Auditory Representation

Wu Xihong

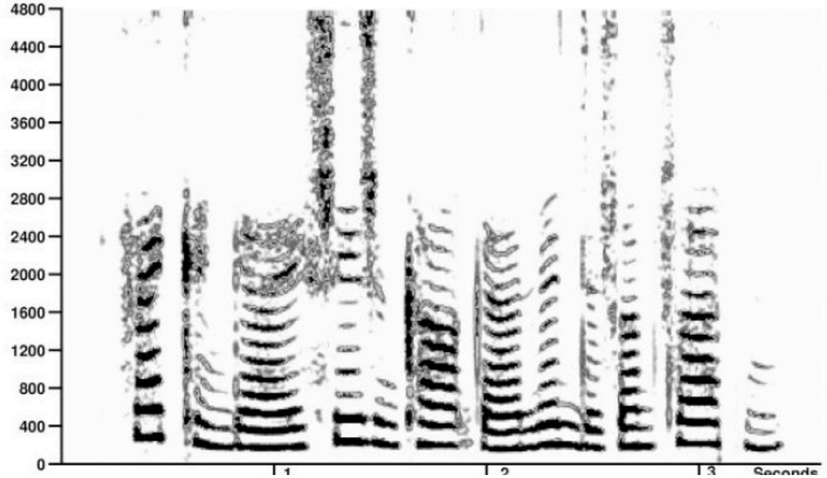
School of Artificial Intelligence

Computational Goals in Audition

- Orient to sounds
- Identify sounds
- Communication

Different points from visual perception:

- wavelength
- sound objects



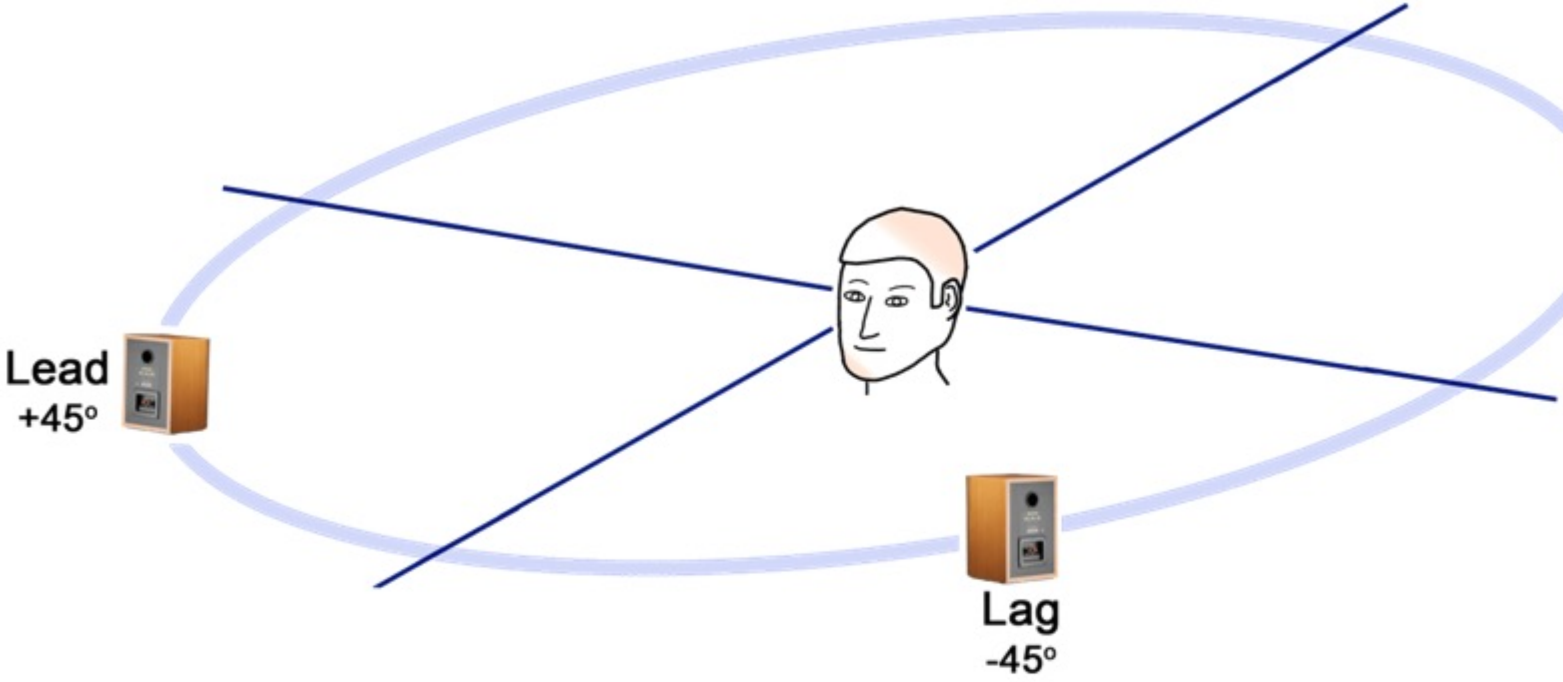
What makes these hard in a natural environment

- background noise
- multiple sounds
- environmental reflection

- Auditory representation
- Pitch and timbre of speech and music
- Auditory scene analysis
  - Where
    - sound source localization
  - What
    - speech separation
- Speech Perception

Some examples:

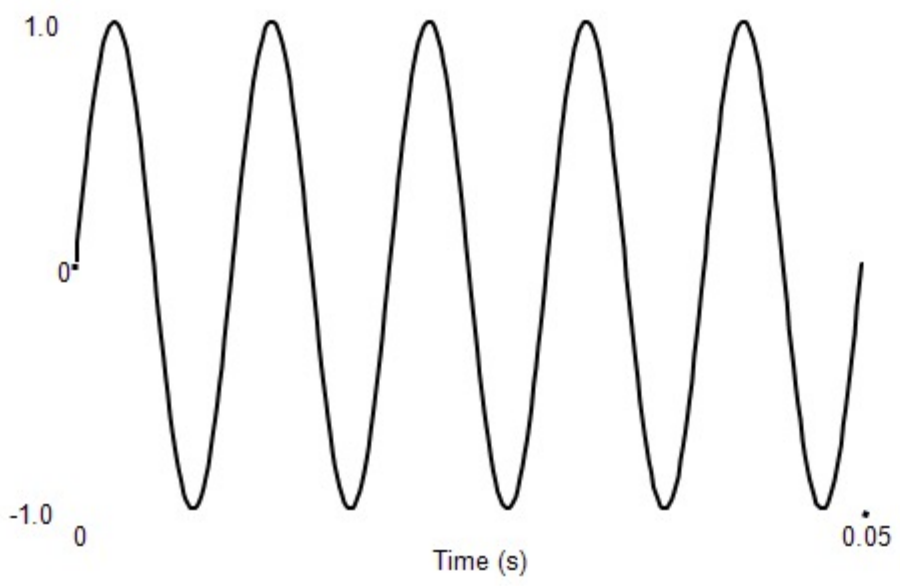
- Precedence effect
- Headphone, recording
- Nonlinear system



Applications

- Environmental surveillance
- Music codec
- Virtual 3D audio
- music retrieval
- Machine speech recognition, speech translation, human-machine dialog
- Hearing rehabilitation

