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).
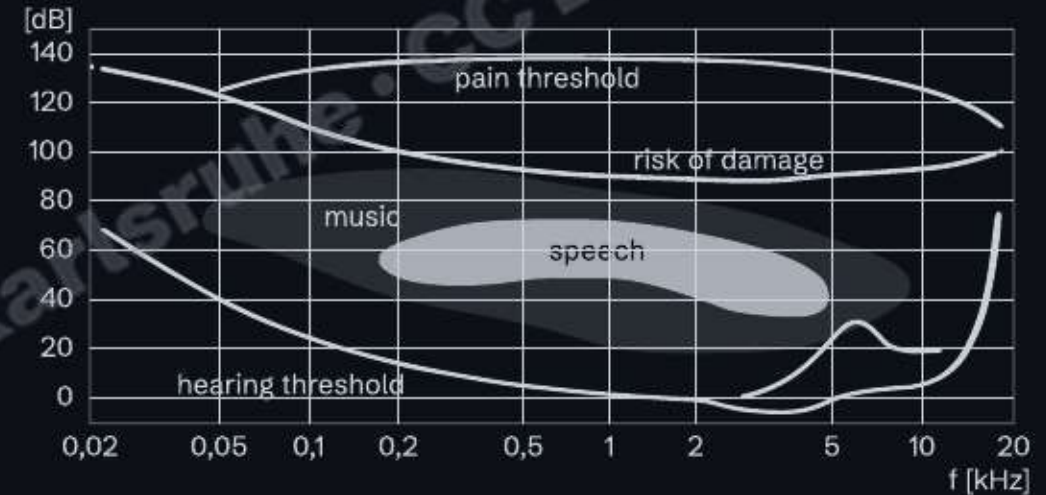- **Above 20 kHz:** ultrasonic, inaudible but used in technology.

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.

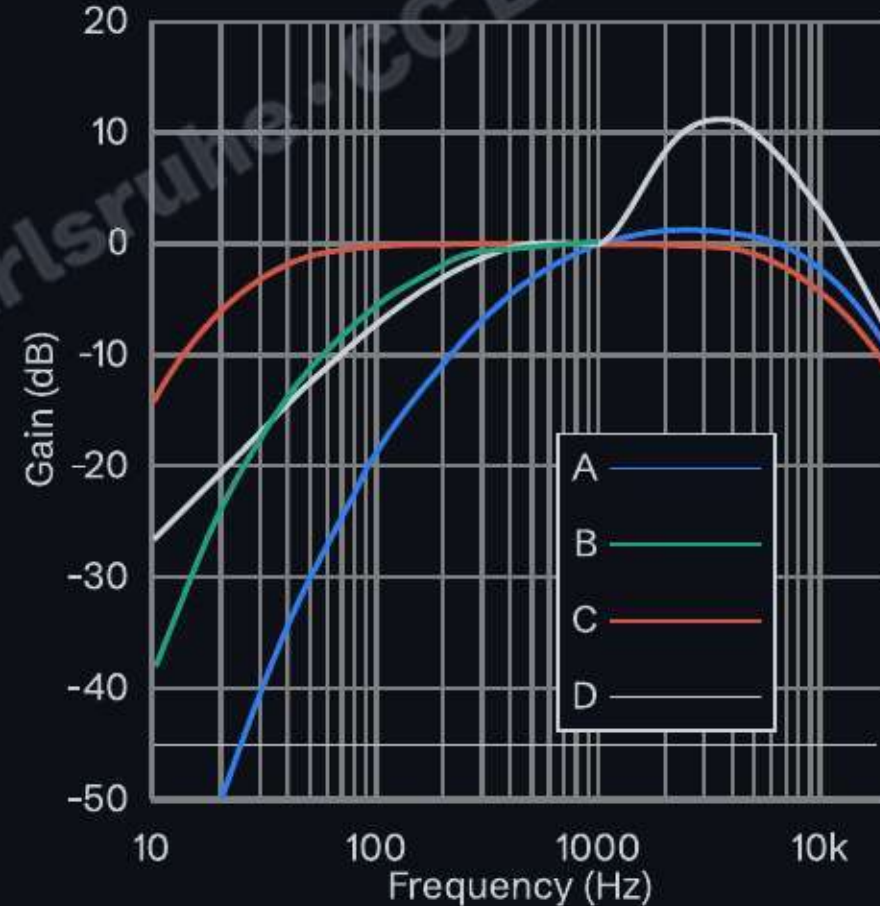Two tones with identical physical amplitude can be perceived as having different loudness:

▶ 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

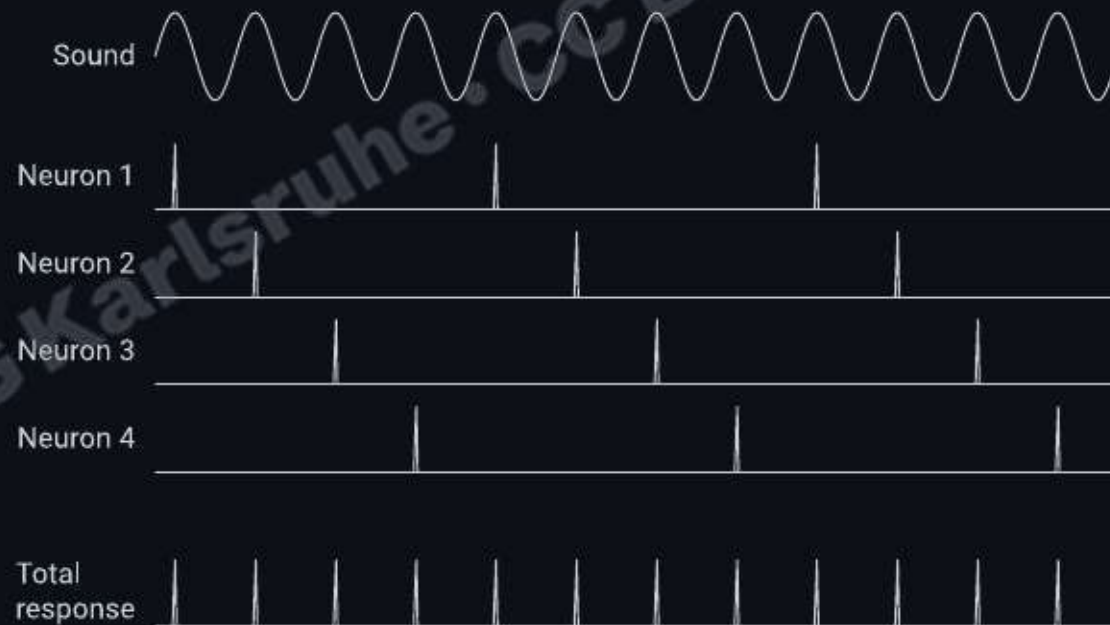# Place and temporal theories

- **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

18

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.
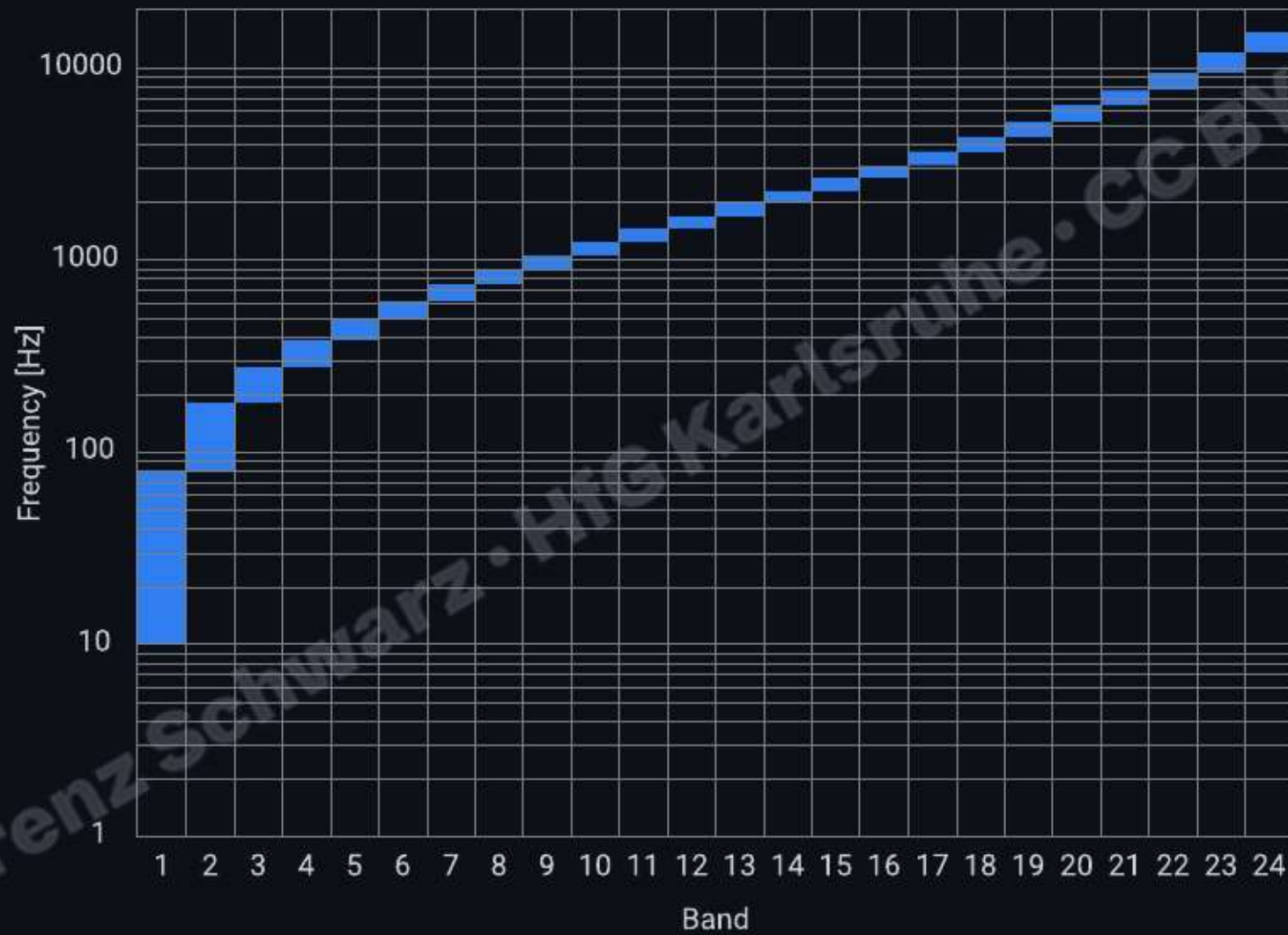
The JND for pitch varies with frequency:

- 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

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

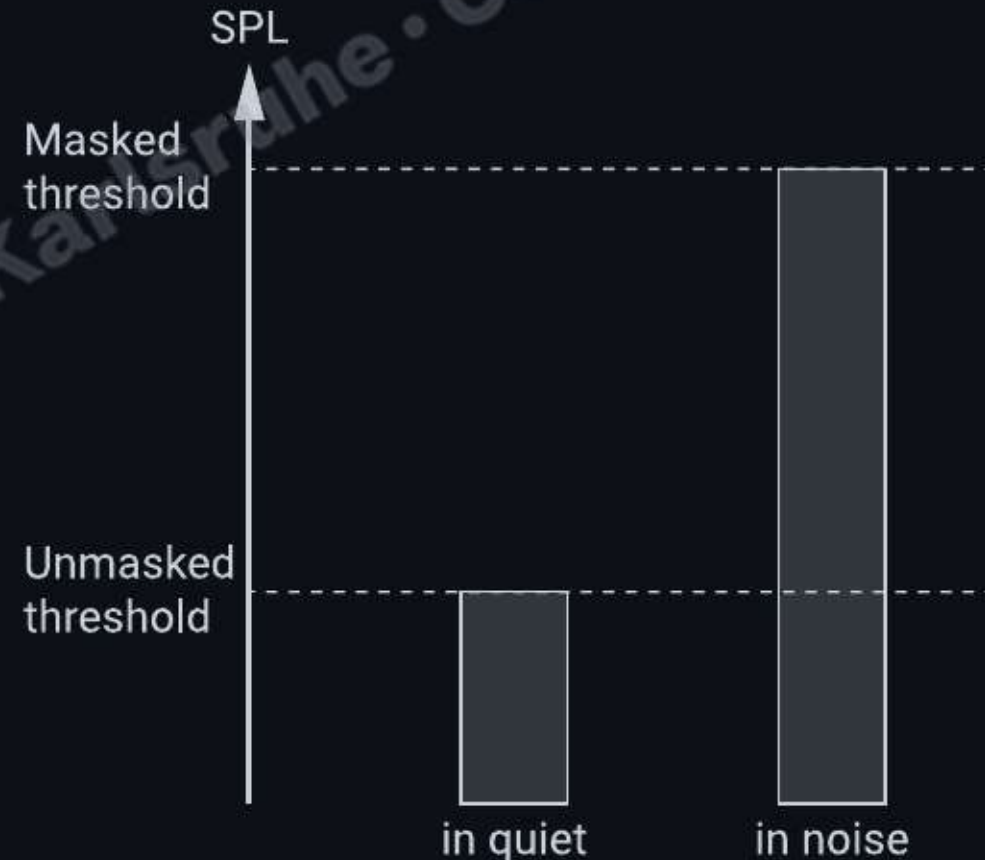# MASKING · CRITICAL BANDS

