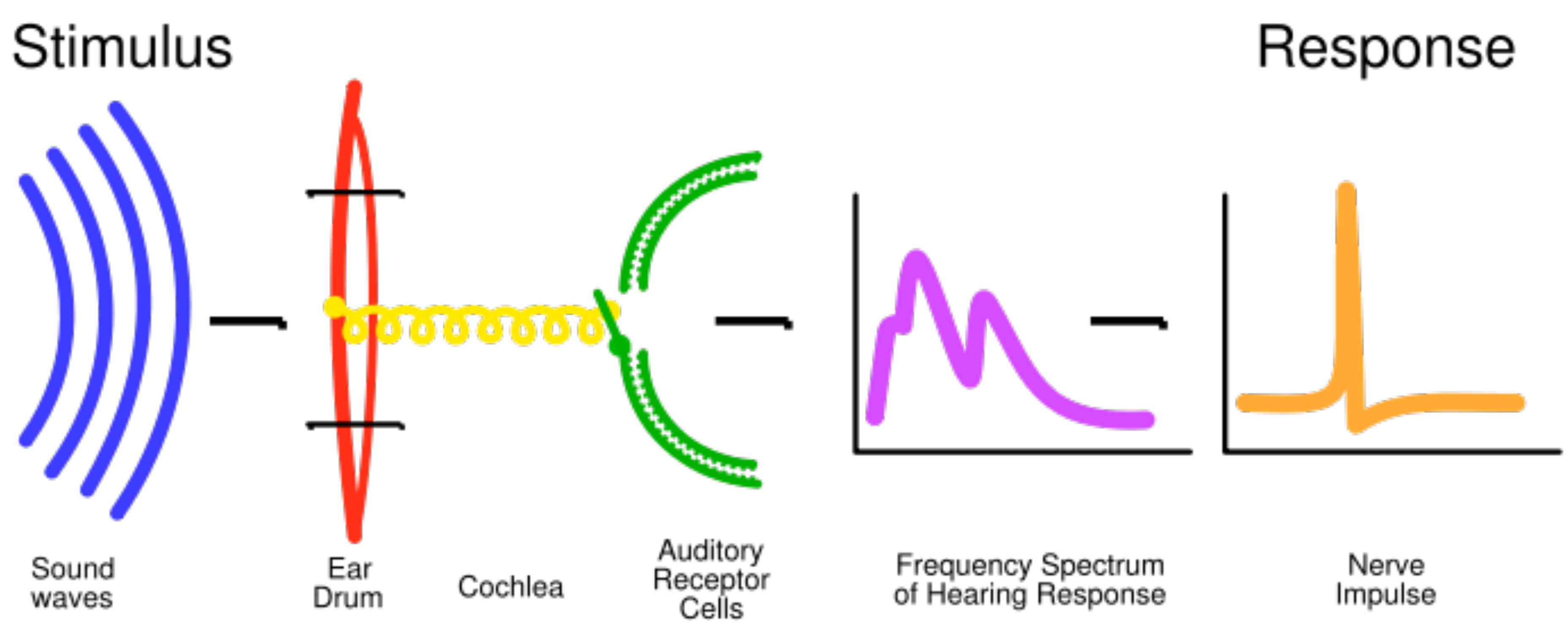
BSc Degree in Physics

Sound perception: human ear, loudness, pitch & timbre

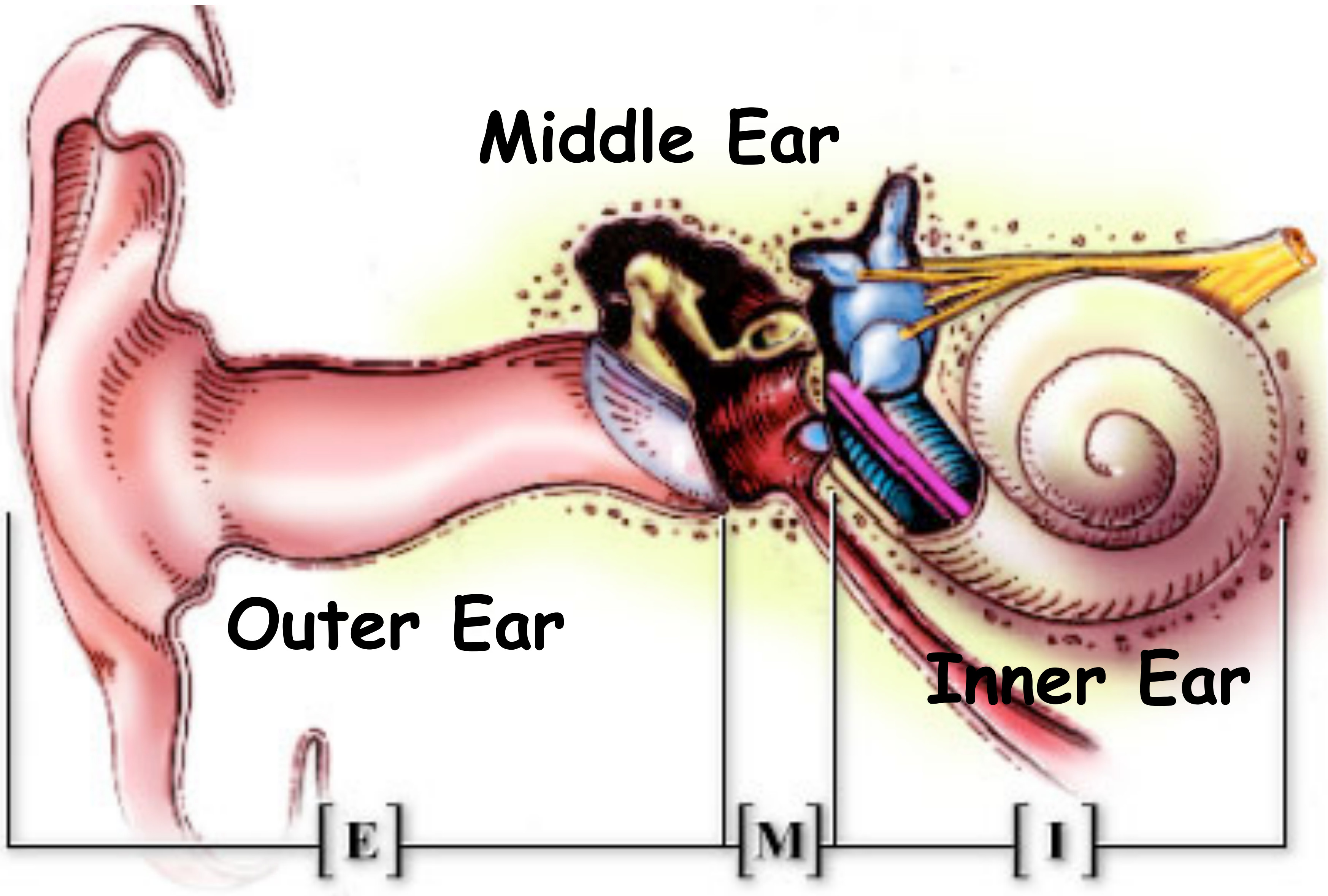
Fabio Romanelli
Department of Mathematics & Geosciences
University of Trieste
Email: romanel@units.it

The human ear

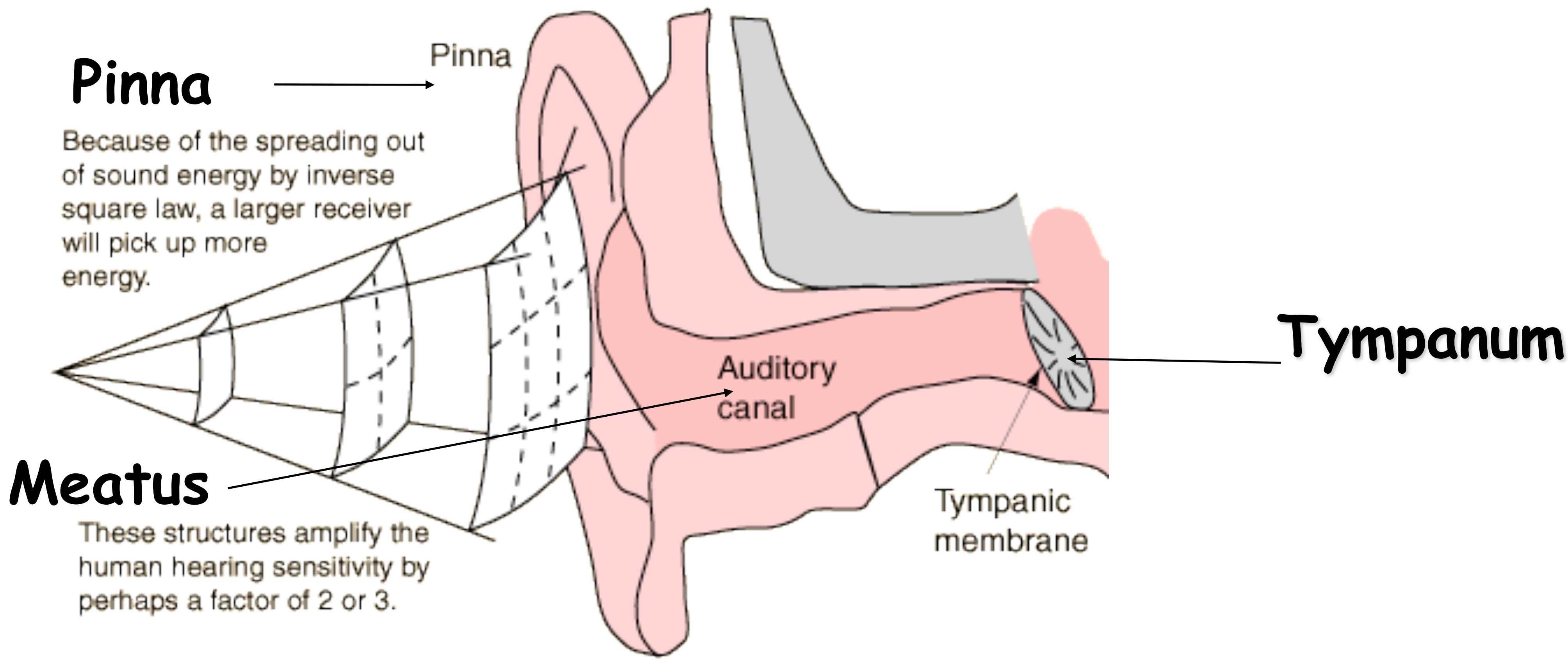
The human ear is a highly sensitive sound receptor in which **pressure fluctuations** in the outer ear are transformed into **vibrations of small bones** (the ossicles) in the middle ear that are ultimately communicated to the cochlea located in the inner ear, where the vibrations are further transformed by **stereocilia** (hair cells) into **neural impulses** distributed by frequency.



Anatomy of ear



Outer ear

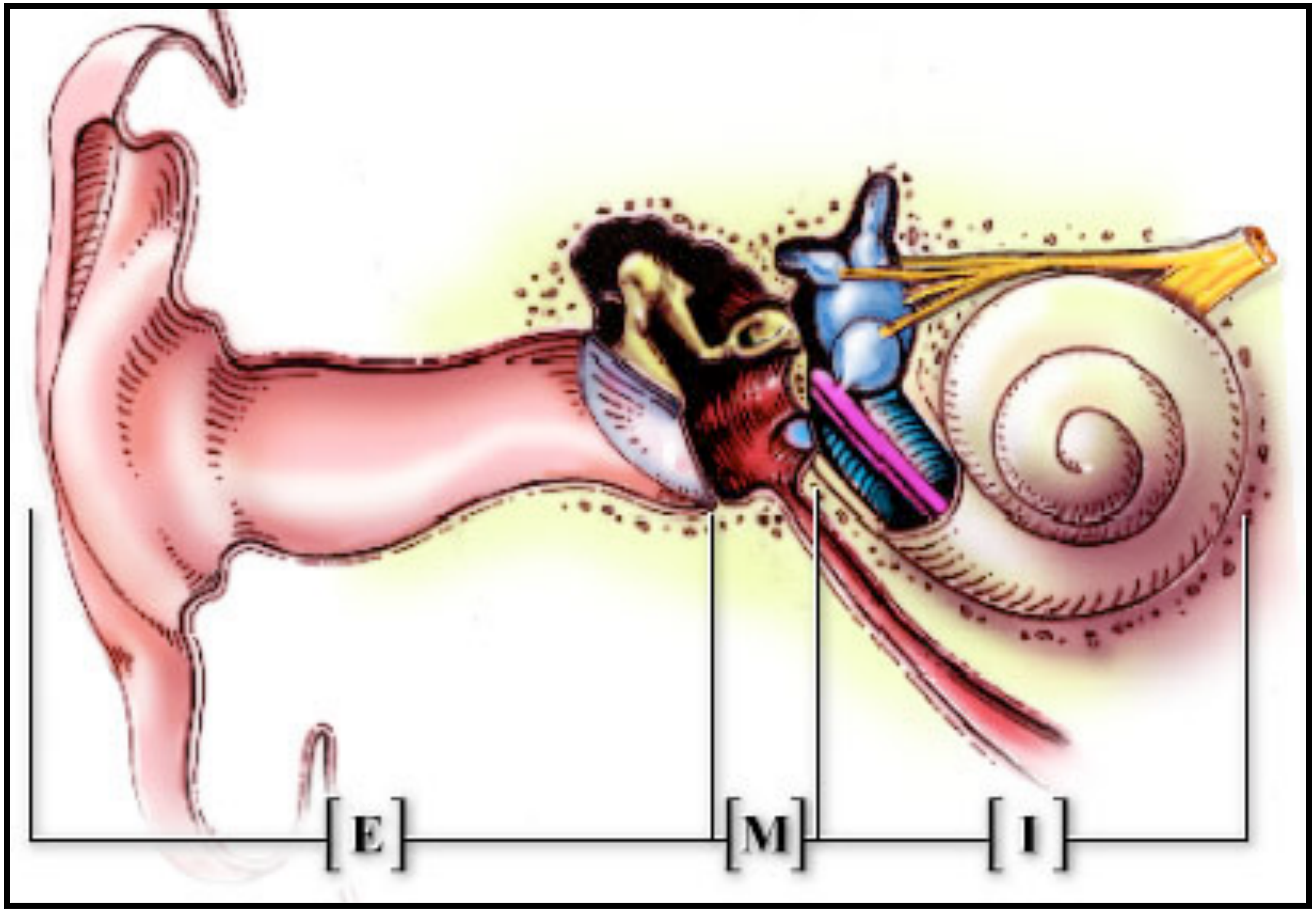


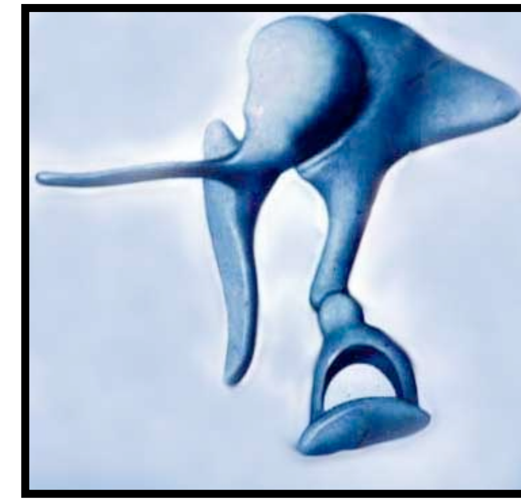
Pinna – (the feather) matches ear canal to outside world.

Meatus - (the passageway) conducts sound into head.

Tympanum –(the drum) transforms pressure fluctuations into displacement.

Middle ear: the ossicles



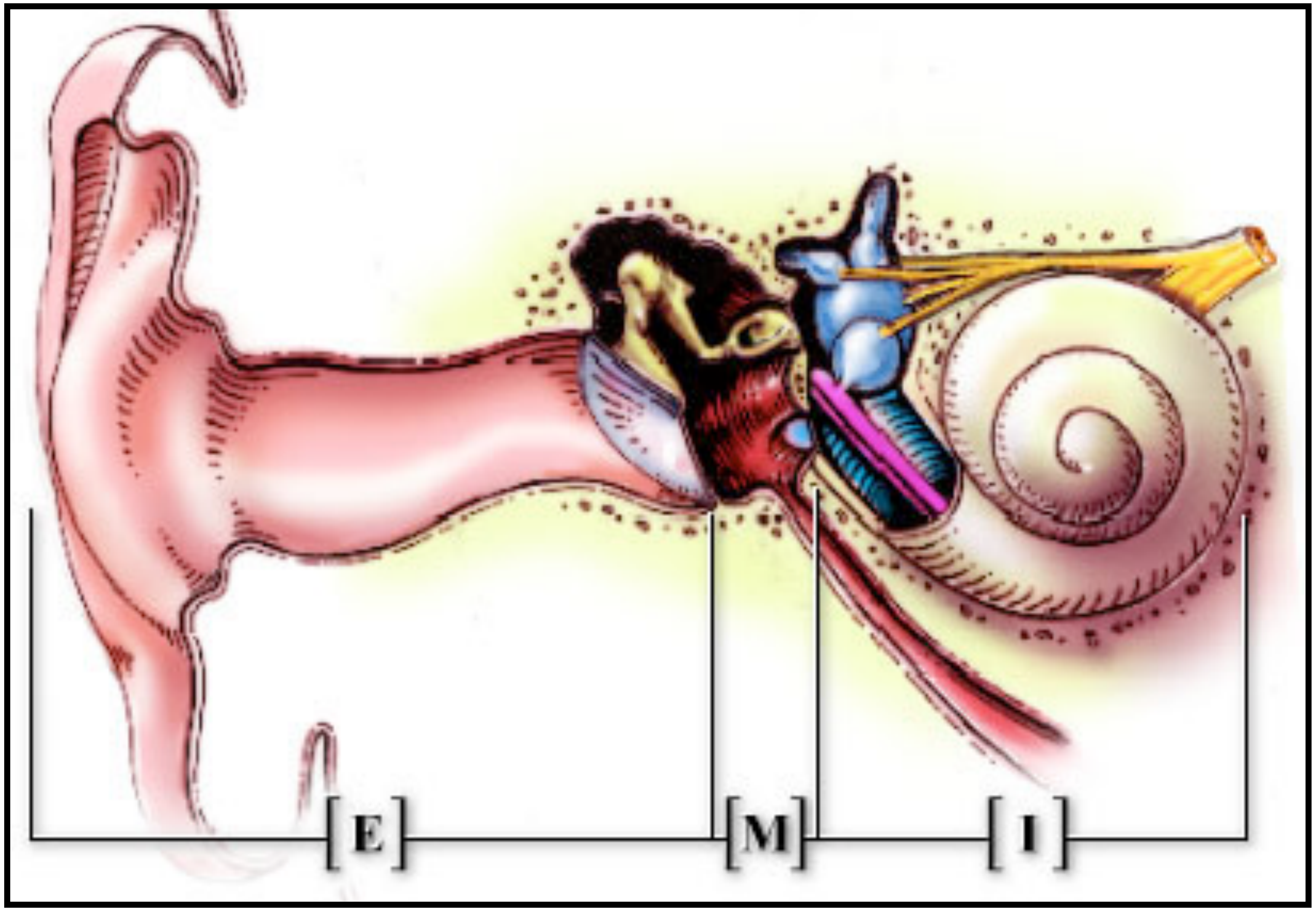


Malleus — (the hammer) moved by Tympanium.

Incus — (the anvil) supported by ligaments that protect against loud percussion.

Stapes — (the stirrup) force multiplied by 1.3 because of lever action.

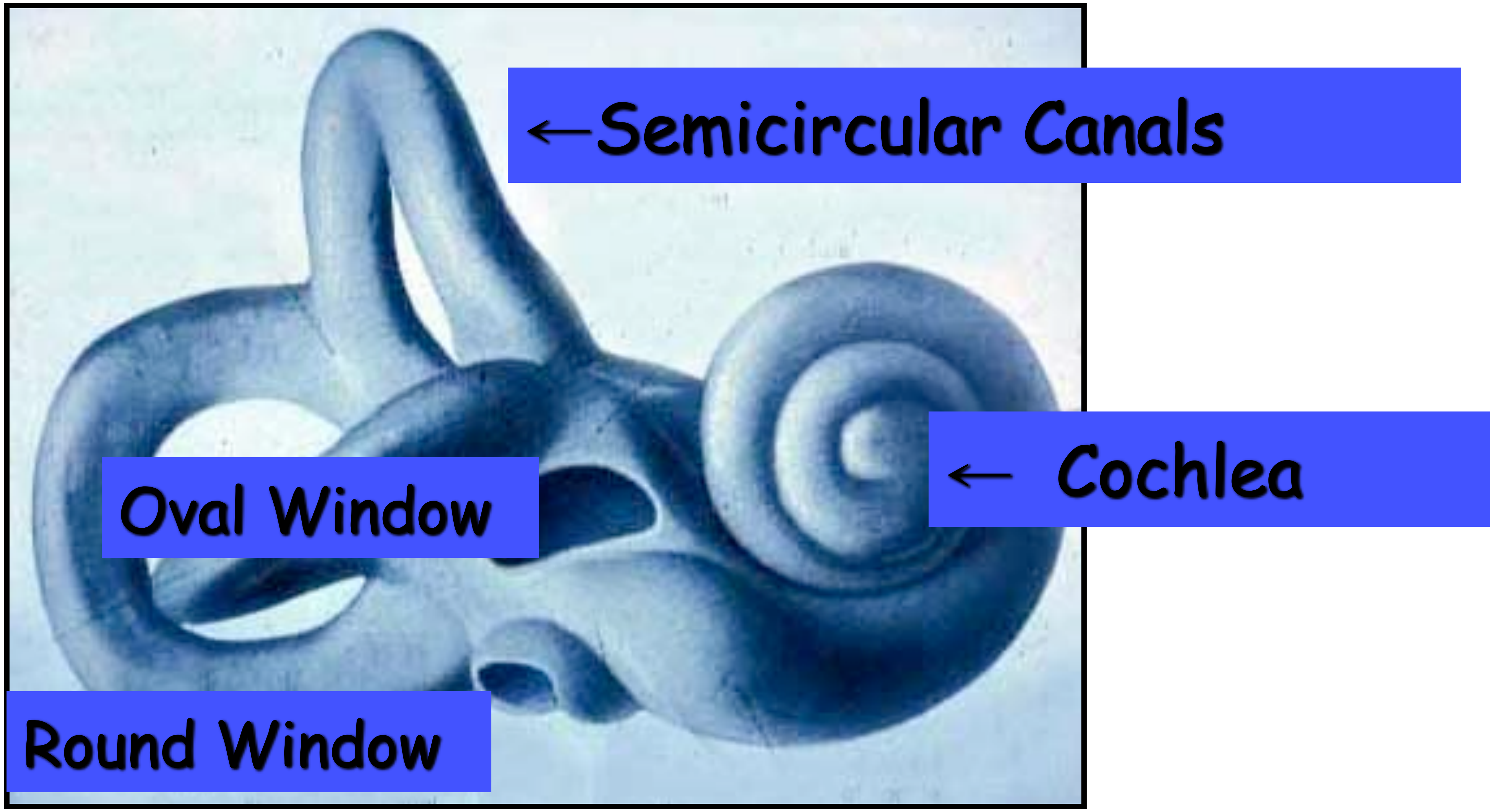
Inner ear: the cochlea



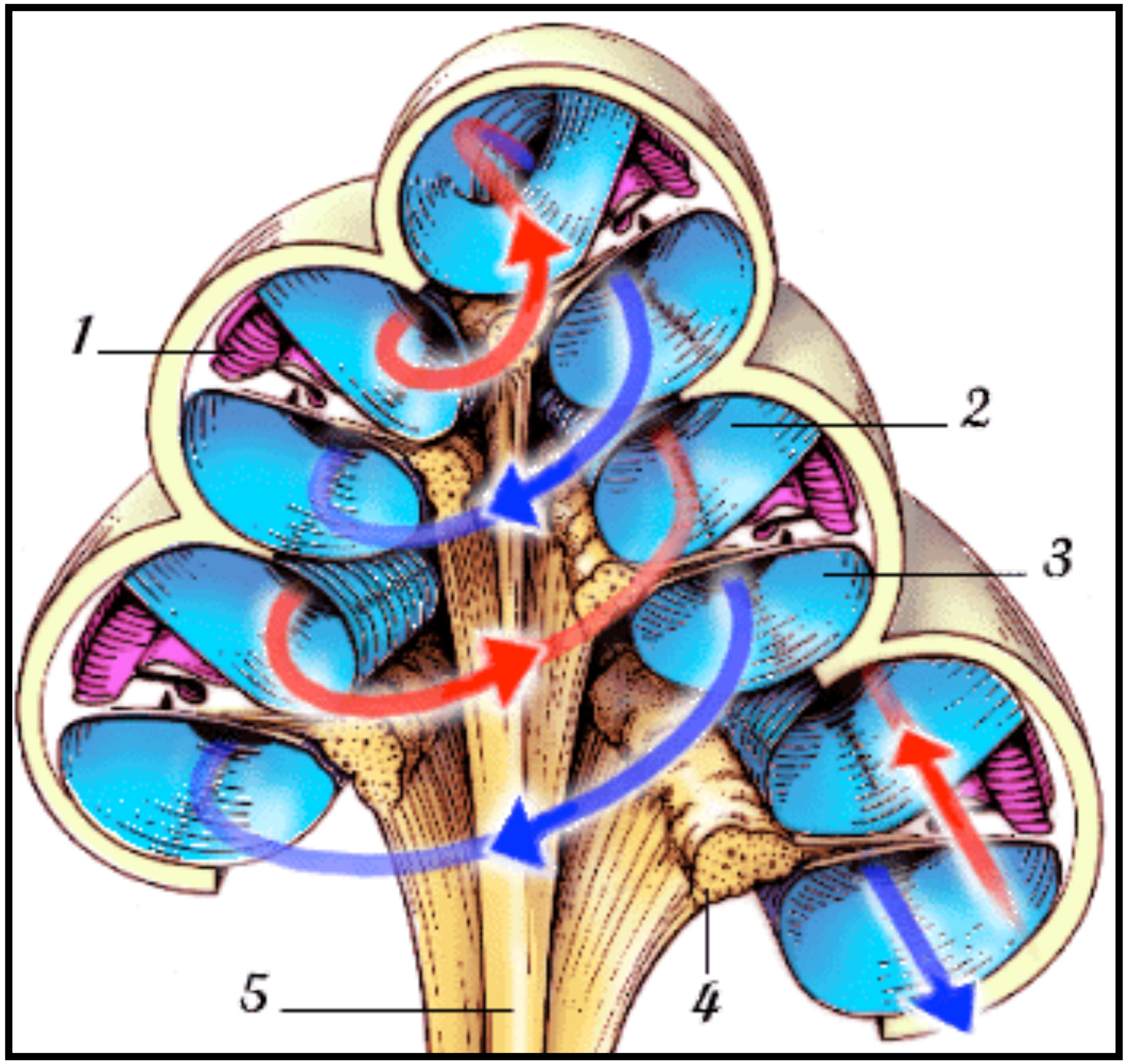
Cochlea — (the snail) converts displacements into neural impulses

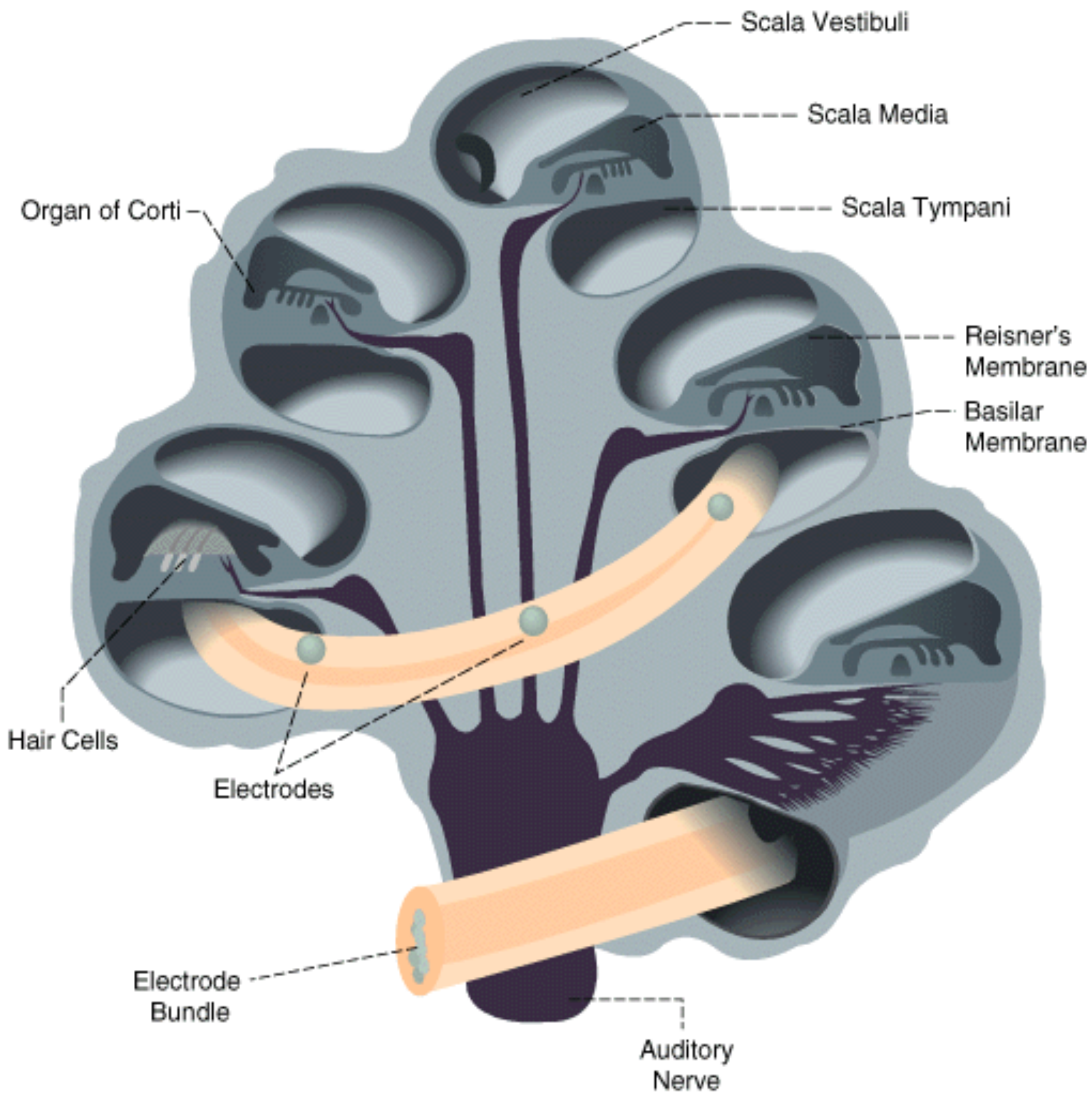
Auditory nerve — neural impulses to brain

