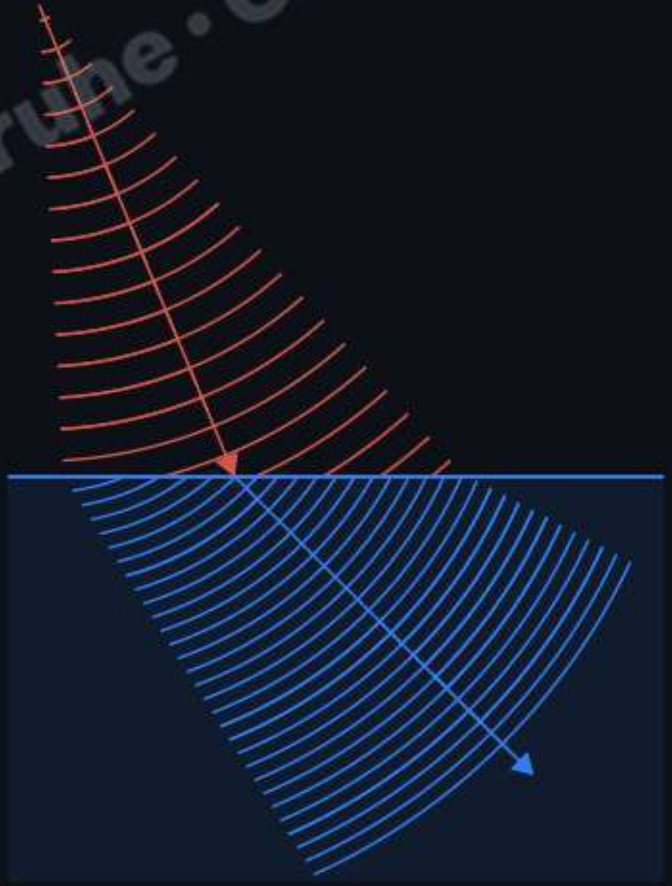
where d is the obstacle size, λ is wavelength, c is speed of sound, and f is frequency.



Acoustic shadow behind an obstacle

Refraction

Changes in the medium (e.g., temperature, air density) or variations in surface curvature cause sound waves to bend, altering their direction and intensity.



REVERBERATION · MODES · RESONANCE

Sound in enclosed spaces

When sound propagates within an enclosed space, it interacts repeatedly with multiple surfaces.
These interactions give rise to characteristic acoustic effects that shape how a room sounds.

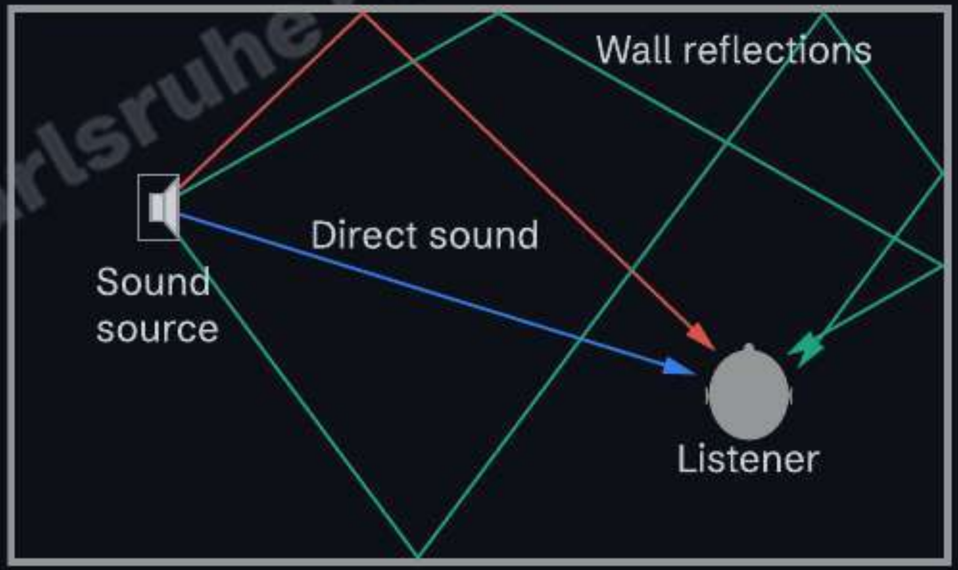
Typical phenomena include:

- **Reverberation** – persistence of sound due to repeated reflections
- **Standing waves** – resonances caused by boundary conditions

Reverberation

Reverberation is the persistence of sound in a space, caused by multiple reflections of sound waves off surfaces. It is the sum of sound reflections in an enclosed space that arrive after the direct sound.

▶ IRs of various spaces



Reverberation in an enclosed space

Reverberation

- No reverberation (e.g., in an anechoic chamber) sounds unnatural and artificial.
- Excessive reverberation reduces speech intelligibility (e.g., measured by Alcons).
- Can make certain sound structures muddy and unclear.
- Can mask imprecise playing by a musician.
- Enhances and amplifies sound, making it richer and more resonant.

Temporal structure of reverberation

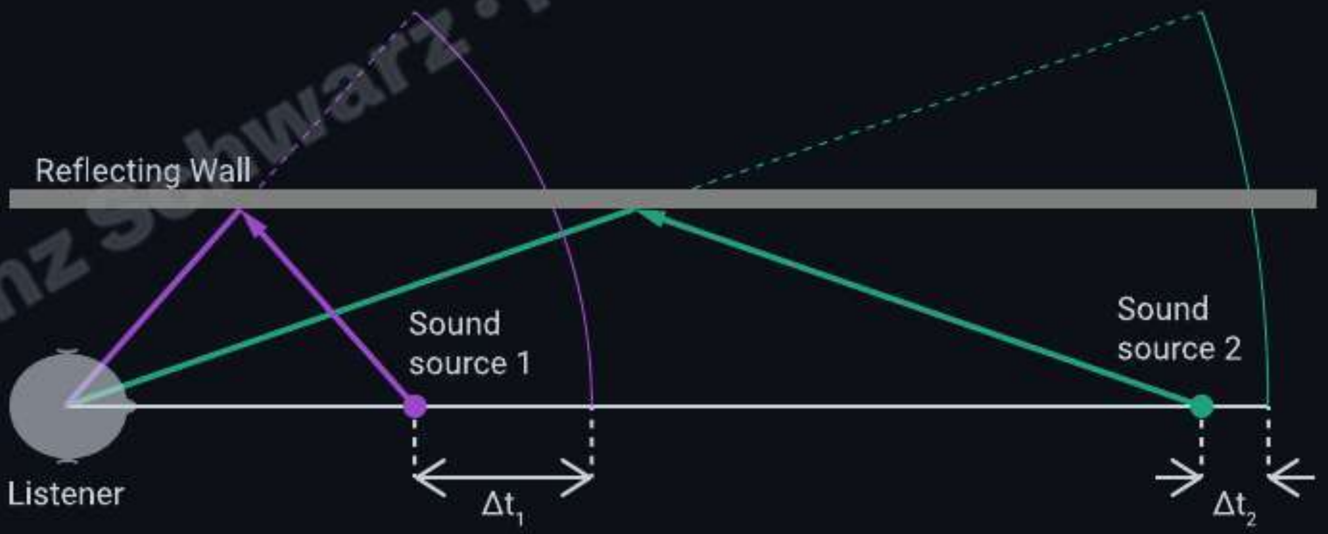
- **Direct sound:** The initial sound wave that travels straight from the source to the listener without any reflections.
- **Predelay:** The time gap between the arrival of the direct sound and the first early reflection.
- **Early reflections:** First set of reflected sound waves that arrive shortly after the direct sound.
- **Reverberation tail:** Overlapping reflections that decay gradually over time.

Temporal structure of reverberation

Initial time delay gap (ITDG)

The Initial Time Delay Gap (ITDG) is the time interval between the direct sound and the first early reflection at the listener's position.

→ *Close sound sources result in a longer ITDG, while distant sound sources result in a shorter ITDG.*



Echo

An echo is a distinct, repeated sound reflection heard after the original direct sound.

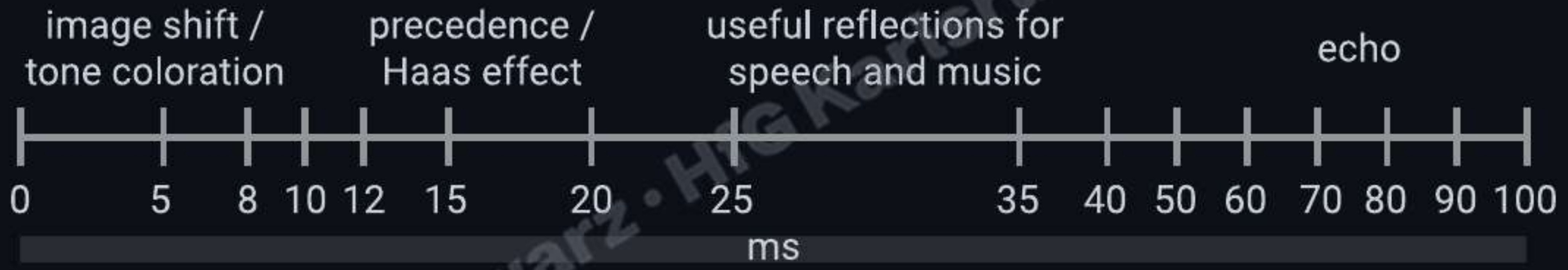
→ Whether a sound is perceived as an echo depends on the nature of the sound and the number of reflecting surfaces.

Precedence effect (Haas effect)

The precedence effect describes how spatial localization of a sound is dominated by the "first wavefront", even if subsequent copies of the sound (reflections) arrive within a short delay window (a few ms to ~30 ms).