Frequency: 100-Hz Sine Wave

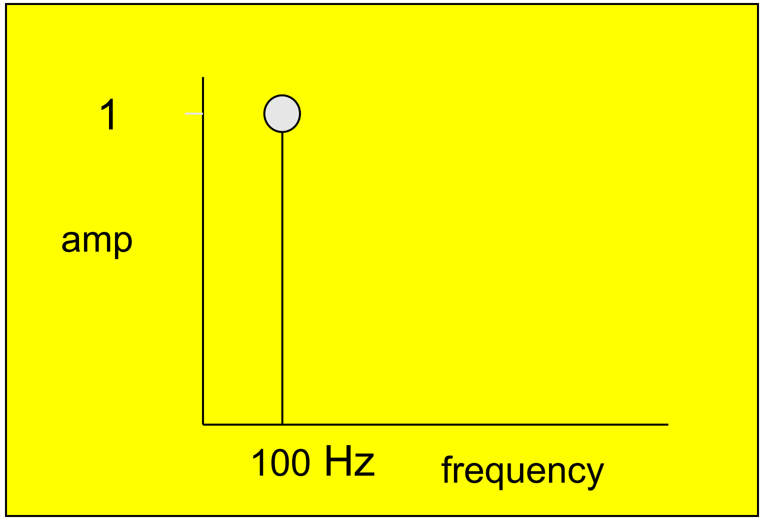


Waveform

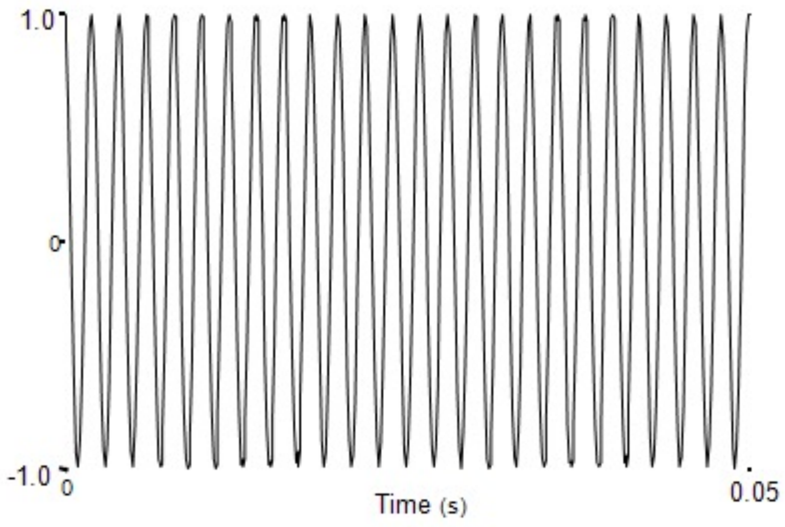
Amplitude against time

Spectrum

Amplitude against frequency



Frequency: 500-Hz Sine Wave

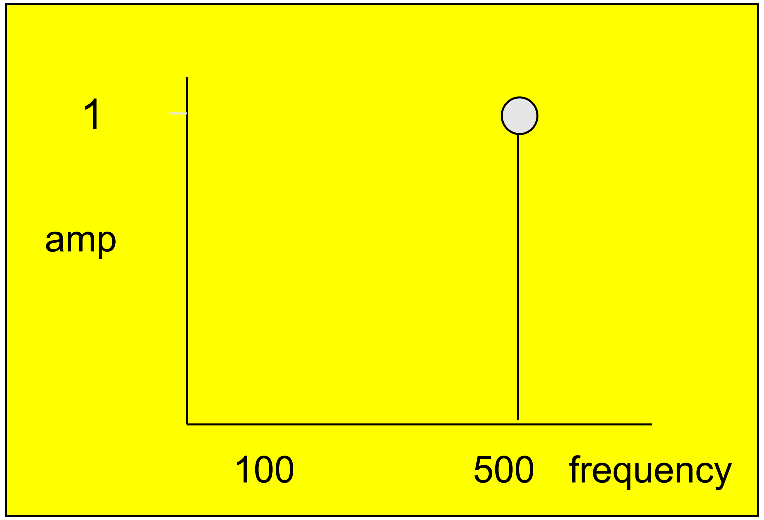


Waveform

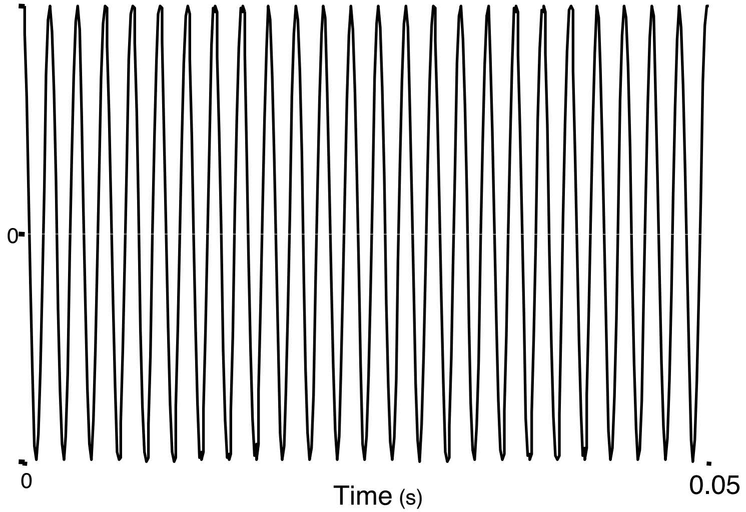
Amplitude against time

Spectrum

Amplitude against frequency

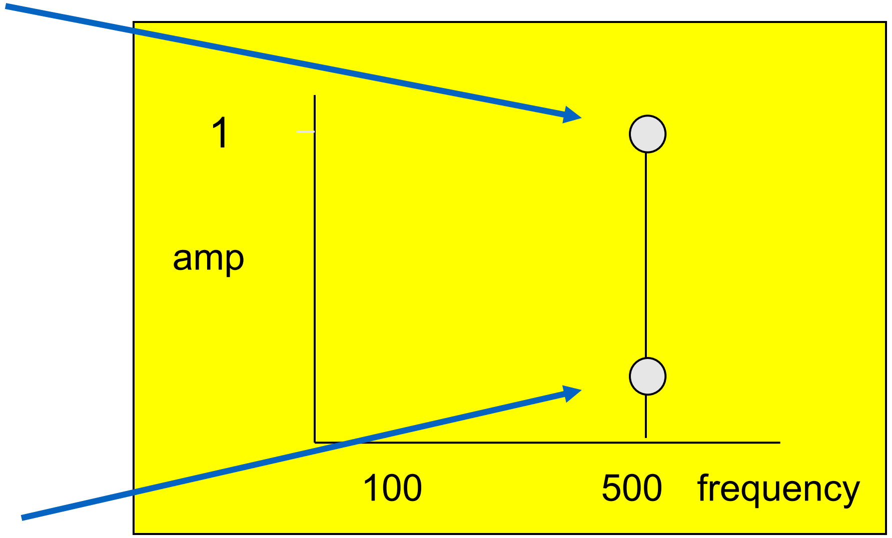