Semicircular canals — detect motion and orientation


←Semicircular Canals
← Cochlea
Oval Window
Round Window

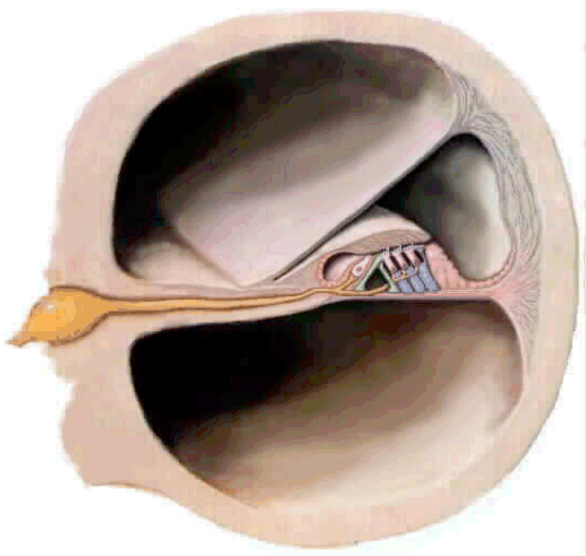
Structure of cochlea

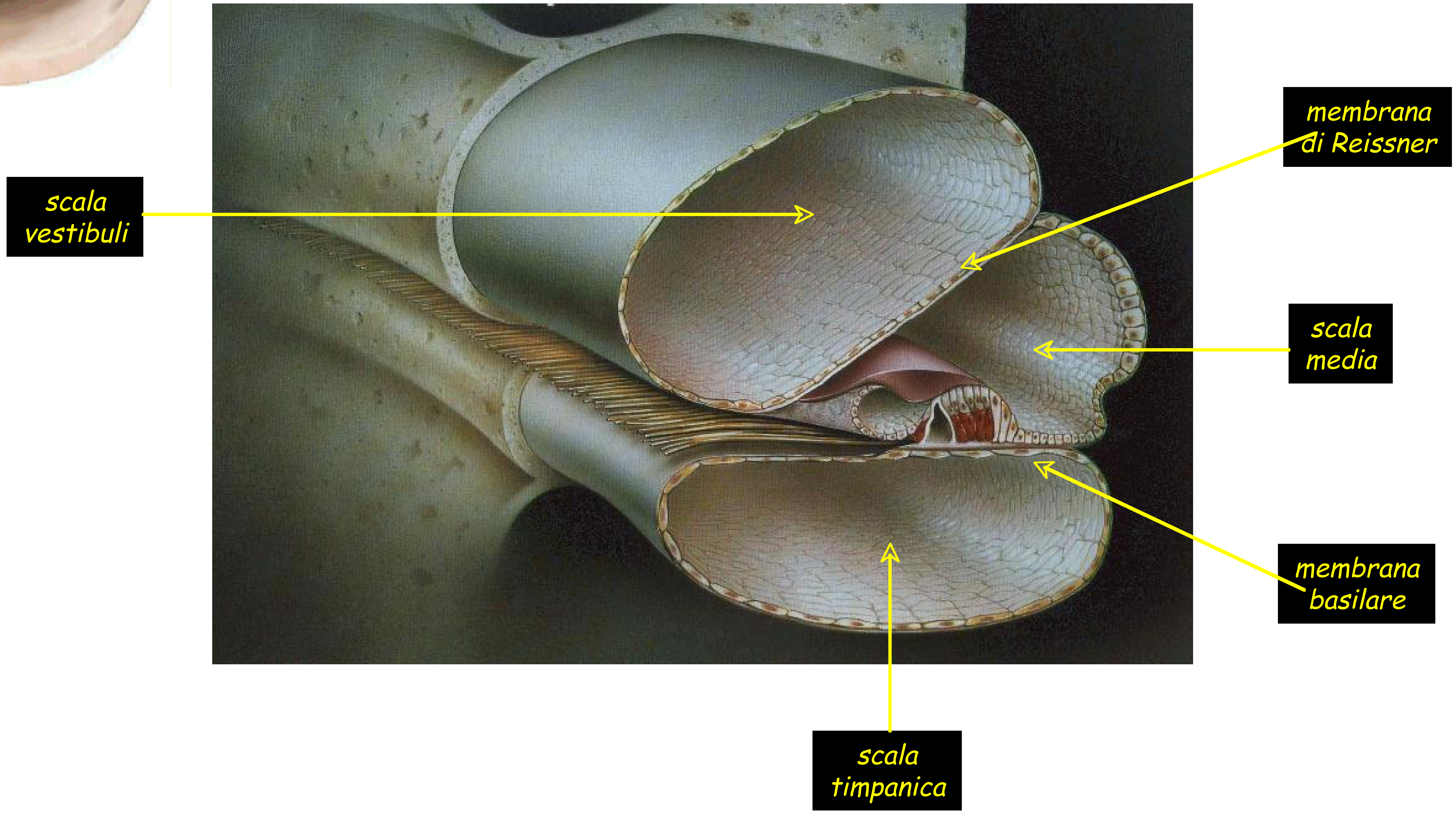




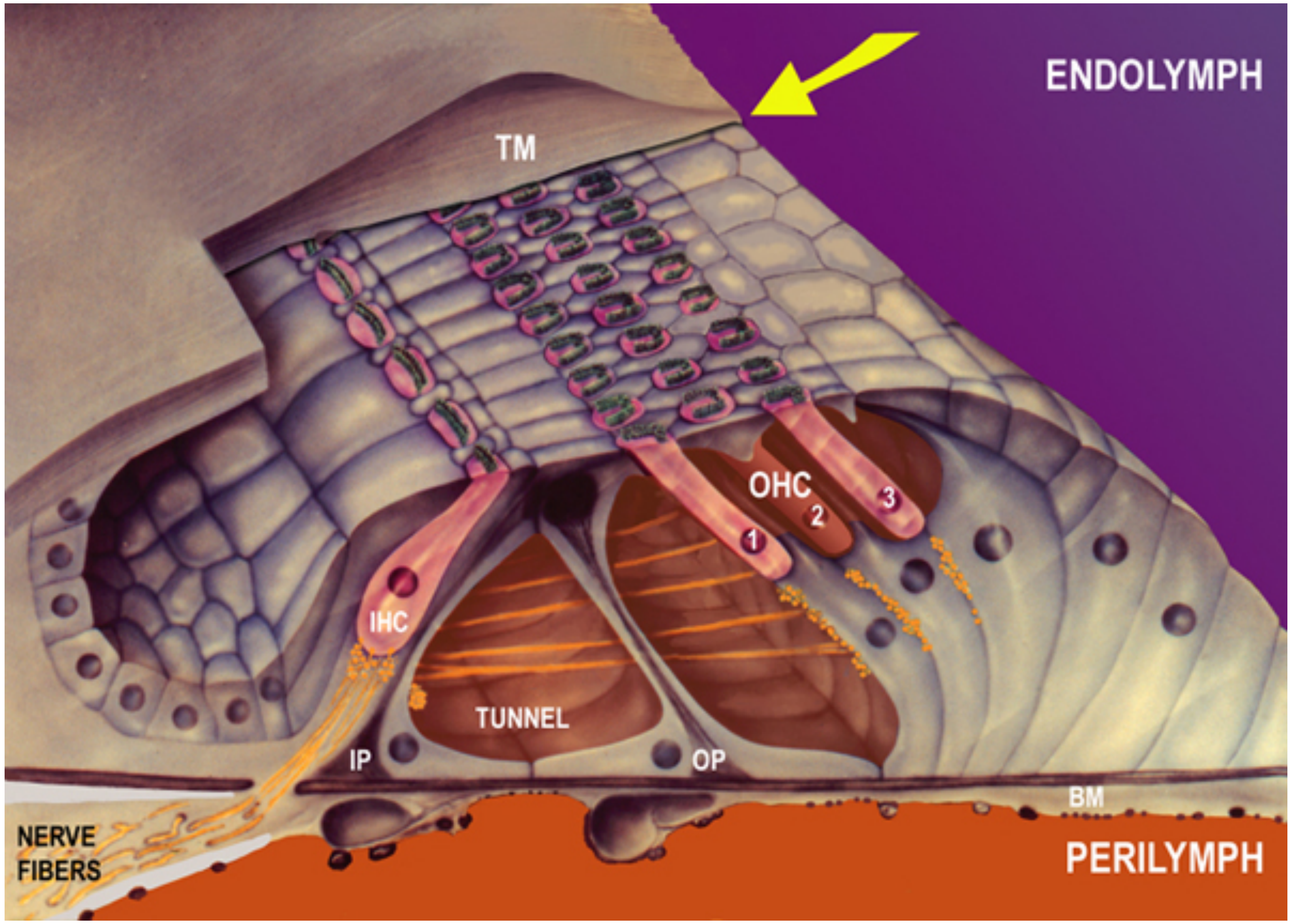
Cochlear duct (1) which contains **endolymph**, the **scala vestibuli** (2) and **scala tympani** (3) which contain **perilymph**. The red arrow is from the oval window, the blue arrow points to the round window. Within the modiolus, the **spiral ganglion** (4) and **auditory nerve fibres** (5) are seen.

Microstructure of cochlea



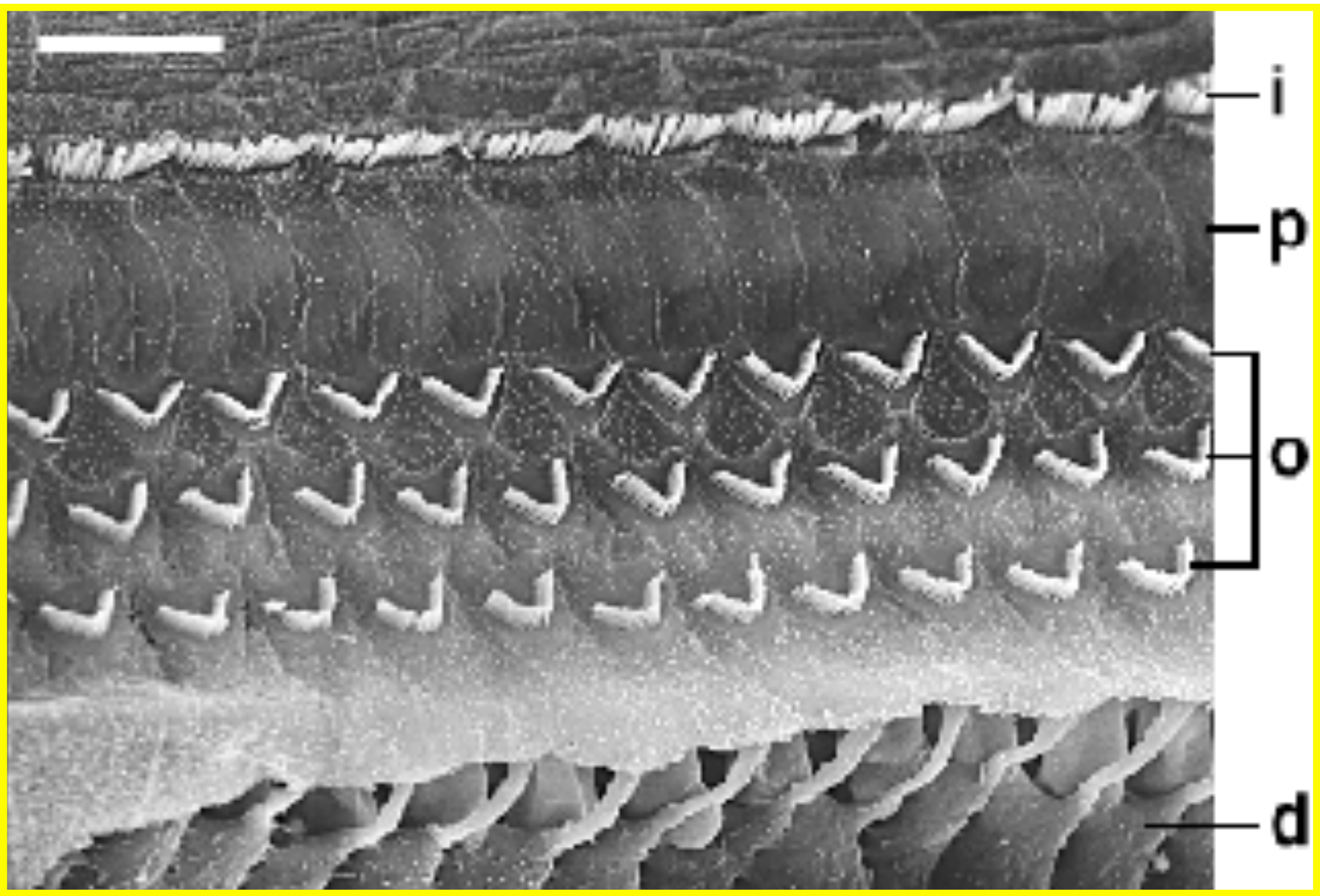


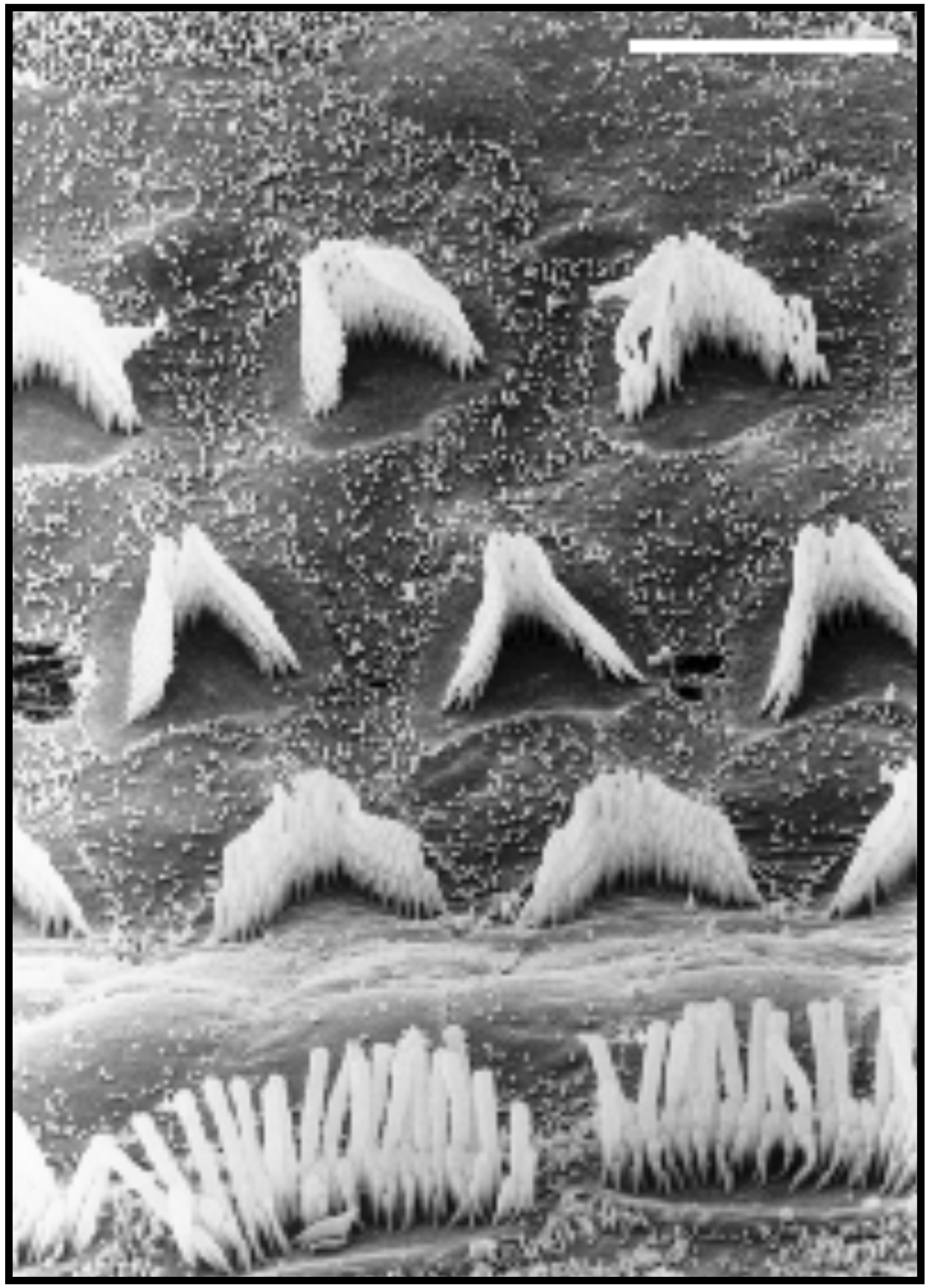
Organ of Corti and Basilar Membrane



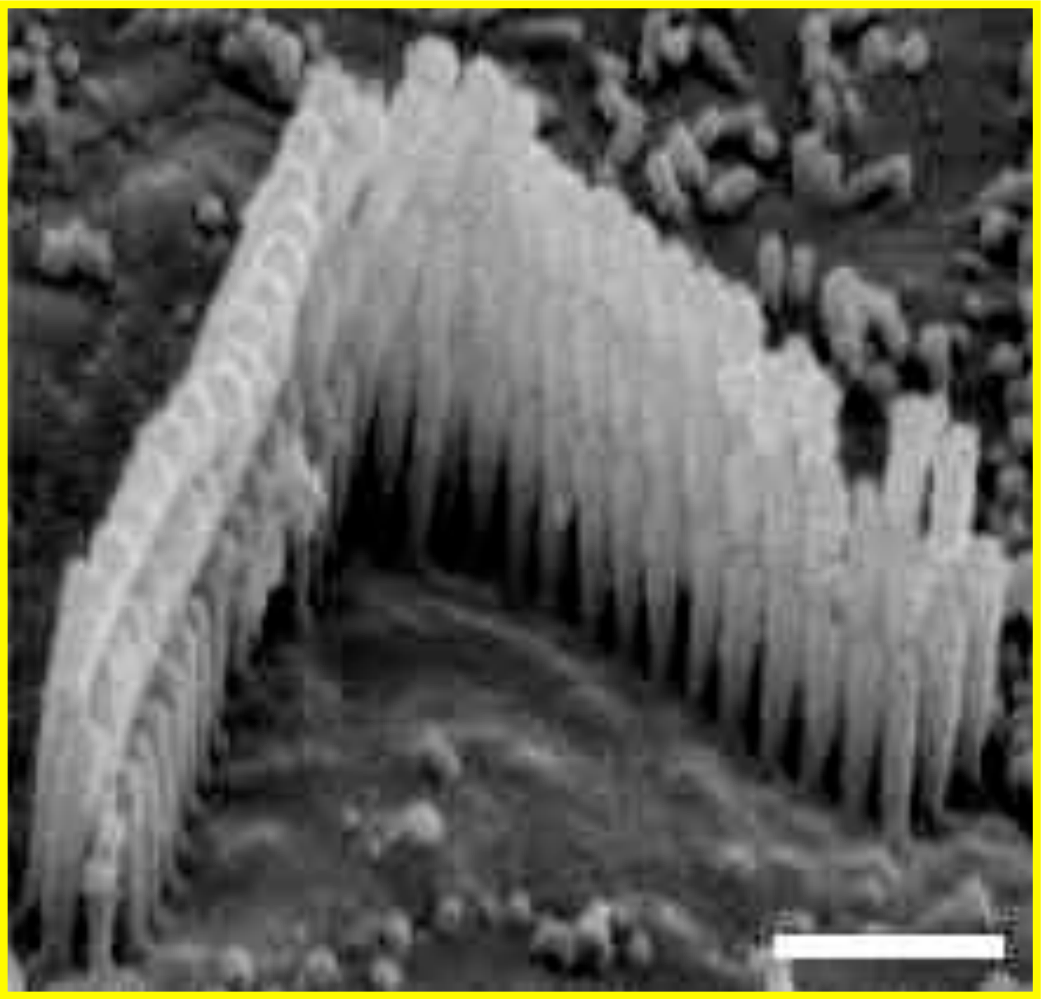
This drawing illustrates the normal organ of Corti (i.e., hearing organ) in cross-section. The organ is attached to the vibratory basilar membrane (**BM**) and is surrounded by fluid spaces that contain endolymph (purple) and perilymph (orange). The tectorial membrane (**TM**) covers the surface (i.e., reticular lamina) of the organ of Corti and is closely applied to supporting cells laterally (yellow arrow). The organ of Corti contains two types of sensory cells (pink) [a single row of inner hair cells (**IHC**) & three rows of outer hair cells (**OHC 1, 2, 3**)], supporting cells (e.g., **IP**; **OP** - blue-purple), nerve fibers (yellow-orange) and fluid spaces (e.g., tunnel - transparent orange). Elongated microvilli (i.e., stereocilia - green) project into the TM from the apical surface of each OHC. The IP, OP and three rows of Deiters' cells (one below each OHC) contain parallel bundles of microtubules that maintain the shape of the organ of Corti and keep the stereocilia in contact with the TM. Ordinarily, the reticular lamina provides a barrier to prevent the intermixing of endolymph and the fluid (i.e., Cortilymph) within the organ of Corti.