- Law of the first wavefront

Audible effects of delayed signals of equal level

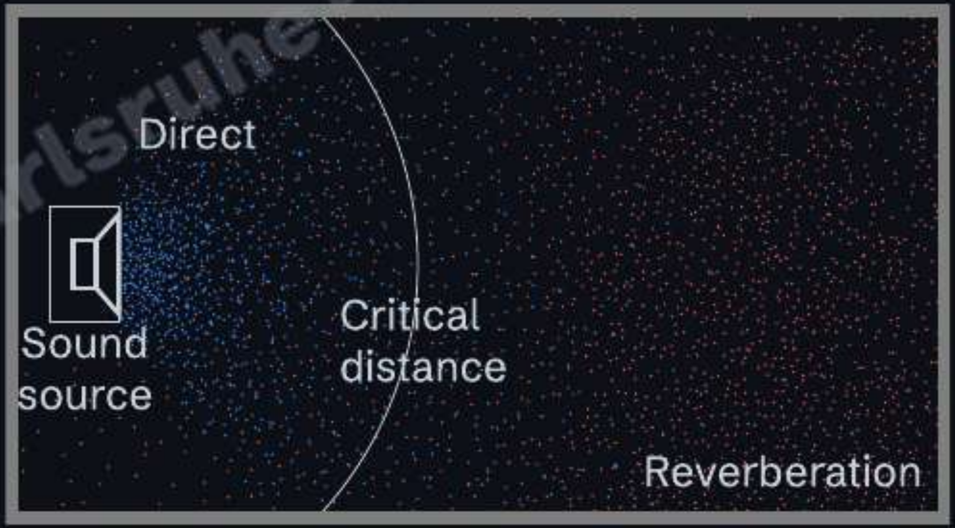


▶ Effects of different delays 0, 1, 20, 45, 100 ms

Critical distance

The critical distance is the point in space where the combined amplitude of all reflected sound (R) equals the amplitude of the direct sound (D) from the source (D = R).

→ Sound reflections = direct sound

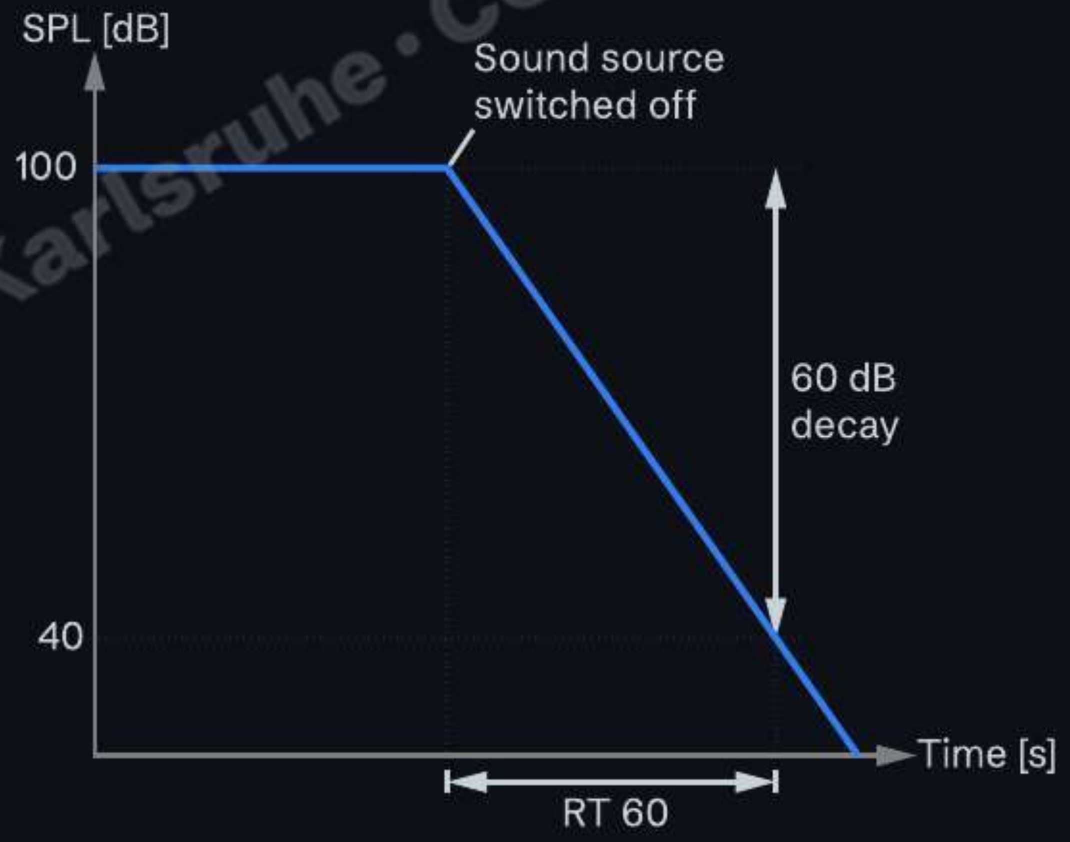


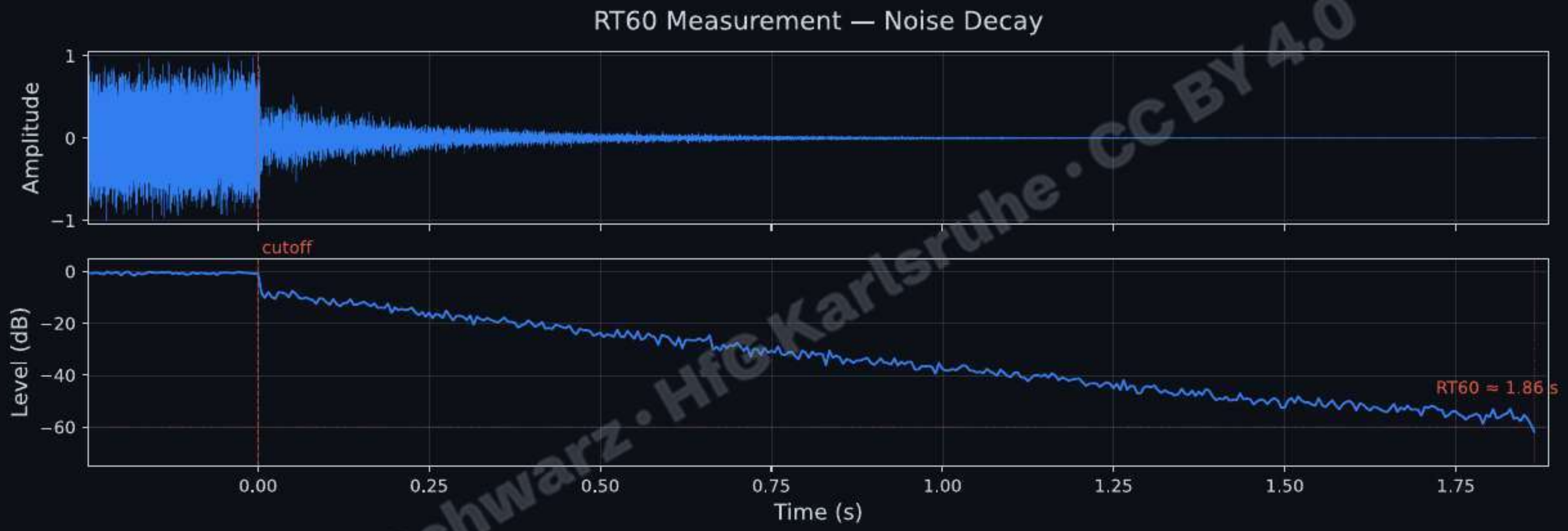
Critical distance: D = R

Reverberation time (RT_{60}) measurement

T_{60} measurement:

- The time required for the sound pressure level to decrease by 60 dB after a test signal is abruptly stopped.





▶ RT60 measurement of noise decay

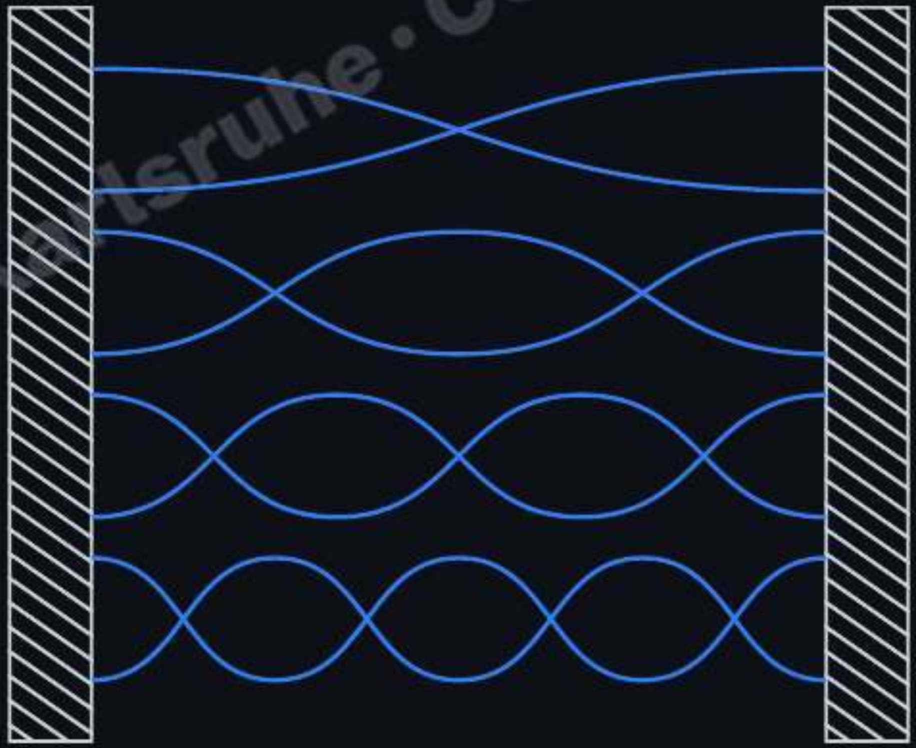
Room modes

Resonant frequencies in a room create standing waves with zones of high and low sound pressure, shaping the room's frequency response.

- **Node:** Point where sound waves cancel each other out (minimum pressure)
- **Antinode:** Point where sound waves reinforce each other (maximum pressure)

$$f_{res} = \frac{c}{2L}$$

L — longest distance in meters between boundary surfaces.