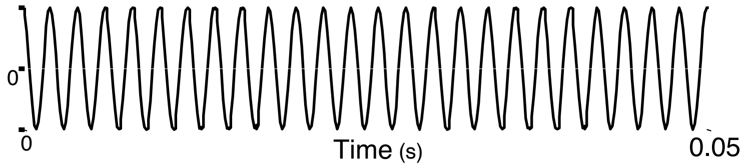
Amplitude: 500-Hz Sine Wave



Spectrum

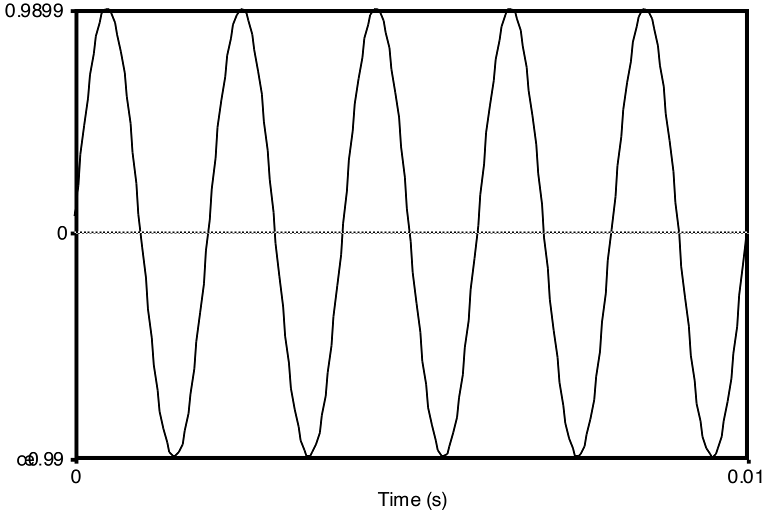
Amplitude against frequency



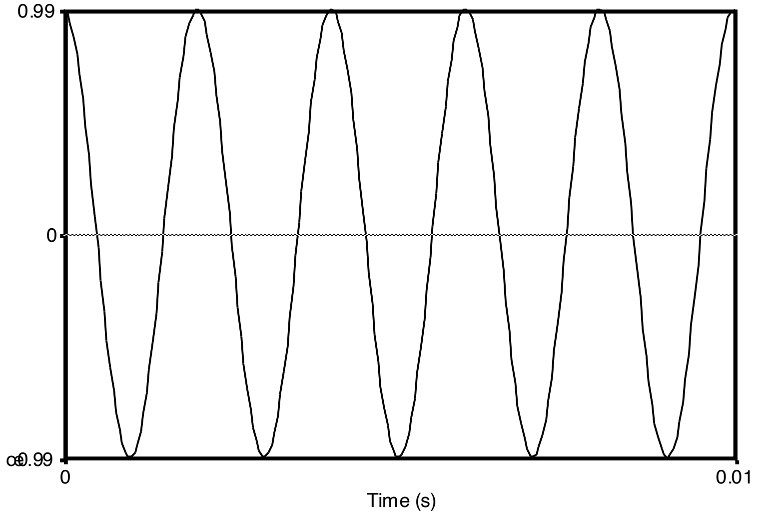


Phase: 500-Hz Sine Wave

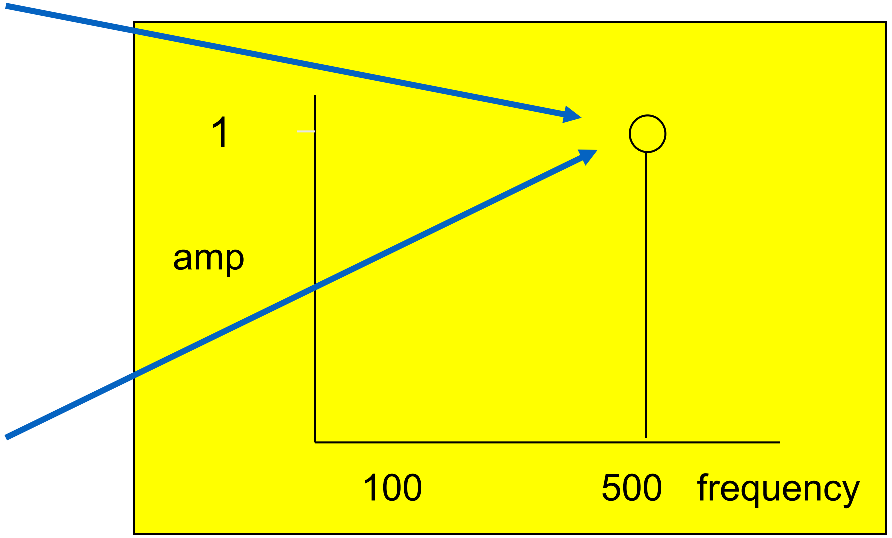
sine



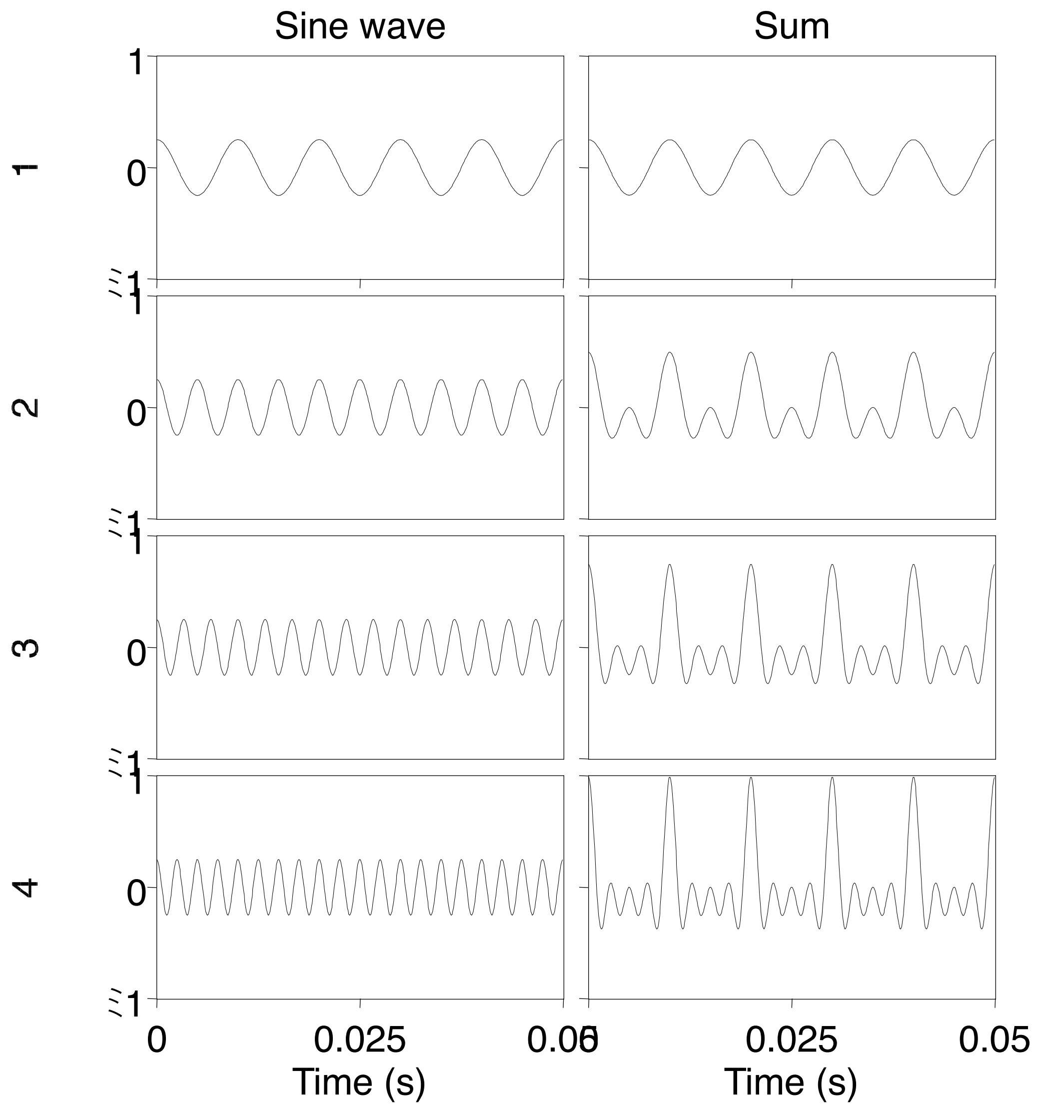
cosine



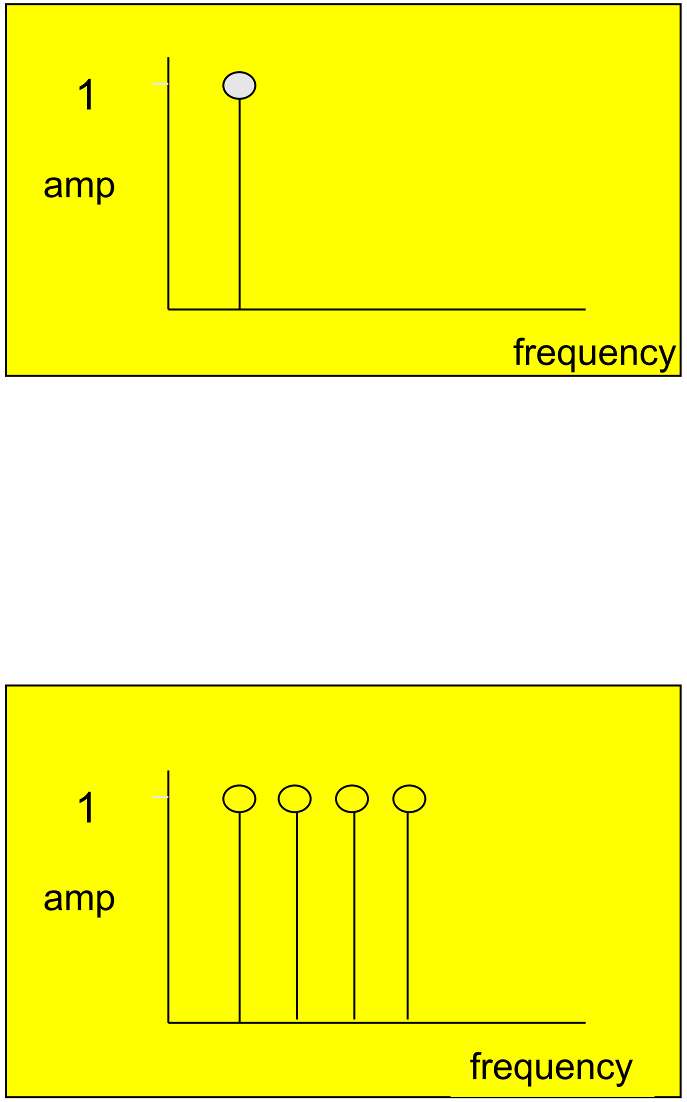
The amplitude spectrum does not show phase