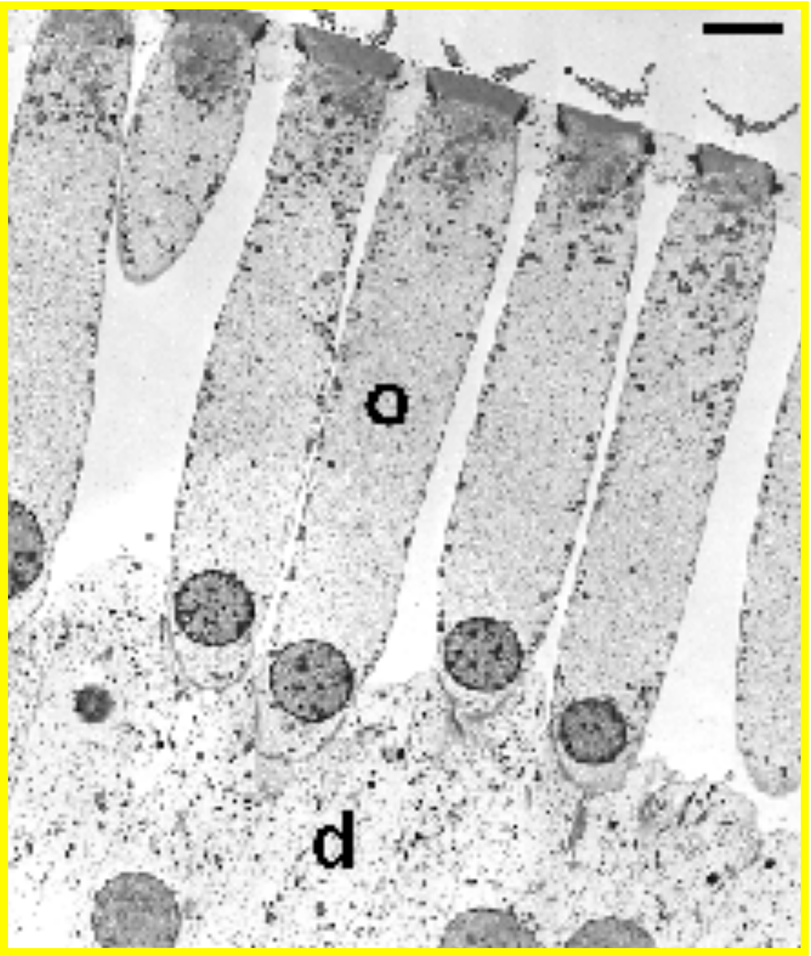
Hair cells

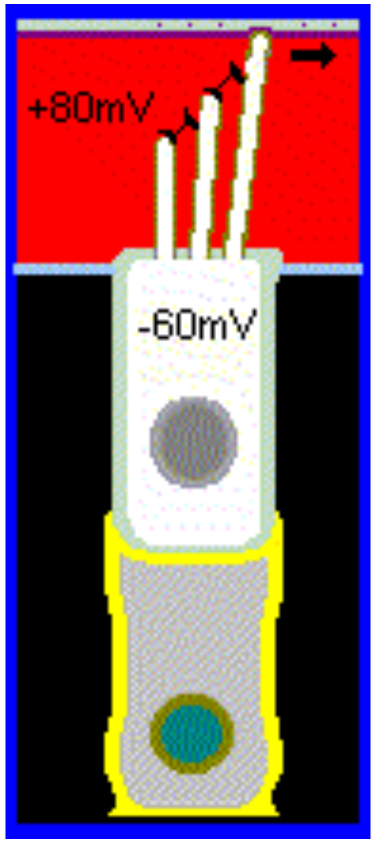




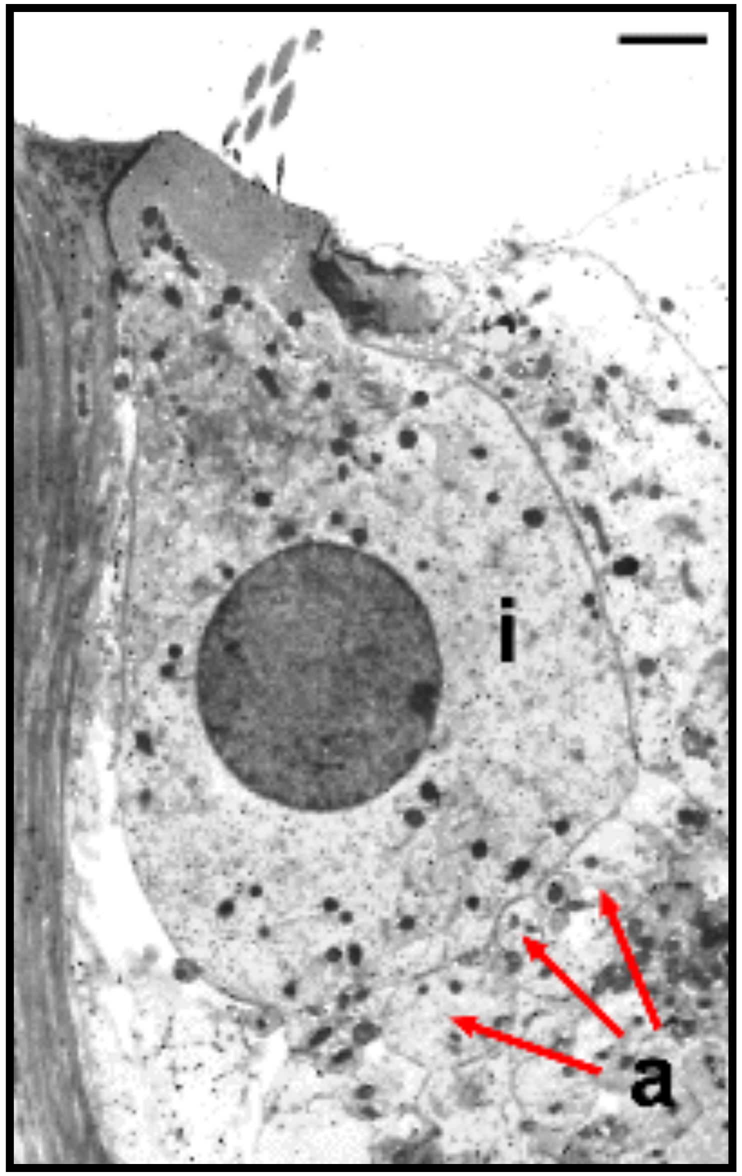
Outer hair cells

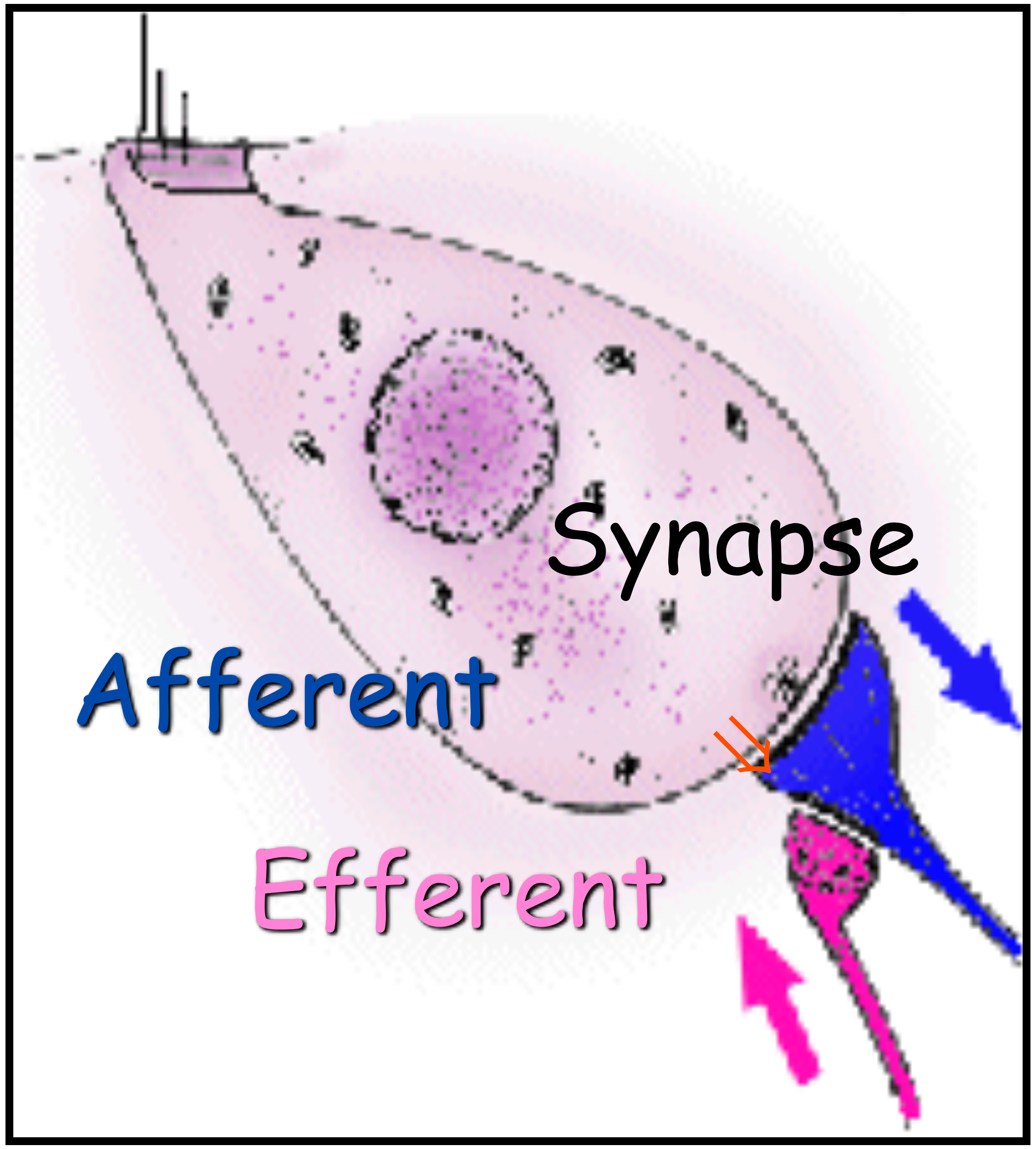


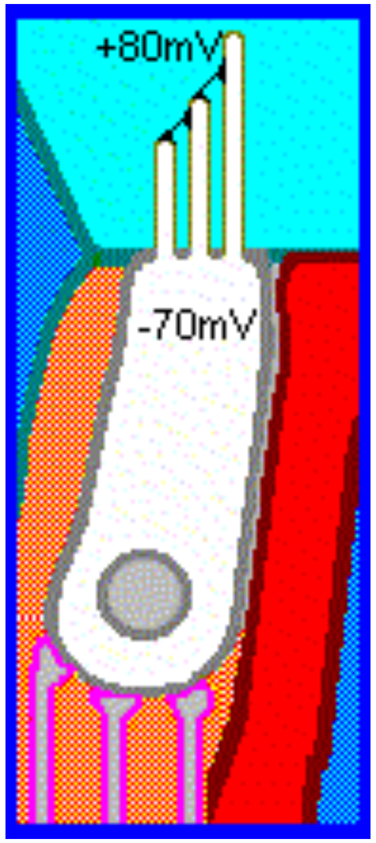




Inner hair cells

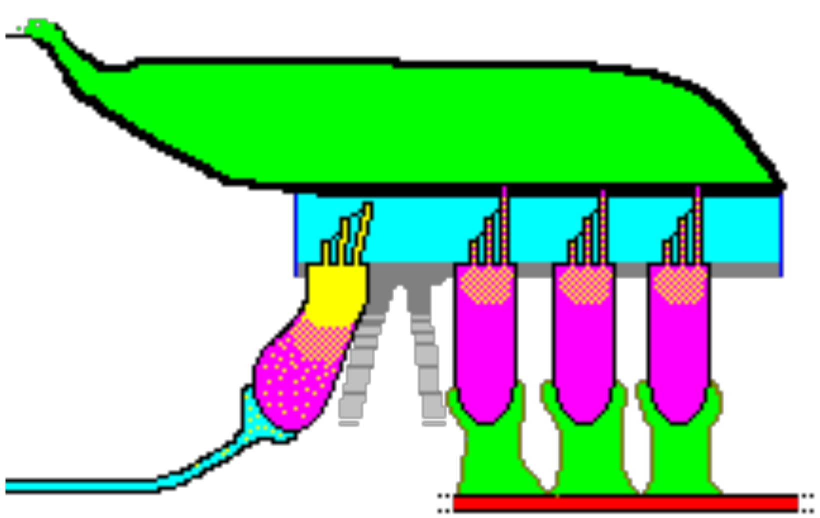


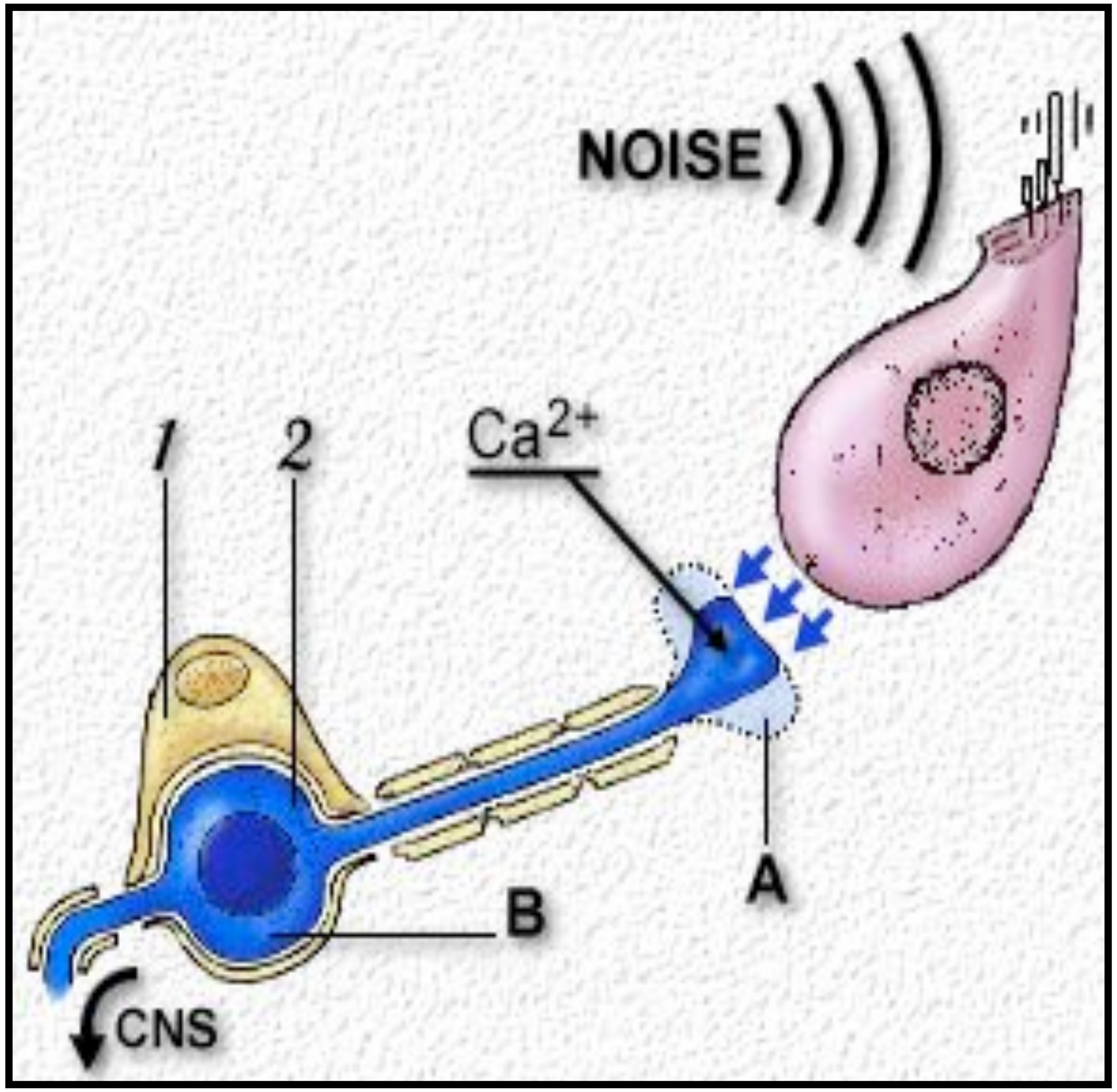




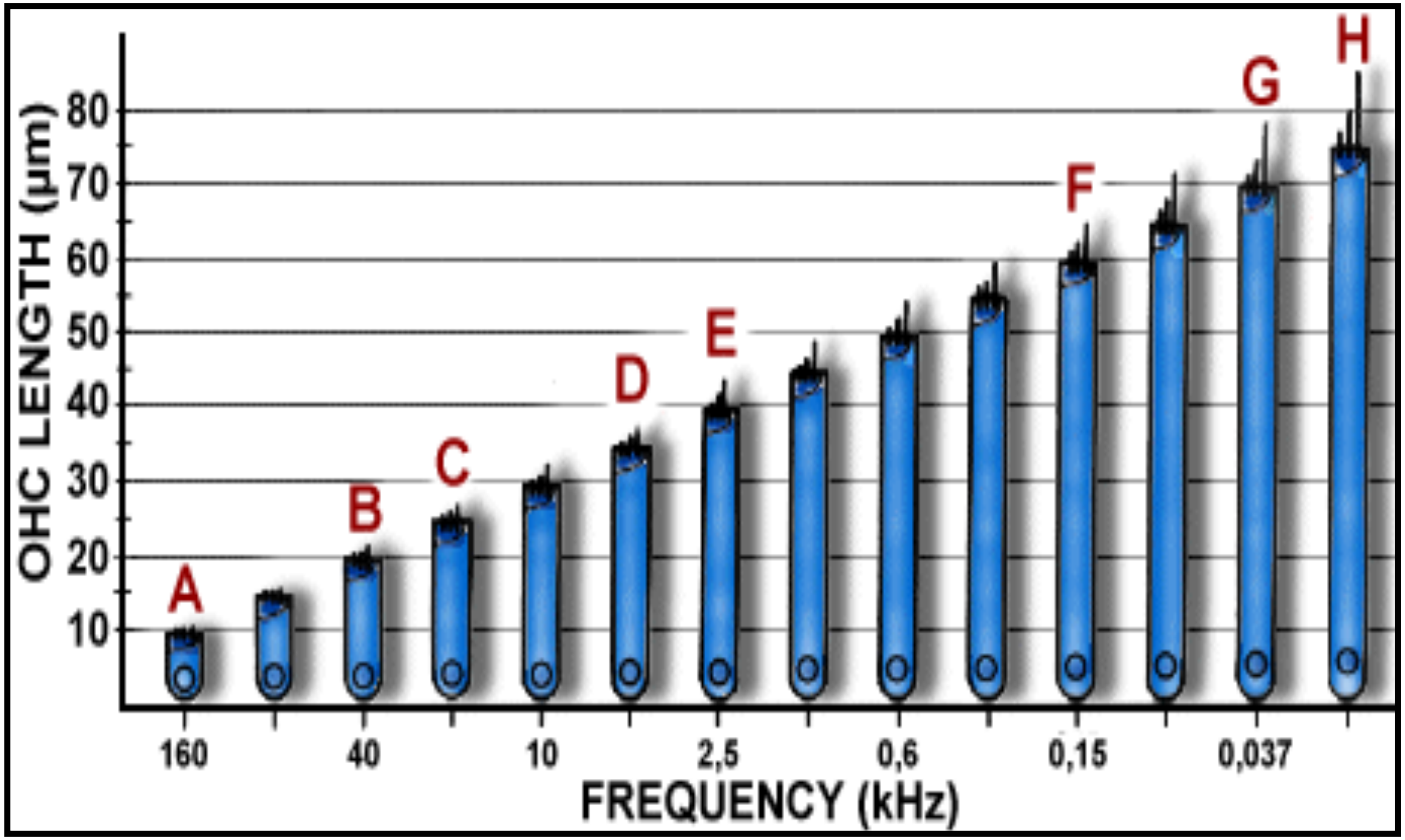
Function of stereocilia

Stimulation causes neurotransmitter to stimulate neuron in Auditory Nerve





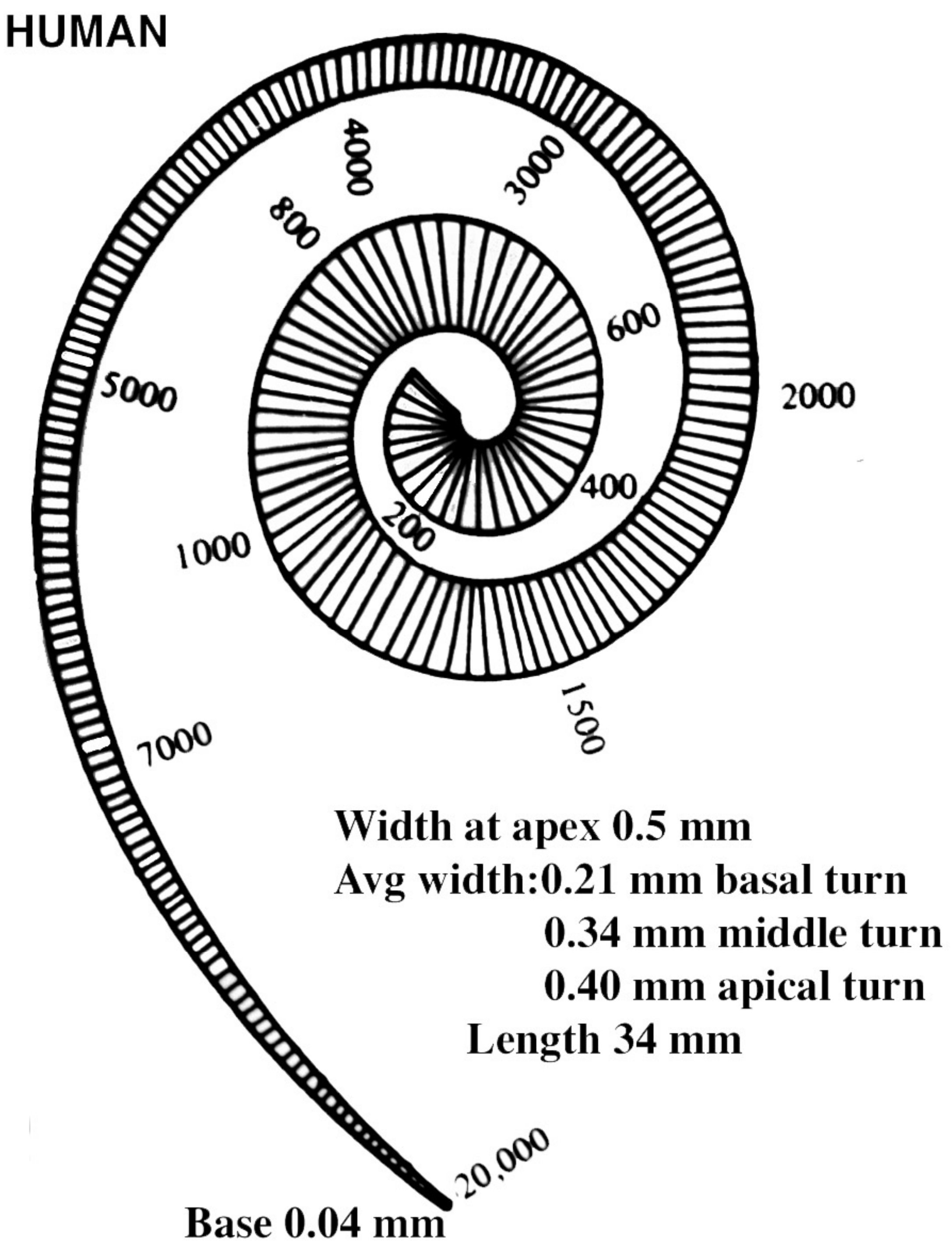
Frequency Response of Hair Cells



- 20 Hz to 20 kHz (typical in Humans)
- Resonances in Basilar membrane and in HC cause spatial separation by frequency.
- Differential movement of membranes stimulate HC.
- Minimum stimulation required for response.
- Inhibition of neighbors causes non-linear response.

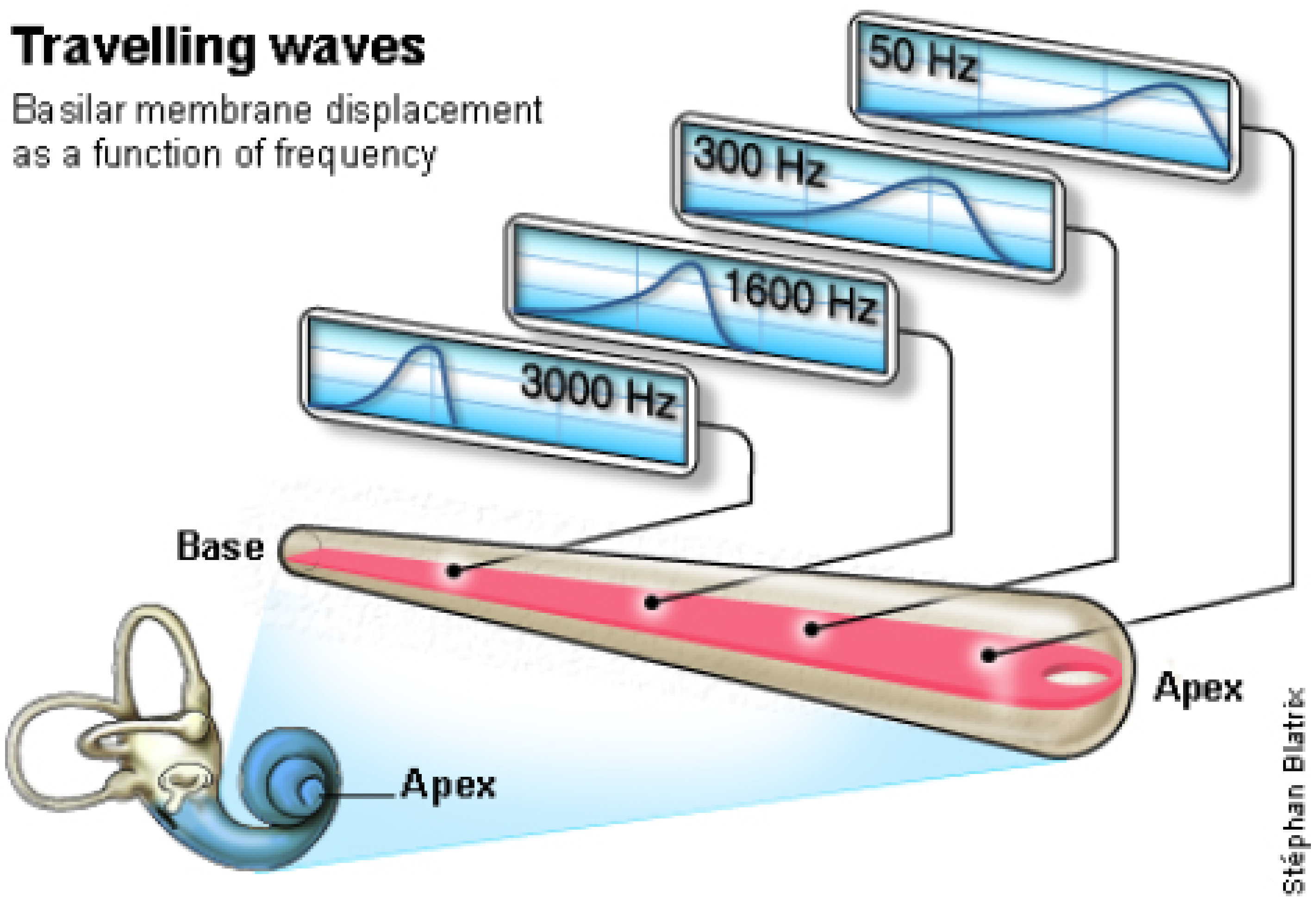
Neuronal Response to Sound

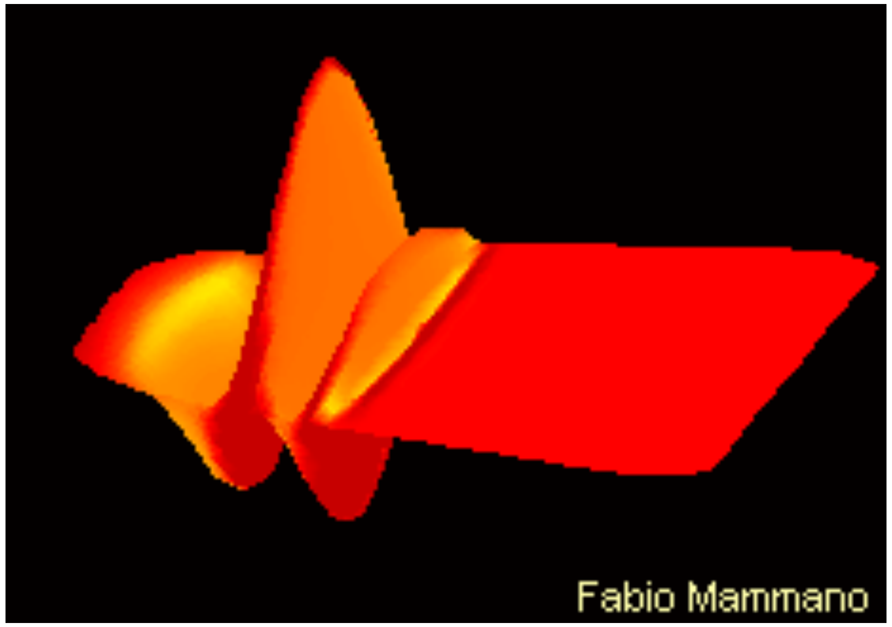
- Frequency → <u>Where</u>?
  The location where in the Cochlea the stereocilia are stimulated.
- Intensity → <u>How many</u>?
  The number of HC that are stimulated by the sound determines the perceived loudness.