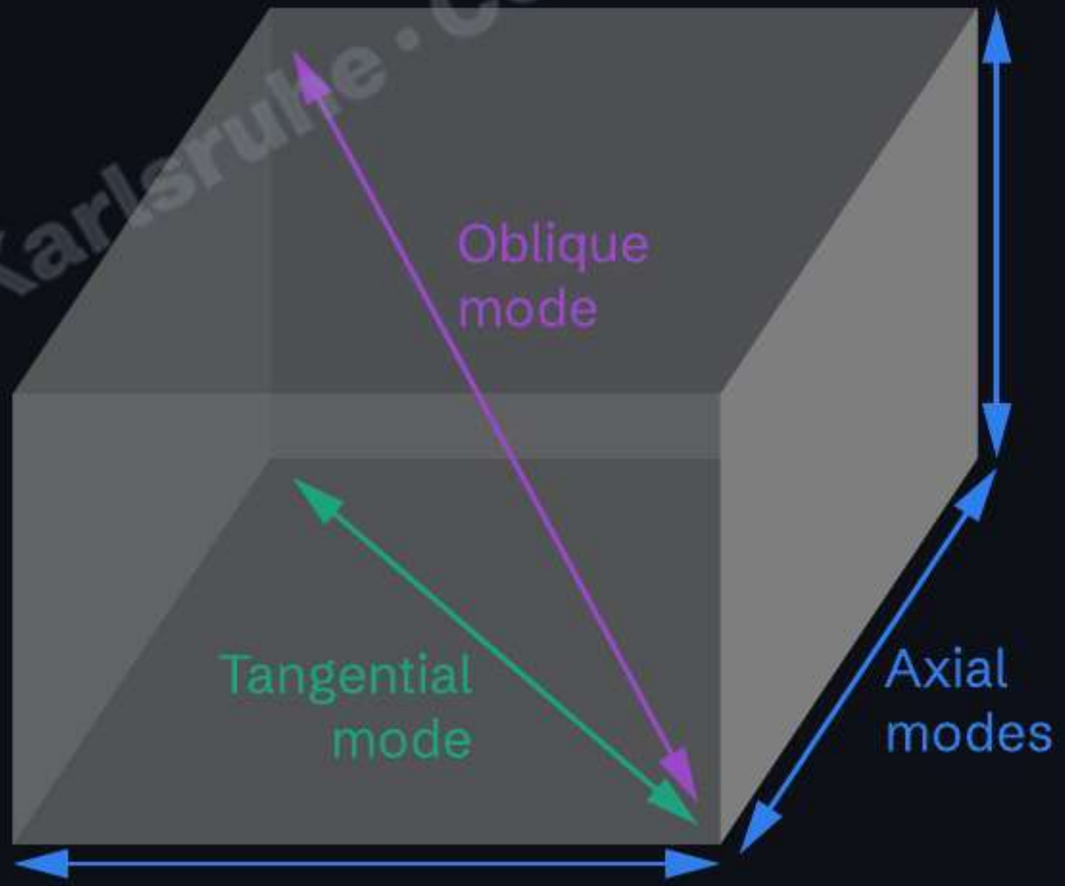
Room modes refer to standing waves that occur in an enclosed space.

Room modes

- **Axial:** along one dimension (strongest)
- **Tangential:** between two surfaces (≈3 dB weaker)
- **Oblique:** across three surfaces (≈6 dB weaker)

→ *Room modes can lead to uneven bass response, with certain frequencies being amplified or diminished at specific locations.*

▶ Sine sweep and room modes



Anechoic chamber

A controlled environment for acoustic measurements under free-field conditions.

- Insulated against structure-borne sound.
- Constructed using sound-absorbing materials.
- Provides an approximation of free-field conditions.

Reverberation time guidelines

A general guideline for roughly evaluating the quality of auditory conditions in a typical multi-purpose auditorium:

- **Below 1 second (lecture hall):** Good for speech, too dry for most music
- **1 to 1.5 seconds (concert hall):** Good for speech and chamber music
- **1.5 to 2 seconds:** Fair for speech, good for orchestral, choral, church music
- **Over 2 seconds (church):** Poor for speech, good for large organ, liturgical choir

Artificial reverberation

Artificial reverberation simulates the natural persistence of sound in a space, adding richness and spatial depth.

- **Physical approaches:**
  - **Echo chambers:** Utilizing physical spaces to record natural reverberation
  - **Convolution reverb:** Recreating real acoustics by applying impulse response recordings
  - **Acoustic raytracing / finite element method:** Computational modeling of sound propagation and reflections
- **Synthetic approaches:**
  - **Plate and spring reverb:** Vibrating metal plates or coiled springs to emulate reflection patterns
  - **Digital delay lines:** Creating reverb through digital signal processing algorithms

DIGITAL REPRESENTATION OF SOUND

Sound (physical domain)

Sound is a physical phenomenon consisting of pressure variations in a medium (e.g. air) over time.

- Exists as acoustic energy
- Described by sound pressure, particle velocity, and intensity
- Continuous in time and amplitude

Transduction of sound

A transducer (e.g. a microphone) converts sound pressure into a corresponding electrical voltage.

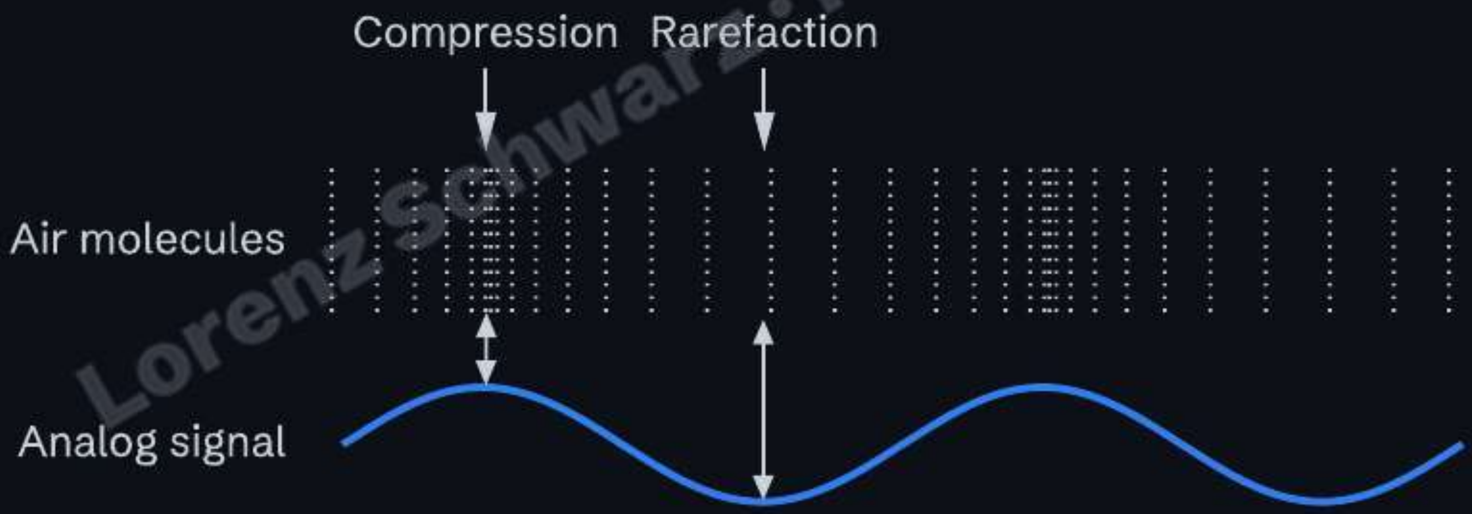
- Energy changes from acoustic to electrical

acoustic sound - electrical signal - acoustic sound

Analog signal

An analog signal is a continuous-time electrical signal whose voltage variations correspond directly to sound pressure variations.

- Continuous in time and amplitude
- Proportional to the original acoustic waveform



Digital representation of sound

Digital audio systems convert the analog audio signal into a stream of numbers.

Advantages of digital audio

- Cost-efficient: storage, duplication, and processing are inexpensive
- Compact and scalable: minimal cabling, no complex analog signal paths
- Low noise floor: no tape hiss or cumulative analog noise
- Deterministic processing: sample-accurate timing and repeatability
- Look-ahead processing: enabled by buffering and latency
- Visual feedback: waveforms, meters, and spectral displays
- Non-destructive editing: undo/redo, versioning, and recall
- Automation: precise recording and playback of parameter changes

Digital representation of sound

In digital audio, the acoustic wave is converted into numerical values representing its amplitude at discrete points in time.

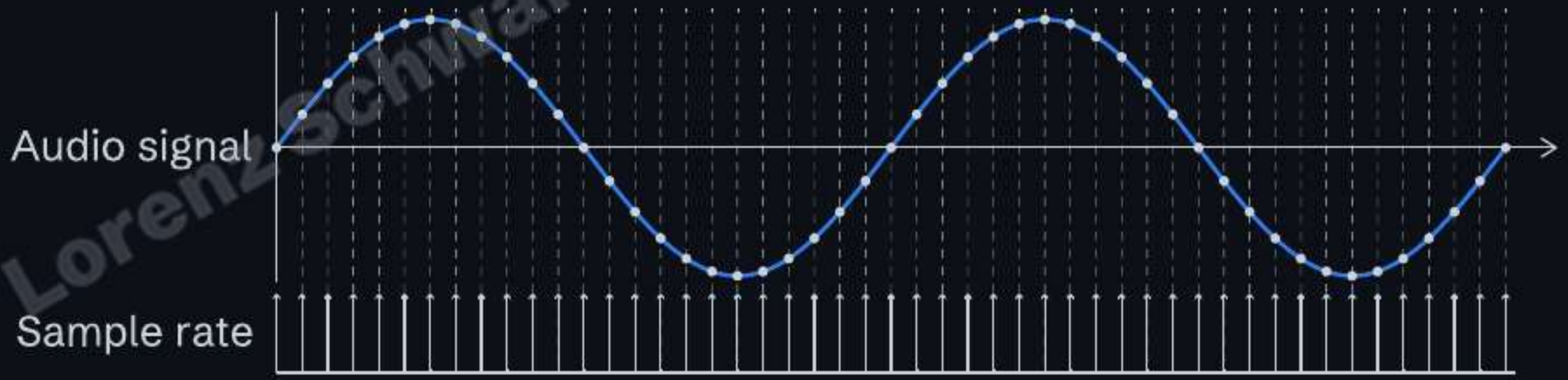
Digital systems can store and represent only:

- Discrete-time values (sampling in time)
- Discrete-amplitude values (quantization in amplitude)

Sample Rate

The sample rate defines how often per time period (s) the discrete voltage levels are measured and stored.

- Defines maximum recordable frequency (Nyquist = ½ rate)
- Common sample rates: 44.1 kHz (CD standard), 48 kHz (professional/video standard), 96 kHz+ (high-resolution audio)



Calculation of the sample interval

$$T_{\text{sample interval}} = \frac{1}{f_{\text{sample rate}}}$$

- $T_{\text{sample interval}}$ is the time between consecutive samples (e.g., 22.67 µs for 44.1 kHz).
- $f_{\text{sample rate}}$ is the sampling rate (e.g., 44.1 kHz for CD audio).