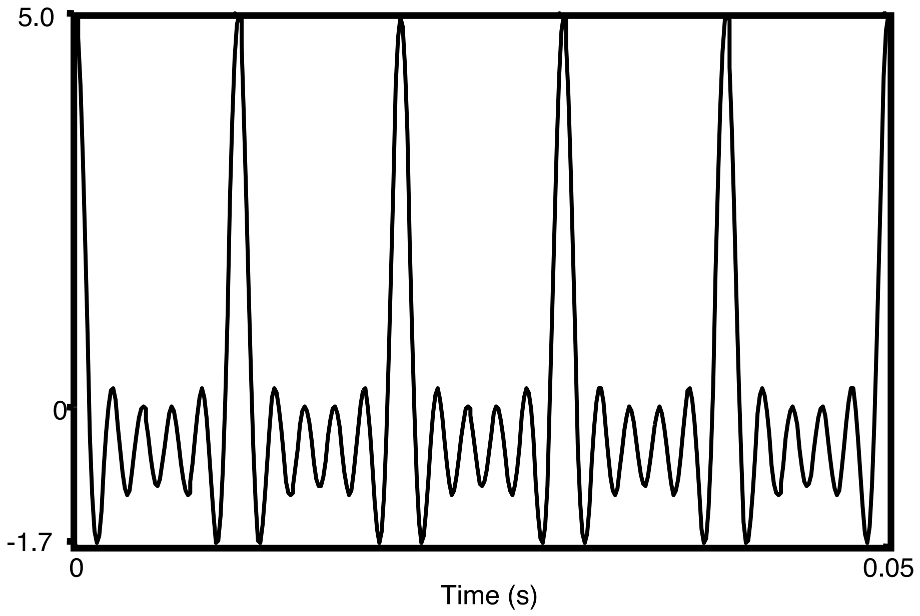
adding sine waves



Spectrum of Sum



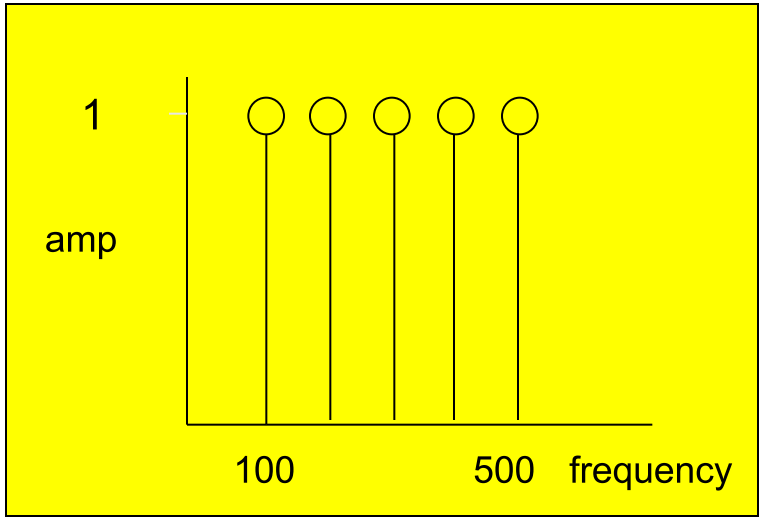
100-Hz fundamental Complex Wave



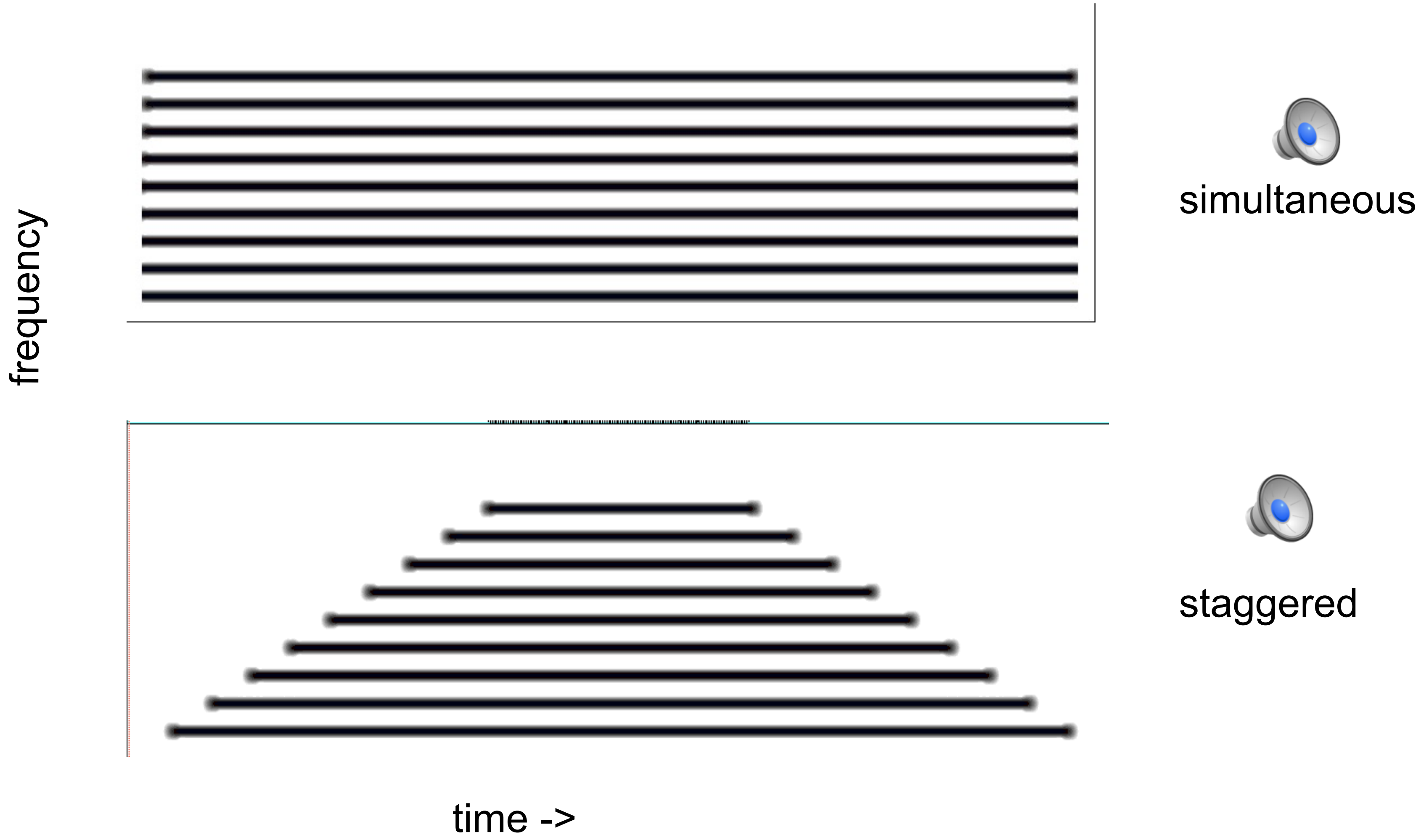
Waveform

Amplitude against time