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.
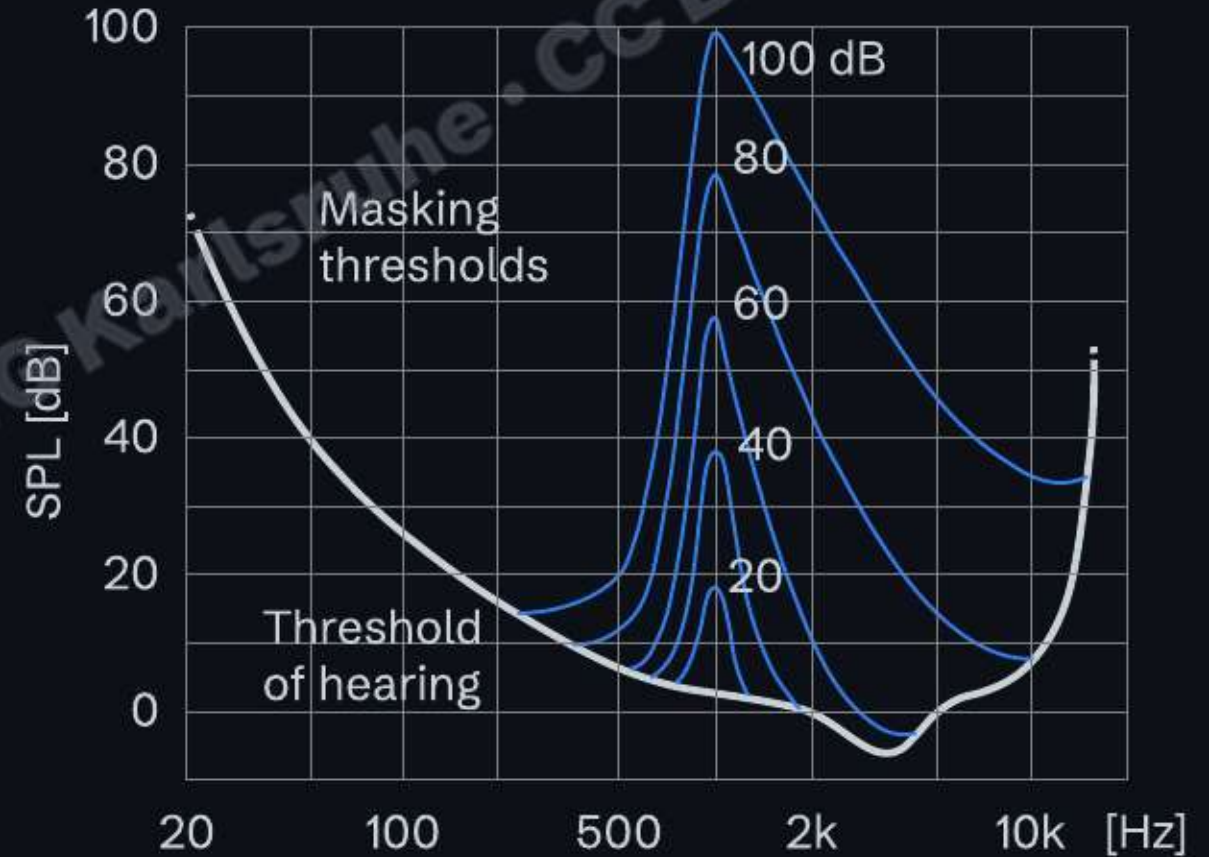
▶ Broadband noise masking sine tone
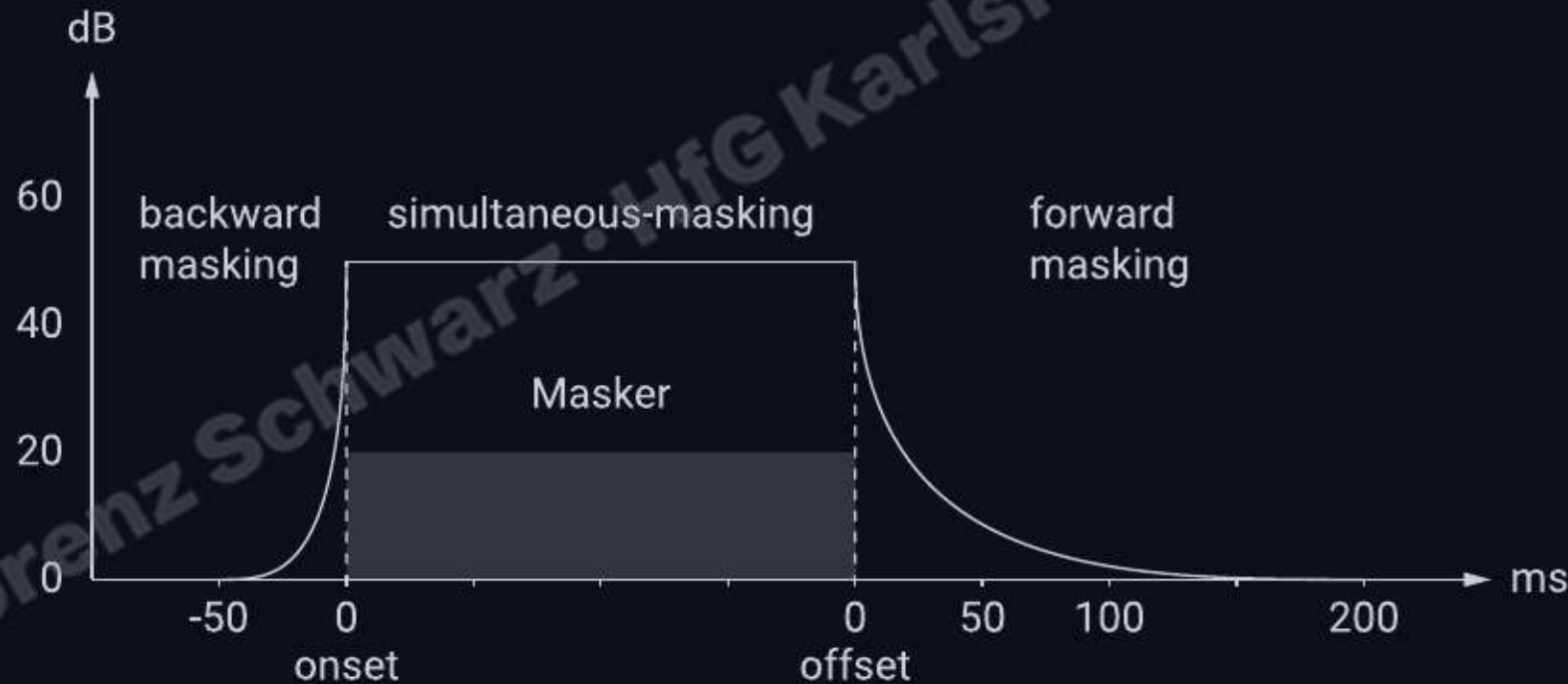
Spectral masking (or frequency masking)