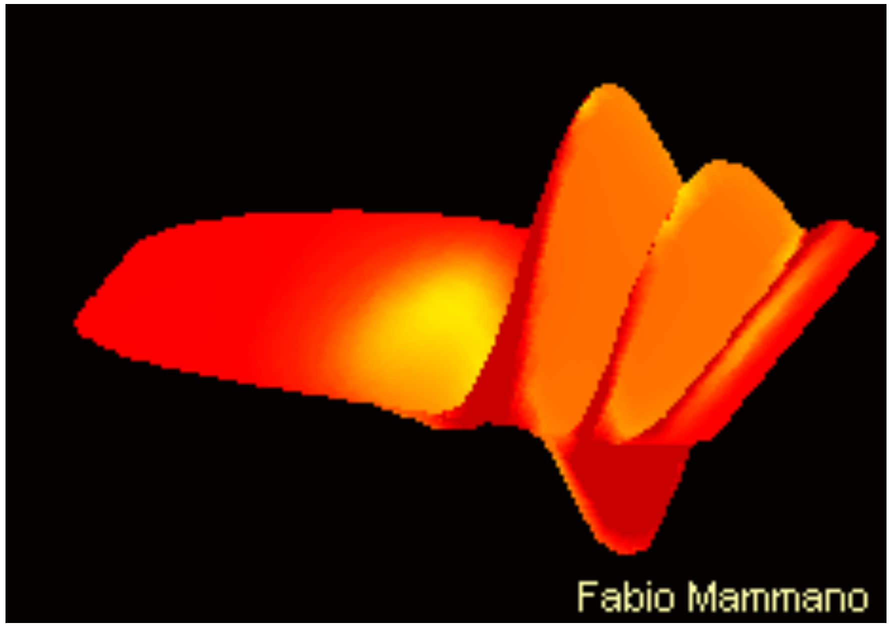


Tonotopic theory

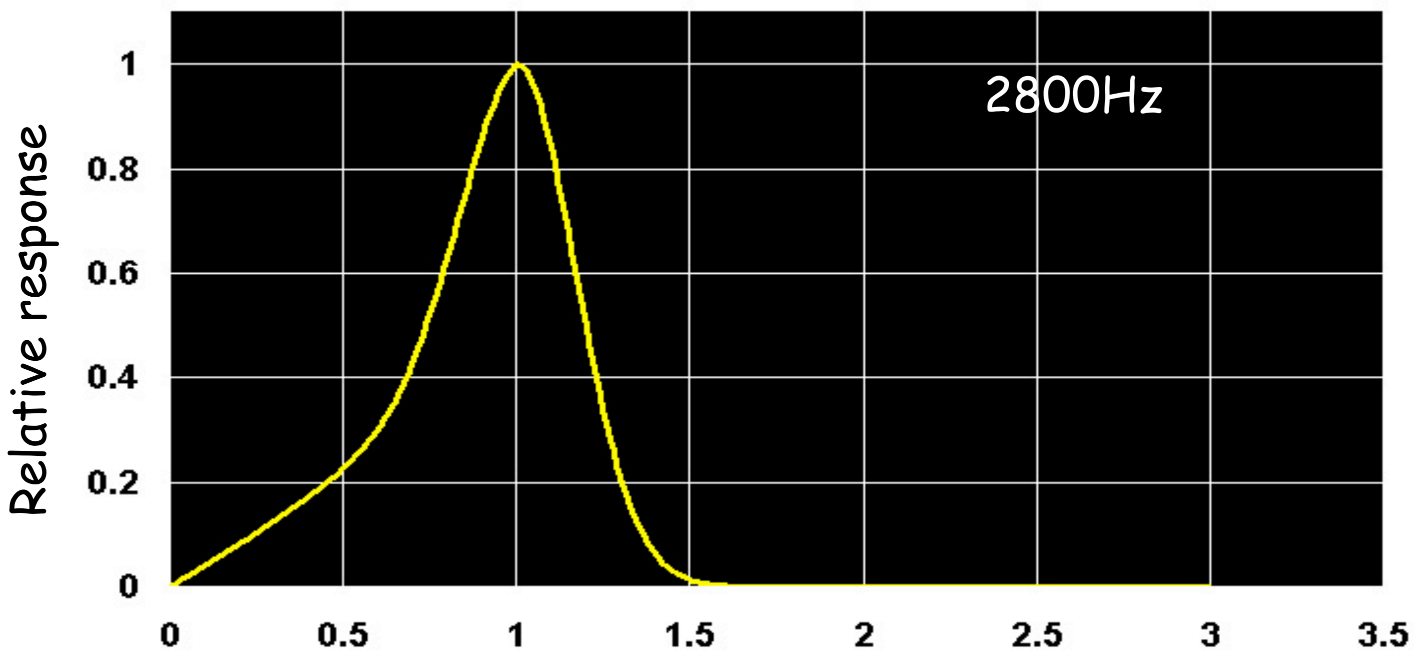
Place Coding of Frequency

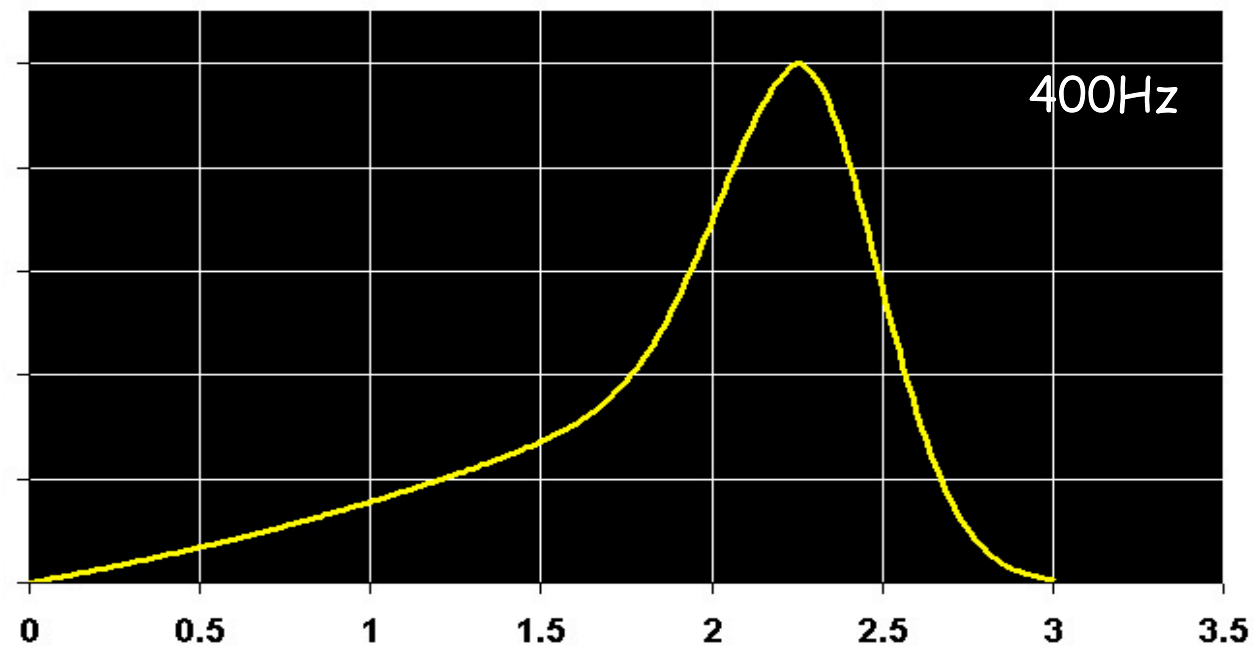


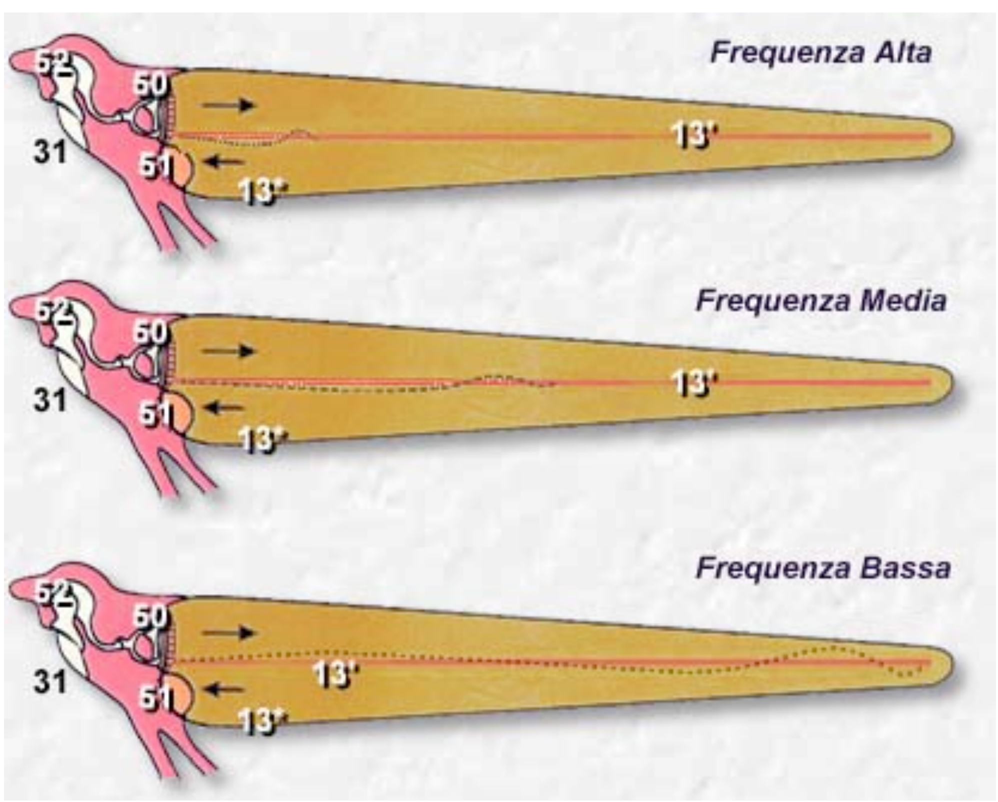


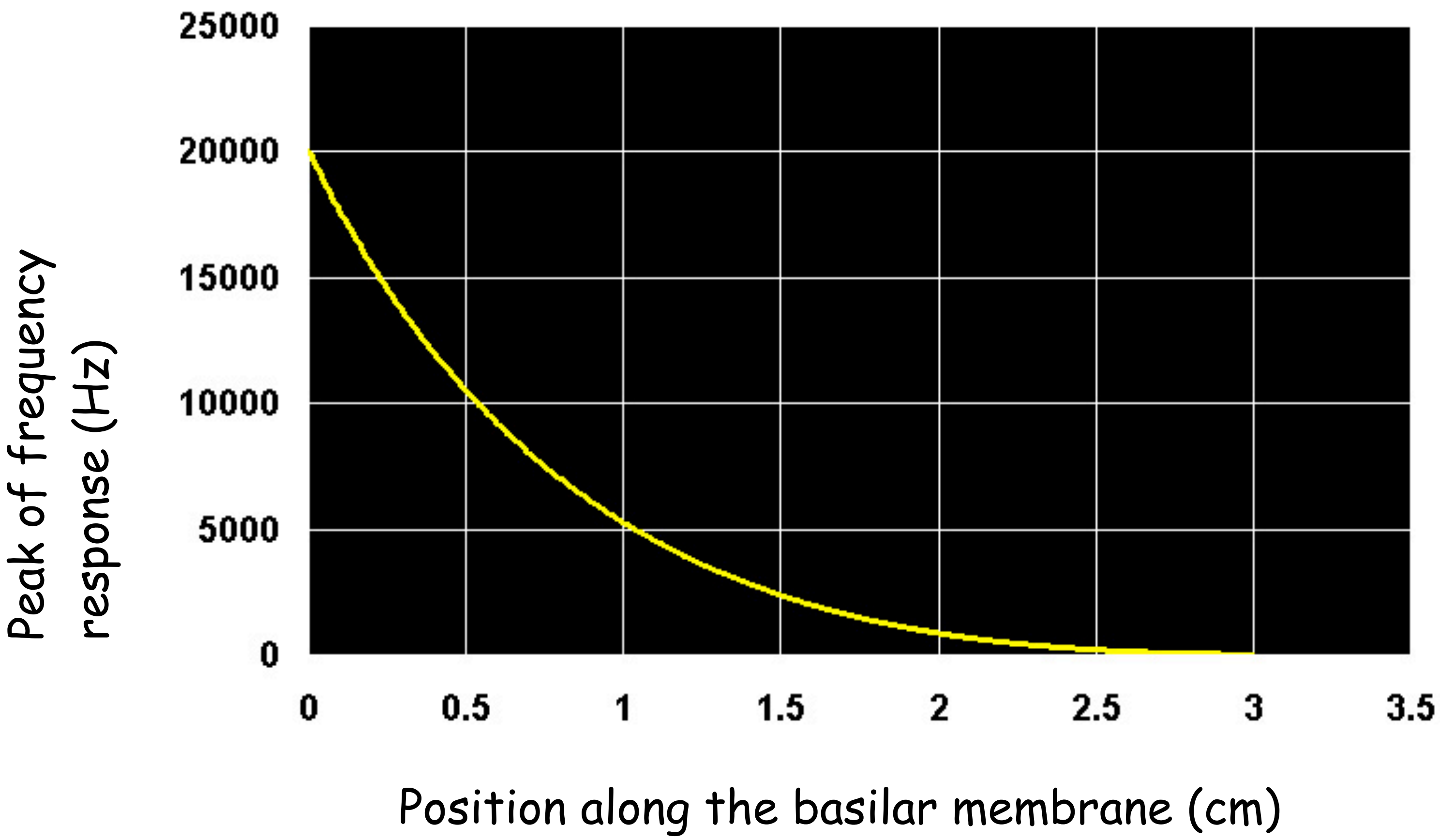


Frequency response

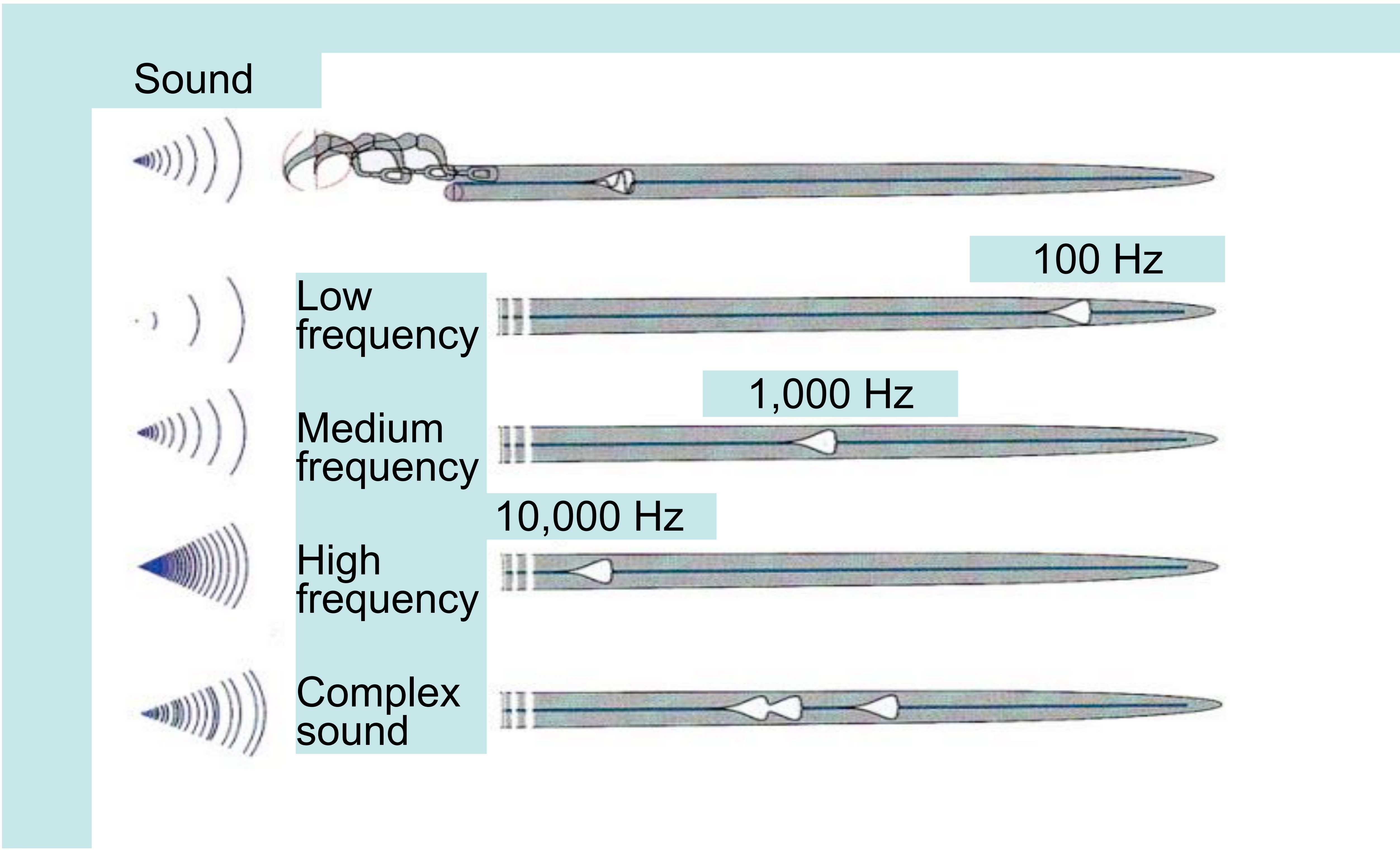






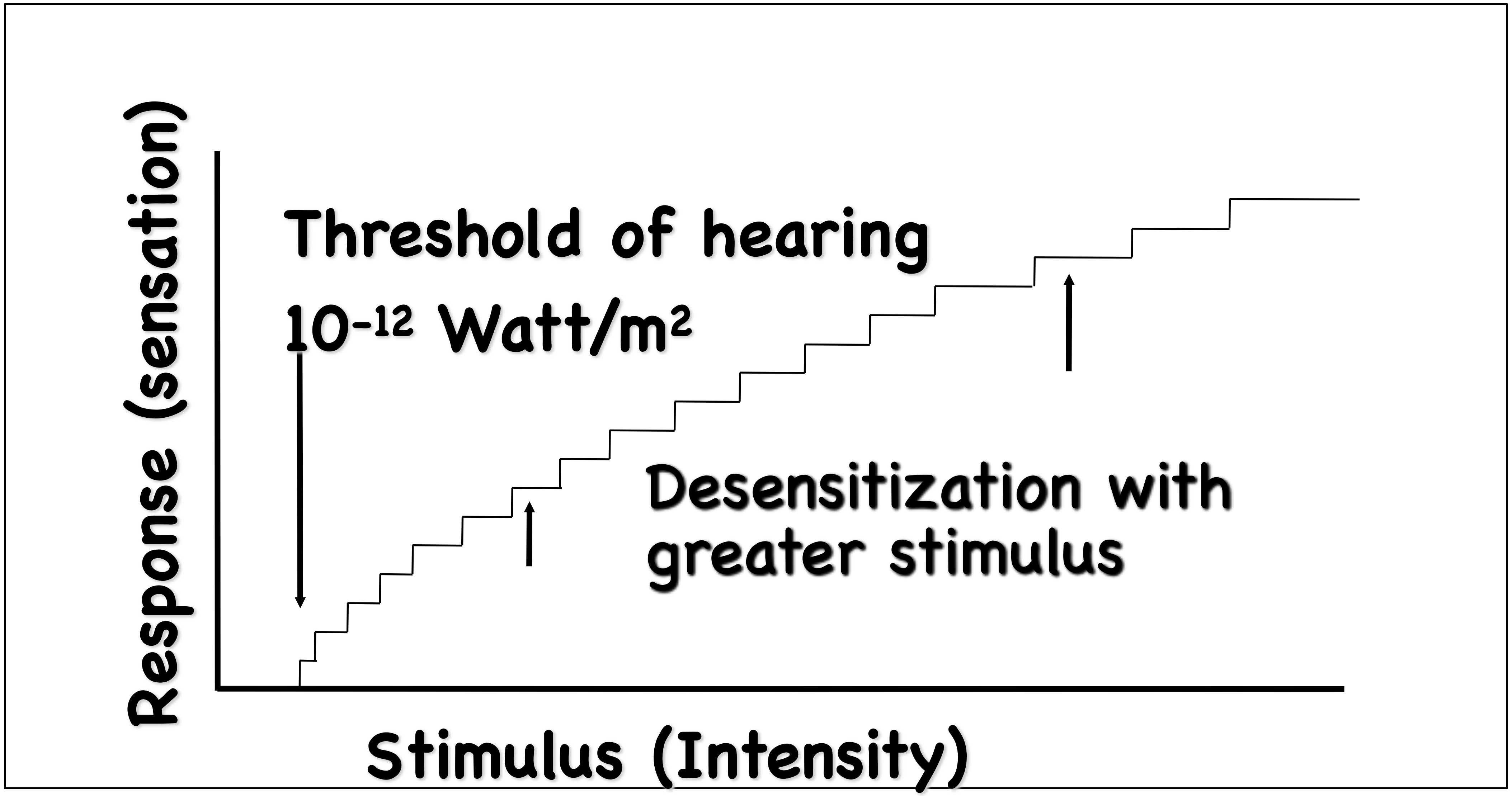


Place theory of pitch perception

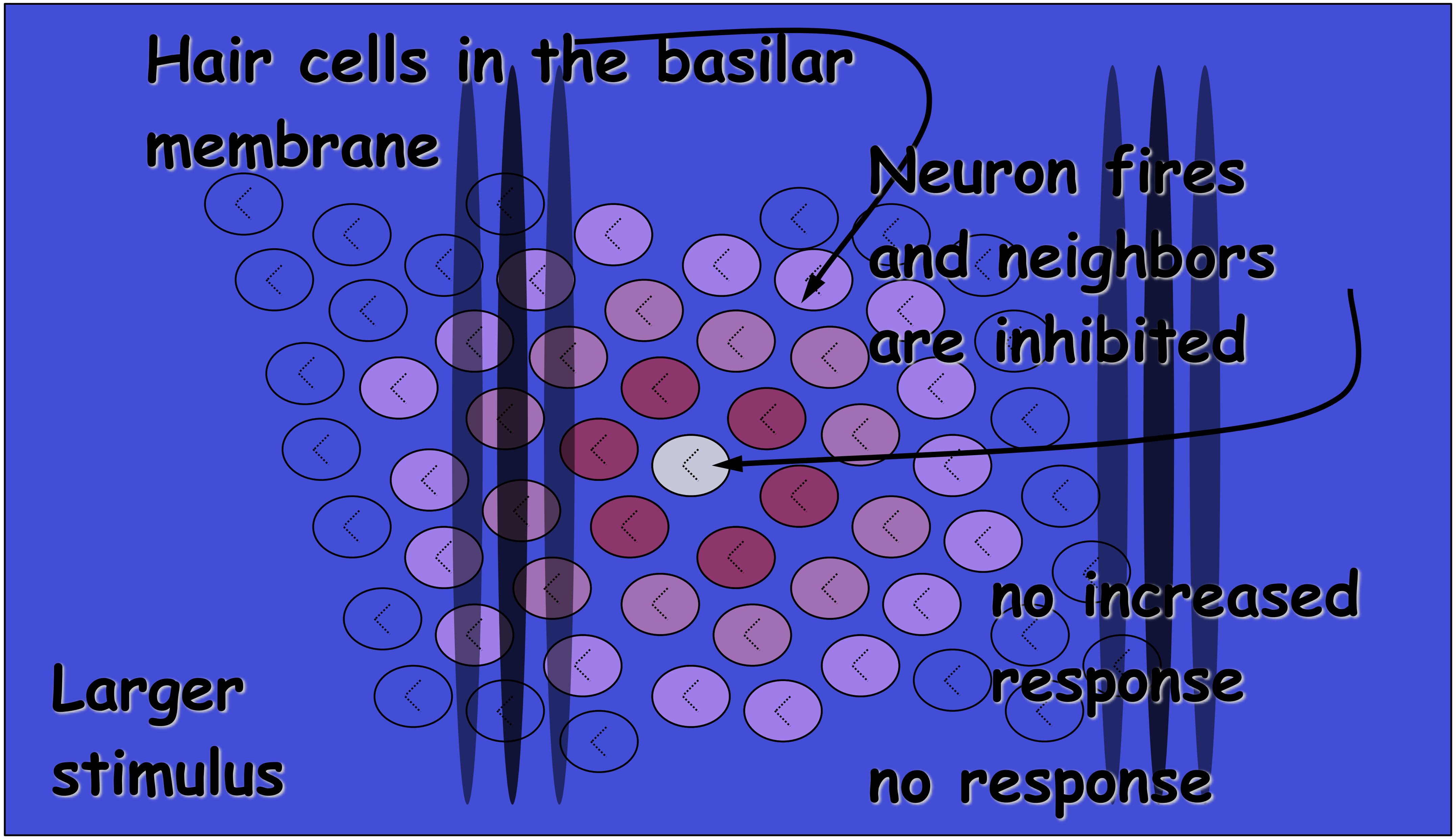


Neuronal response

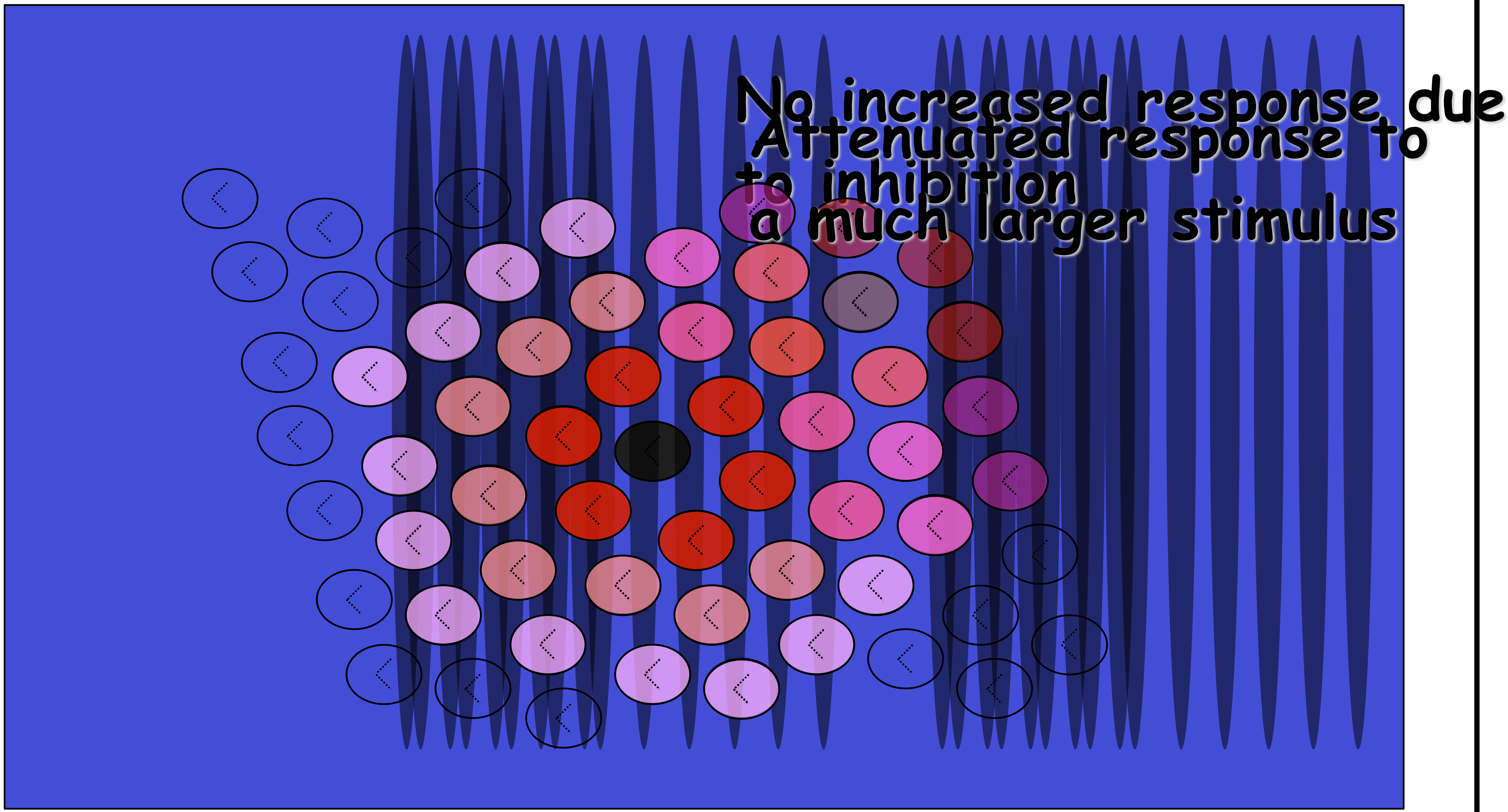
The stereocilia behave like neurons, i.e. can be idealized as a switches network. They have an intensity threshold to be activated and their activation inhibite the neighbours, leading to a non linear response



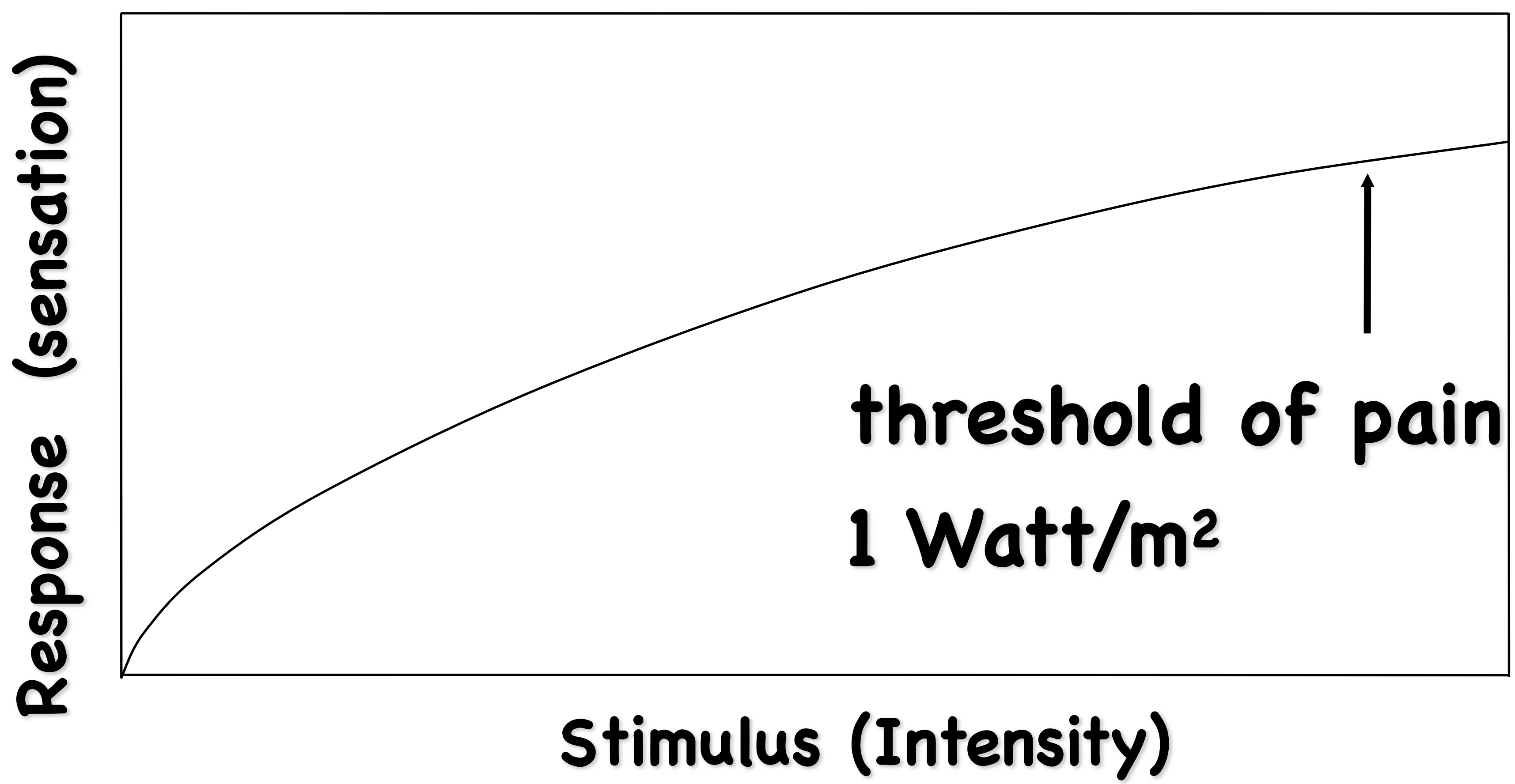
Neuronal response of hair cells



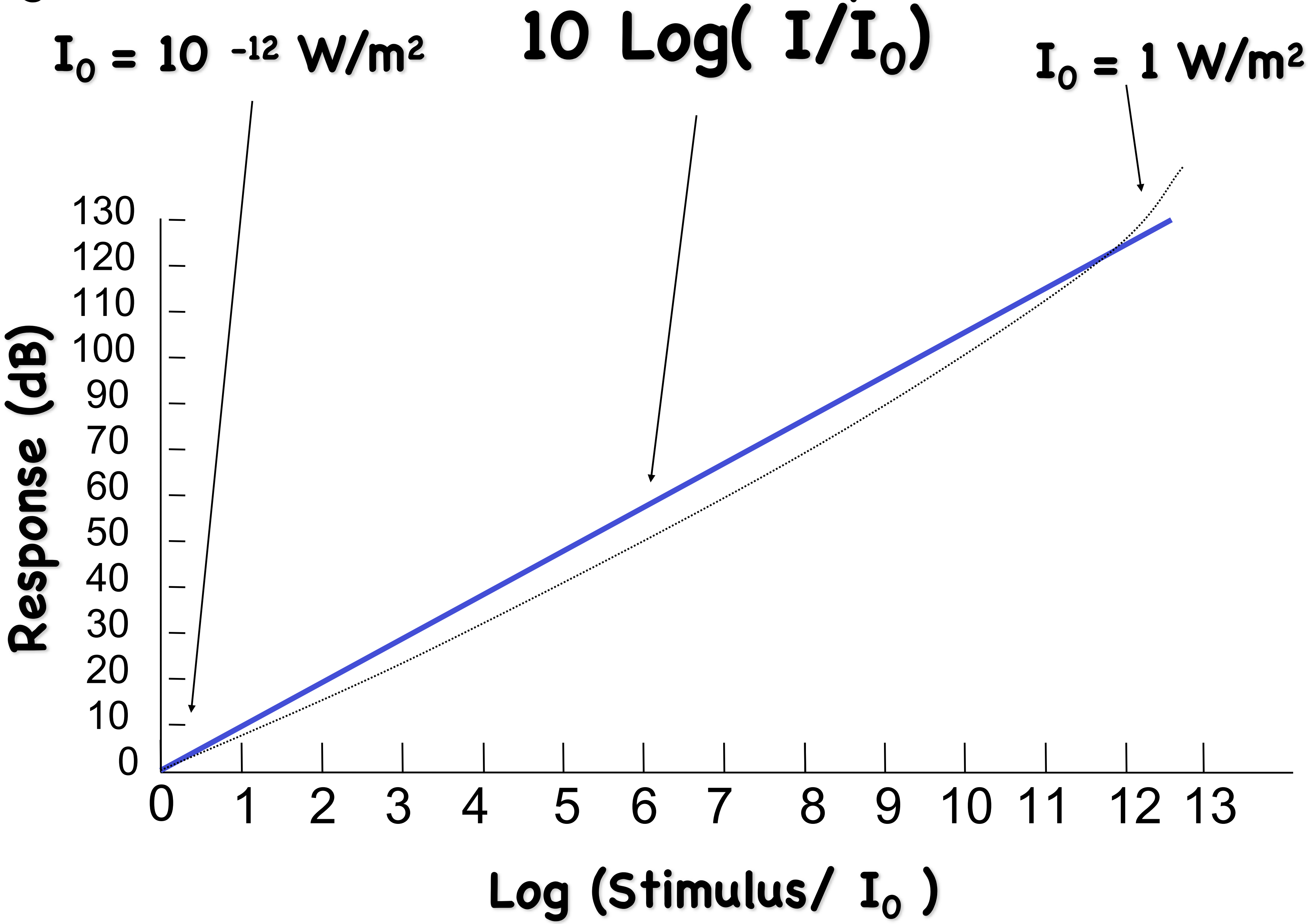
Neuronal response of hair cells



Consequence: Threshold and non linear response



Logarithm of the stimulus versus response



SIL & SPL

That is why the **sound intensity level** (SIL) is defined as:

$$I_{dB} = 10\log\left(\frac{I}{I_0}\right) \approx 20\log\left(\frac{p}{p_0}\right) = p_{dB}$$

where I is the sound intensity, I_o is the audibility limit (~10^{-12} Wm^{-2}), and is measured in decibel (dB).

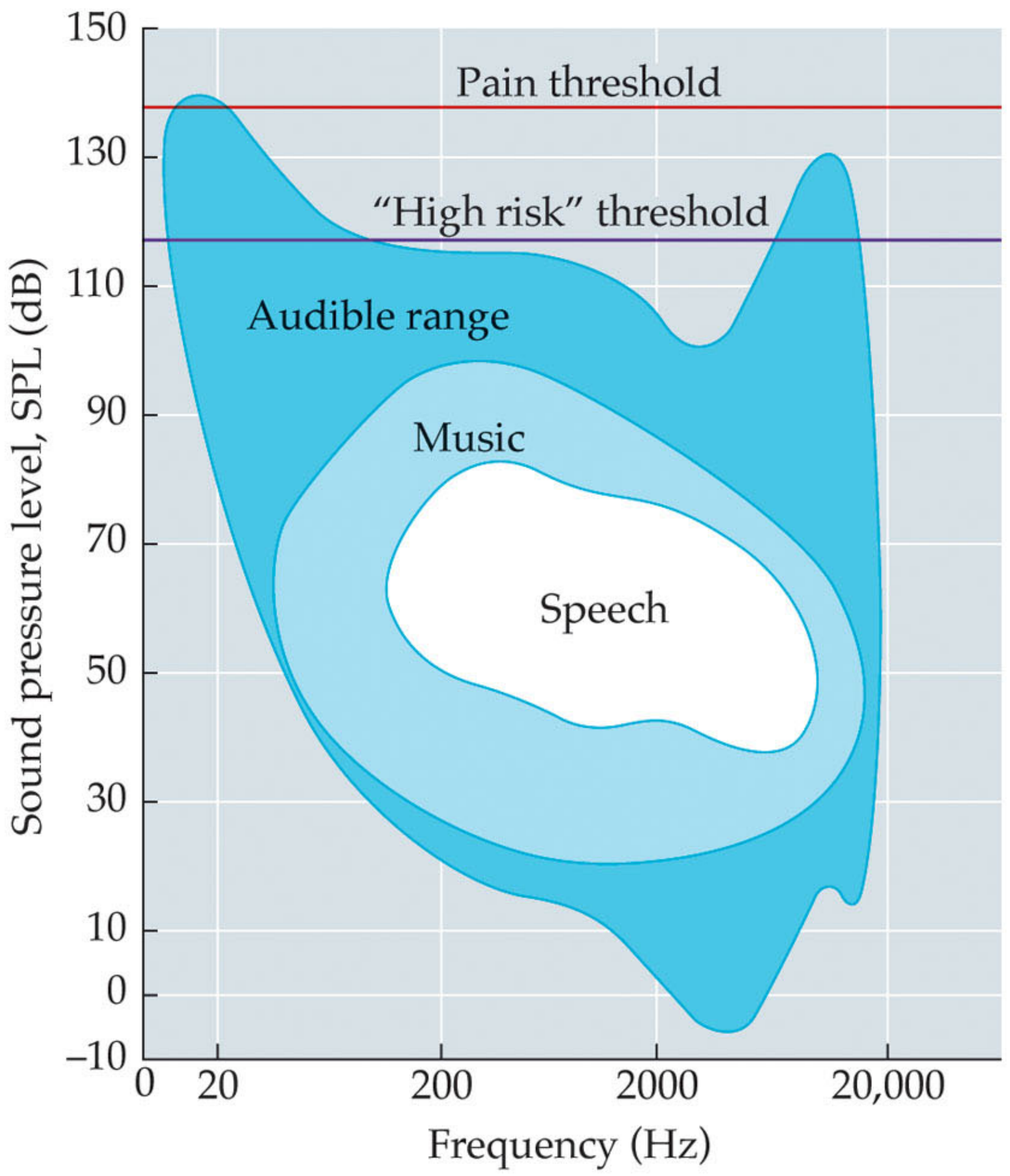
The **sound pressure level** (SPL) coincides with it for perfectly spherical fronts,

where p is the sound pressure, p_o is the audibility limit (~$2x10^{-5}$ Pa), and is measured in decibel (dB).

