Perceived stamp

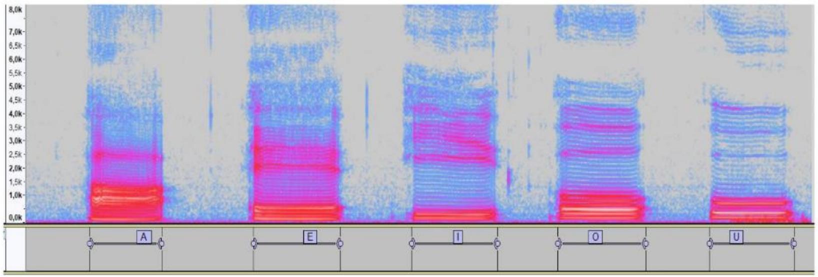
The timbre describes the quality of a sound. This is the parameter that distinguishes two sounds with the same height and volume. The main physical determinant of timbre is the waveform, that is the harmonic content of the sound (envelope, transients, and vibrato / tremolo phenomena).

The harmonic content is particularly important for timbre especially for sounds that remain constant (sustained).

Perceived stamp - The formants of the vowels

Vowels (unlike consonants) are sounds that can be supported in spoken language.

The harmonic content of the vowels is characterized by the formants - Specific energy distributions on the frequencies, which characterize each vowel.



- A: 800-1200 Hz
- E: 400-600 Hz e 2200-2600 Hz
- I: 200-400 Hz e 3000-3500 Hz
- O: 400-600 Hz
- U: 200-400 Hz

Tremolo and Vibrato

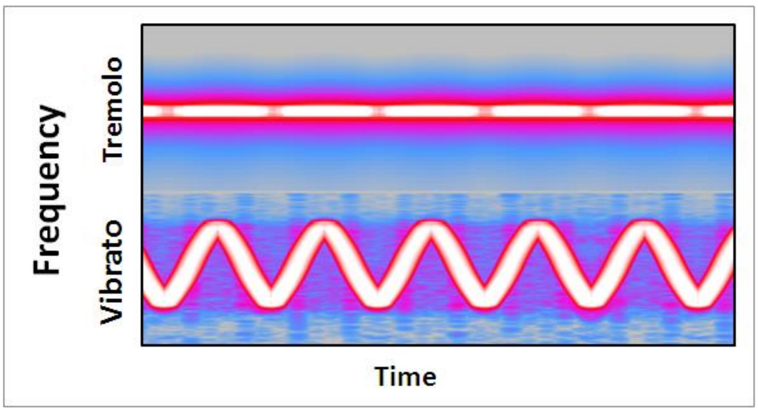
In addition to the transients with harmonic content, the fundamental contributions to the timbre can be modified by the possible presence of vibrato / tremolo

Tremolo:

- Periodic variation of the width of a note (amplitude modulation).

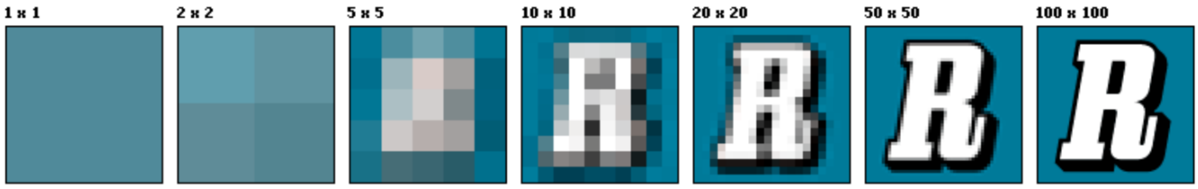
Vibrato:

- Periodic variation of the height of a note (frequency modulation)



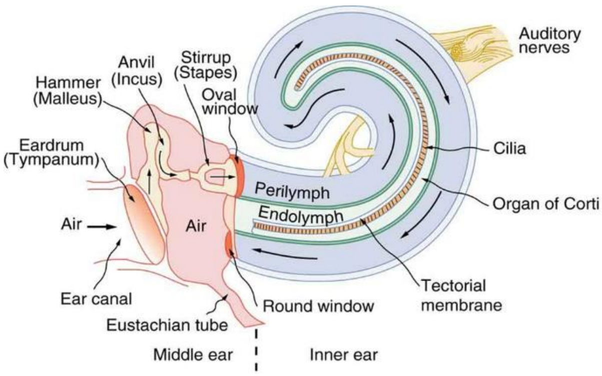
Resolution

Starting with an analogy, a known problem helps to understand the complex topic: The graphic resolution in images indicates the density of pixels. Greater resolution means that we will be able to distinguish more details.