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.
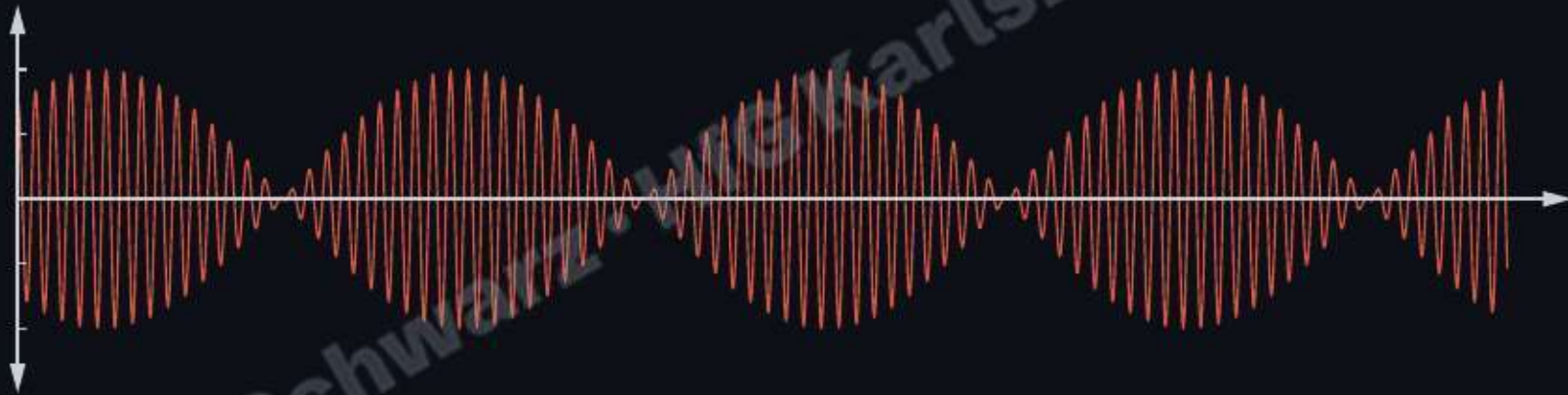
INTERFERENCE PHENOMENA

Beatychoacoustics

Beat

Interference of sound waves with slightly different frequencies.