→ *Practical sampling rate must exceed* $2 \cdot f_{max}$ *by margin due to anti aliasing filter roll-off.*

Nyquist-Shannon Sampling Theorem

To accurately capture all frequencies in a signal, the sampling rate must exceed twice the highest frequency:

$$f_{\text{sampling}} > 2 \cdot f_{\max}$$

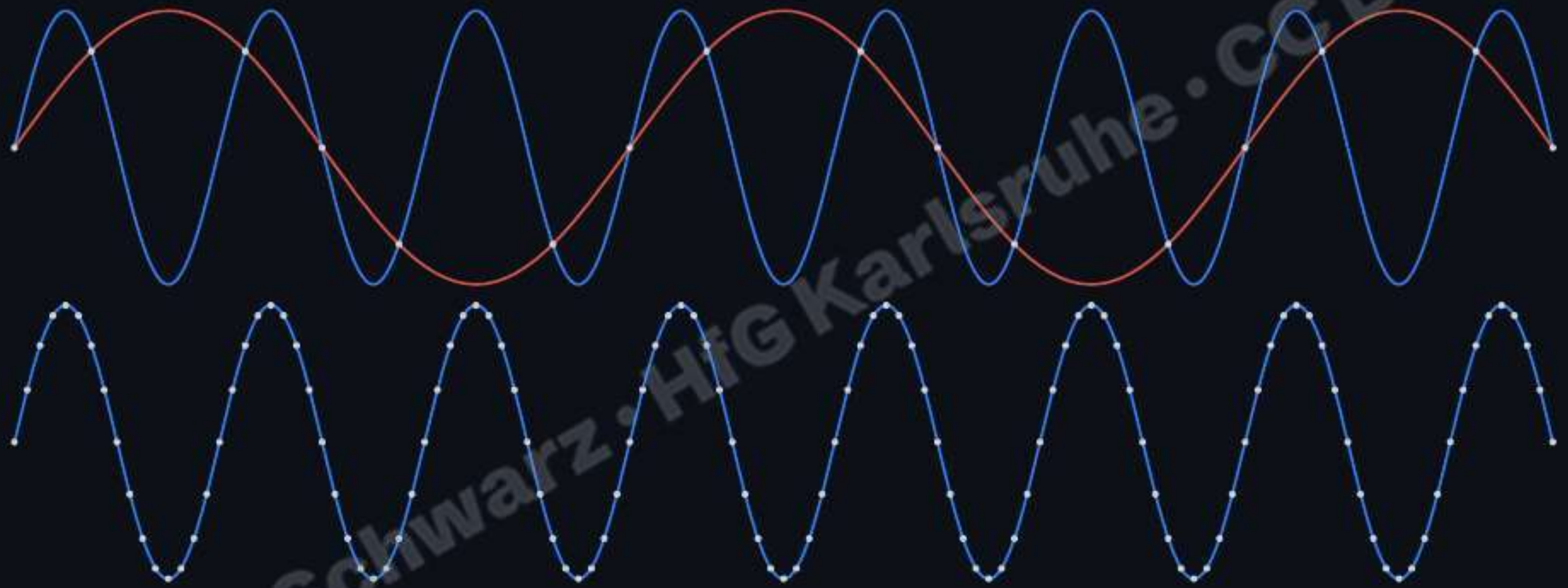
where $f_{\max}$ is the highest frequency component in the signal.

Maximum capturable frequency (Nyquist frequency):

$$f_{\text{Nyquist}} = \frac{\text{Sample rate}}{2}$$

→ *Signal frequencies exceeding half the sample rate cause aliasing.*

Aliasing



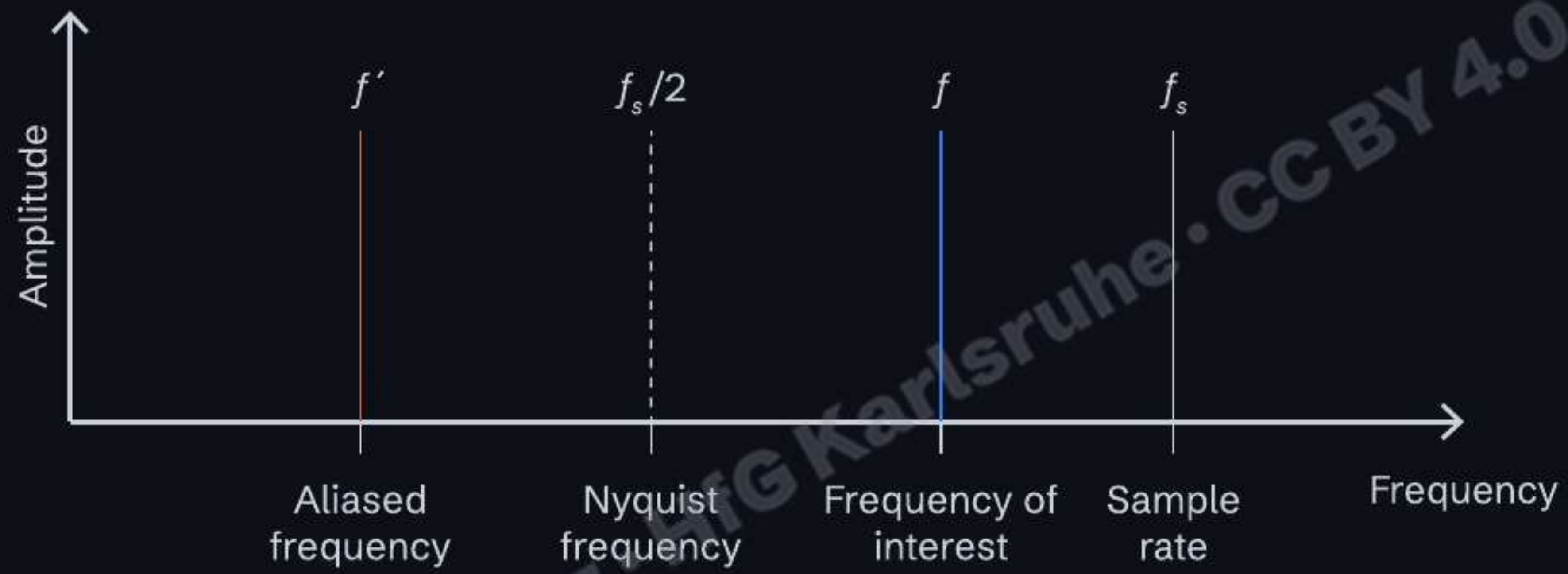
Undersampling (top) vs. correct sampling (bottom) of a sine wave

Contraints of Sampling: Aliasing

When sampling a signal, frequencies above the Nyquist frequency are reflected back into the audible range, creating unwanted artifacts.

Example:
30 kHz tone sampled at 44.1 kHz (Nyquist = 22.05 kHz) appears as 14.1 kHz, which is the difference between the frequency being sampled and the Nyquist frequency.

The tone is mirrored to the Nyquist frequency and folded back into the useful spectrum.



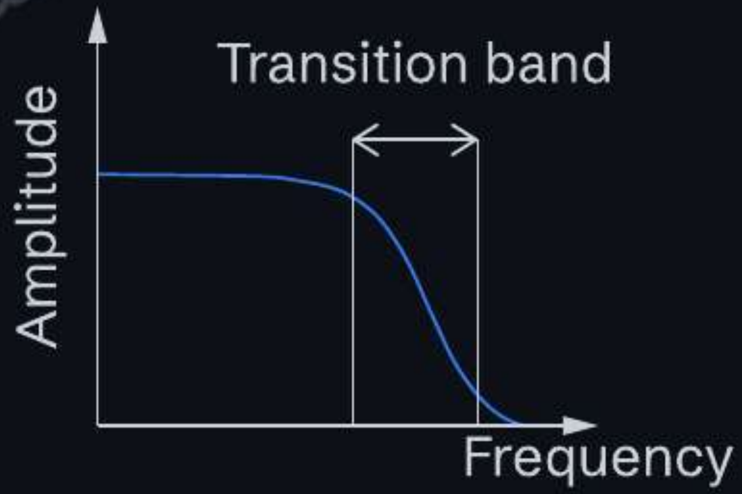
Sine sweep exceeding Nyquist frequency is mirrored back into the audible range.

▶ Play aliased sine sweep

Minimizing aliasing: Anti-aliasing filter

Low-pass filter is located before the ADC, in the analog domain that attenuates frequencies above Nyquist to prevent them from folding back into the desired signal band.

Ideal filter



Practical filter

Ideal vs. Real Anti-Aliasing Filters

Ideal brick-wall filter:

- Infinite slope at Nyquist frequency
- Perfect pass/stop separation
- Impossible to build in analog domain

Real analog filters:

- Gradual roll-off (typically 12-18 dB/octave)
- Require transition band between passband and stopband
- Introduce phase shifts

Practical implications of anti-aliasing filters

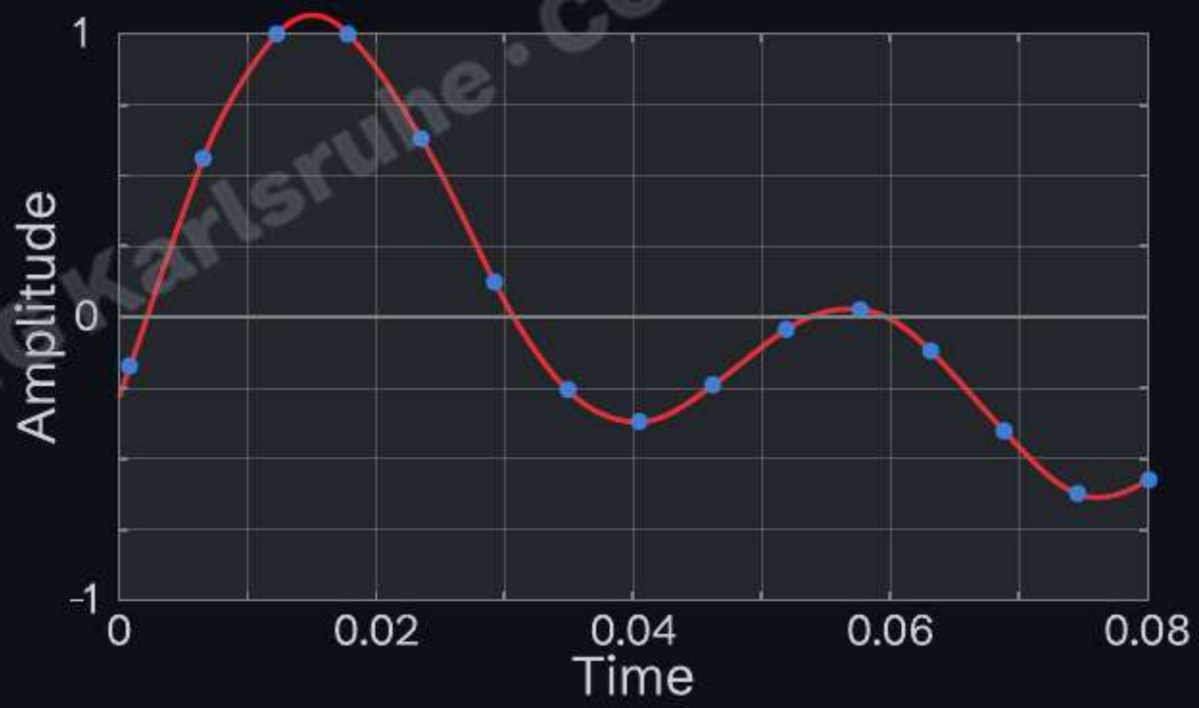
Sample rate must exceed 2× by sufficient margin to accommodate filter slope.