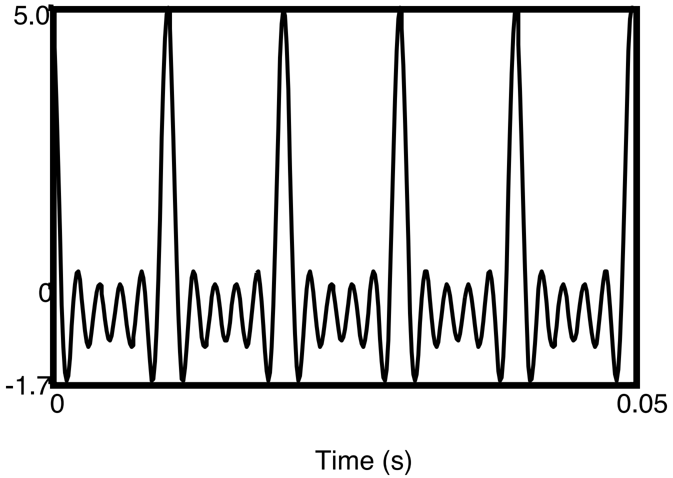
Spectrum Amplitude against frequency

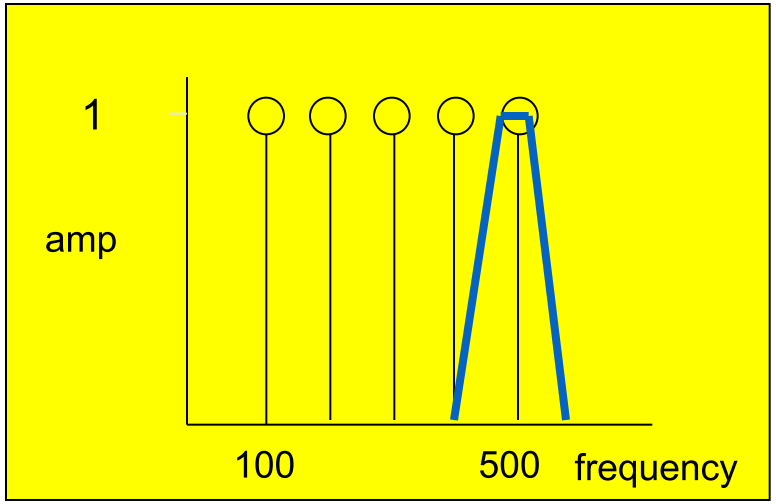


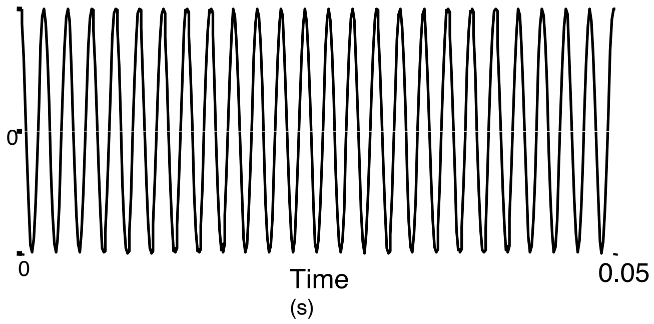
Staggered sine-waves adding

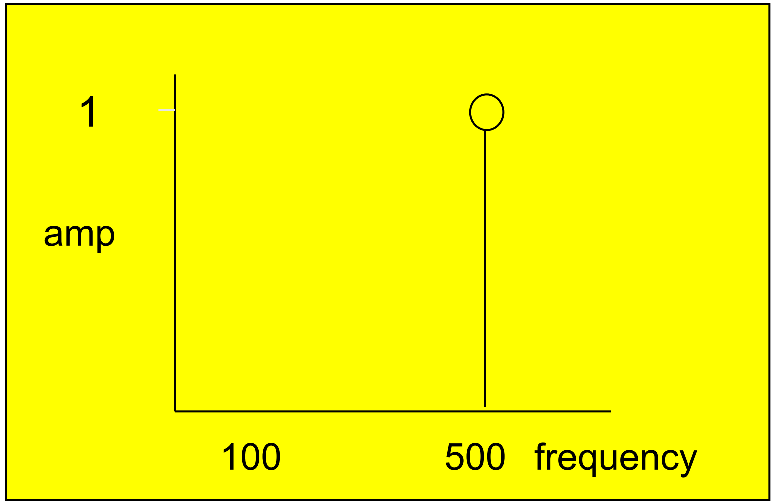


Bandpass filtering (narrow)