Summary

How does Anatomy affect perception?

- Frequency response
- Loudness perception
- Phase insensitivity
- Large "non-linear" range of 12 orders of magnitude in intensity
- Deafness
  - Disruption of "acoustic chain."
  - Nerve death.



JND in intensity

Just Noticeable Difference (JND) is the <u>limen</u> of difference that elicits 75% correct answers in a Two Alternative Forced-Choice test (2AFC test).

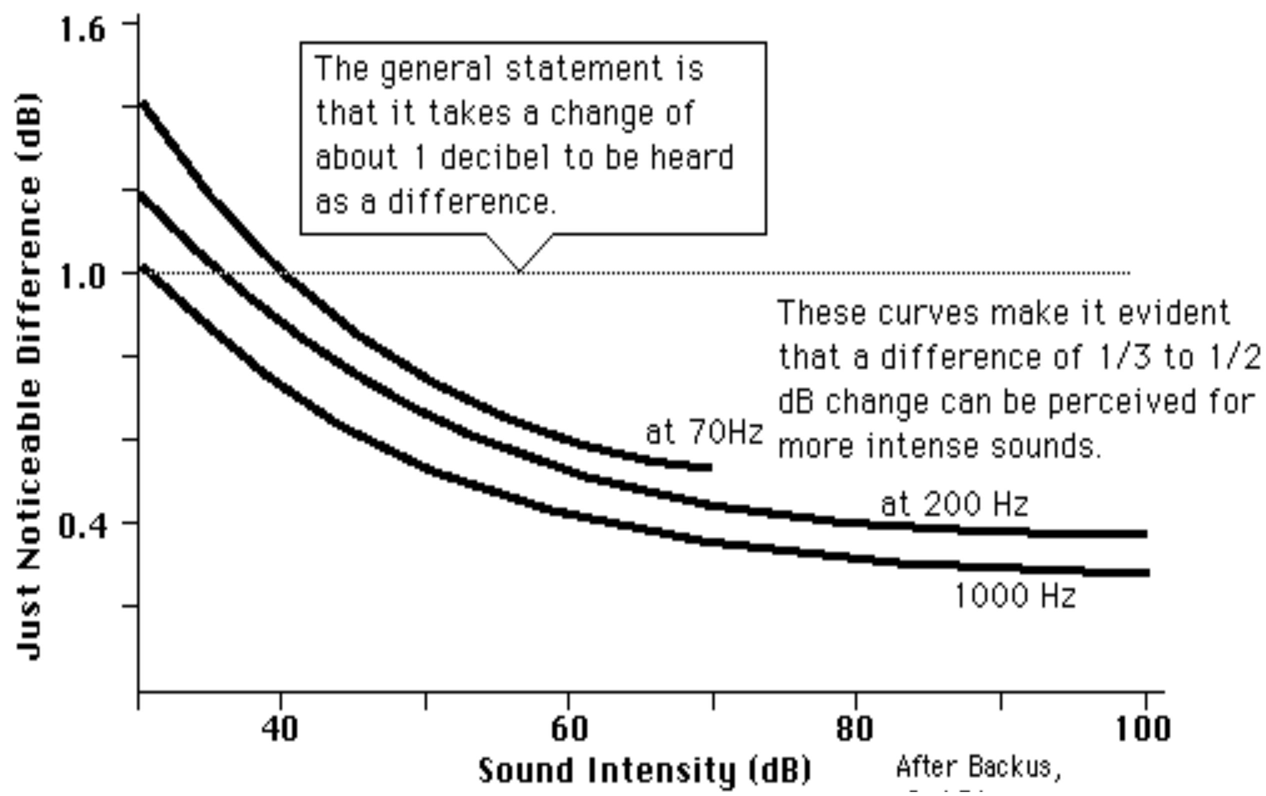
Why 75%?

In 2 Alternative Forced Choice:

50% correct means random choice

100% means can always tell the difference.

Thus, 75% is halfway between random and certainty.



The limen of intensity is a ratio of about 1.26 which corresponds to a SIL difference of 1 dB.

$$10 \,\mathrm{Log}(1.26) = 1.0$$

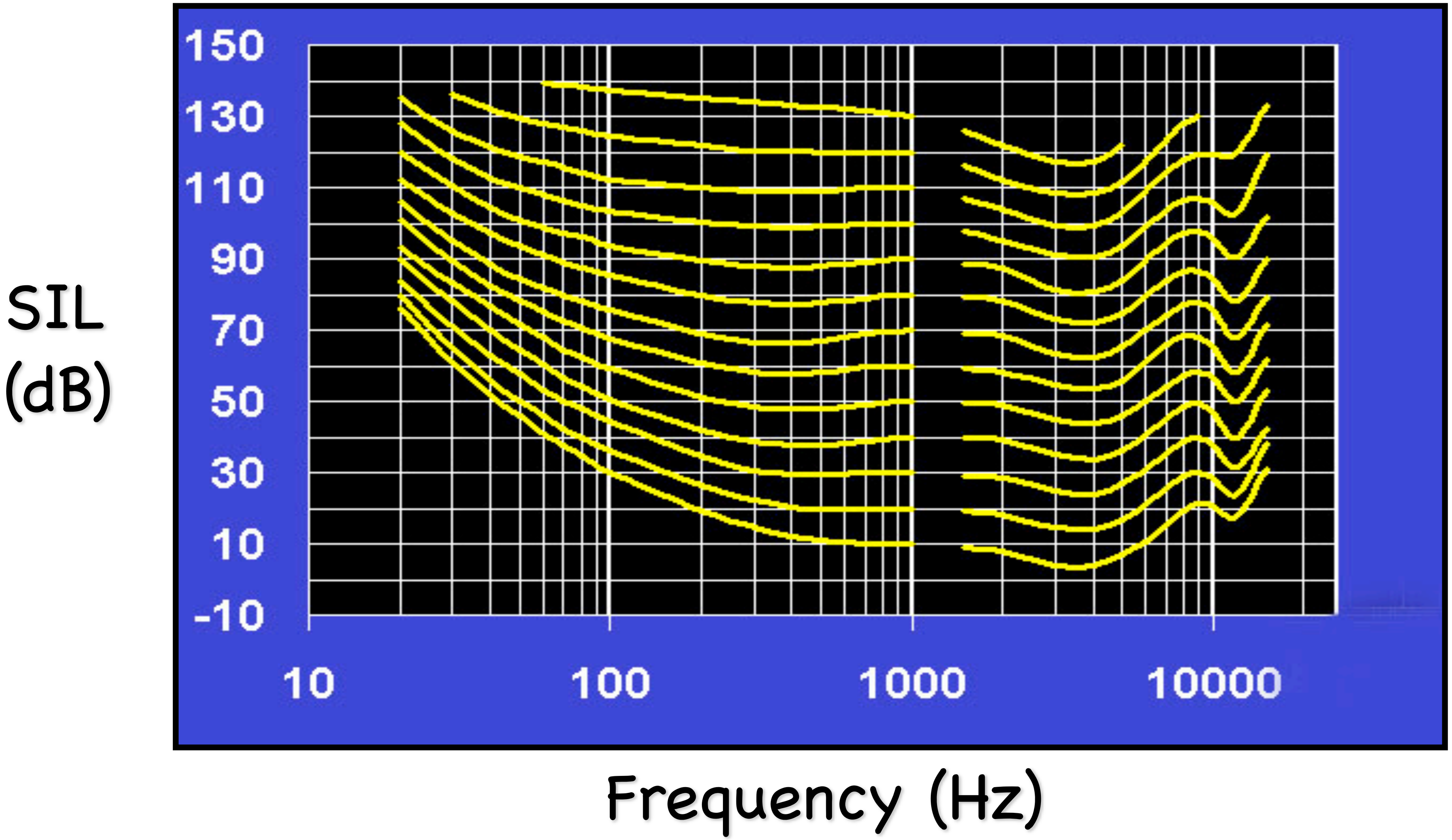
Musical dynamics

Musical Dynamics

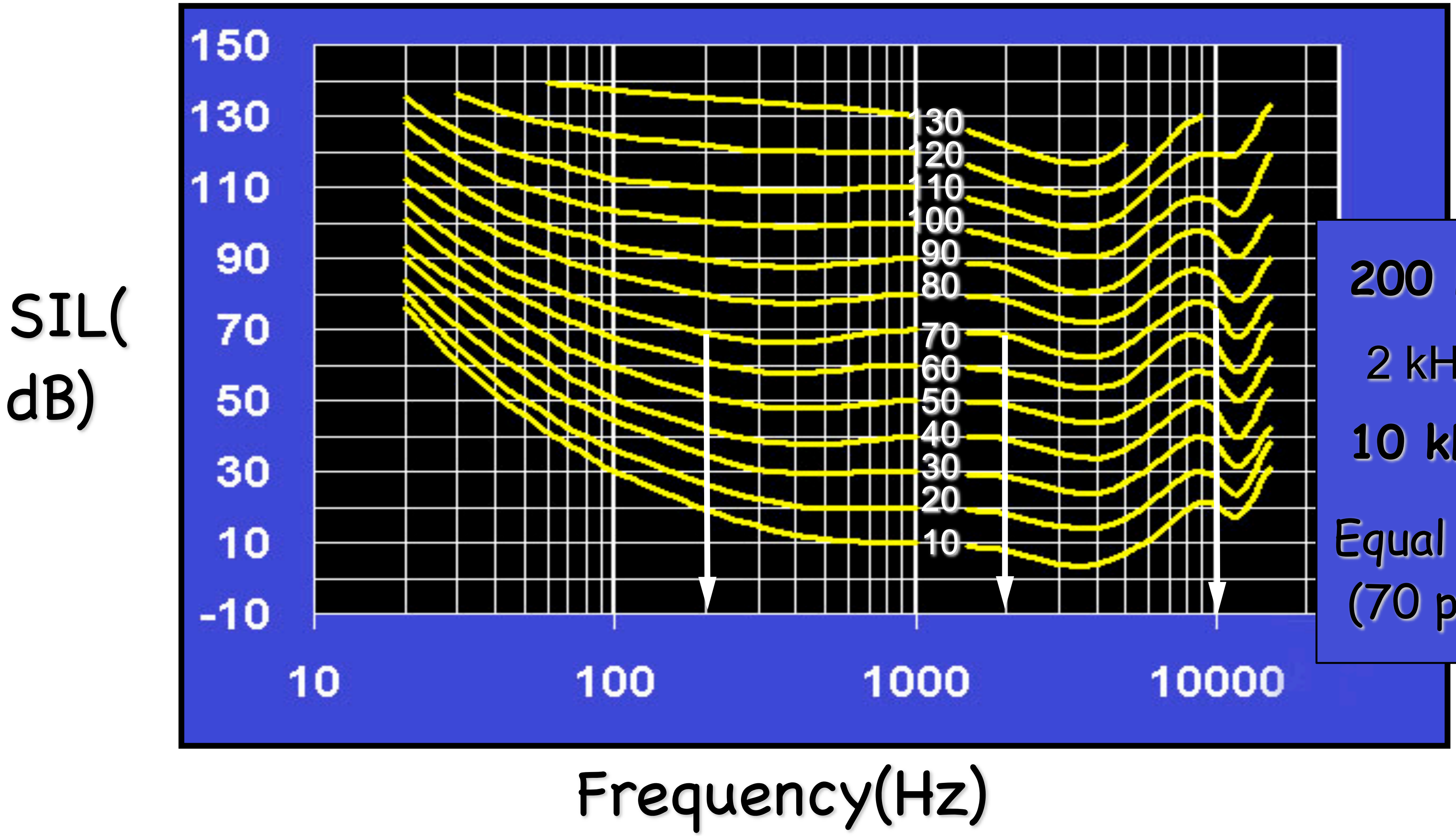
Pianissimo:	pp	very soft:	50 dB
Piano:	p	soft:	60 dB
Mezzopiano:	mp	medium soft:	66 dB
Mezzoforte:	mf	medium loud:	76 dB
Forte:	f	loud:	80 dB
Fortissimo:	ff	very loud:	90 dB
Fortississimo:	fff	Very, very loud:	100 dB

Loudness

The **Fletcher-Munson Diagram** is a plot of the SIL (in dB) versus frequency for the SIL required to produce an equal sensation as that produced at 1000 Hz. The contours are of equal **loudness level**. The unit of loudness level is the **phon**.



Determine the necessary SIL to feel a 70 phon level at 200 Hz, 2000 Hz and 10,000 Hz.



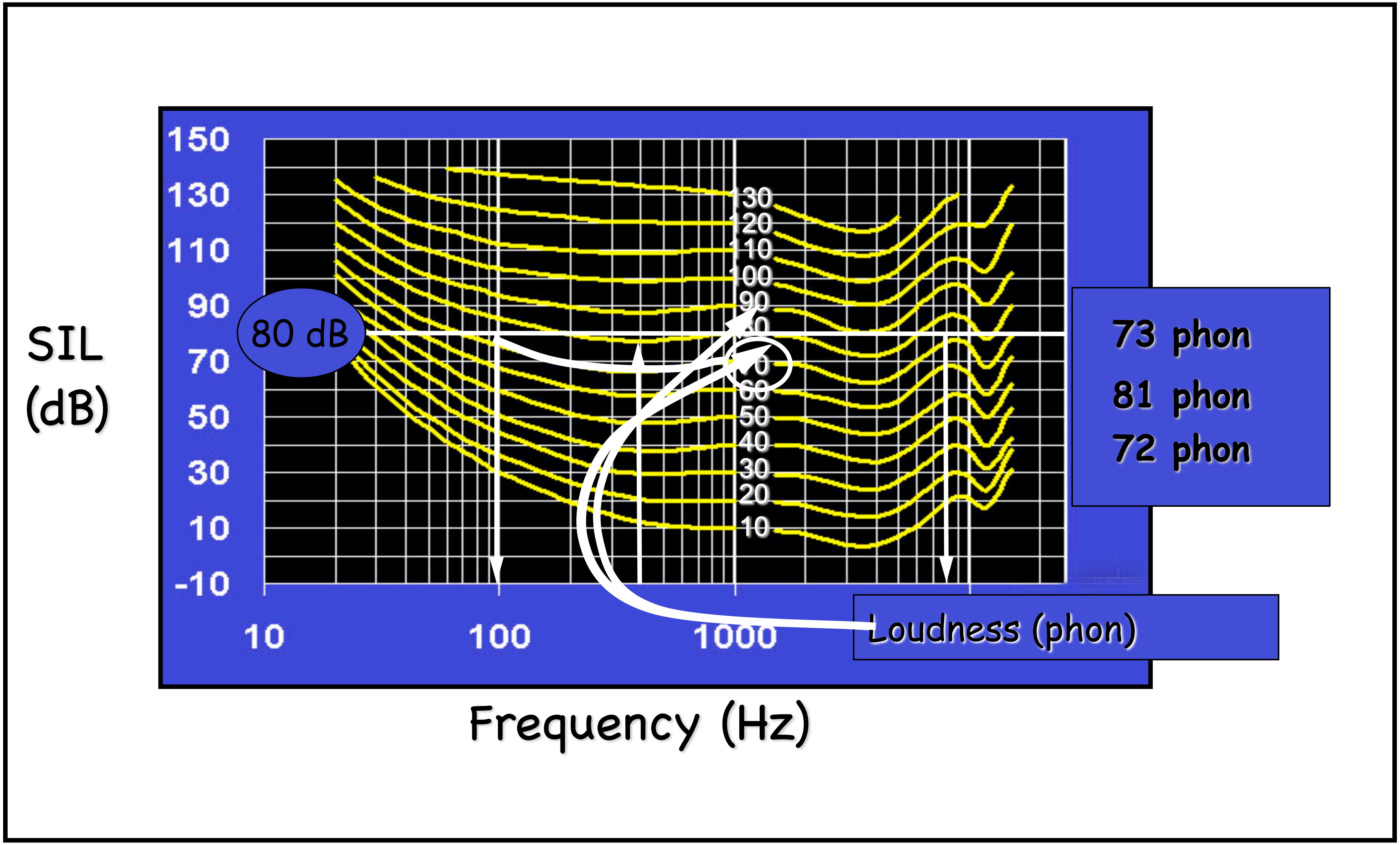
200 Hz - 70 dB

2 kHz - 70 dB

10 kHz - 75 dB

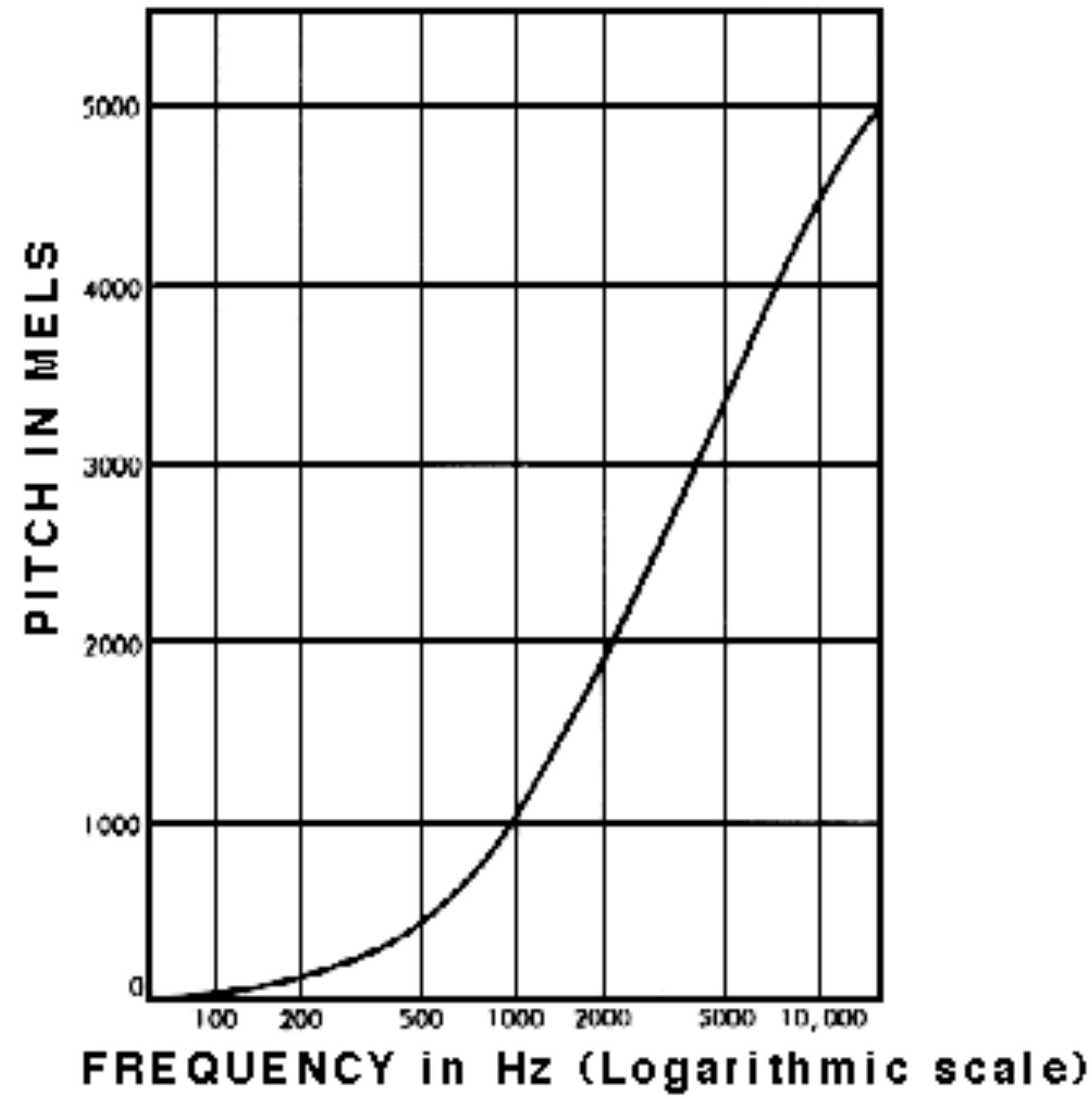
Equal loudness (70 phon)

Determine the loudness of a sound with 80 dB at 100, 400 and 8000 Hz.



Frequency perception: PITCH

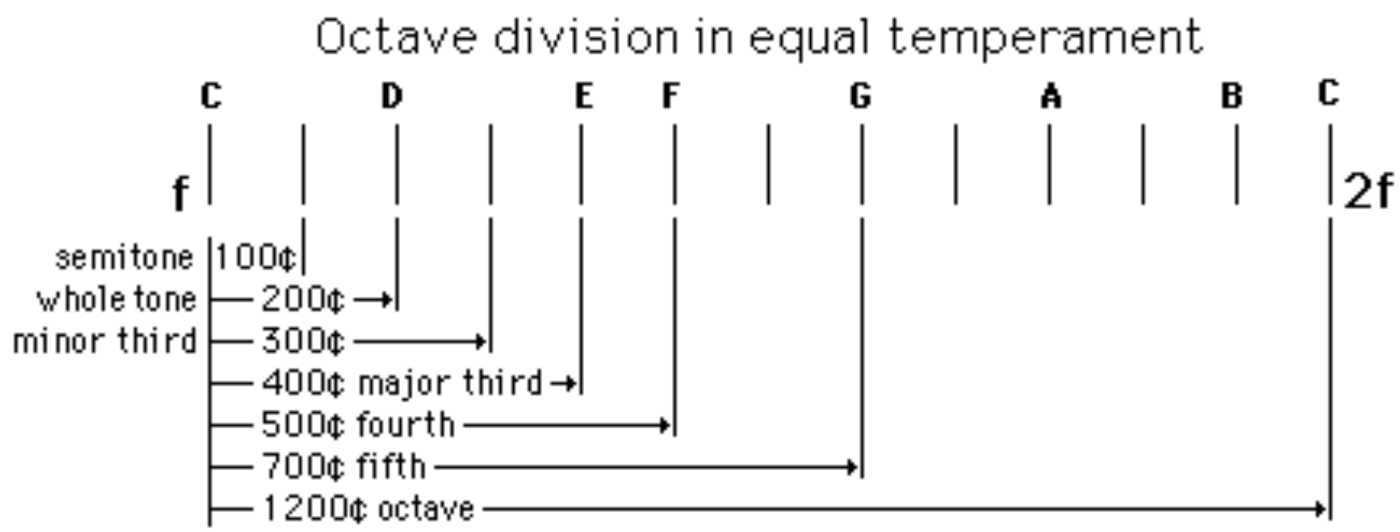
- The perception of the frequency of a sound wave (**pitch**) depends on the response of the hearing ear.
- Absolute perception (out of a context) of absolute pitch is extremely rare (0.01%), while the **relative** perception (e.g. musical interval) is common.
- There is a **non-linear** correspondance: definition of **mel**.



Unit of subjectively estimated pitch. A sine wave with a frequency of 1000 hertz, 40 decibels above the listener's threshold of hearing, has by definition a pitch of 1000 mels. A sound that a listener judges to be 2 times the pitch of a sound with a pitch of 1000 mels has a pitch of 2000 mels...

Pitch intervals

- An equal frequency ratio is perceived as an equal difference (interval) of pitches.
- One **OCTAVE** is the pitch interval corresponding to a frequency ratio = 2/1.
- One **SEMITONE** is 1/12 of an octave.
- One **CENT** (¢) is 1/100 of a semitone or 1/1200 of an octave. 1 octave = 1200¢



To find a pitch interval in cents

$$\frac{f_2}{f_1} = 2^{\frac{¢}{1200}} \qquad ¢ = \frac{1200}{\log(2)} \log\left(\frac{f_2}{f_1}\right)$$

JND in Pitch

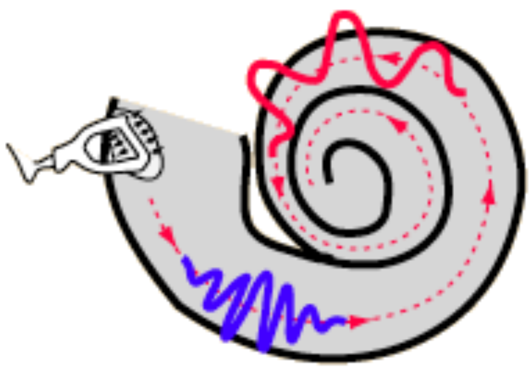
There is a natural limit to an individuals ability to establish a relative order of pitch when two pure tones (sine tones) of the same intensity are presented one after another. When the difference in frequency between the two tones is too small, both tones are judged as having the same pitch. This is true for order judgments for all psychophysical magnitudes: whenever the variation of an original physical stimulus lies within a certain difference limen or just noticeable difference (**JND**) the associated sensation is judged as remaining the same. As soon as the variation exceeds the JND, a change in sensation is detected.

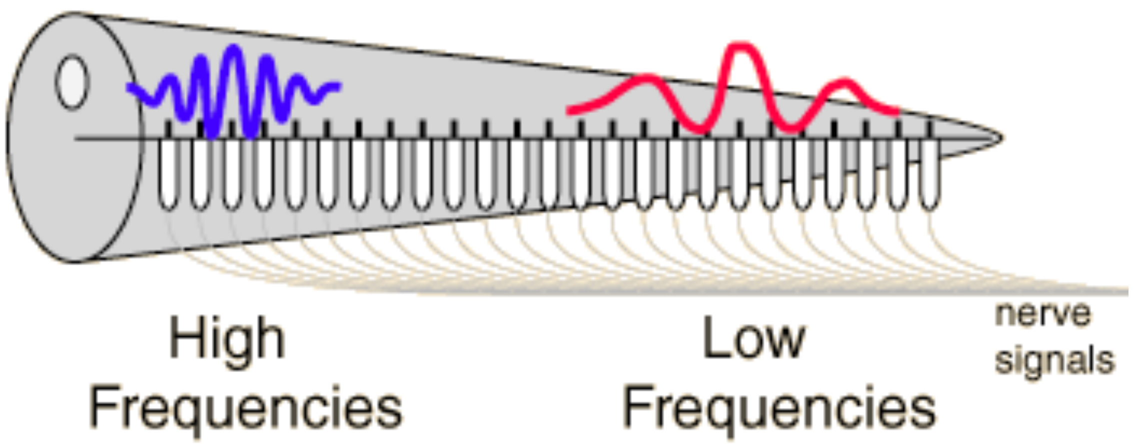
The degree of sensitivity of the primary pitch mechanism to frequency changes (frequency resolution capability) depends on the **frequency, intensity**, and **duration** of the tone in question. It varies greatly from person to person, is affected by musical training, and depends considerably on the method of measurement employed. On average it is **8¢** and the number of frequencies resolved by a human is:

$$\frac{\frac{1200}{\log(2)}\log\left(\frac{20000}{20}\right)}{8} = \frac{3600}{8\log(2)} \approx 1500$$

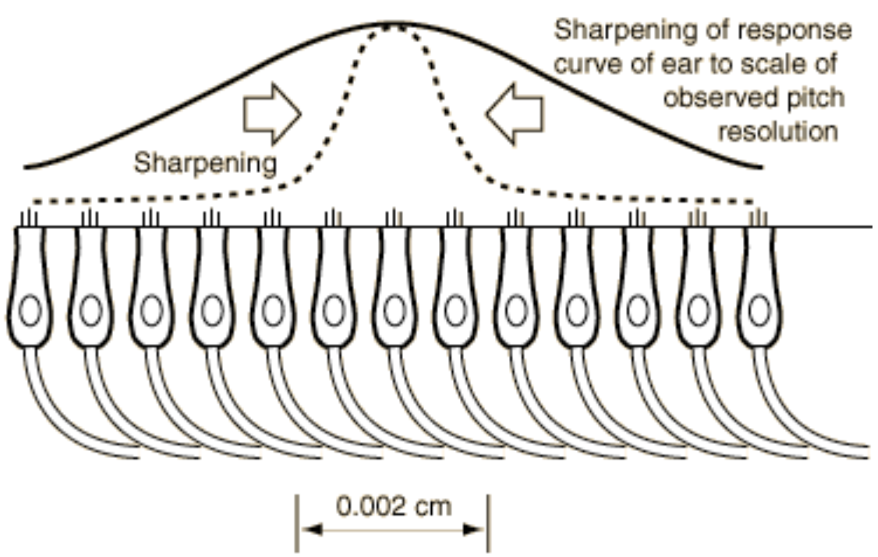
Place theory and resolving power

Thus the human hear can resolve 1500 pitches with 16000-20000 stereocylia in 3.5 cm, i.e. a pitch every 0.002 cm!