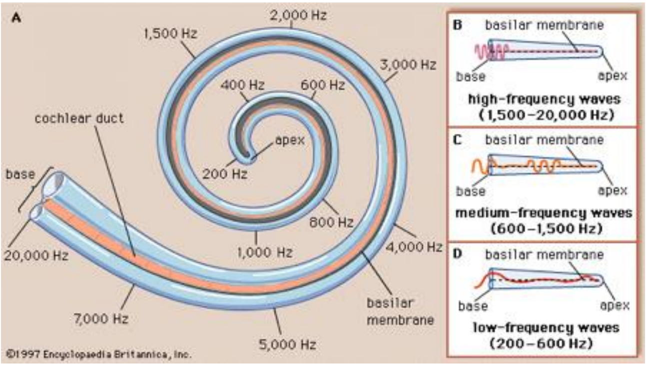


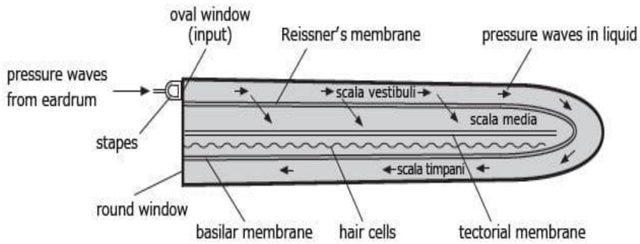
Frequency Resolution

The ear has a typical sound function. Theoretically, every area of the ear should capture a certain frequency, but the sounds that reach the organ of Corti are never completely pure. The activation area on the basilar membrane is not punctiform and therefore more frequencies fall in the same area.



The discriminating ability of the auditory system is called frequency resolution. As with image resolution, higher frequency resolution means we can distinguish more details.





The mechanics of the inner ear are designed in such a way that high tones lead directly to the beginning of the basilar membrane to resonances and thus to the excitement of nerve cells. After the resonance point, they are strongly attenuated and no longer affect the nerve cells responsible for deeper tones. However, low frequencies must first "walk along" the entire length of the basilar membrane before they cause resonance and excitement of the nerve cells and before they are attenuated. This means that even nerve cells for high and medium pitches get the bass vibrations. Medium tones must be at least as strong in the presence of low notes that they "drown out" the co-arousal through the bass.

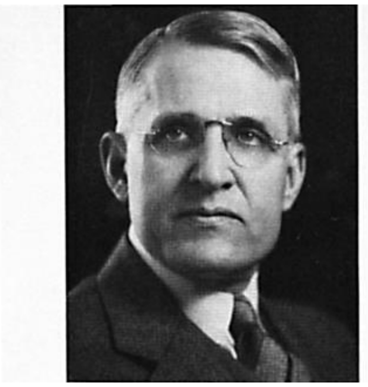
Masking and Critical Band

The **Critical Band** describes the frequency bandwidth of the "auditory filter" created by the cochlea.

In 1940 Harvey Fletcher expounded the concept of critical band, which states that the force of masking depends on the amount of noise energy within the critical band.

Approximately, the critical band is the band of audio frequencies within which a second tone interferes with the perception of the first tone by auditory masking.

The more intense the noise inside the two cutting frequencies, the stronger the masking will be.



How to calculate the bandwidth of the auditory filters?