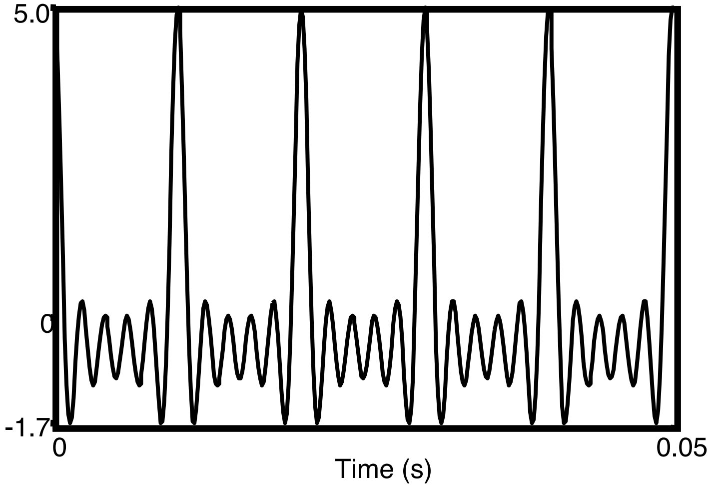


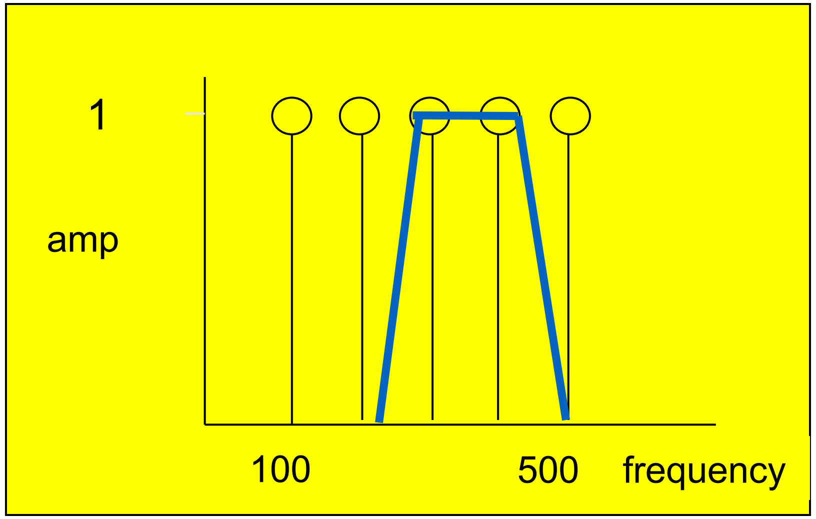


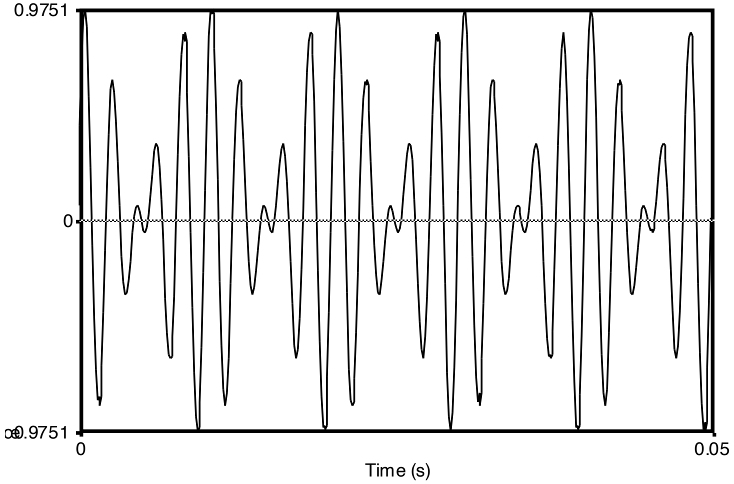


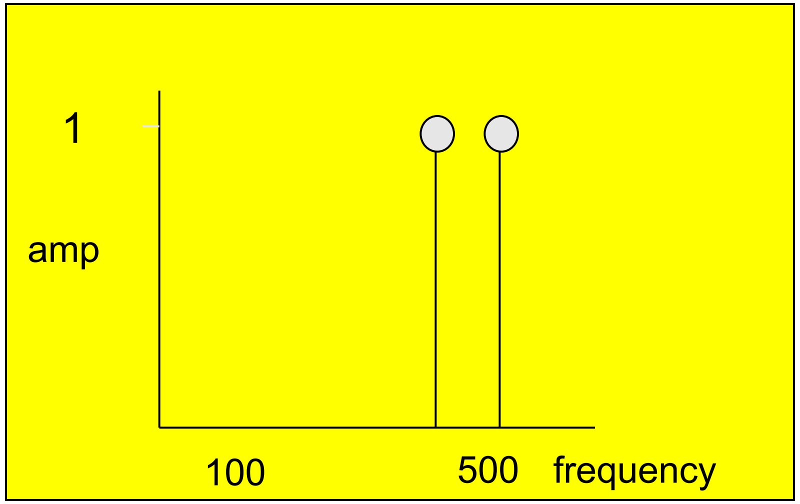


Bandpass filtering (wide)



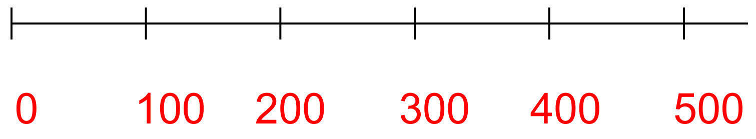




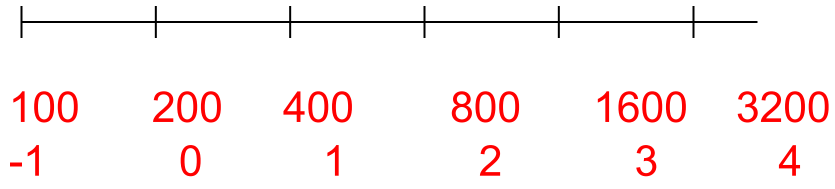


The linear vs log scales

Linear： equal distances
represent equal differences



Log： equal distances
represent equal ratios



e.g. Piano keyboard frequencies

Octave = doubling of frequency

deciBel (dB) scale

Sound A is x dB more intense than sound B when:

$$x = 10*\log_{10} (\text{energy of A / energy of B})$$

or $$x = 20*\log_{10} (\text{amp of A / amp of B})$$

So if A is 20 watts and B is 10 watts

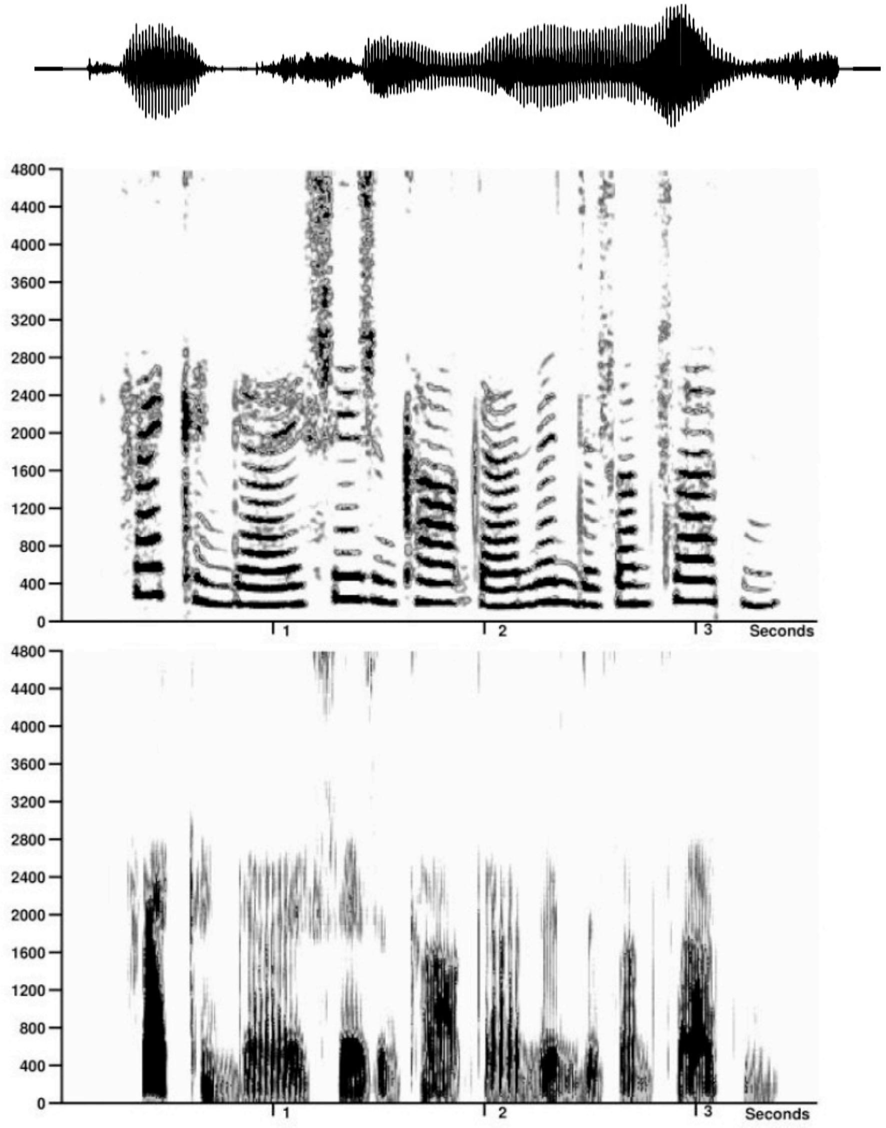
$$x = 10*\log_{10} (20/10) = 10*0.3 = 3\text{dB}$$

You can usually just hear a difference of 1dB (jnd)