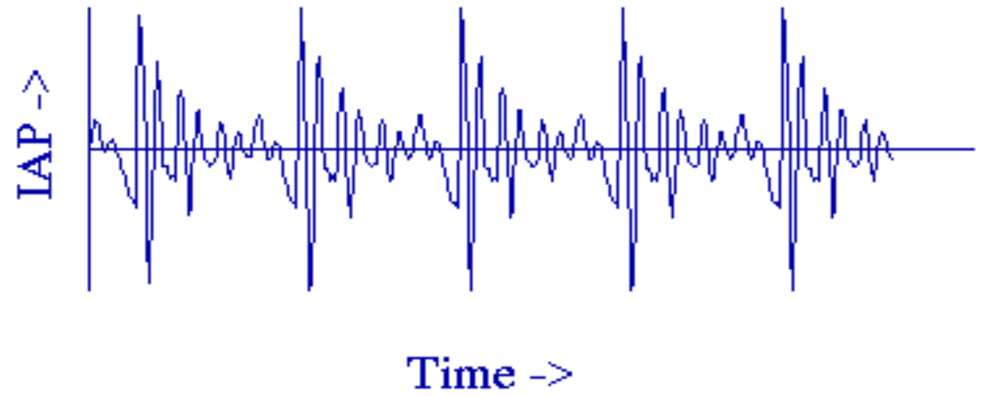


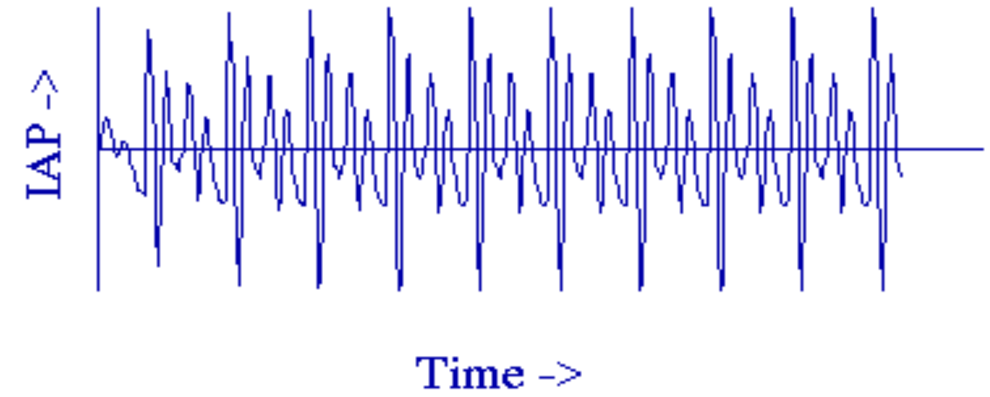
Following the place theory, such an high resolving power implies that only 4 cellular groups are used and this means that should exist a mechanism that is sharpening Corti's organ response to the perturbations in the basilar membrane.

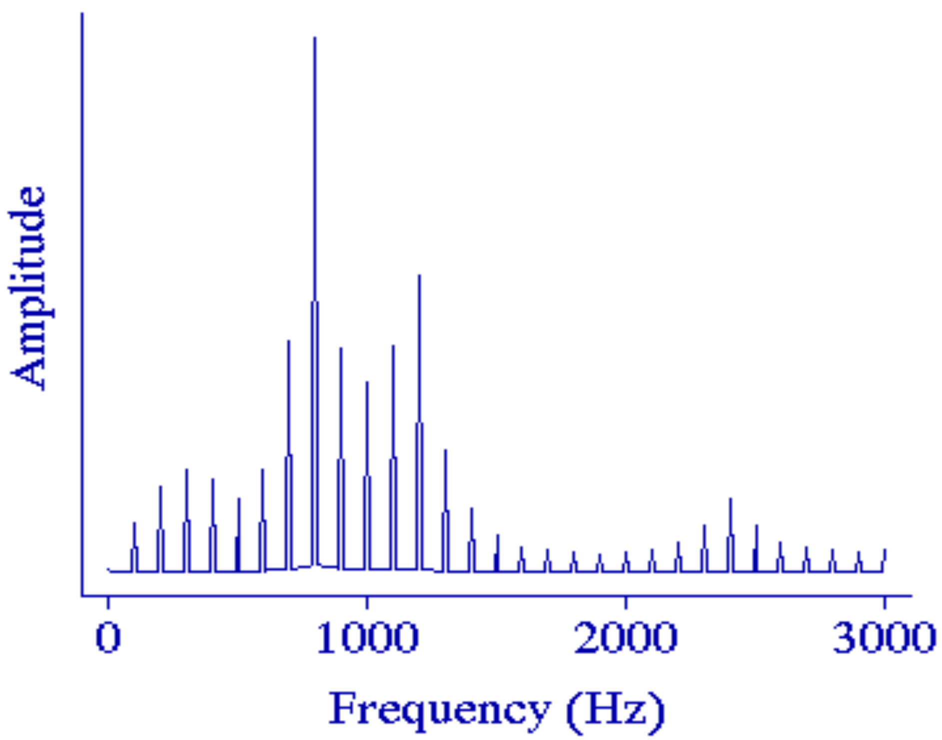


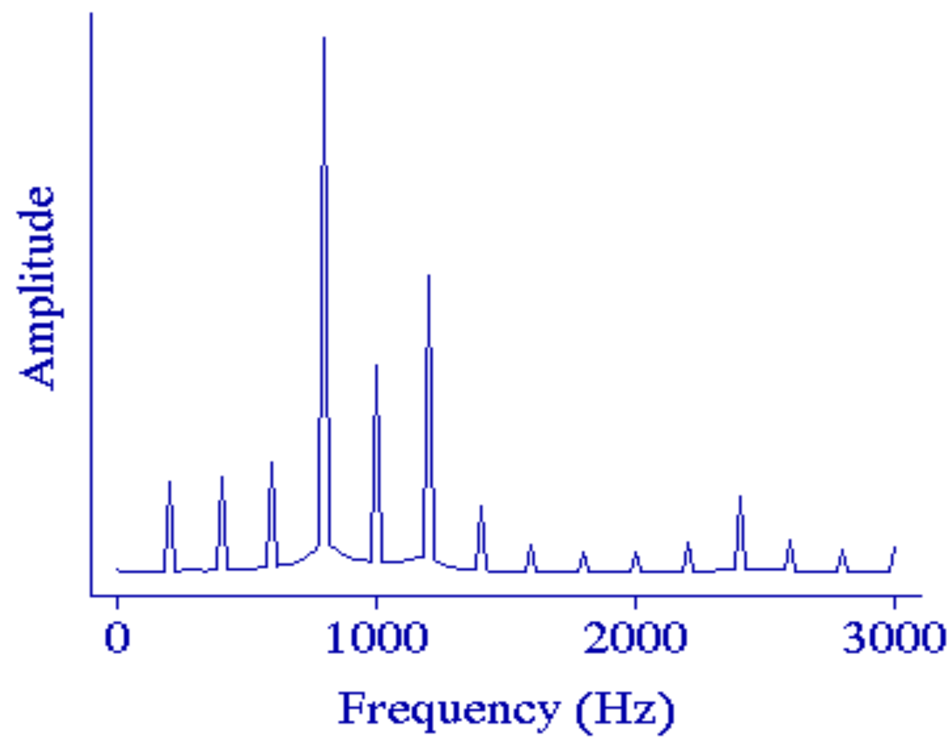
A possible mechanism is the inhibition of the neighboring cellula

Pitch and fundamental frequency









Lower F_0, lower pitch

Higher F_0, higher pitch

Pitch Perception

The ear is more sensitive to F_0 differences in the low frequencies than the higher frequencies. This means that:

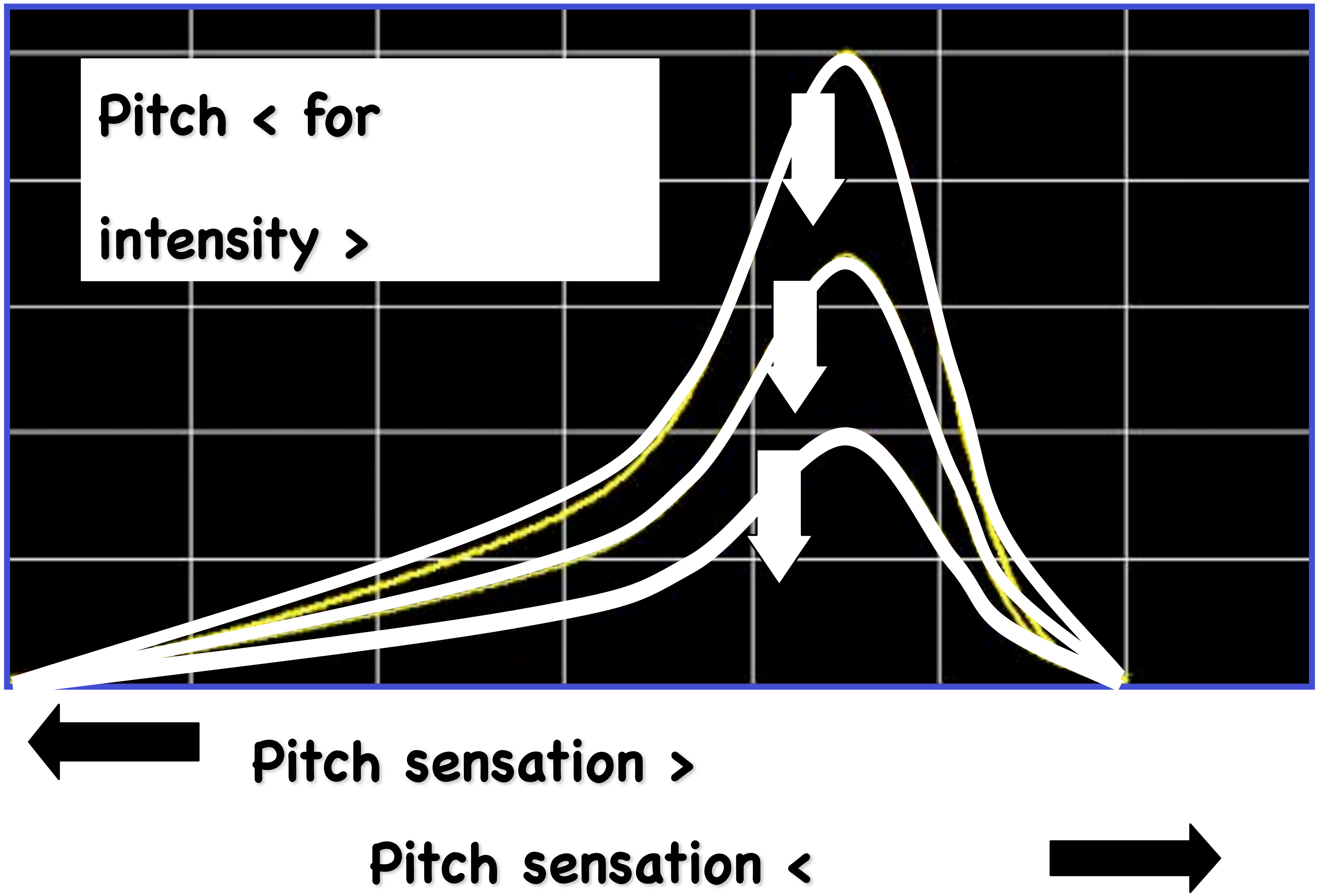
300 vs. 350 $\neq$ 3000 vs. 3050

That is, the difference in perceived pitch (not F_0) between 300 and 350 Hz is NOT the same as the difference in pitch between 3000 and 3050 Hz, even though the physical differences in F_0 are the same.

300-350: 3000-3050:

Pitch & loudness

- Changing pitch with intensity



Loudness and Pitch

- More sensitive to loudness at mid frequencies than at other frequencies
  - intermediate frequencies at [500hz, 5000hz]
- Perceived loudness of a sound changes based on the frequency of that sound
  - basilar membrane reacts more to intermediate frequencies than other frequencies

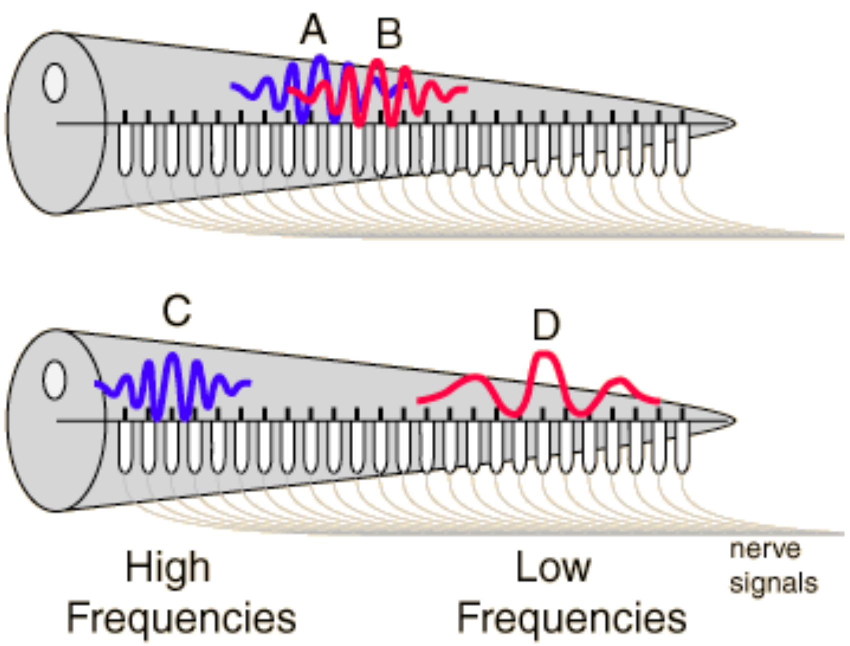
Loudness Perception

Loudness is strongly affected by the frequency of the signal. If intensity is held constant, a mid-frequency signal (in the range from ~1000-4000 Hz) will be louder than lower or higher frequency signals.

125 Hz, 3000 Hz, 8000 Hz

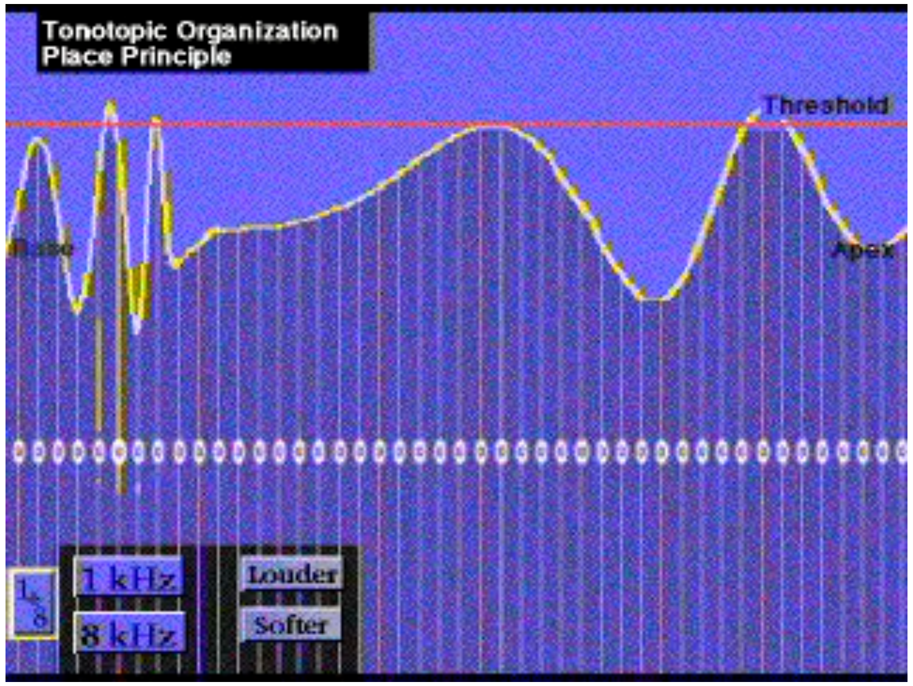
The 3000 Hz signal should appear louder than the 125 or the 8000 signal, despite the fact that their intensities are equal.

Pitch and loudness

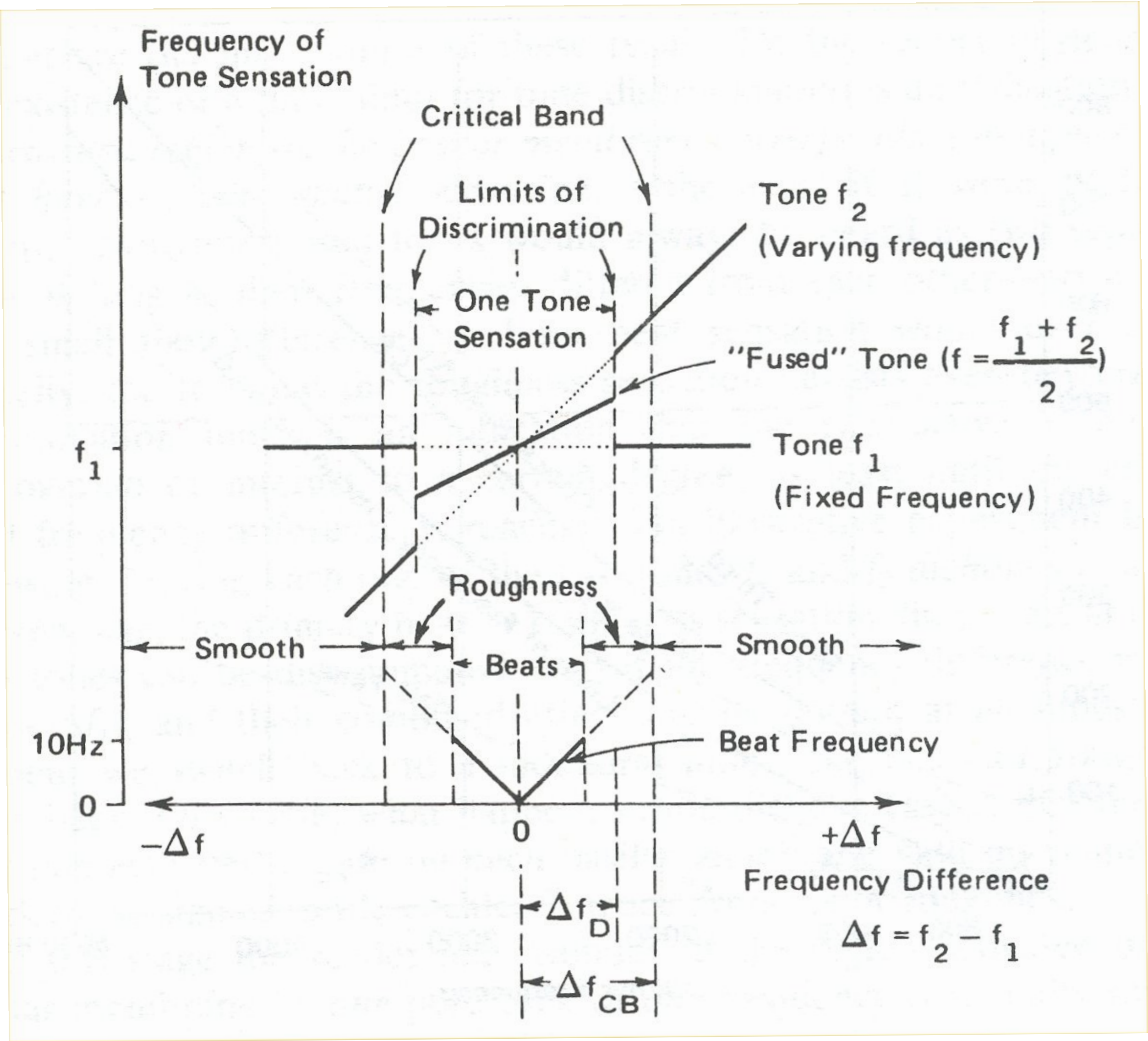


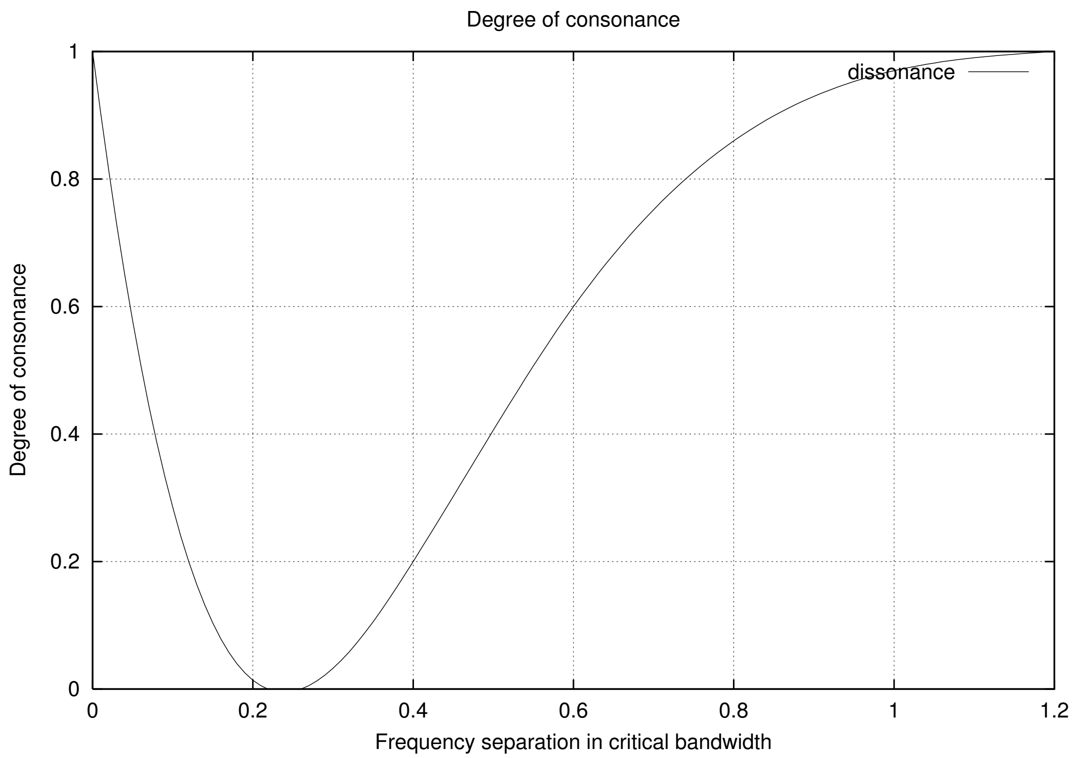
Let us suppose to add two sounds. The loudness perception will depend on the frequency separation.

C+D will be characterized by a greater loudness since it interests different parts of the basilar membrane. It is outside the **critical band,** i.e. the frequency interval corresponding to the specific part of the basilar membrane, and thus to the specific number of neural cells, interested by a sound with a given pitch.



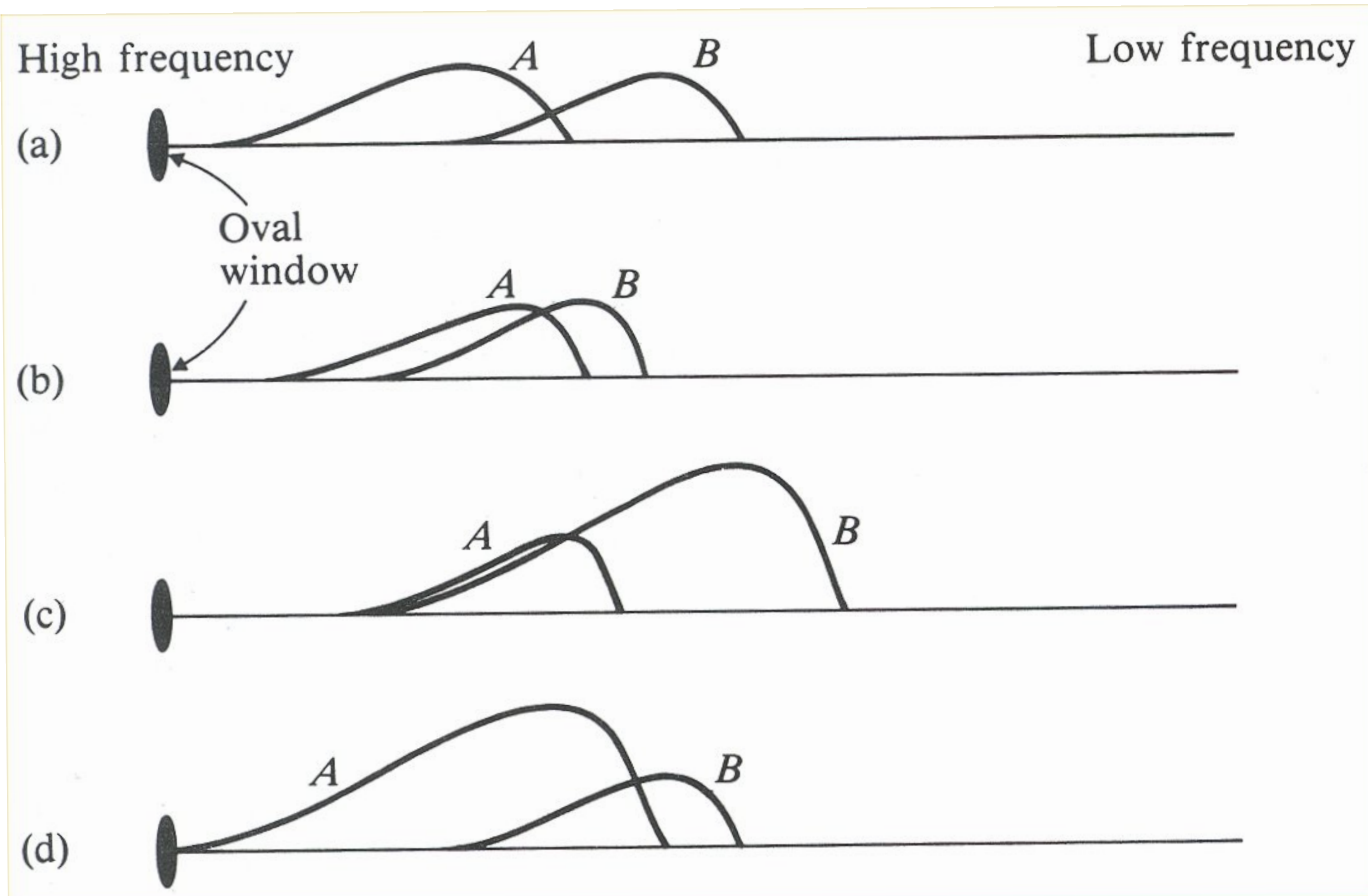
Critical band



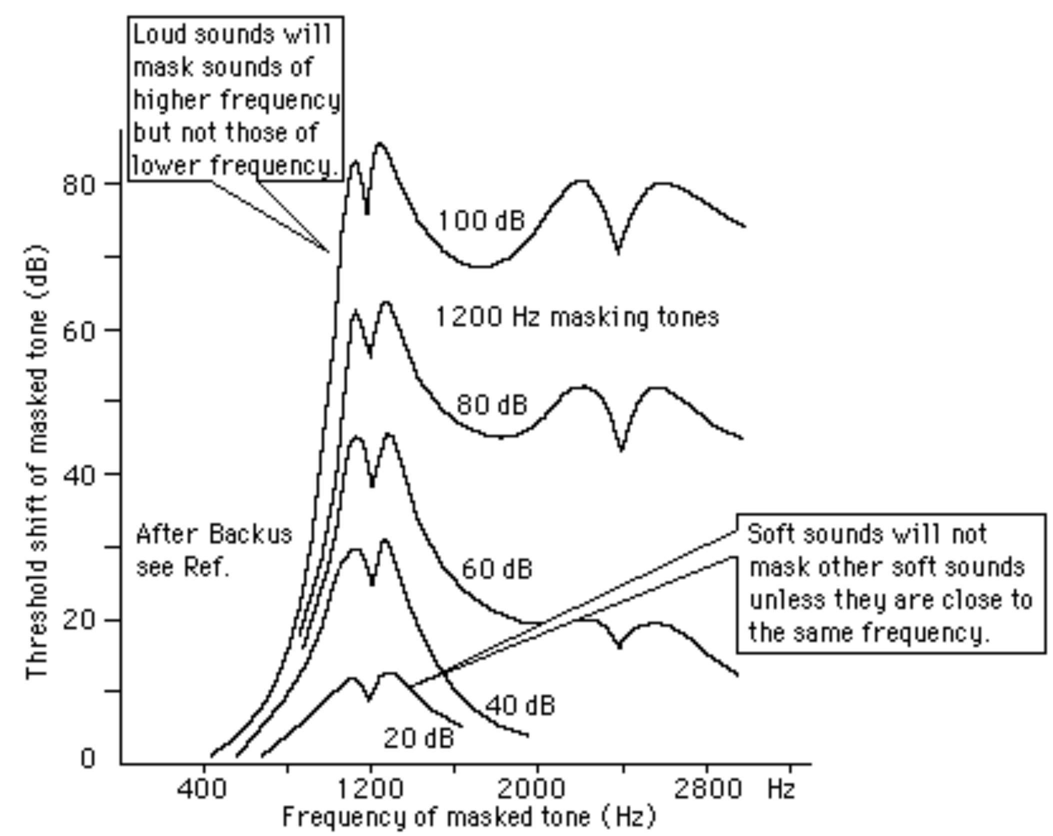


The region where roughness gets in defines a critical band, and that frequency region roughly corresponds to the segment of basilar membrane that gets excited by the tone at frequency f_1 . The sensation of roughness is related with that property of sound quality that is called consonance, and that can be evaluated along a continuous scale. The maximum degree of dissonance is found at about one quarter of critical band-width.