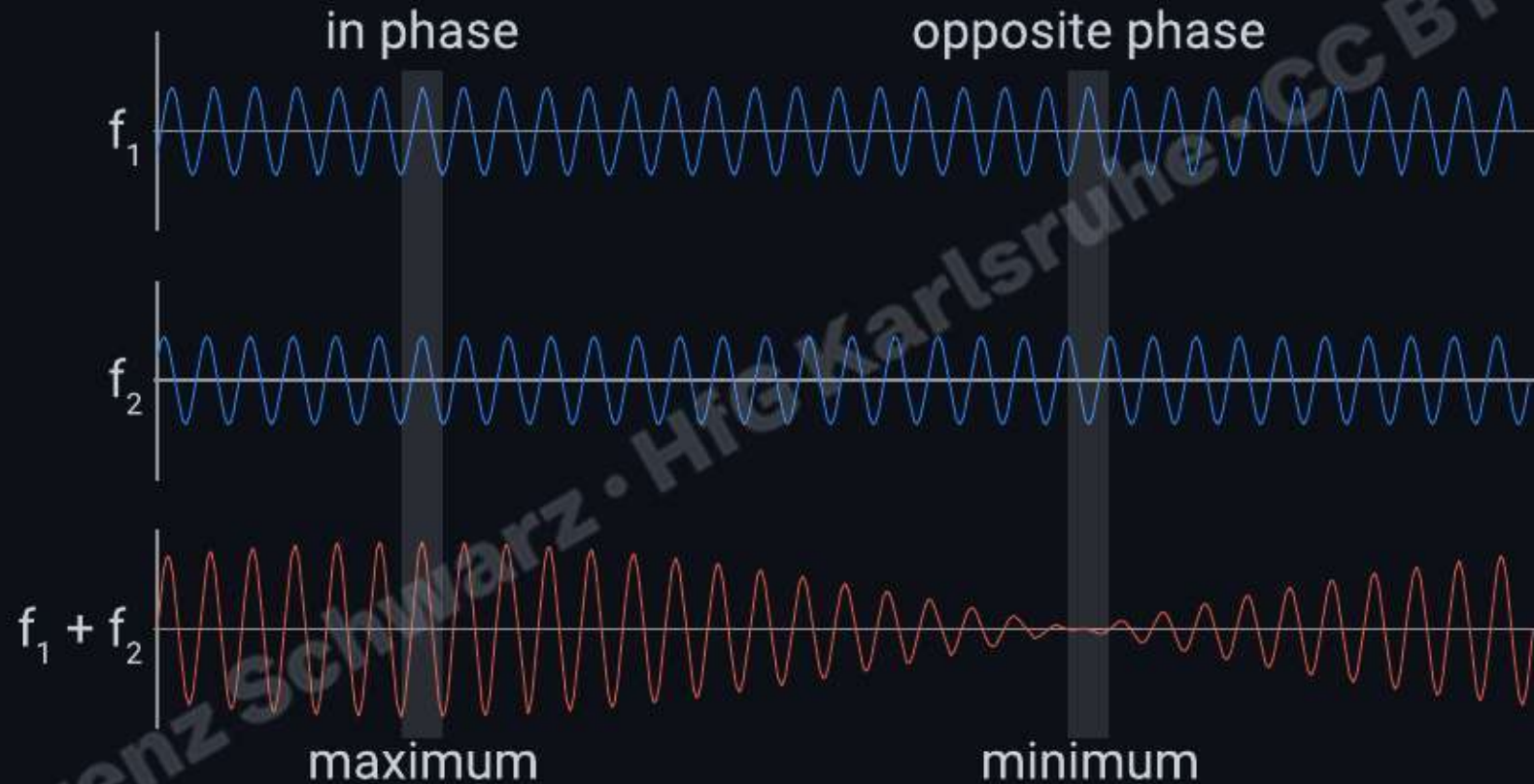Fundamentals of Sound | Lorenz Schwarz | WS 2024/2025

34

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)