Relative Amplitudes and Decibels for Environmental Sounds

SOUND	RELATIVE AMPLITUDE	DECIBELS (DB)
Barely audible (threshold)	1	0
Leaves rustling	10	20
Quiet residential community	100	40
Average speaking voice	1,000	60
Express subway train	100,000	100
Propeller plane at takeoff	1,000,000	120
Jet engine at takeoff (pain threshold)	10,000,000	140

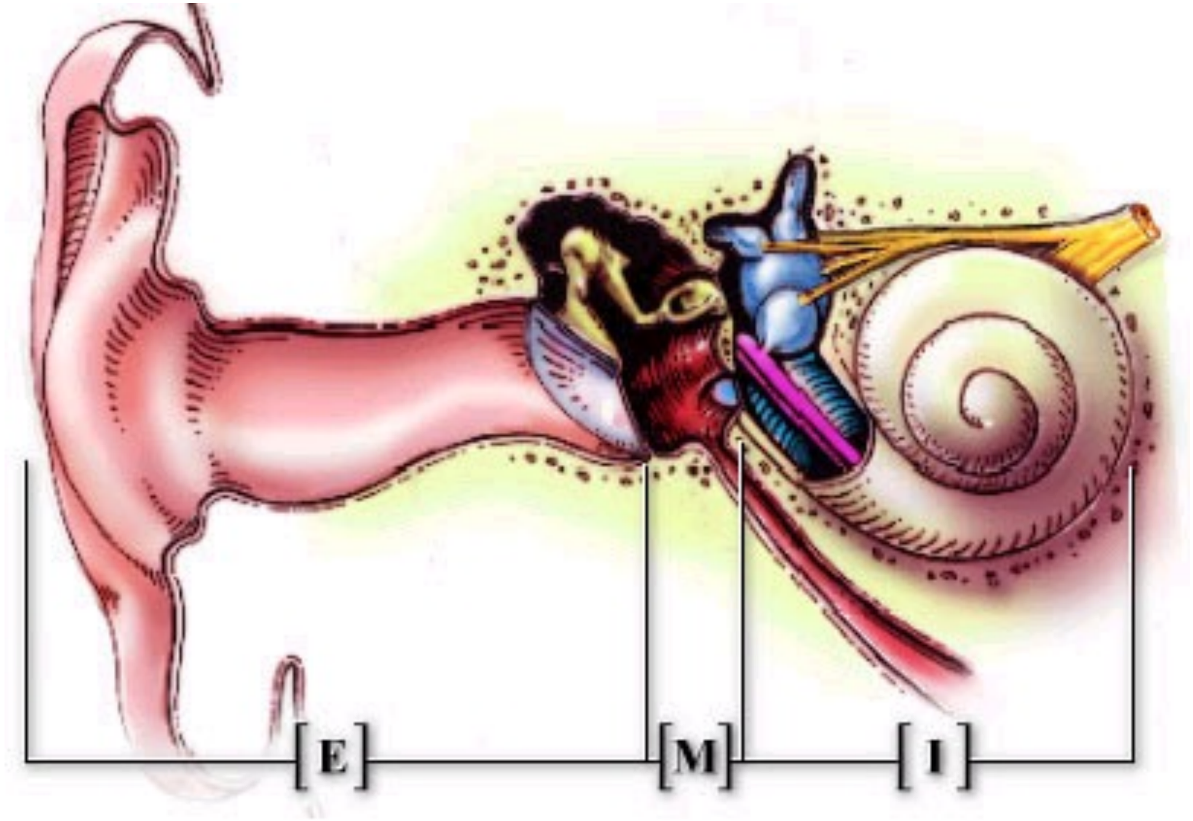
What auditory structure should be represented?



Conventional spectrograms of speech:

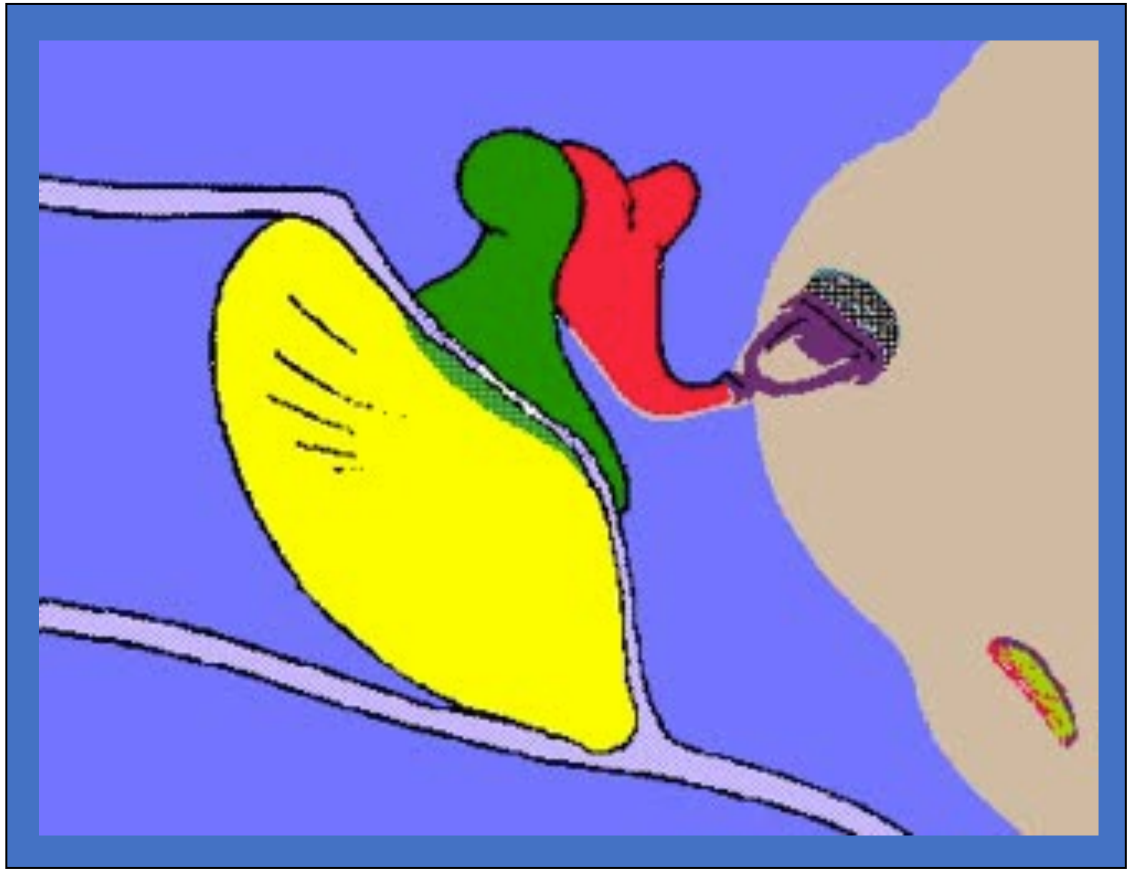
- upper time window: 20Hz
- lower time window: 200Hz