Masking



Masking by a 1200 Hz sound with different intensity



Asymmetry of the impulses traveling on the basilar membrane explains why low frequency sounds can mask higher frequency ones

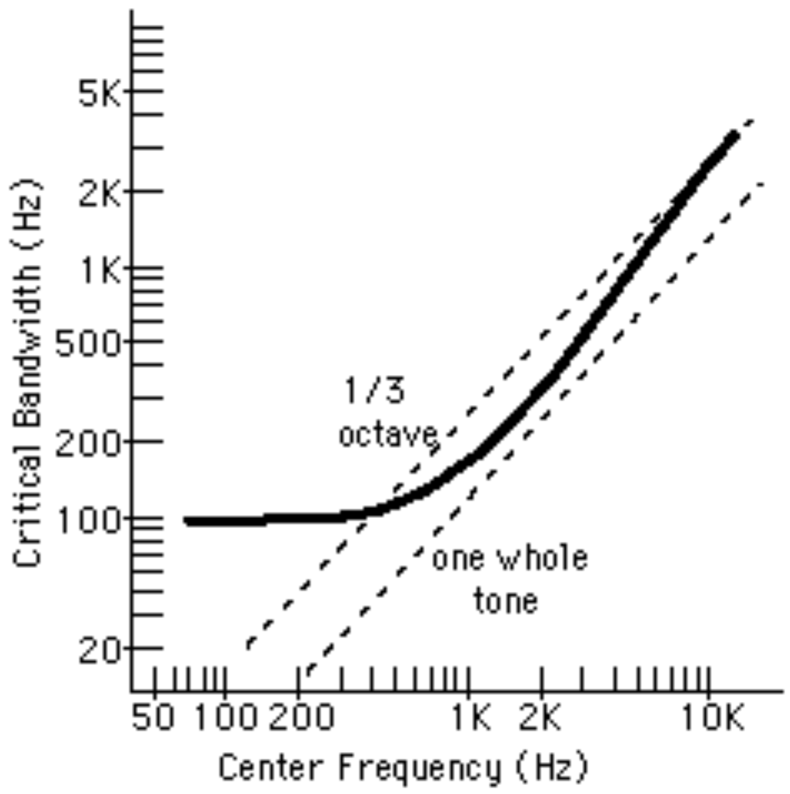
Masking

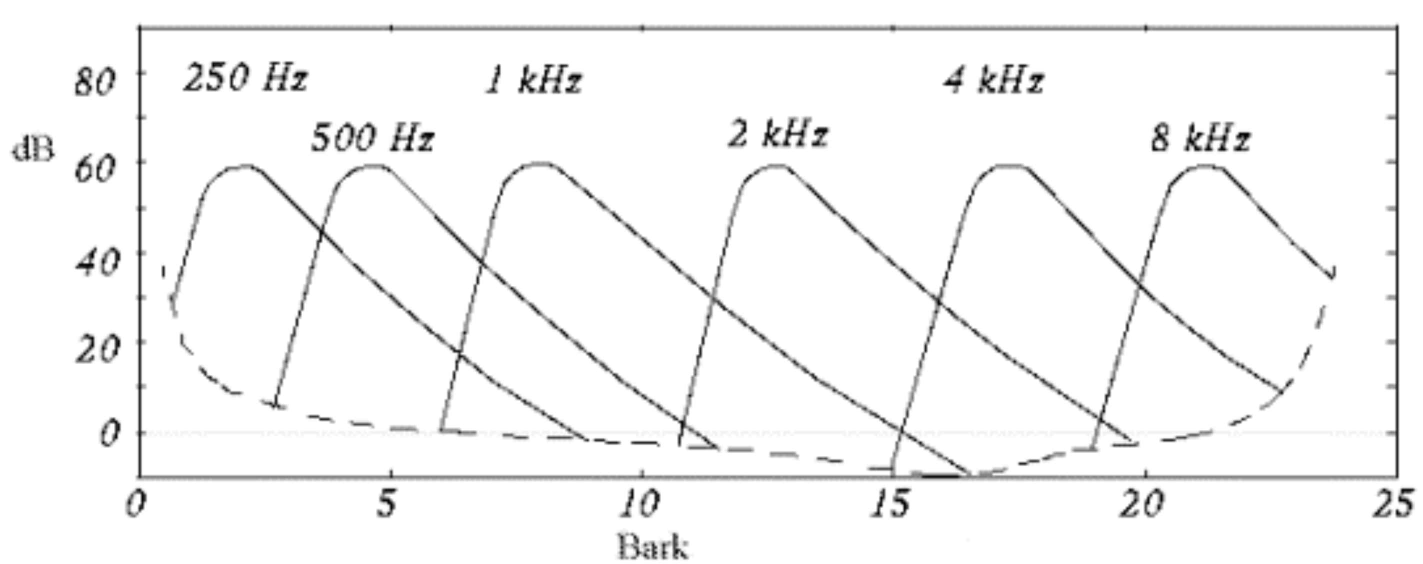
Masking depends on the width of the critical band at different frequencies.

1bark = width of critical band
1bark~ f/100 (for f < 500 Hz)
1bark~9+4log(f/100) (for f>500 Hz)

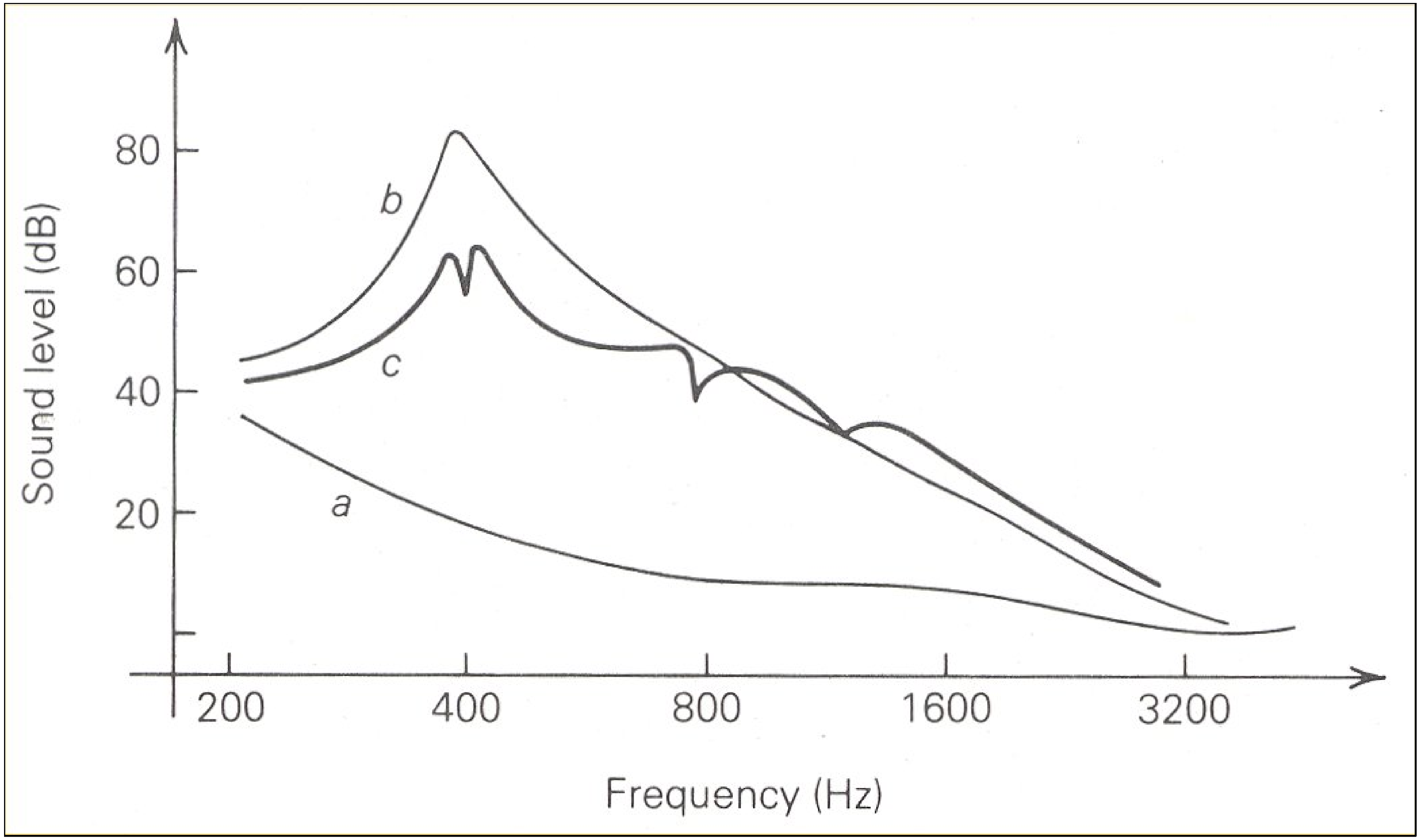
Masking depends in frequency also:
if one considers sound at 60dB we obtain the following masking curves:

Measure of the critical band





Masking curves



a. Threshold of hearing for pure tone

b. Masking by 365 Hz to 455 Hz 80 dB noise band

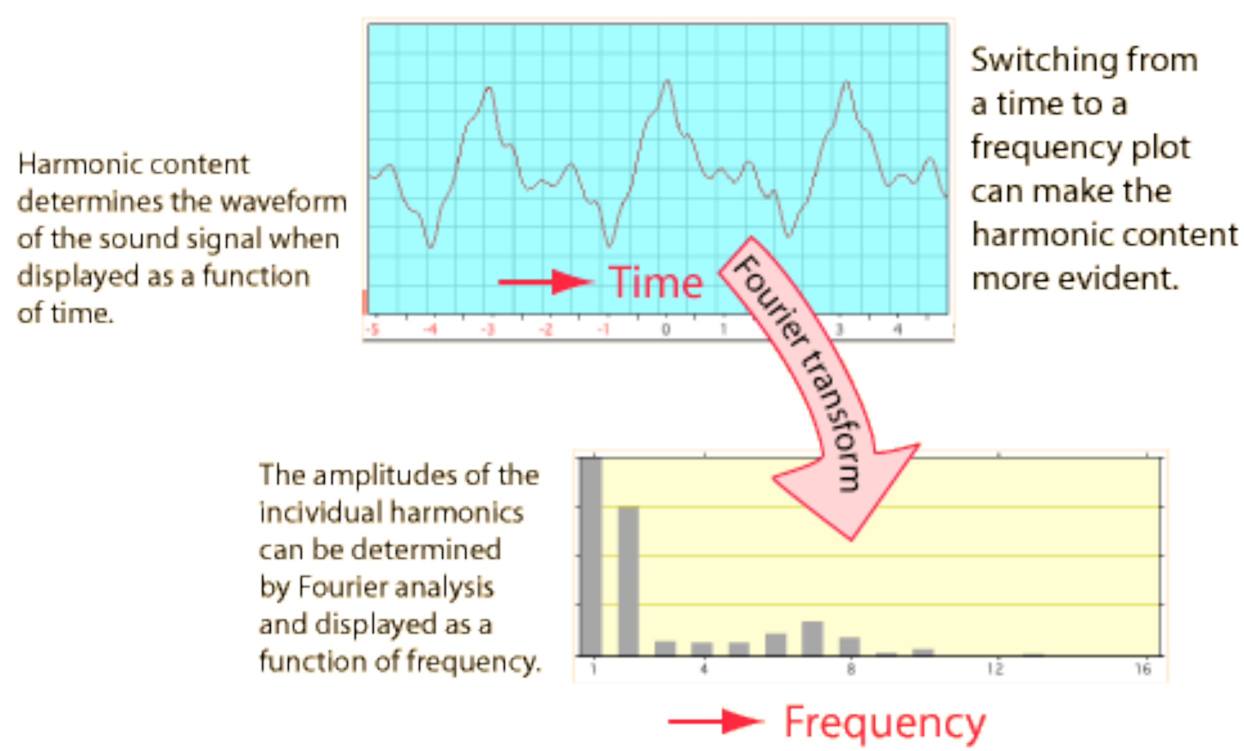
c. Masking by 400 Hz 80 dB tone

Complex sounds: timbre

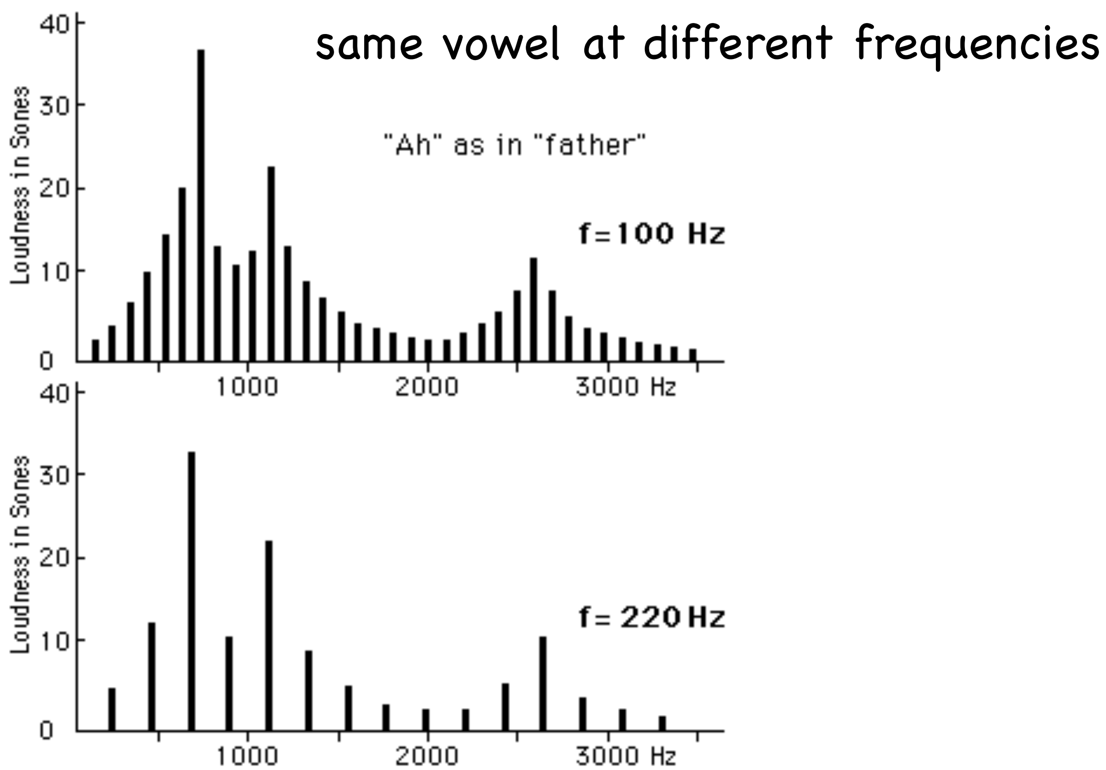
Sound "quality" or "timbre" describes those characteristics of sound which allow the ear to distinguish sounds which have the same pitch and loudness. Timbre is then a general term for the distinguishable characteristics of a tone.

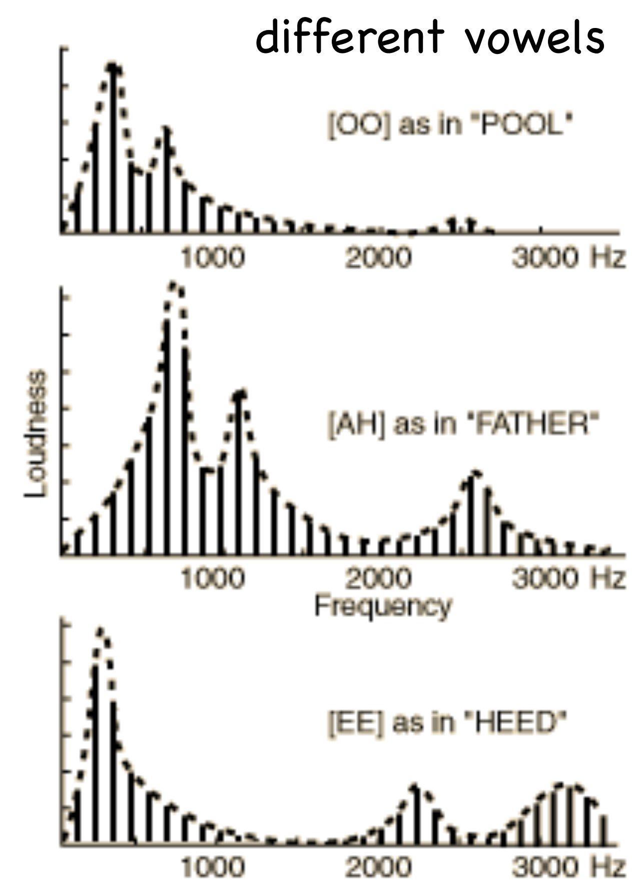
Timbre is mainly determined by the **harmonic content** of a sound and the dynamic characteristics of the sound such as **vibrato** and the **attack-decay** envelope of the sound.

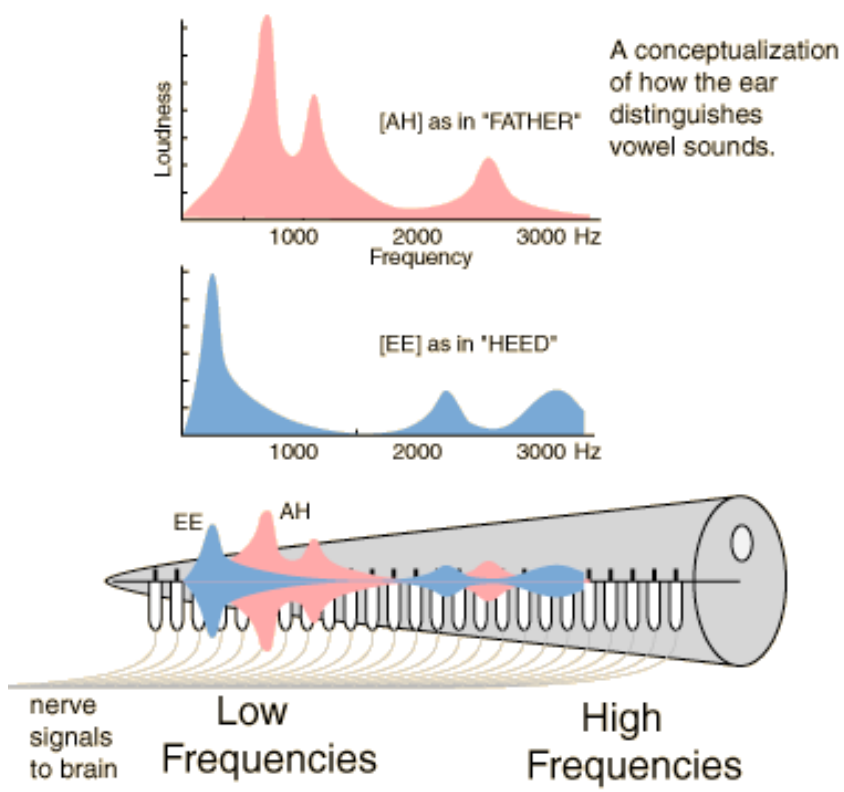
It is always possible to characterize a periodic waveform in terms of harmonics - such an analysis is called Fourier analysis. It is common practice to characterize a sound waveform by the spectrum of harmonics necessary to reproduce the observed waveform.

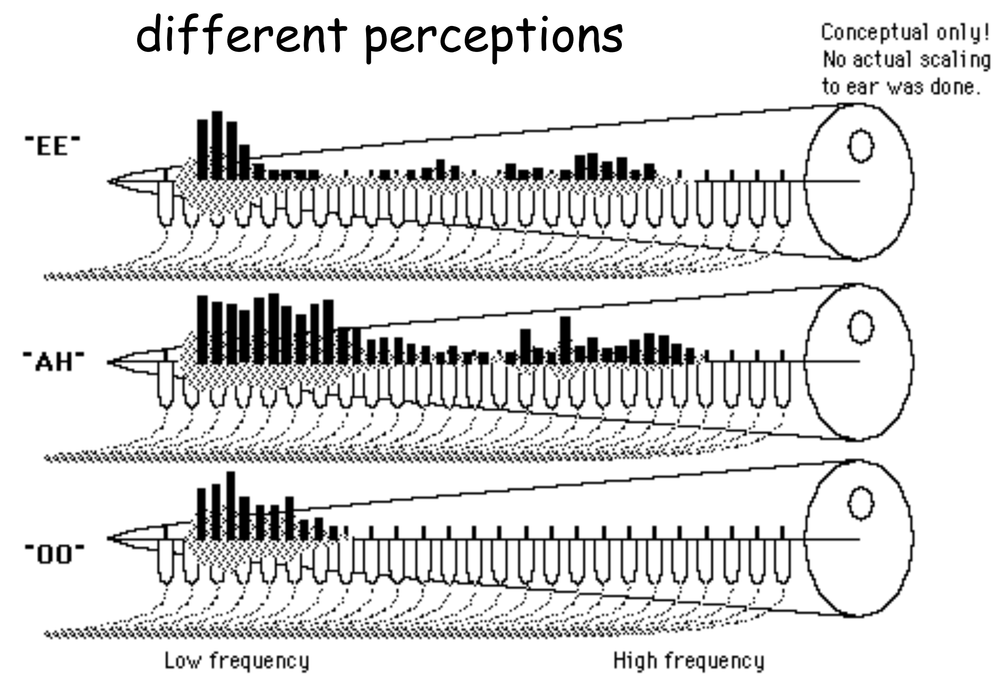


Vowels

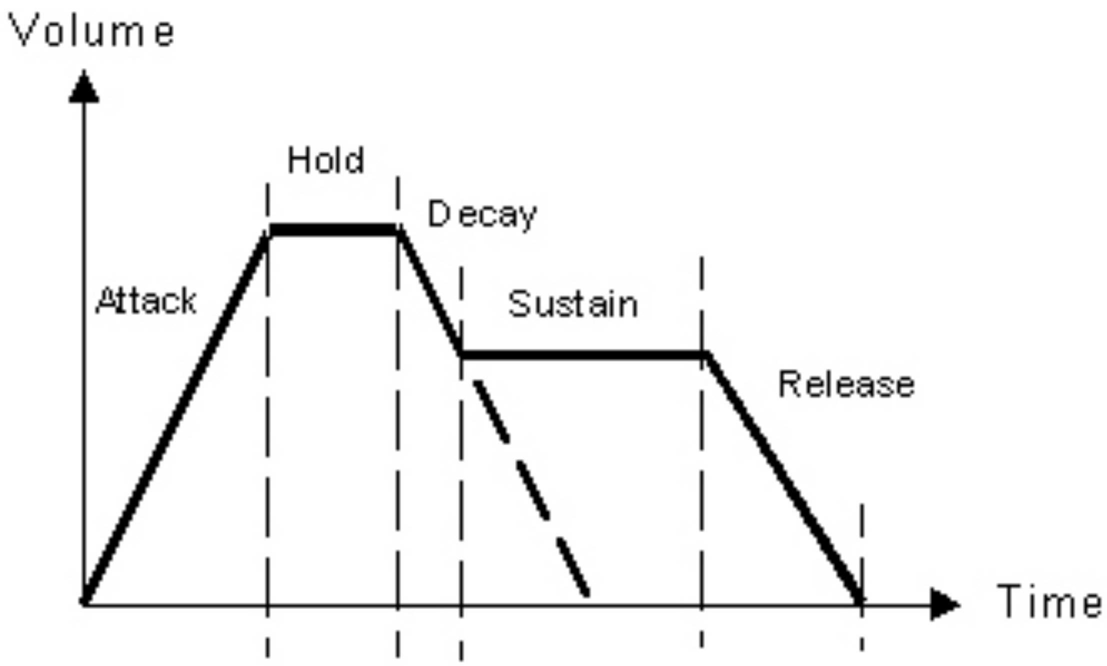

same vowel at different frequencies
"Ah" as in "father"
f=100 Hz
f= 220 Hz
Loudness in Sones
0
10
20
30
40
1000
2000
3000 Hz


different vowels
[OO] as in "POOL"
[AH] as in "FATHER"
[EE] as in "HEED"
Loudness
Frequency
1000
2000
3000 Hz


A conceptualization of how the ear distinguishes vowel sounds.
[AH] as in "FATHER"
[EE] as in "HEED"
Loudness
Frequency
1000
2000
3000 Hz
EE
AH
nerve signals to brain
Low Frequencies
High Frequencies


different perceptions
Conceptual only! No actual scaling to ear was done.
"EE"
"AH"
"OO"
Low frequency
High frequency

Sound envelope



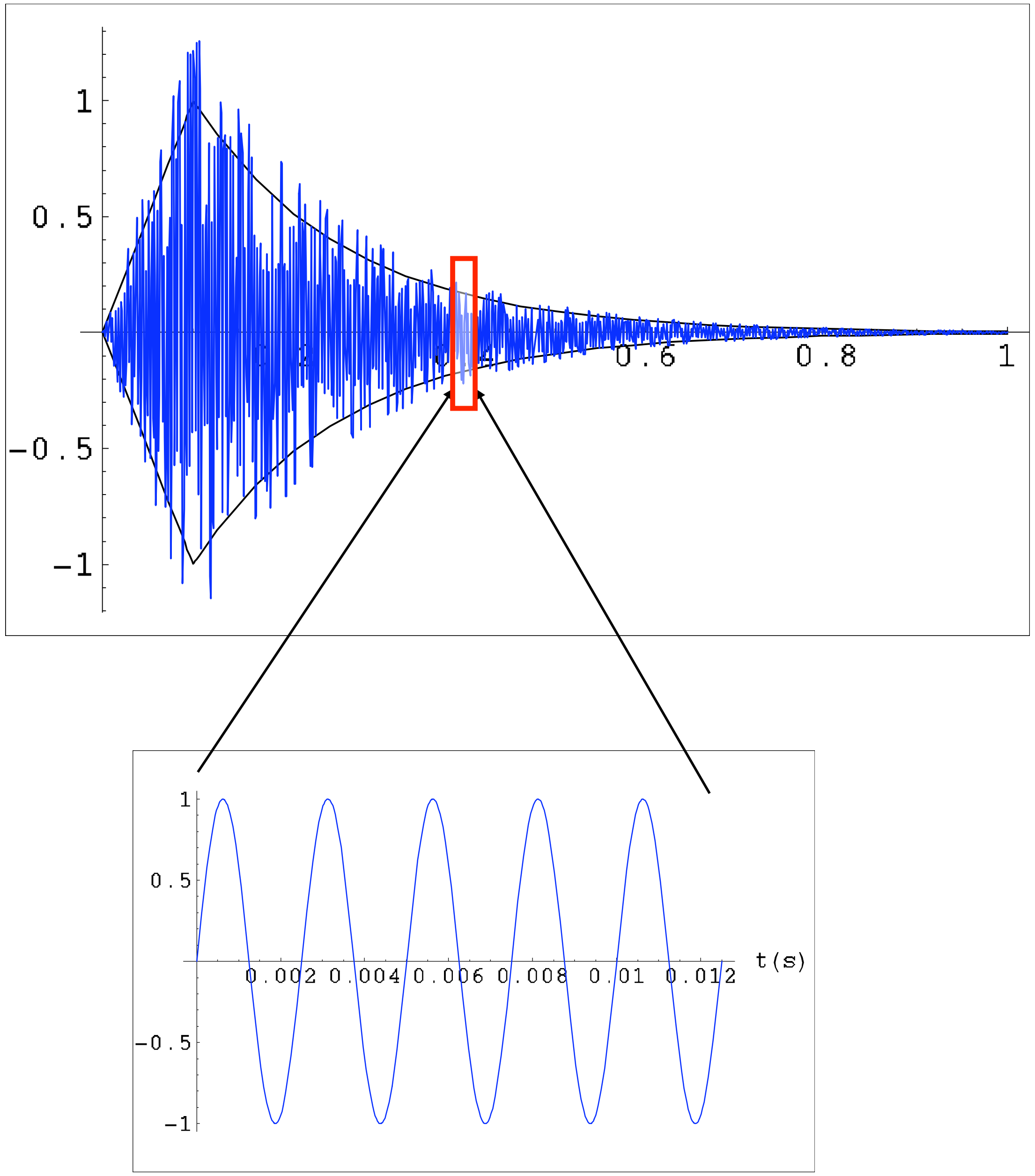
The volume envelope

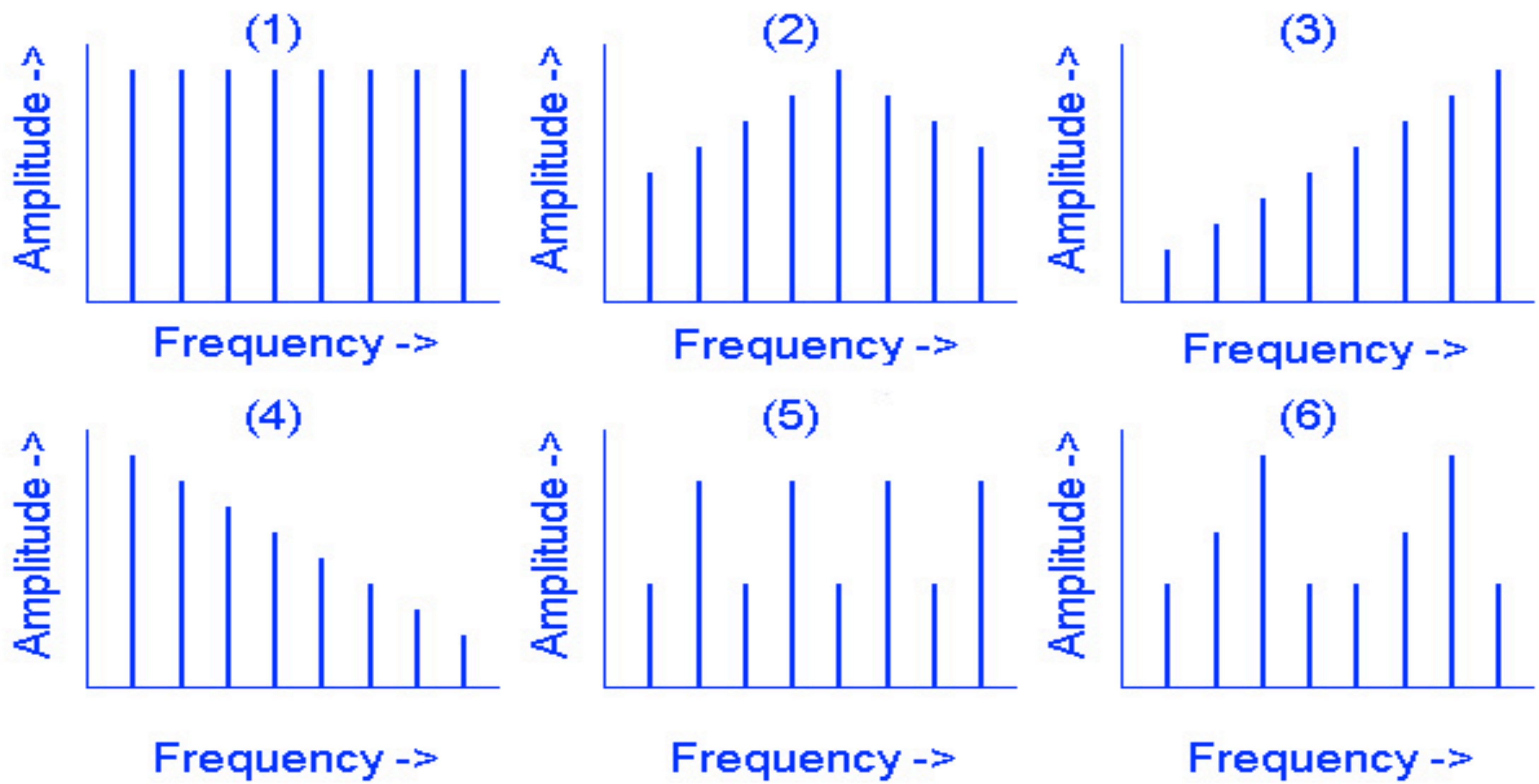
plucked guitar string

striking a cymbal with a stick

- **Attack:** How quickly the sound reaches full volume after the sound is activated
- **Decay:** How quickly the sound drops to the sustain level after the initial peak.
- **Sustain:** The "constant" volume that the sound takes after decay until the note is released. Note that this parameter specifies a volume level rather than a time period.
- **Release:** How quickly the sound fades when a note ends (the key is released). Often, this time is very short. An example where the release is longer might be a percussion instrument.