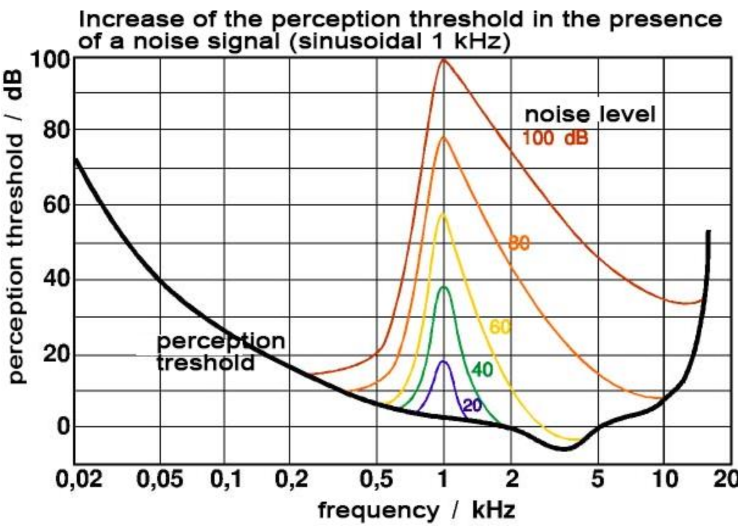
The psychoacoustic phenomenon that enables detection is called **masking**. A strong signal masks a weak signal. A similar effect occurs in radio reception when a strong station prevents the reception of a weak station.

The bandwidth with which the auditory filters work is named critical band.

Example

The hearing is not able to perceive very quiet sounds in the middle frequency range at very loud basses.

The basses mask the middle here. The minimum level from which these mids are perceived in this example depends on the level of the bass signal and the frequency spacing between bass and mid-tones.



The picture shows the effect of masking effects. For example, if a 1 kHz tone with a sound level of 80 dB is present, a 2 kHz tone of 40 dB can no longer be perceived. That is, the 2 kHz tone can be omitted without a human hearing this difference. If a 2 kHz tone of 60 dB occurs together with a 1 kHz tone of 80 dB, you can hear both tones. But you can transmit this 2 kHz sound with very poor quality: Even noises of 40 dB can no longer be perceived by humans.

Critical Band

It is the range (= interval) of frequencies within which masking phenomena occur. Sounds can be discriminated only when they fall into different critical bands.

Masking

We distinguish two types of masking:

Tonal: → Masking is done with one tone (simple or complex tone, i.e. with simpler tones)

Non-Tonal: → masking occurs with noise (broadband or narrow)

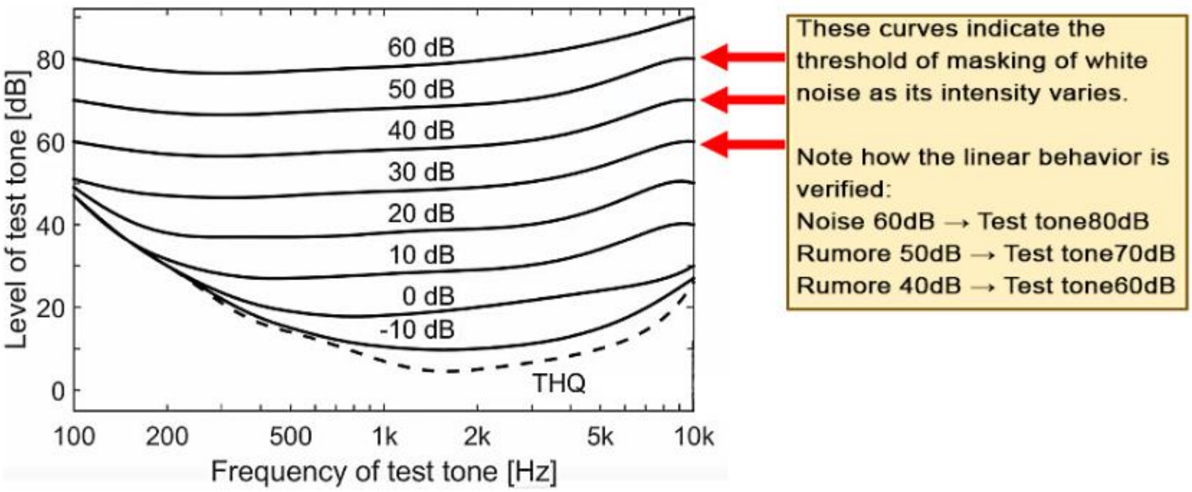
<u>Non-Tonal Masking</u>

White broadband noise

Non-tonal masking is defined as a perceptive interaction, where the masking tone turns out to be a noise, which partially or totally prevents the perception of an auditory signal.