- 44.1 kHz allows ~20 kHz audio with practical filter design
- Filter begins attenuating around 20 kHz
- Full attenuation reached above 22.05 kHz (Nyquist)

Inter-sample peaks

Sample values might be at 0 dBFS, but the reconstructed waveform between (inter-sample) them can exceed this, potentially causing clipping or distortion.

→ *Provide a headroom buffer of 1 to 2 dBFS during mastering/export*



Inter-sample peak

Oversampling

Oversampling processes audio at a higher internal sample rate than the project rate.

- Easier anti-aliasing filter design (gentler slope, less phase shift)
- Prevents aliasing when plugins generate new harmonics (distortion, saturation)

Quantization

Besides the time-domain sampling, the second important step to digitize a signal is amplitude-domain quantization:

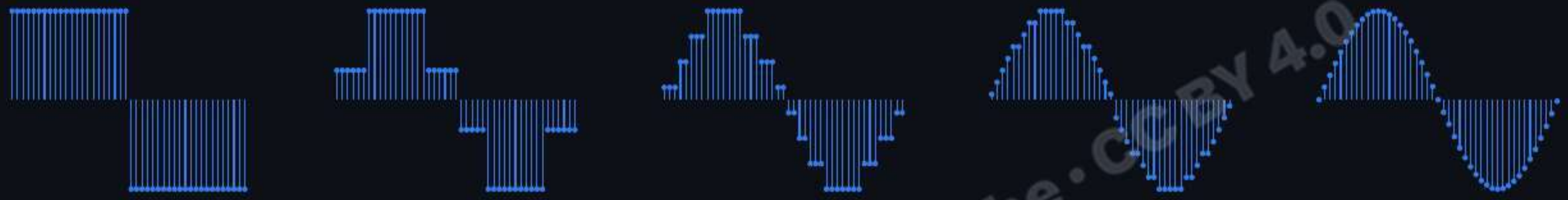
Each sample is rounded to the nearest amplitude value set by the bit depth, introducing small quantization errors. (Quantization and bit depth)

Quantization

Digital systems can only represent numbers with finite (limited) precision.

Sampling requires mapping each sample to the nearest value within a finite set of amplitude levels.

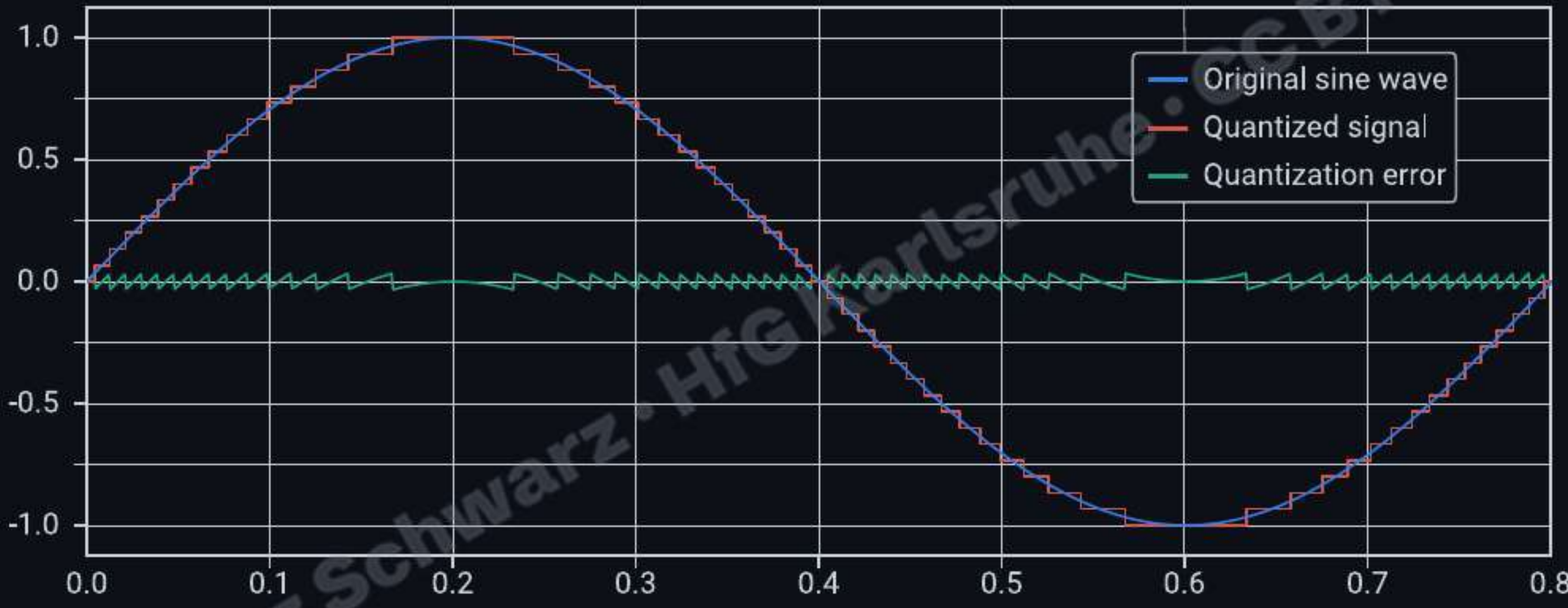
- Bit depth defines resolution, rounding/truncation error, and dynamic range
- Quantization introduces small errors (noise) in the signal
- Common Bit Depths: 16, 24, 32-bit float



Examples (from left to right):

1-bit quantization (2 levels) ▶
2-bit quantization (4 levels) ▶
3-bit quantization (8 levels) ▶
4-bit quantization (16 levels) ▶
8-bit quantization (256 levels) ▶

Quantization error



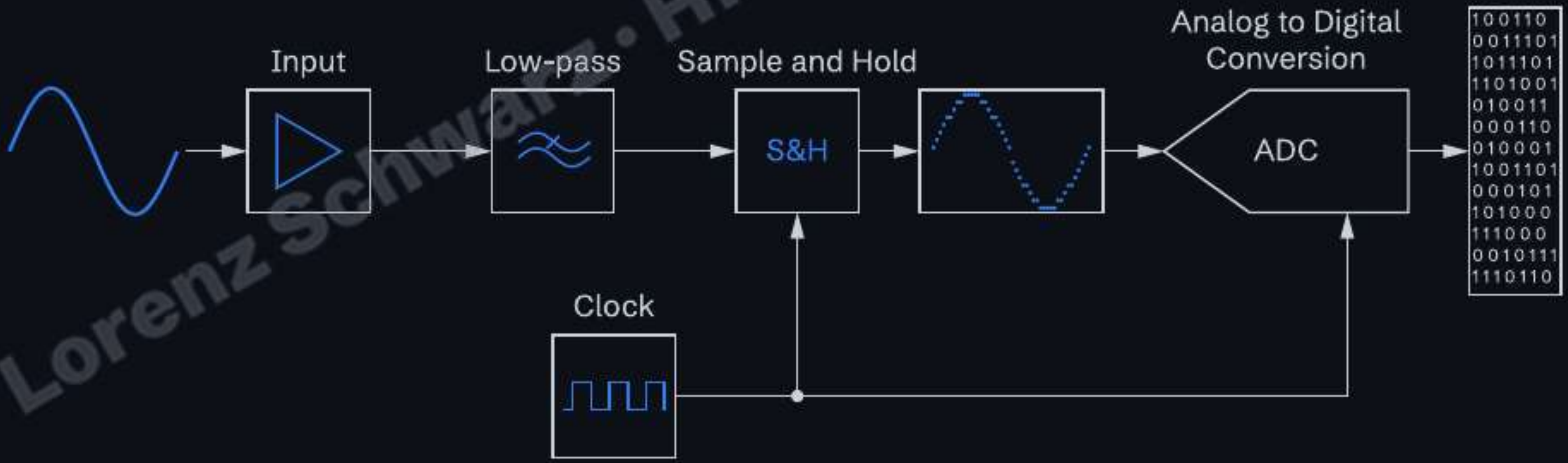
The difference between the actual amplitude (blue) and the quantized value (stepped red line) is the quantization error (green).

Summary: Sampling and quantization

- **Horizontal resolution (time):** determined by sample rate
- **Vertical resolution (amplitude):** determined by bit depth

Analog-to-Digital Conversion (ADC)

1. **Anti-aliasing filter (analog):** remove frequencies > Nyquist
2. **Sampling:** capture instantaneous voltage
3. **Quantization:** round to nearest digital value
4. **Encoding:** convert to binary data



Digital-to-Analog Conversion (DAC)

A **DAC** transforms digital signals (binary data) back into continuous analog waveforms.

1. **Decode:** convert binary to amplitude values
2. **Digital-to-analog conversion:** create stepped analog signal
3. **Reconstruction filter (analog):** smooth steps into continuous wave
4. **Amplification:** boost to line level

Dither

Quantization creates systematic rounding errors that produce audible distortion at low signal levels.