Attack and decay transients

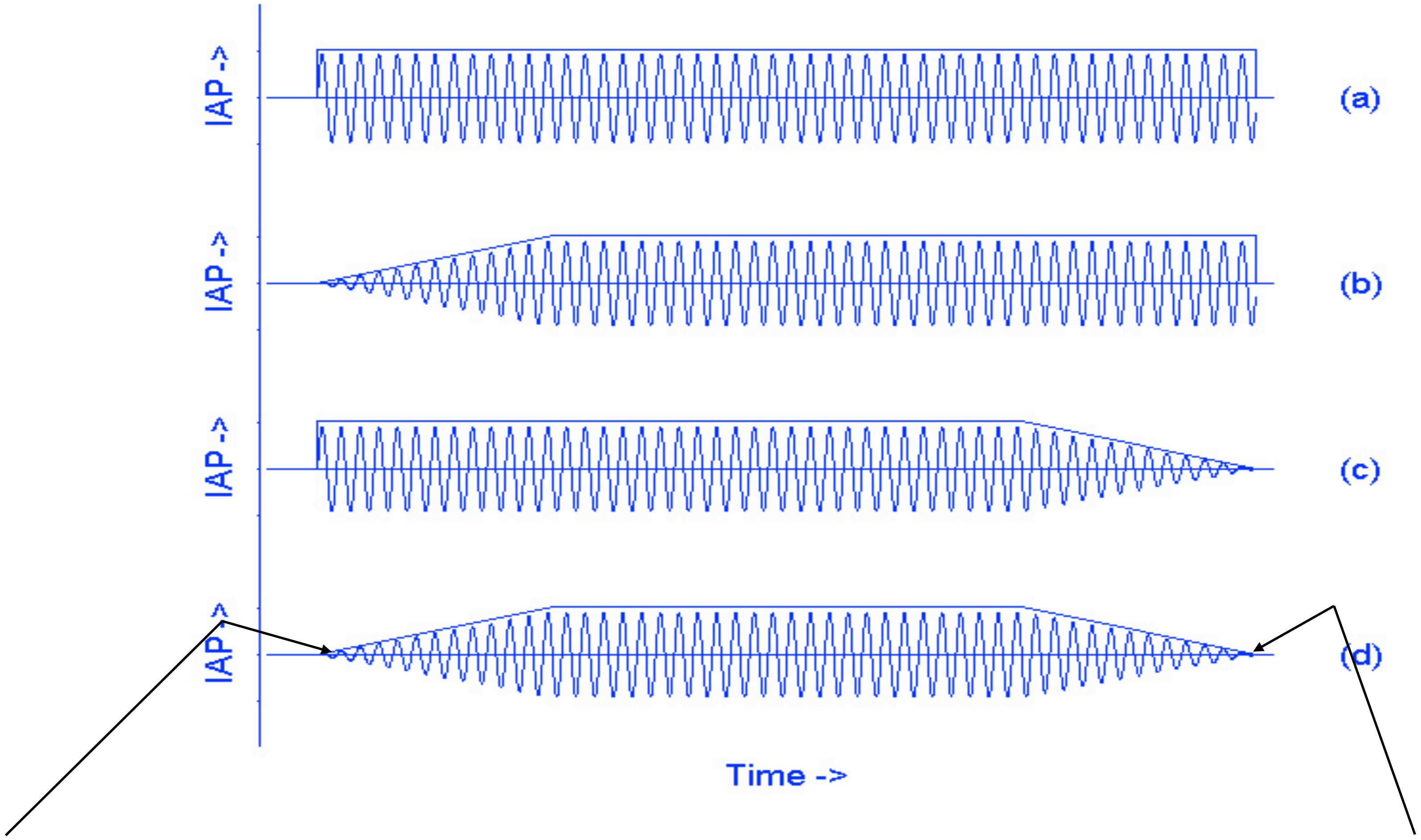




Note the similarities in pitch (due to constant F_0/harmonic spacing) and the differences in **timbre** or **sound quality**.

Amplitude envelope

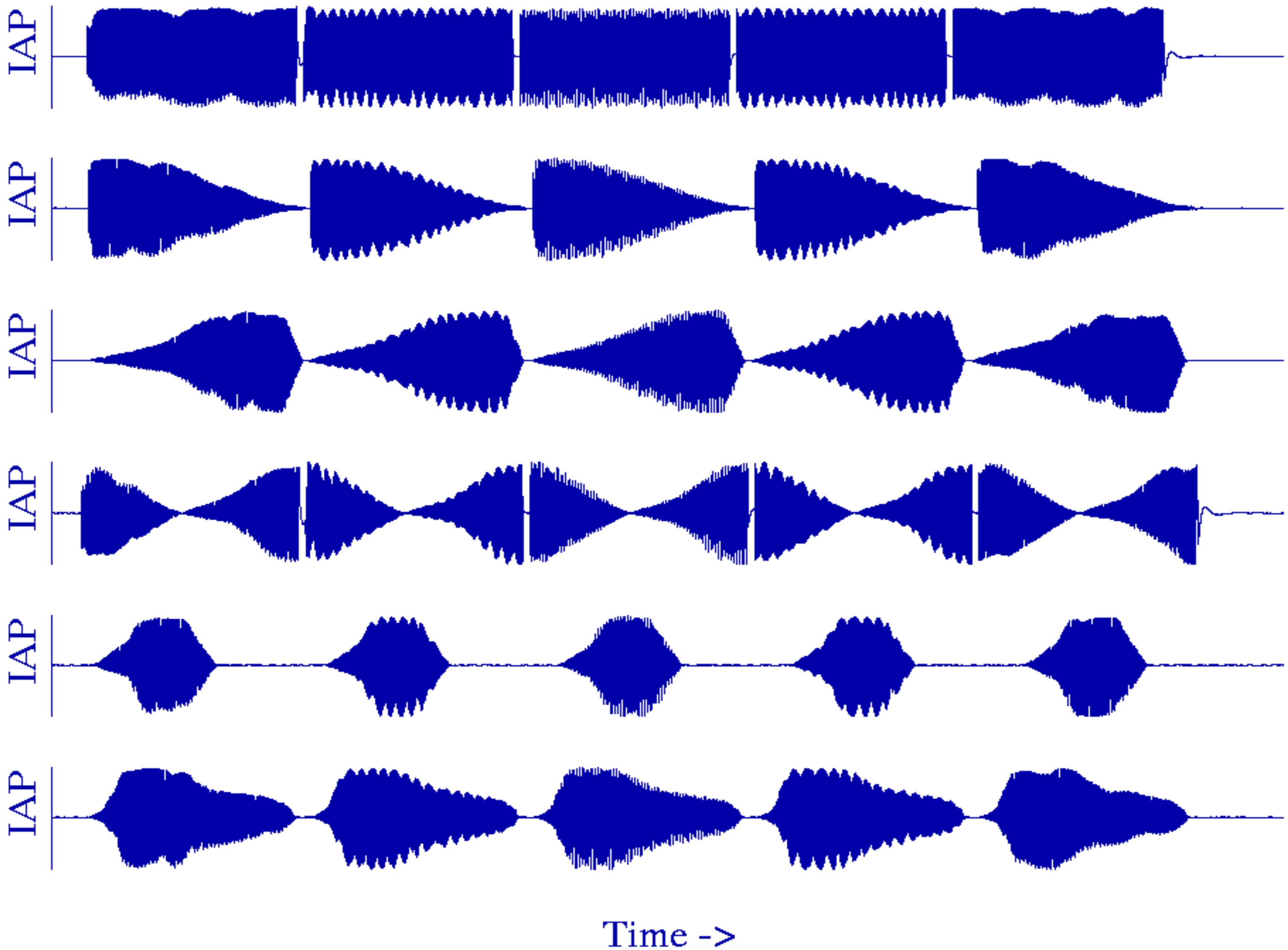
- Timbre also affected by **amplitude envelope**
- sometimes called the **amplitude contour** or **energy contour** of the sound wave
- the way sounds are turned on and turned off



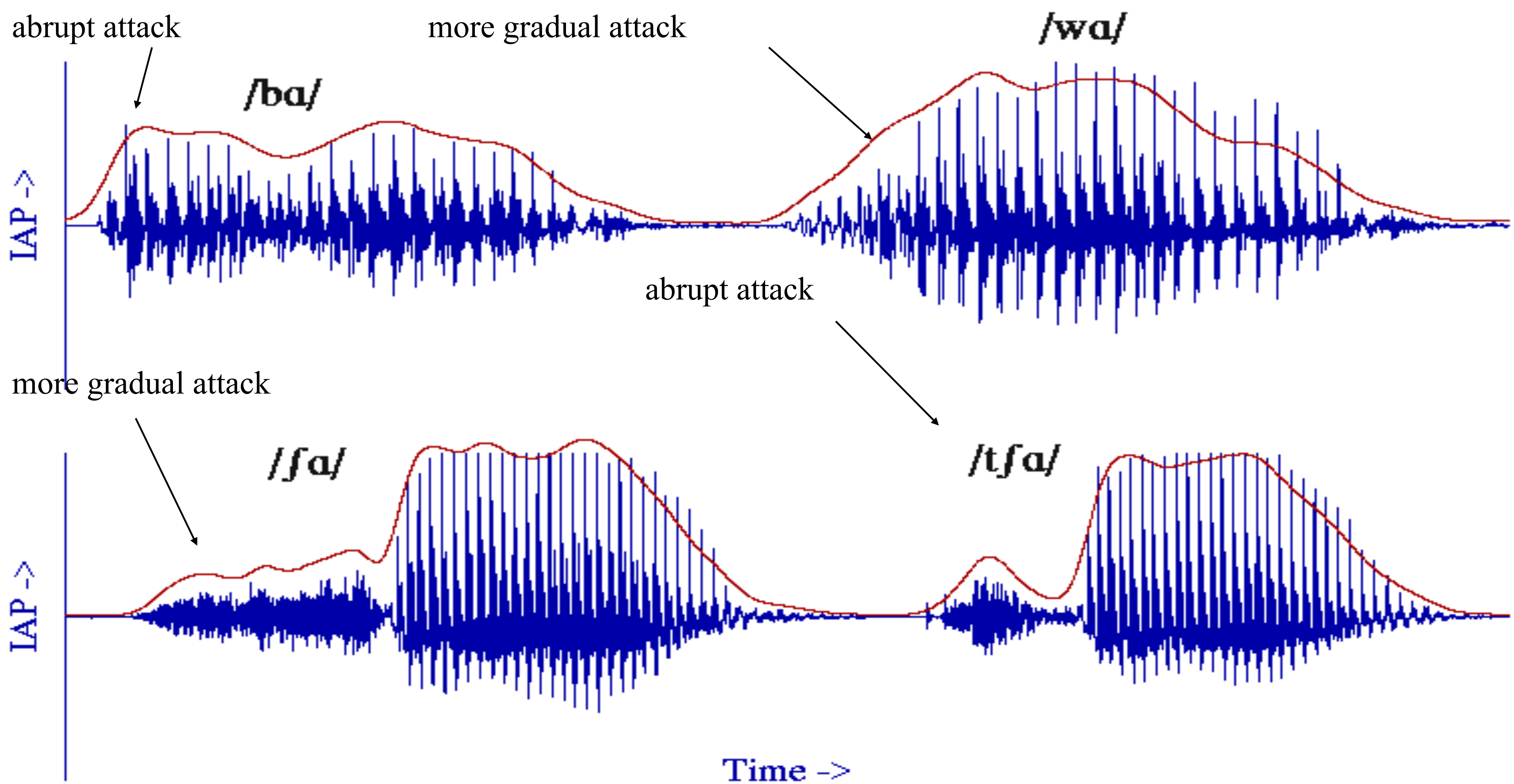
The attack especially has a large effect on timbre.

Same melody, same spectrum envelope (if sustained), different amplitude envelopes (i.e., different attack and decay characteristics).

Note differences in timbre or sound quality as the amplitude envelope varies.



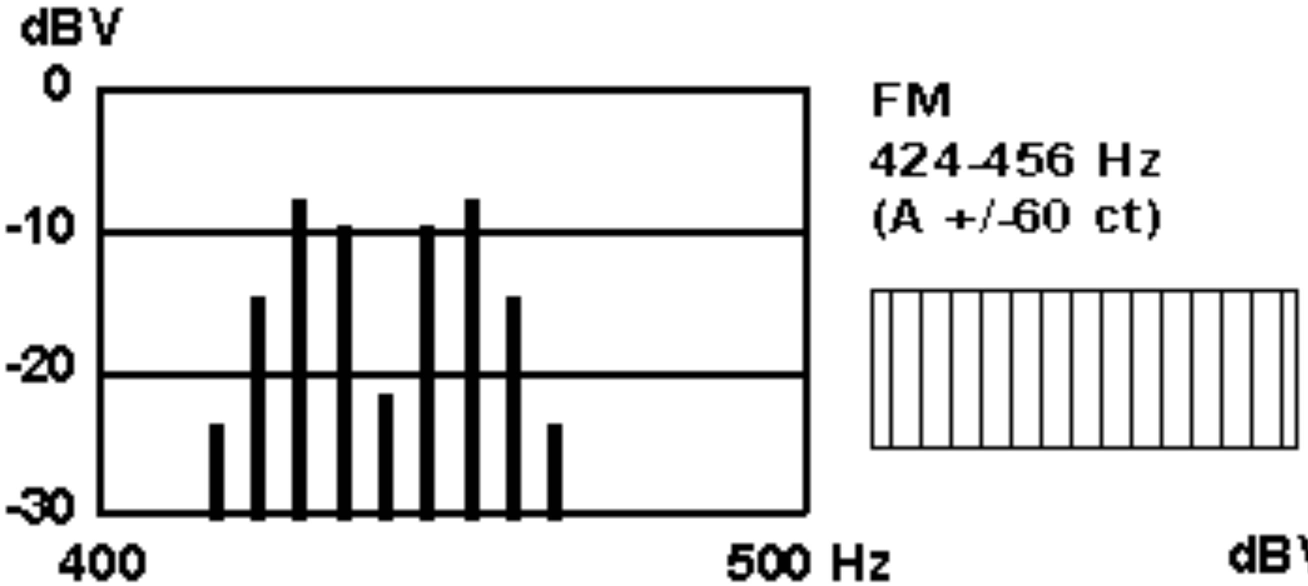
Timbre differences related to amplitude envelope also play a role in speech. Note the **differences in the shape of the attack** for /b/ vs. /w/ and /S/ vs. /tS/.

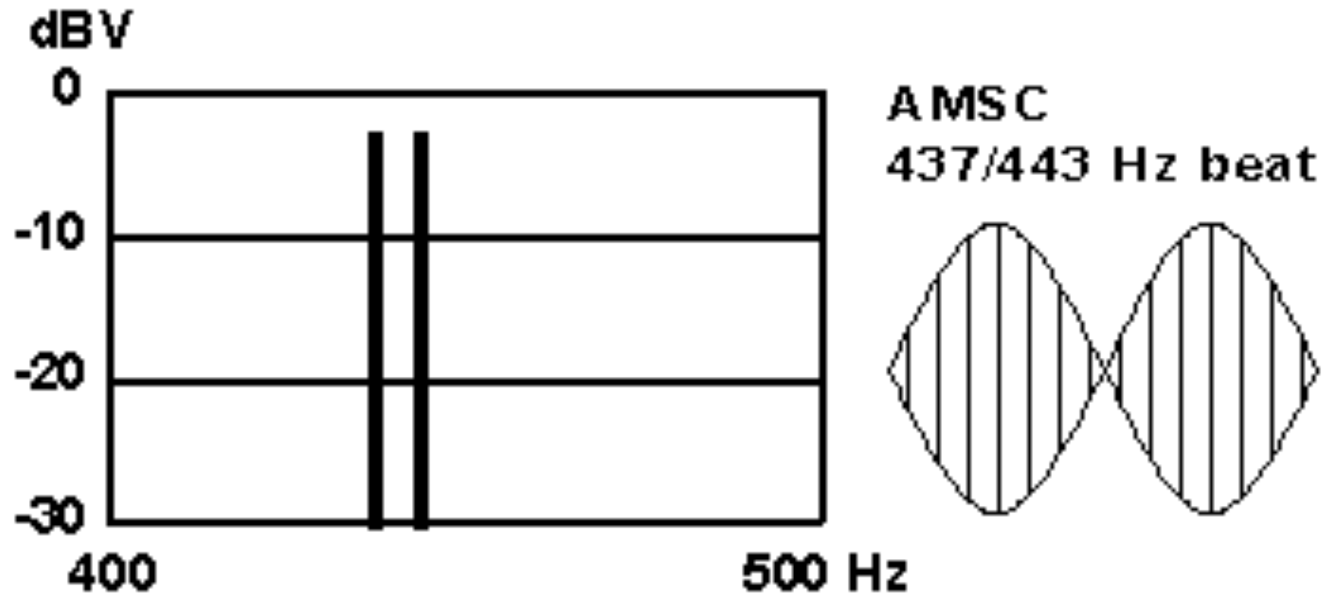


Vibrato & Tremolo

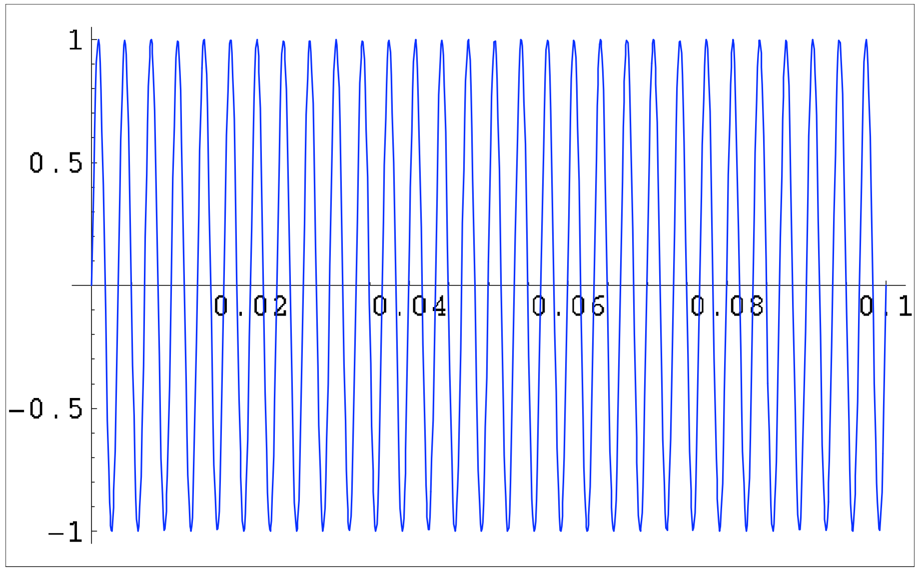
- The ordinary definition of **vibrato** is "periodic changes in the pitch of the tone", and the term **tremolo** is used to indicate periodic changes in the amplitude or loudness of the tone.

- So vibrato could be called FM (frequency modulation) and tremolo could be called AM (amplitude modulation) of the tone.



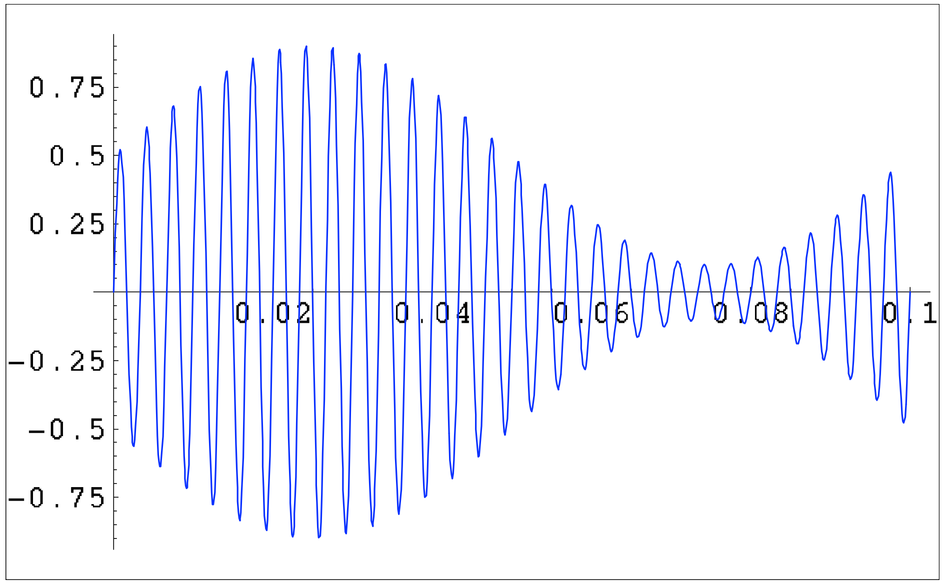


Vibrato



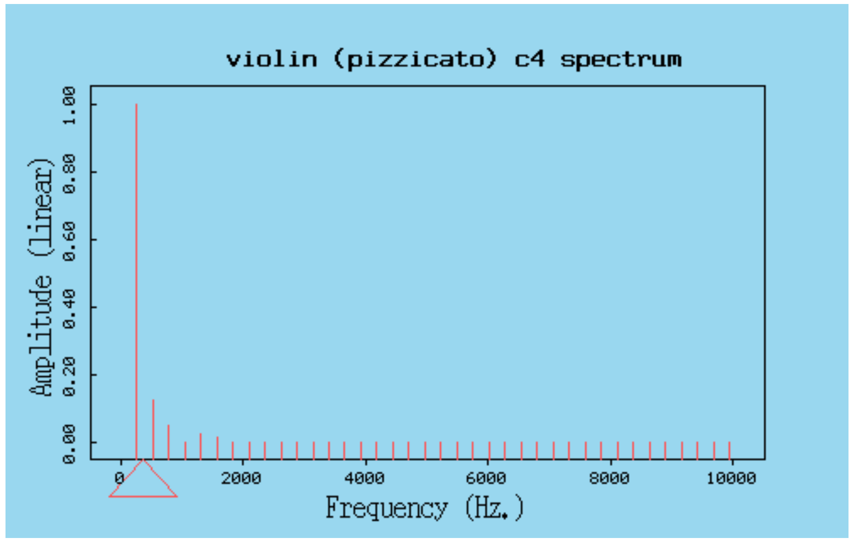
oscillation in frequency

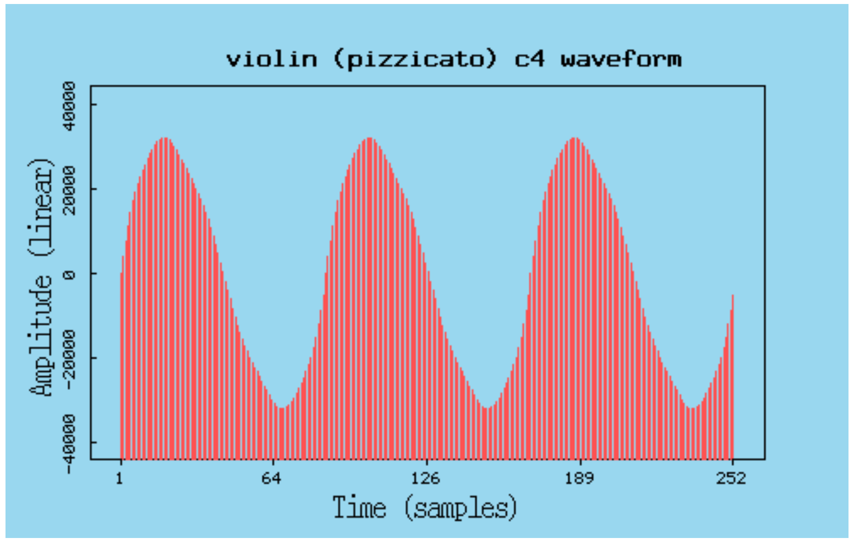
Tremolo

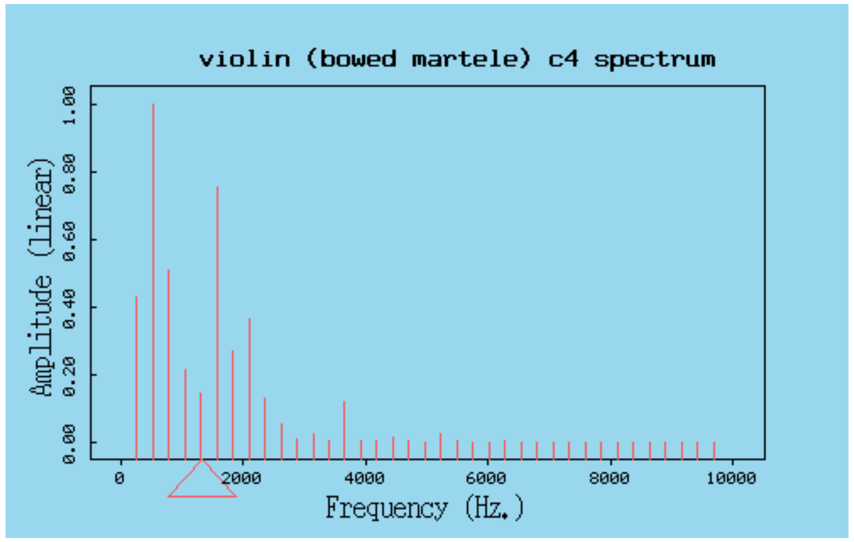


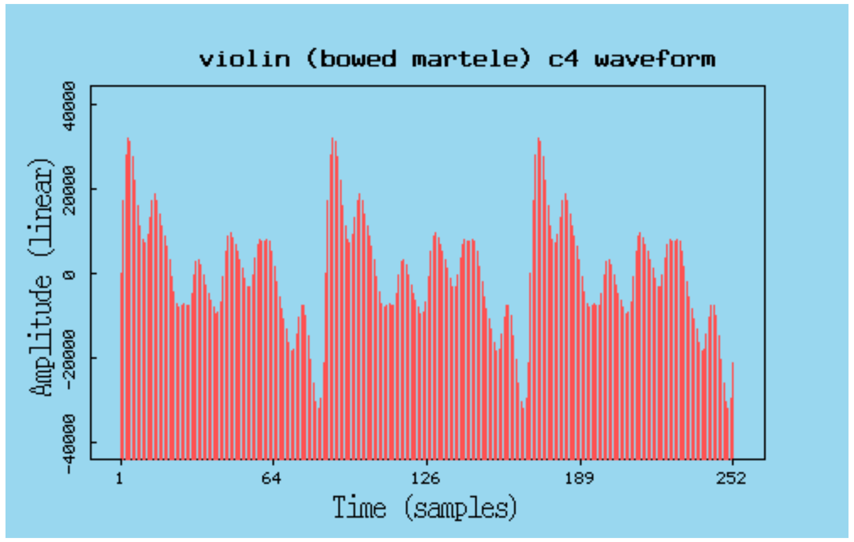
oscillation in amplitude

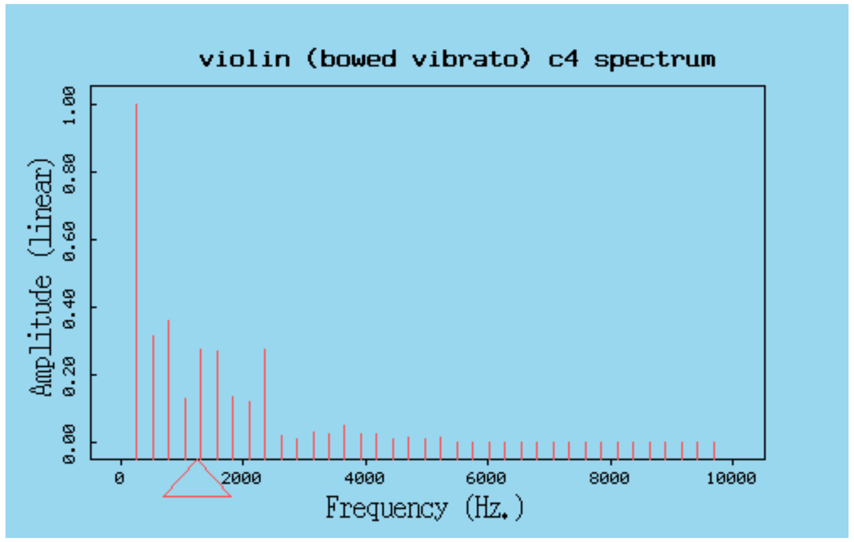
Example: Violin

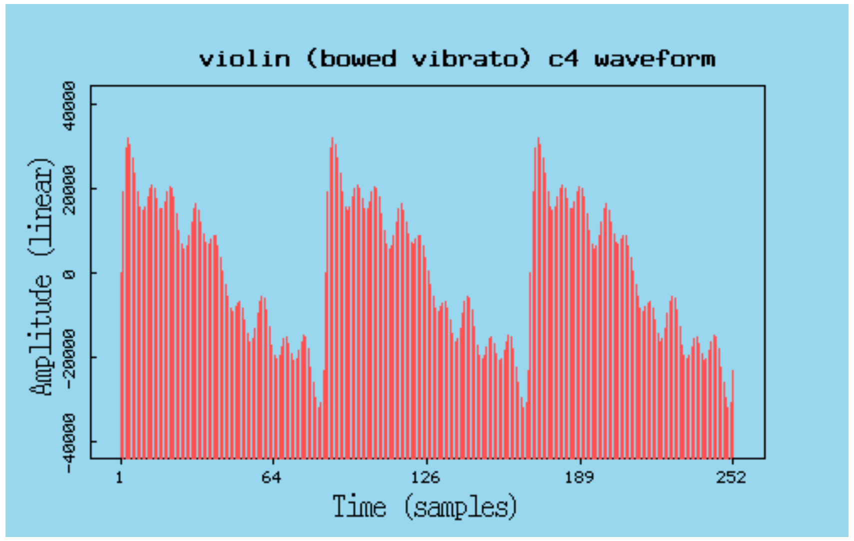












Physical and "acoustic" properties