$\rightarrow$ *Similar to electronic ring modulation*

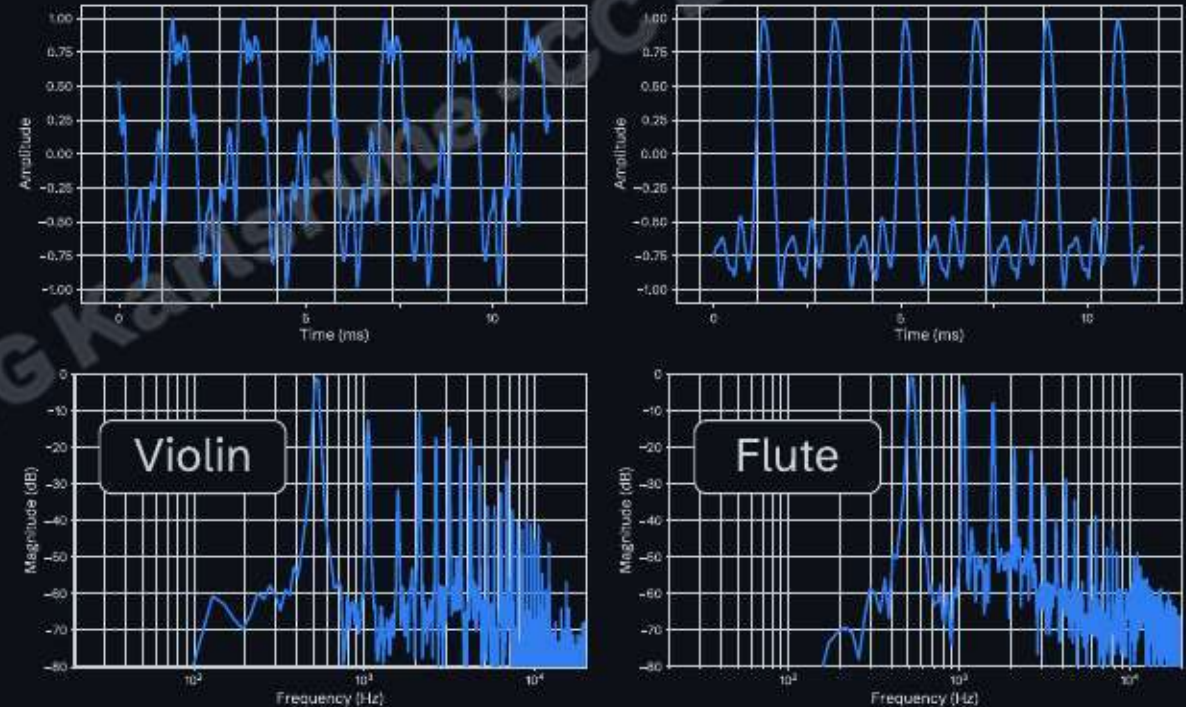▶ Combination tone 1500 Hz and 1000 Hz

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
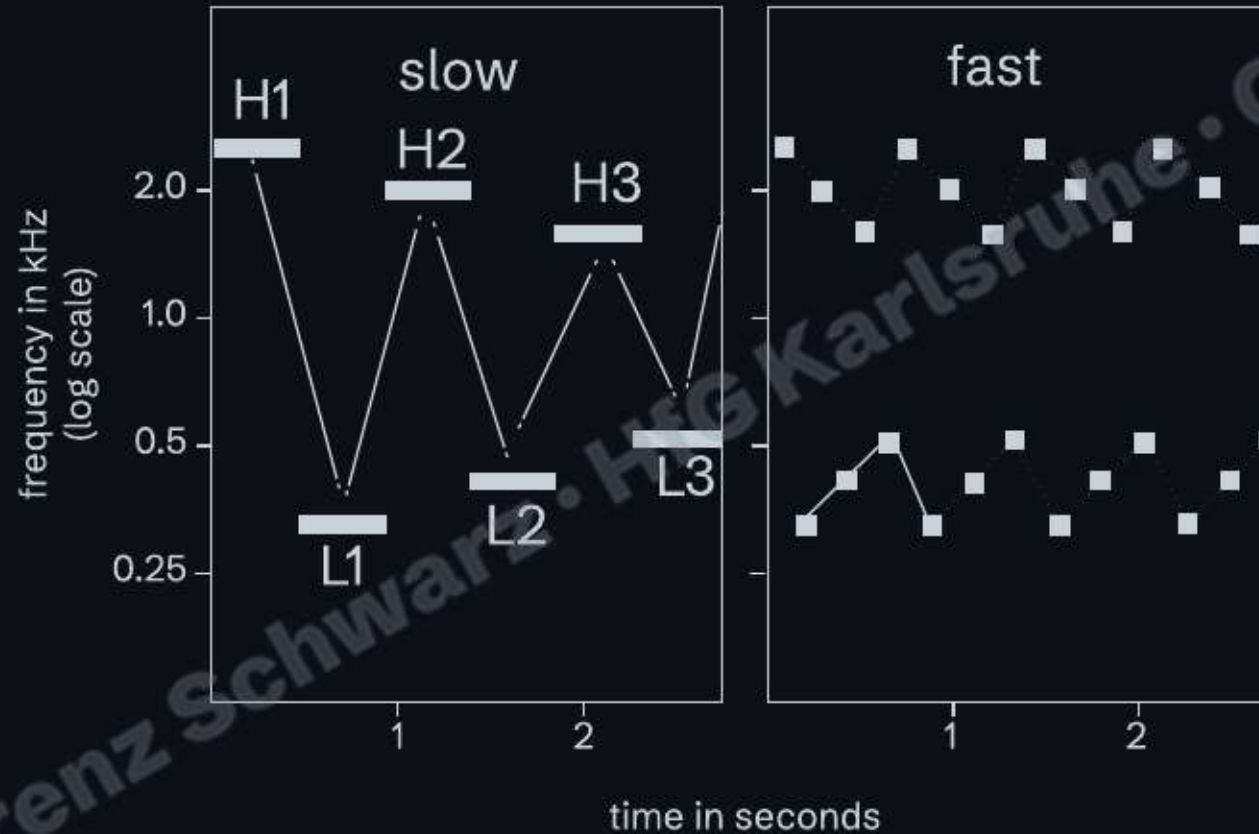- Segregation