Dither adds very low-level noise before quantization to randomize these errors, transforming harsh distortion into low background noise.

→ *Always dither when exporting to lower bit depth to preserve low-level detail.*

Dynamic range

Dynamic range (DR): theoretical maximum determined by bit depth calculation

$$DR \approx 6.02N + 1.76 \text{ dB}$$

- **N = bit depth**
- 6.02 dB per bit + offset (1.76 dB)

→ 24-bit audio enables a 146 dB dynamic range, corresponding to the span from whisper (minimum) to jet engine at close range (maximum).

Dynamic range of various bit rates

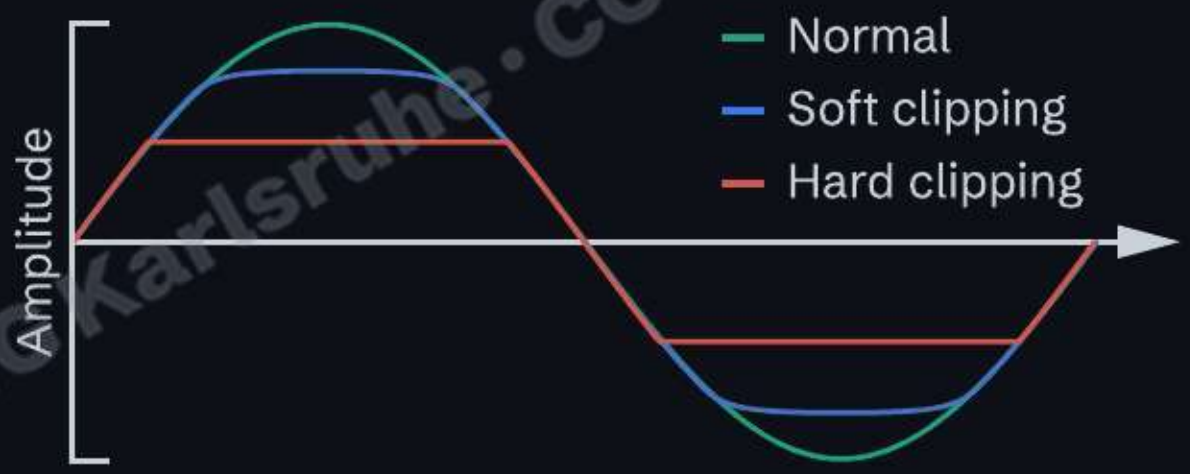
## bits	SNR (Audio)	Minimum amplitude step (dB)	possible values per sample
8	49.93 dB	0.1948 dB	256
16	98.09 dB	0.00598 dB	65,536
24	146.26 dB	0.00000871 dB	16,777,216
32	194.42 dB	0.0000000452 dB	4,294,967,296

→ *Dynamic range of humans: threshold of hearing to threshold of pain ≈ 120 dB*

Clipping

Clipping is a change of the waveform due to electronic or digital limitations.

- Introduces new frequencies (distortion)
- Digital clipping: abrupt flattening (hard clipping)
- Analog clipping: gradual saturation (soft clipping)



Hard clipping vs. soft clipping

▶ Clipping of a sine wave

Floating Point vs. Fixed Point

- **Fixed Point (16/24-bit):**
  - Fixed range, limited dynamic range
  - Used for recording and final delivery
- **Floating Point (32/64-bit):**
  - Audio range: -1.0 to +1.0 represents 0 dBFS at output
  - Internal processing can exceed 1.0 (e.g., value 2.0 ≈ +6 dBFS above 0 dBFS)
  - Prevents clipping during summing (e.g., 0.8 + 0.9 = 1.7, no clip yet)
  - Must be brought back ≤ 1.0 before D/A conversion or file export
  - Used for internal DAW processing

→ *DAWs usually process at 32-bit float.*

Calculating bit rate and file size

Bit rate (amount of data per second):

$$\text{Bit rate} = \text{Sample Rate} \times \text{Bit Depth} \times \text{Channels}$$

File size (total data for duration):

$$\text{File Size (bytes)} = \frac{\text{Bit Rate} \times \text{Duration (s)}}{8}$$

Example: 1 minute stereo, 48 kHz, 24-bit
= (48,000 × 24 × 2 × 60) / 8 = 8,640,000 bytes ≈ **8.64 MB**

Example sizes for WAV file

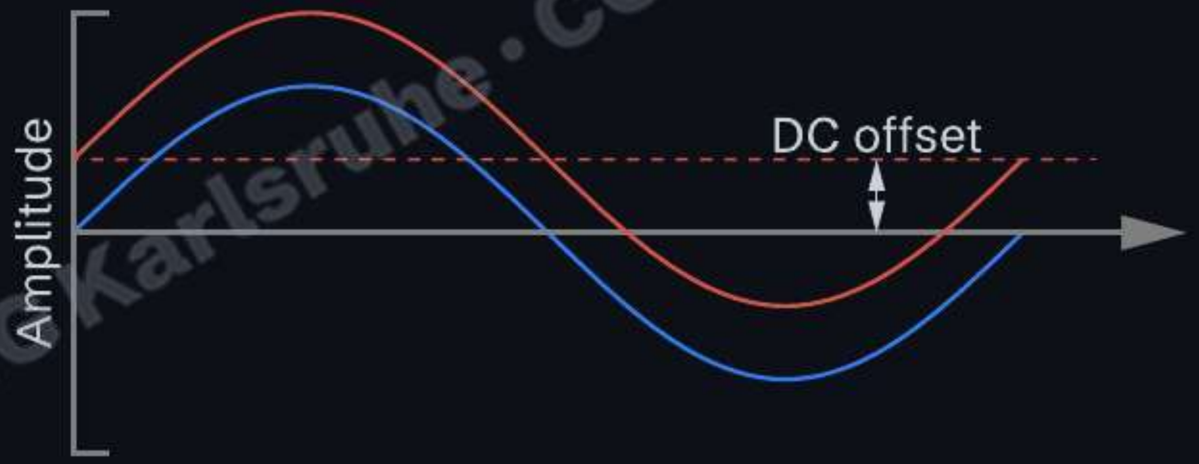
for a one minute long file:

Channels	Sample Rate (kHz)	Bit Depth	File Size (MB)
1	44.1	16	5.29 MB
1	44.1	24	7.94 MB
1	48	24	8.64 MB
1	48	32 float	11.52 MB
1	96	24	17.28 MB
1	96	32 float	23.04 MB
2	**48**	**24**	**17.28 MB**

DC offset

DC offset occurs when a waveform has a non-zero average value, shifting the entire signal away from the zero line.

- Reduces available dynamic range
- Can cause clipping during processing
- Creates clicks/pops when starting/stopping playback
- May damage speakers



Waveform with DC offset

→ *Apply DC offset removal / high-pass filter (e.g., 20 Hz).*

Practical DAW settings

Every project defines two key parameters that determine audio quality and file size:

- **Sample rate (kHz):** how often the sound is measured (e.g. 44.1, 48, or 96 kHz).
- **Bit depth (bits):** how precisely each sample is stored (e.g. 16, 24, or 32-bit float).

→ *Recommended setting: 48 kHz / 24-bit*

Digital processing: latency and buffers

Latency is the time between an audio signal entering and leaving the system.

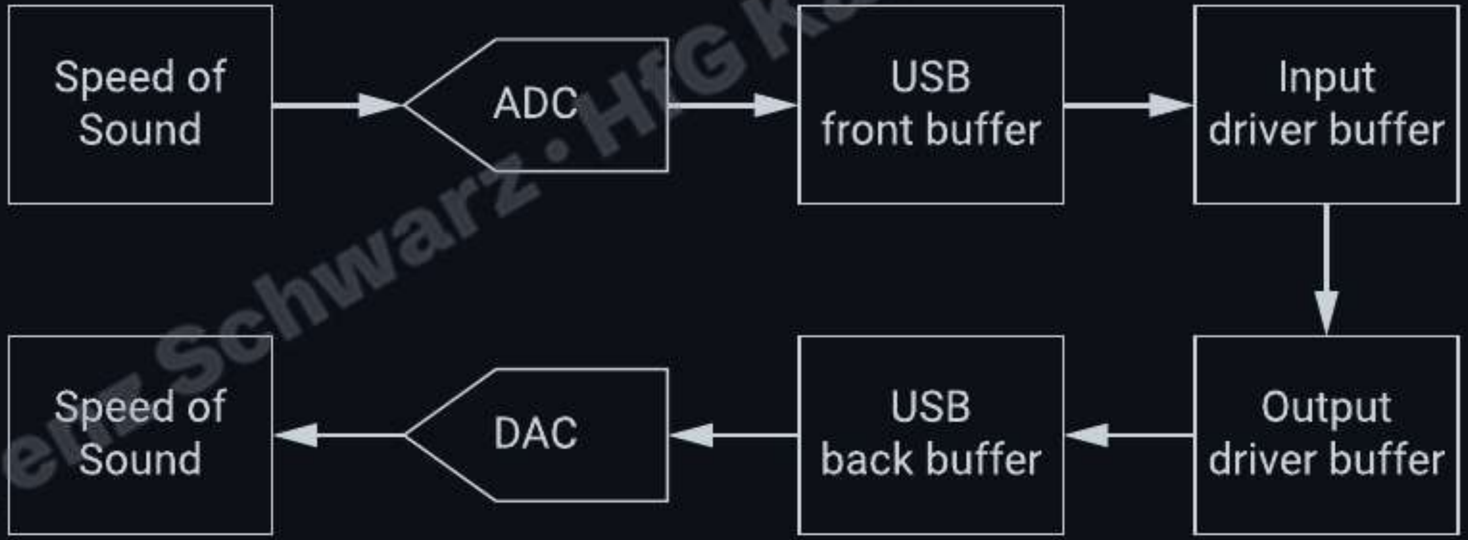
Main causes:

- Buffering in the computer or audio interface
- A/D and D/A conversion
- Digital processing (plugins, effects)

Some latency is unavoidable (conversion, processing) while buffer latency is adjustable.

Buffers

Digital audio systems use buffers (small sections of temporary memory) to process audio in blocks. The buffer size affects both latency (delay) and system stability.



Buffer size and latency

Buffer size defines how many audio samples the system processes at once and directly affects latency and stability.