The noise interrupts auditory communication more or less uniformly throughout the frequency spectrum.

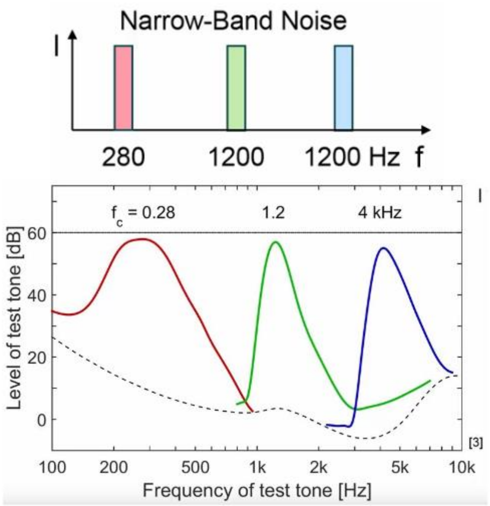
For every 10dB of increase to the amplitude of the noise, the level (in terms of amplitude) of the tone must be increased by 10dB to be audible (→ linear behaviour).

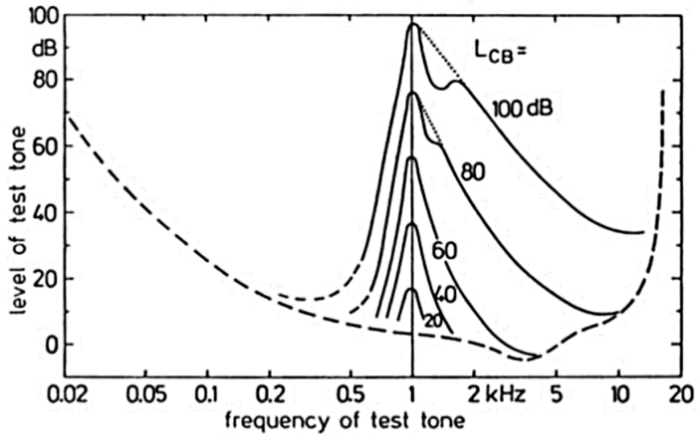


Generally, the noise guarantees a clearly stronger masking effect than a tone.

Narrow-band noise

The masking sound is a form of noise more or less narrow in which it is not possible to identify a specific tone.





1000 Hz masking sound with 200 Hz band

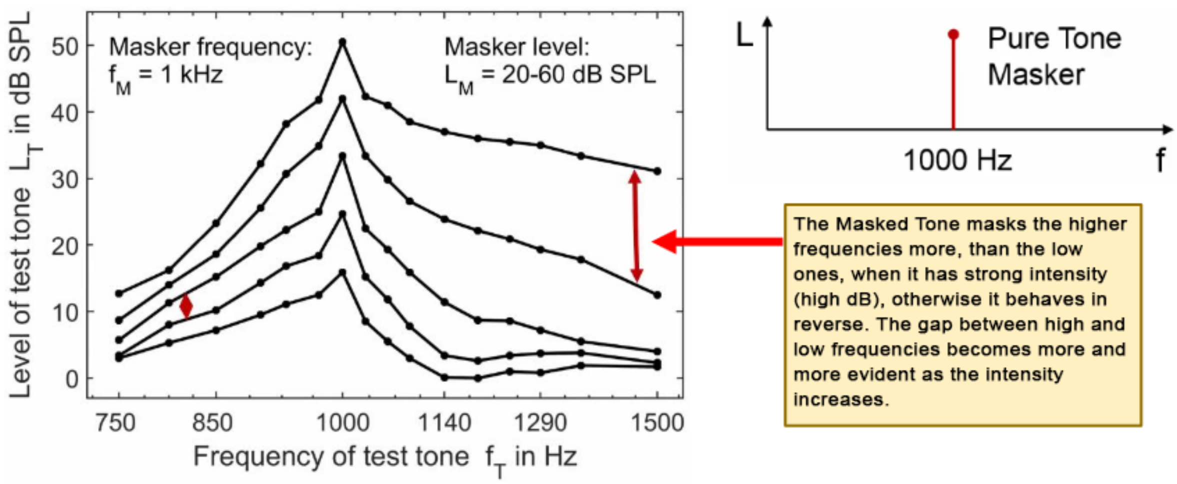
Non tonal masking imposes a higher tonal threshold but is less effective for high frequencies.