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)
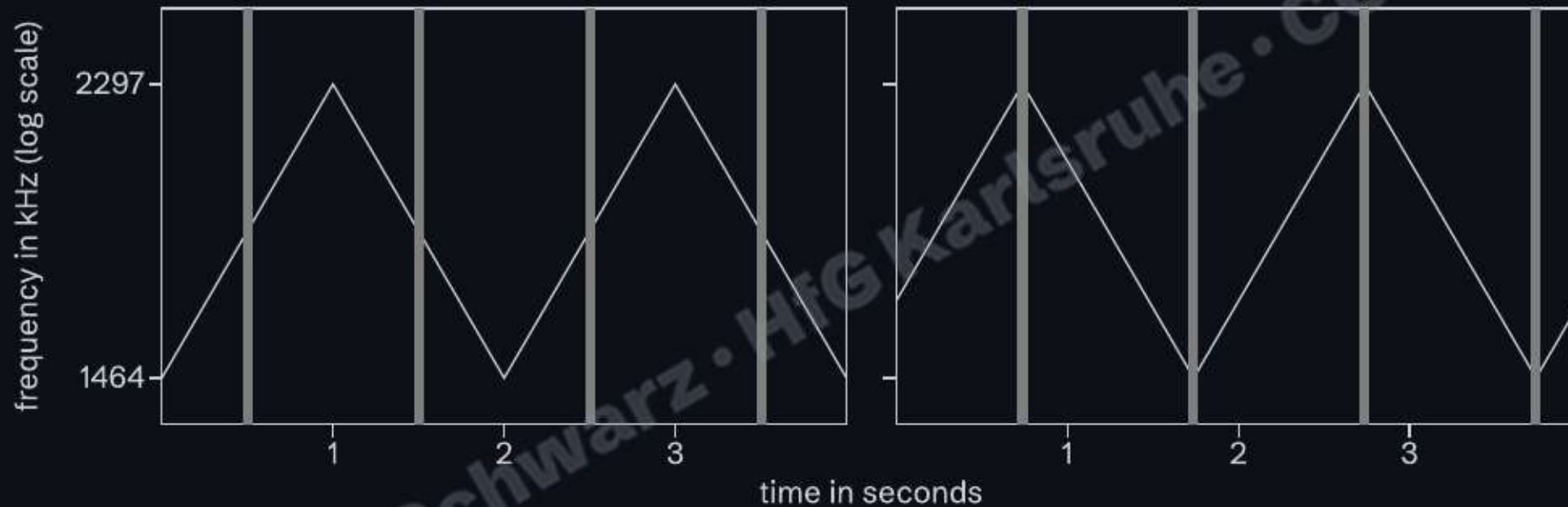- Neural cross-correlation extracts the target source

▶ 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

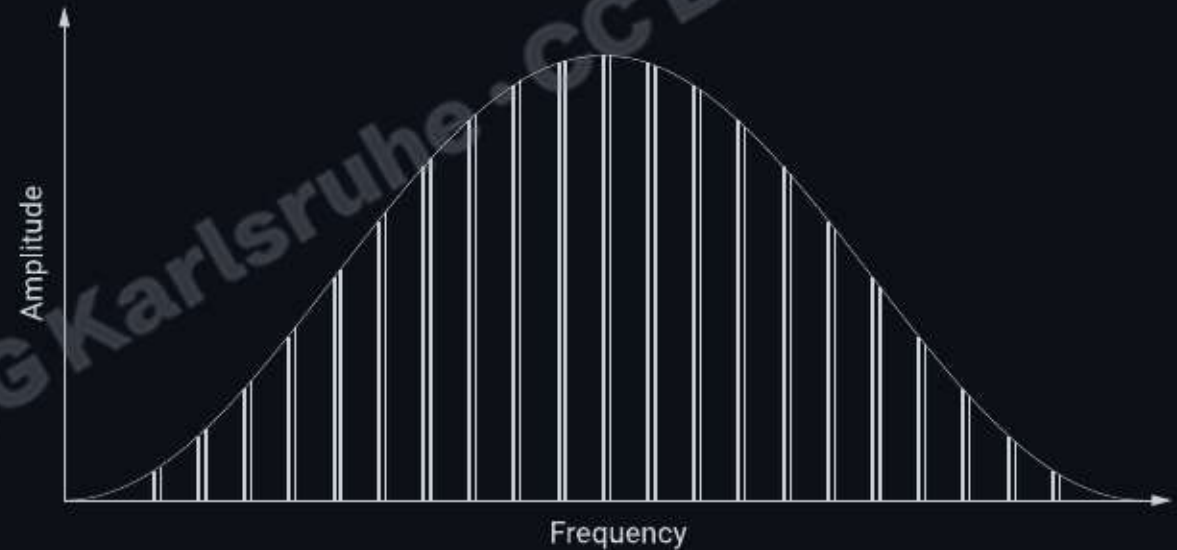▶ Perceptual continuation of a gliding tone through a noise burst

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

Copyright and Licensing

Original content: © 2025 Lorenz Schwarz
Licensed under CC BY 4.0. Attribution required for all reuse.

Includes: text, diagrams, illustrations, photos, videos, and audio.

Third-party materials: Copyright respective owners, educational use.

Contact: lschwarz@hfg-karlsruhe.de