Small buffers (64–128 samples): Low latency for recording and live monitoring, higher CPU load, risk of dropouts

Large buffers (512–1024 samples): Higher latency, lower CPU load, stable playback for mixing

→ *Use the smallest buffer size that avoids dropouts for the given task.*

Buffer Size & Delay (Latency)

$$\text{Buffer (samples)} = f_s\,[\text{Hz}] \times t\,[\text{s}]$$

$$\text{Delay (ms)} = \frac{\text{buffer}_{\text{samples}}}{f_s} \times 1000$$

Example:

Buffer = 128 samples at 48 kHz:
Delay = (128 / 48000) × 1000 = 2.67 ms

Human perception of latency

humans can detect a silent gap between two sounds of about 2–3 ms.

If sounds are less similar, or in noise / lower intensity, or onsets with less pronounced attack phase, threshold increases (≥ 4-5 ms).

→ *Buffer settings around 128 samples (≈3ms at 48kHz) feel immediate to most musicians during recording*

Buffer size and delay

Buffer Size in samples	Delay in ms for 44.1kHz	Delay in ms for 48kHz
32	0.72	0.66
64	1.45	1.33
128	2.9	2.6
256	5.8	5.3
512	11.6	10.6
1024	23.2	21.3
2048	45.9	42.1

Physical vs. Digital Latency

Well-optimized digital systems introduce less latency than the physical distance between performers and their monitoring systems (speed of sound ≈343 m/s in air).

Acoustic propagation delay:

- 1 meter: ≈3 ms
- 3 meters: ≈9 ms (typical distance to studio monitors)
- Distance between band members on stage: 3–6 meters (≈9–18 ms)

→ *Digital audio latency (3–10 ms) is comparable to or shorter than acoustic delays musicians naturally encounter.*

File formats and storage

Pulse-Code Modulation (PCM):

Analog signal amplitude is **sampled at uniform intervals** and each sample is **quantized to the nearest digital step.**

→ *PCM = sampling + quantization.*

Audio File Formats

- **WAV / AIFF (PCM):** standard uncompressed, full quality.
- **FLAC / ALAC:** lossless compression.
- **MP3 / AAC / OGG:** lossy, smaller size but reduced fidelity.

Audio file formats and compression

- Uncompressed (PCM)
  - WAV / AIFF: full quality, standard for recording and production
- Lossless compression
  - FLAC / ALAC: identical audio quality, reduced file size
  - → suitable for storage and archiving
- Lossy compression
  - MP3 / AAC / OGG: smaller files with irreversible quality loss
  - → suitable only for final delivery/distribution

Wordclock

Word clock is used when multiple digital devices (interface and converters) are connected:

- One device as "master clock," others as "slave"
- A clock signal that synchronizes sampling across digital audio devices
- Ensures all devices sample at the same time
- Prevents clicks, jitter, and drift

Jitter and synchronization

Jitter is unwanted timing variation in the digital audio clock, causing samples to be processed at incorrect times and potentially introducing distortion.

Synchronization aligns multiple devices to a common clock to minimize jitter and ensure stable audio transfer.

Jitter

I. Logarithmic Scales in Acoustics

The human hearing range

Human hearing range

The human ear **perceives** sound pressures from **20 µPa** (threshold of hearing) to **20 Pa** (pain threshold), a ratio of **1 : 1,000,000**.

$$\frac{20}{20 \times 10^{-6}} = 1 \times 10^{6}$$

The pain threshold is about a million times louder than the weakest audible sound.

Scaling human hearing

The human hearing is too large for a linear scale, so the logarithmic decibel scale is used, matching human hearing.

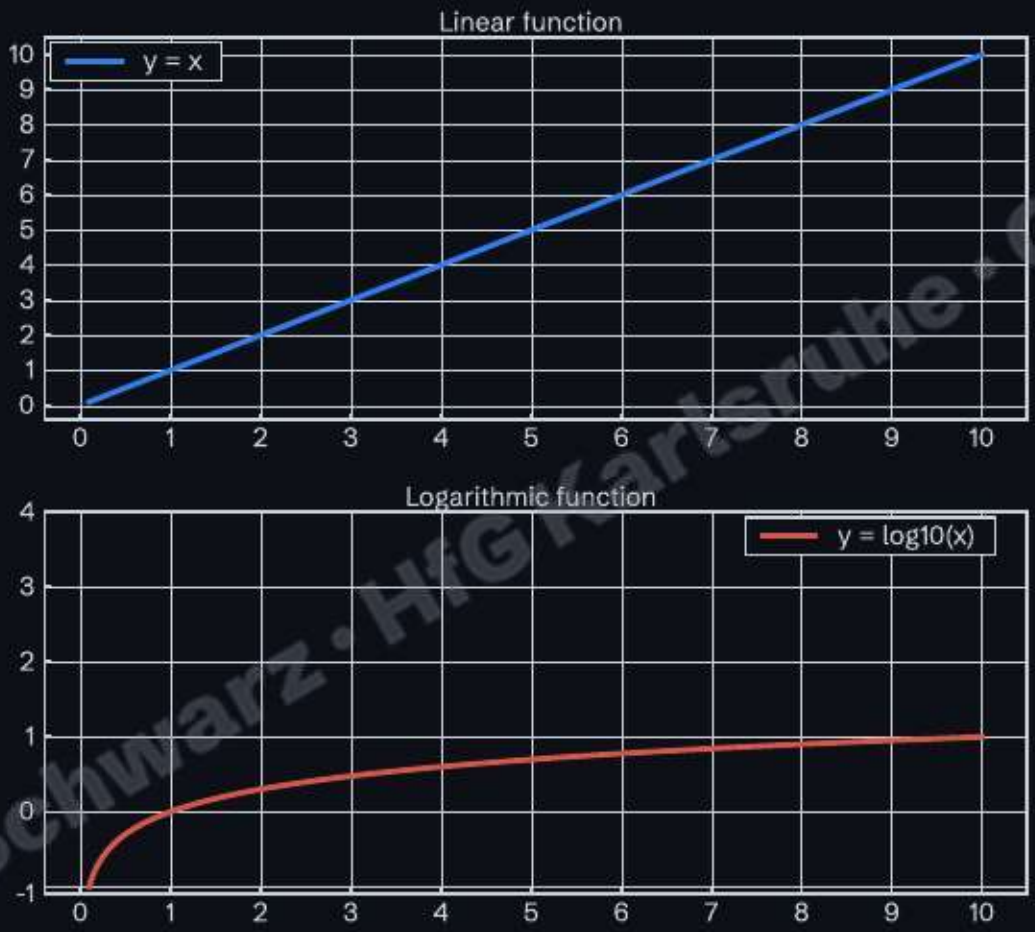
$$(0.00002 \rightarrow 20Pa)$$

$$\Downarrow$$

$$(0 \rightarrow 120dBSPL)$$

Examples in air at standard atmospheric pressure

Sound source	Distance	Pa	dB_{SPL}
Eruption of Krakatoa	165 km	—	172
Jet engine	1 m	632	150
Trumpet	0.5 m	63.2	130
Threshold of pain	At ear	20–200	120–140
Risk of instantaneous noise-induced hearing loss	At ear	20.0	120
Jet engine	30–100 m	6.32–200	110–140
Traffic on a busy roadway	10 m	0.20–0.63	80–90
Hearing damage (long-term exposure)	At ear	0.36	85
TV (home level)	1 m	6.32×10^{-3}–0.02	50–60
Normal conversation	1 m	2×10^{-3}–0.02	40–60
Light leaf rustling, calm breathing	Ambient	6.32×10^{-5}	10
Hearing threshold	—	20×10^{-6}	0



Logarithmic compression makes the enormous range of human hearing manageable.

II. Logarithmic Principles

Decibels and ratios

Decibel

A decibel is one-tenth of a Bel, a unit named after Alexander Graham Bell, expressing power ratios logarithmically.

- Expresses **relative change** (ratio between two physical quantities)
- A **logarithmic unit** (compresses large ranges into manageable numbers)

Logarithm

The logarithm is the inverse function of exponentiation:

$$b^x = a \quad \Longrightarrow \quad x = \log_b(a)$$

Example:

$$2^5 = 32 \quad \Longrightarrow \quad 5 = \log_2(32)$$

Note: $\log_x(1) = 0 \;\; \Rightarrow$ ratio of 1 gives 0 dB

Properties of logarithmic scales

Human perception of intensity follows an approximately logarithmic relationship.

Benefits:

- Matches nonlinear human perception
- Represent ratios (×2, ×10, etc.) consistently
- Display data spanning many orders of magnitude on one scale
- Reveal details at both low and high ends

→ Used for frequency and magnitude displays

Linear and logarithmic scales

With a logarithmic scale, the values of the tick marks increase by the same factor over equal distances (e.g., a base value of 10 raised to the powers 0, 1, 2, 3, etc.)

With a linear scale, the values of the tick marks increase by the same amount over equal distances.

Decibel relationships

Change (dB)	Power	Amplitude	Perception
+1 dB	~1.26×	~1.12×	Barely noticeable
+3 dB	~2×	~1.41×	Clearly noticeable
+6 dB	~4×	~2×	Noticeably louder
+10 dB	10×	~3.16×	About 2× as loud
-1 dB	~0.79×	~0.89×	Barely noticeable
-3 dB	~½	~0.71×	Clearly quieter
-6 dB	~¼	~½	Noticeably quieter
-10 dB	1/10	~0.32×	About ½ as loud

Audio demonstration: Decibel changes

Pink noise demonstrating relative dB changes (each compared to reference):

▶ Play demonstration

Three A/B comparisons:

1. Reference → **+3 dB** (2× power, clearly noticeable)
2. Reference → **+6 dB** (4× power, 2× amplitude, noticeably louder)
3. Reference → **+10 dB** (10× power, about 2× as loud)

Absolute and relative dB

- **With suffix** → absolute level (SPL, dBV, dBm, dBFS)
- **Without suffix** → relative ratio (input/output)

III. Decibel Formulas

Power and magnitude ratios

Decibel formulas

Power quantities:

$$X_{dB} = 10\log_{10}\left(\frac{X}{X_{\text{ref}}}\right)$$

Field quantities (magnitude):

$$X_{dB} = 20\log_{10}\left(\frac{X}{X_{\text{ref}}}\right)$$

- $10\log_{10}$ → Power, Intensity
- $20\log_{10}$ → Amplitude, Voltage, Current, Pressure