<u>**Tonal masking**</u>

Simple tone (single pure tone)

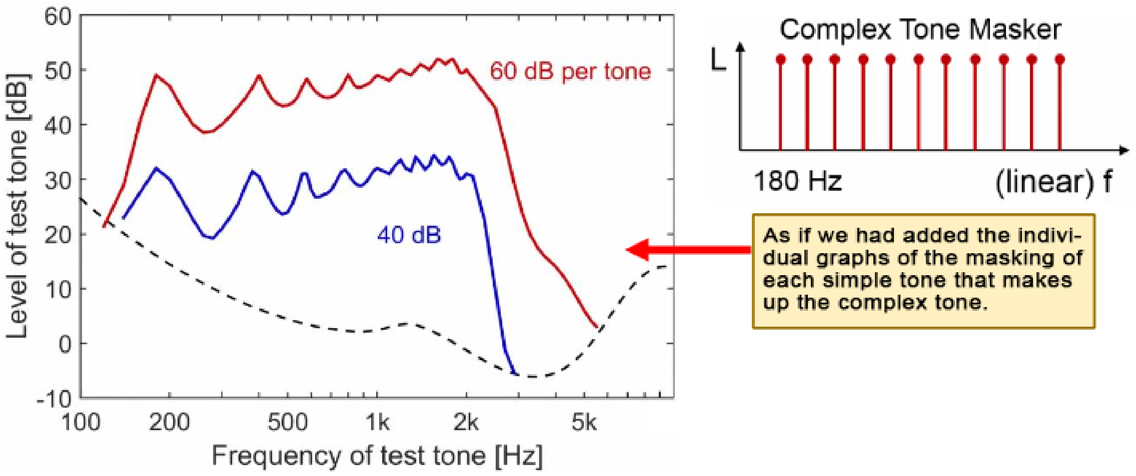
The tonal masking is a perceptible interaction between two sounds which causes them to be indistinguishable.

<u>We distinguish between</u>: Masked Tone and Masked Tone Test



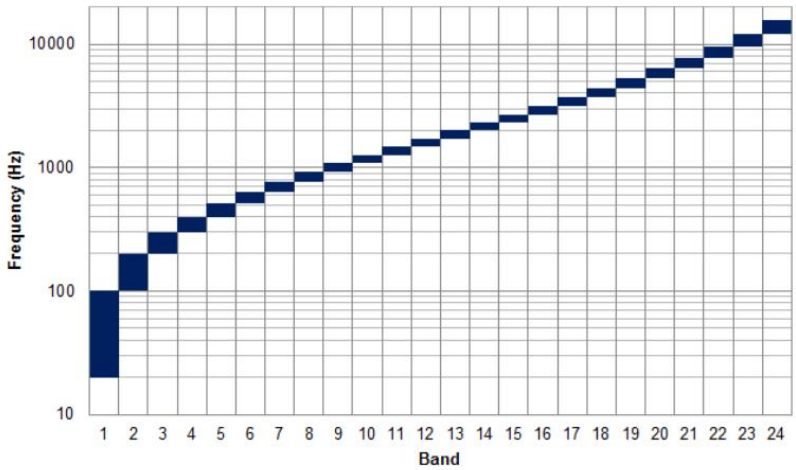
Complex tone (more simple tones)

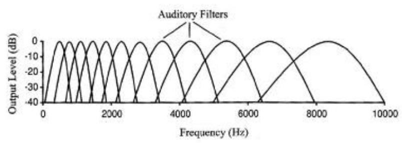
An example of a complex tone is given by the formants in the vowels

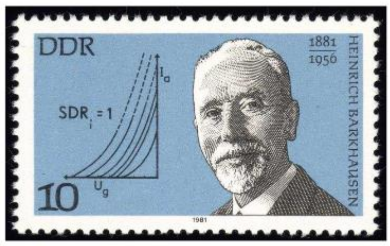


Barkhausen scale

The entire range of audible frequencies is divided into 24 critical bands (or auditorium filters, or cochlear filters). Human hearing divides the audible frequencies into approximately 24 frequency groups.