Power and magnitude relationship

Power is proportional to the square of field quantities:

$$P = U \cdot I = R \cdot I^2 = \frac{U^2}{R}$$

When converting field quantities to decibels, this square relationship means:

$$20 \log_{10}\left(\frac{U}{U_{\text{ref}}}\right) = 10 \log_{10}\left(\frac{U^2}{U_{\text{ref}}^2}\right) = 10 \log_{10}\left(\frac{P}{P_{\text{ref}}}\right)$$

→ *The factor of 2 comes from the square relationship between power and field quantities.*

Converting between linear values and decibels

Power ratio:

$$x\,\mathrm{dB} = 10\log_{10}(a) \iff a = 10^{\frac{x}{10}}$$

Amplitude ratio:

$$x\,\mathrm{dB} = 20\log_{10}(a) \iff a = 10^{\frac{x}{20}}$$

Example:

Two signals whose levels differ by 6 dB have an amplitude ratio of $10^{\frac{6}{20}} \approx 2$

Example: doubling voltage and power in dB

- Voltage doubles (1 V → 2 V):

$$20\log_{10}(2/1)\,\mathrm{dB} \approx 20 \cdot 0.3010\,\mathrm{dB} = +6.02\,\mathrm{dB}$$

- Power doubles (1 W → 2 W):

$$10\log_{10}(2/1)\,\mathrm{dB} \approx 10 \cdot 0.3010\,\mathrm{dB} = +3.01\,\mathrm{dB}$$

dB conversion table

Decibels (dB)	Magnitude (ratio, 20·log)	Power (ratio, 10·log)
-20	0.1	0.01
-12	0.25	0.06
-6	0.5	0.25
-3	0.7	0.5
0	1	1
+3	1.4	2
+6	2	4
+12	4	16
+20	10	100

dB conversion graph

IV. Reference Systems

Absolute measurements

Decibel reference values

To express an **absolute value**, the suffix specifies the reference X_{ref}:

$$X_{dB} = k \cdot \log_{10}\left(\frac{X}{X_{\text{ref}}}\right), \quad k = \begin{cases} 10 & \text{power-like quantities} \\ 20 & \text{field-like quantities} \end{cases}$$

Unit	Quantity	Reference value X_{ref}
dB SPL	Sound pressure level	20 µPa
dBm	Power	1 mW
dBV	Voltage	1 V
dBu	Voltage	0.775 V
dBFS	Digital full scale	Maximum quantizing level

Example: sound pressure (SPL)

$$L_p = 20 \log_{10} \left(\frac{p_{\mathrm{rms}}}{p_{\mathrm{ref}}} \right) \mathrm{dB\ SPL}$$

p_{rms} is the root mean square of the measured sound pressure
p_{ref} is the standard reference sound pressure of 20 micropascals in air

$$p_{\mathrm{ref}} = 20 \times 10^{-6}\,\mathrm{Pa} = 0.00002\,\mathrm{Pa} = 20\,\mu\mathrm{Pa}$$

Example: calculating dB SPL

Reference: $p_{\text{ref}} = 20\,\mu\text{Pa}$ (threshold of hearing)

$$L_p = 20\log_{10}\left(\frac{p_{\text{rms}}}{p_{\text{ref}}}\right)\text{dB SPL}$$

Example with $p_{\text{rms}} = 0.02\,\text{Pa}$:

$$\Rightarrow \quad 20\log_{10}\left(\frac{0.02\,\text{Pa}}{20\times 10^{-6}\,\text{Pa}}\right)\text{dB SPL} = 20\log_{10}(10^3)\,\text{dB SPL} = 20\cdot 3\,\text{dB SPL} = 60\,\text{dB SPL}$$

Example: dBFS (Full Scale)

Digital audio systems:

- Reference = maximum quantizing level (0 dBFS)
- Values typically negative (counted down from maximum)

16-bit system: Range = -32768 to +32767, so $X_{\text{ref}} = 32767$

Example calculation:

$$X_{dB} = 20\log_{10}\left(\frac{16384}{32767}\right) = 20\log_{10}(0.5) \approx -6\,\text{dBFS}$$

→ *Half maximum amplitude = -6 dBFS*

V. Practical Applications

Using decibels in practice

Example: adding gain stages

Stage →	1	2	3	4	Total
Linear Gain ×	2.0	3.0	1.5	0.5	4.5
Gain in dB +	+6 dB	+9.5 dB	+3.5 dB	-6 dB	+13 dB

→ *Since dB is logarithmic, multiplying ratios in linear terms becomes addition in dB.*

Adding independent sound sources

Convert decibel values to linear form, perform the summation, then reconvert to decibels.

$$L_{\text{total}} = 10\log_{10}\left(10^{L_1/10} + 10^{L_2/10}\right)$$

Independent sources add **powers**, not pressures.

Adding two sound sources 80 and 85 dB

$$10\log_{10}\left(10^{80/10} + 10^{85/10}\right)$$

$$\Downarrow$$

$$10\log_{10}\left(10^{8} + 10^{8.5}\right)$$

$$\Downarrow$$

$$10\log_{10}(4.1623 \times 10^{8})$$

$$\Downarrow$$

$$86.2\,\mathrm{dB}$$

VI. Reference

Conversion factors and relationships

Decibel quick reference

General relationships (relative changes):

- $+3\,\mathrm{dB}$ doubles power/intensity
- $+6\,\mathrm{dB}$ doubles amplitude/pressure
- $+10\,\mathrm{dB}$ doubles perceived loudness
- $0\,\mathrm{dB}$ = no change (ratio = 1)

Applied to sound pressure level (SPL):

- $+3\,\mathrm{dB\ SPL}$ doubles intensity (since $I \propto p^2$)
- $+6\,\mathrm{dB\ SPL}$ doubles pressure
- $+10\,\mathrm{dB\ SPL}$ doubles perceived loudness

Common misconceptions

Adding dB SPL values directly

- Wrong: 70 dB + 70 dB = 140 dB
- Correct: Must convert to linear, add, convert back (≈ 73 dB)

Confusing 10 log vs 20 log

- **20 log** for: Voltage, Current, Pressure, Amplitude
- **10 log** for: Power, Intensity

Example 1

Question: A sound measures 0.2 Pa. What is the SPL in dB?

Solution

$$L_p = 20\log_{10}\left(\frac{0.2}{20\times 10^{-6}}\right) = 20\log_{10}(10^4) = 20\times 4 = 80 \text{ dB SPL}$$

Example 2

Question: Two independent sound sources each produce 70 dB SPL. What is the total SPL when both operate?

$$L_{\text{total}} = 10\log_{10}(10^7 + 10^7)$$

$$= 10\log_{10}(2 \times 10^7)$$

$$= 10[\log_{10}(2) + \log_{10}(10^7)]$$

$$= 10[0.301 + 7] \approx 73 \text{ dB SPL}$$

→ *Independent sources add powers, not dB values.*

- Ballou, Glen M. *Handbook for Sound Engineers*. Elsevier, 2008.
- Beyer, Robert T. Sounds of Our Times: Two Hundred Years of Acoustics. Springer Press, 1998.
- Blauert, Jens. *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT Press, 1997.
- Bregman, Albert S. *Auditory Scene Analysis*. MIT Press, 1999.
- Brixen, Eddy B. *Audio Metering: Measurements, Standards and Practice*. Routledge, 2020.
- Dickreiter, Michael. *Handbuch Der Tonstudiotechnik*. Vol. 1 & 2, De Gruyter Saur, 2022.

- Görne, Thomas. *Tontechnik*. Carl Hanser Verlag, 2006.
- Hartmann, William M. *Signals, Sound, and Sensation*. Springer, 1998.
- Izhaki, Roey. Mixing Audio: Concepts, Practices, and Tools. Routledge, 2018.
- Katz, Robert A. Mastering Audio: Über Die Kunst Und Die Technik. GC Carstensen, 2012.
- Miranda, Eduardo Reck. *Computer Sound Design: Synthesis Techniques and Programming*. Focal Press, Elsevier, 2008.
- Morse, Philip M. *Vibration and Sound*. Acoustical Society of America, 1995.
- Puckette, Miller. *The Theory and Technique of Electronic Music Miller Puckette*. World Scientific, 2011.

- Roads, Curtis. *Composing Electronic Music: a New Aesthetic.* Oxford University Press, 2015.
- Siedenburg, Kai, et al. *Timbre: Acoustics, Perception, and Cognition.* Springer, 2019.
- Smith, Julius O. *Spectral Audio Signal Processing.* W3K, 2011.
- Valimaki, Vesa, et al. "Fifty Years of artificial reverberation." IEEE Transactions on Audio, Speech, and Language Processing, vol. 20, no. 5, July 2012, pp. 1421–1448, https://doi.org/10.1109/tasl.2012.2189567.
- Weinzierl, Stefan. *Handbuch der Audiotechnik.* Springer, 2008.

- Wuttke, Jörg. *Mikrofonaufsätze*. Eigenverlag Schalltechnik Dr.-Ing. Schoeps GmbH. 2000.
- Yost, William A. *Fundamentals of Hearing: An Introduction*. Elsevier/Academic Press, 2007.

Websites

- Bregman, Al. "Auditory Scene Analysis." *Archived Copy of Al Bregman's Personal Website, Created with the Permission of Abigail Sibley and Christopher Lyons*, https://themusiclab.github.io/bregman-archive/asa.htm#. Accessed 27 Jan. 2025.
- Deutsch, Diana. "Illusions and Research." *Diana Deutsch's Web Page*. https://deutsch.ucsd.edu/psychology/pages.php?i=101. Accessed 28 Jan. 2025.
- Hajdu, Gerog "Music Technology Online Repository." Music Technology Online Repository, Hochschule für Musik und Theater Hamburg, https://mutor-2.github.io/. Accessed 23 Feb. 2025.
- Sengpiel, Eberhard. *Forum für Mikrofonaufnahmetechnik und Tonstudiotechnik*. https://www.sengpielaudio.com/ Accessed 6 Nov. 2023.

- Schaedler, Jack. "Circles Sines and Signals" *GitHub*, https://jackschaedler.github.io/circles-sines-signals/. Accessed 9 Oct. 2024.
- Smith, Julius O. III. "Sinusoids." *CCRMA*, 2 April 2024, https://ccrma.stanford.edu/~jos/st/Sinusoids.html. Accessed 9 Oct. 2024.

Copyright and Licensing

Contact: lschwarz@hfg-karlsruhe.de

← Chapters