Critical bands have variable bandwidth, depending on the frequency.

- Frequency < 500Hz: Critical bandwidth is about 100Hz
- Frequency > 500Hz: Critical bandwidth is about 20% of the frequency
- Very high frequencies (> 15kHz): Critical bandwidth is around 6500Hz

Number of the band in Bark	Lower frequency treshold in Hz	Higher frequency treshold in Hz	Width of band in Hz
1	0	100	100
2	100	200	100
3	200	300	100
4	300	400	100
5	400	510	110
6	510	630	120
7	630	770	140
8	770	920	150
9	920	1080	160
10	1080	1270	190
11	1270	1480	210
12	1480	1720	240
13	1720	2000	280
14	2000	2320	320
15	2320	2700	380
16	2700	3150	450
17	3150	3700	550
18	3700	4400	700
19	4400	5300	900
20	5300	6400	1100
21	6400	7700	1300
22	7700	9500	1800
23	9500	12000	2500
24	12000	15500	3500

This classification is described by the Bark scale, on which a frequency group has just a width of 1 bark.

Example:

What is its critical band for a 2kHz tone?

We generate noise composed of a series of frequencies in a range of 2 kHz and variable radius.

This has a variable band around the 2 kHz tone.

Variations in the sound intensity of the original sound can only be perceived with sounds with a bandwidth of less than 250 Hz

Therefore, the critical bandwidth of the 2 kHz signal is 250 Hz.

Masking threshold

The masking threshold of a masked tone is the intensity level at which it can be heard even in the presence of a masking tone.

The amount of energy masked is the difference between the masking threshold and the absolute threshold of audibility.

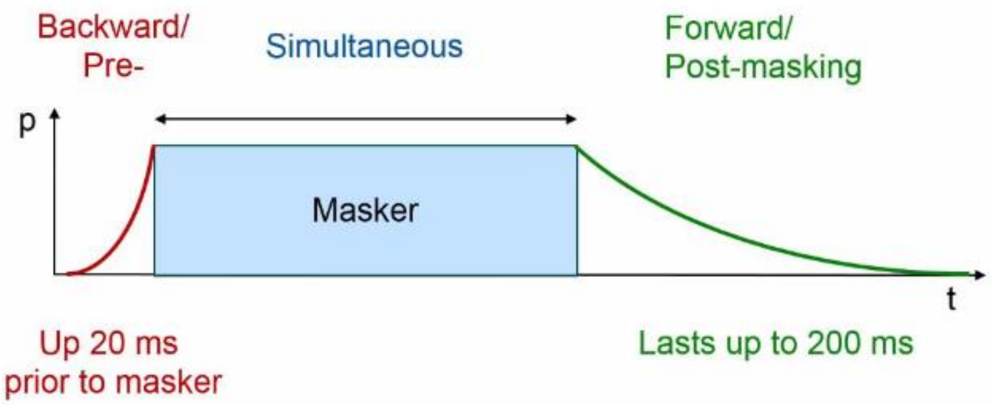
Masking

In the previous examples, we assumed that the masked tone and masking tone were placed at the same moment.

<u>Simultaneous masking</u>: Simultaneous masking occurs when a sound is made inaudible by a noise or unwanted sound of the same duration as the original sound.

<u>There are also phenomena of Temporal Masking</u>: Temporal masking or non-simultaneous masking occurs when a sudden stimulus sound makes inaudible other sounds which are present immediately preceding or following the stimulus. Masking which obscures a sound immediately preceding the masker is called *backward masking* or *pre-masking* and masking which obscures a sound immediately following the masker is called *forward masking* or *post-masking*.

Masking can be used to compress the audio signal, eliminating parts of the signal that would not be perceived by our auditory system.



Each filter needs a certain amount of time to settle. This time is inversely proportional to the bandwidth of the filter.

The frequency group filters of the hearing with a minimum bandwidth of 100 Hz thus require a maximum of 1/100 Hz = 10 ms for settling. That is: only 10 ms after the onset of a sound signal, the sound impression is stable, only then can the hearing rely on the information.