Overview of Human Ear



1）Capture;   2）Amplify mid-freqs;   3）Vertical direction coding

The Ossicular Chain

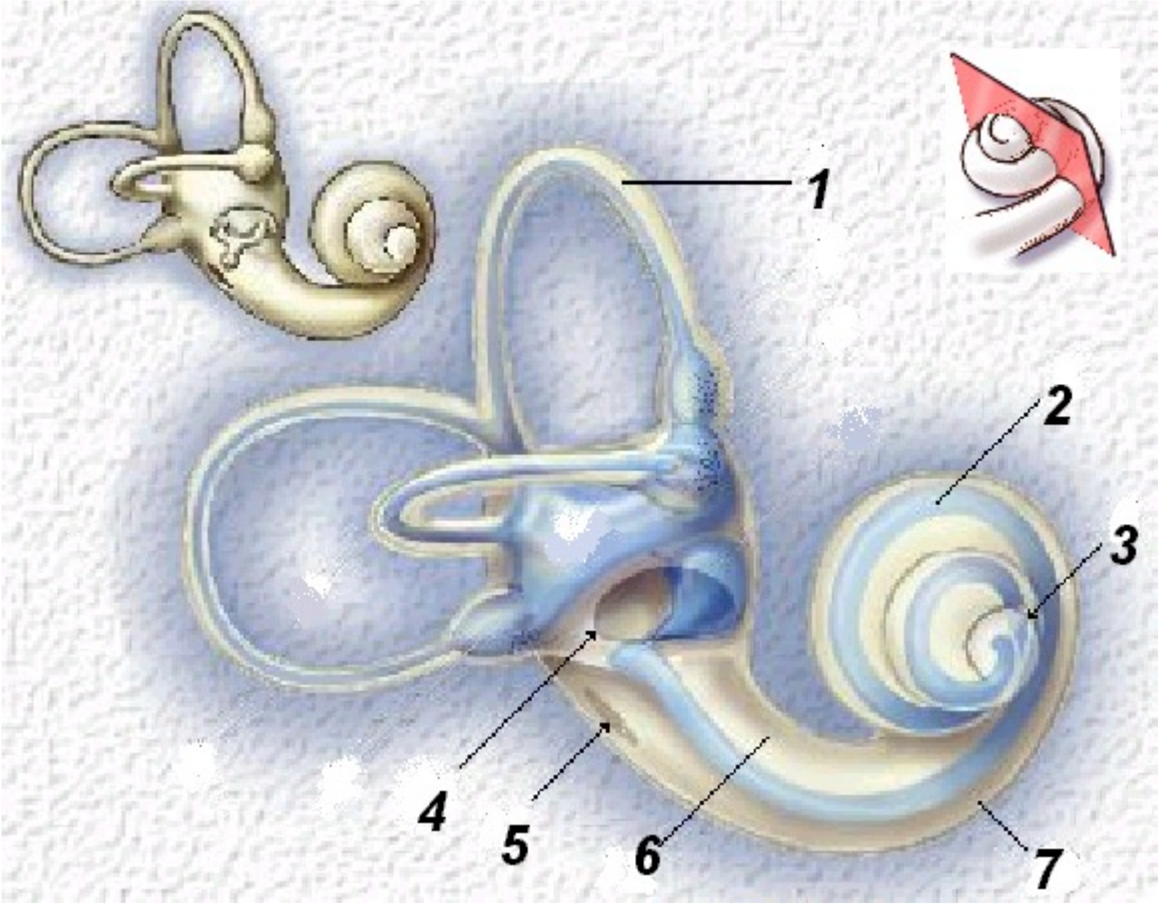


1）Protection;　2）Impedance match

Conductive hearing loss

- Sounds don’t get into cochlea
- Middle ear problems
- Helped by surgery and by amplification

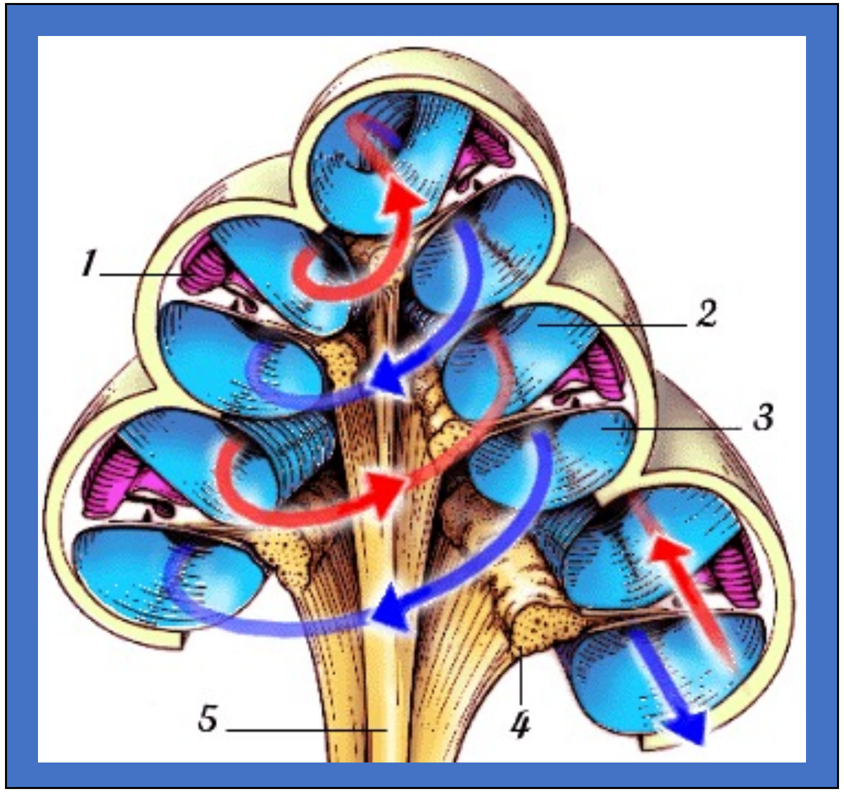
The Inner Ear



1. 1/2 circular canal
2. Cochlear duct
3. helicotrema
4. oval window
5. round window
6. scala vestibuli
7. scala tympani

Frequency analysis; Transduction

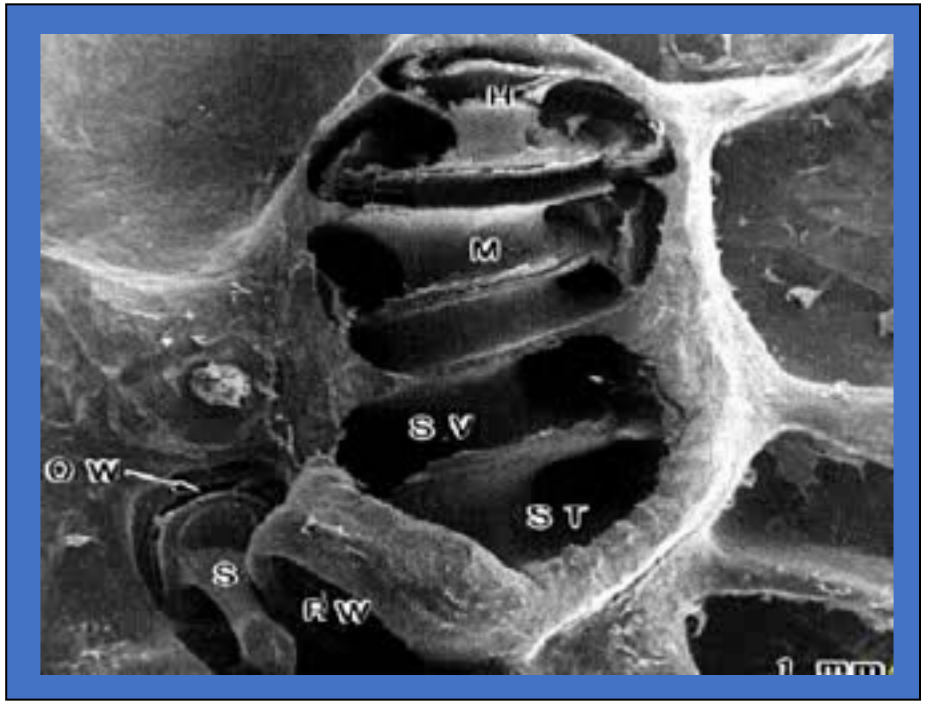
Cross Section of the Cochlea.



1. the cochlear duct
2. the scala vestibuli
3. the scala tympani
4. the spiral ganglion
5. auditory nerve fibers

The red arrow is from the oval window,
the blue arrow points to the round window.

Cross Section of the Cochlea.

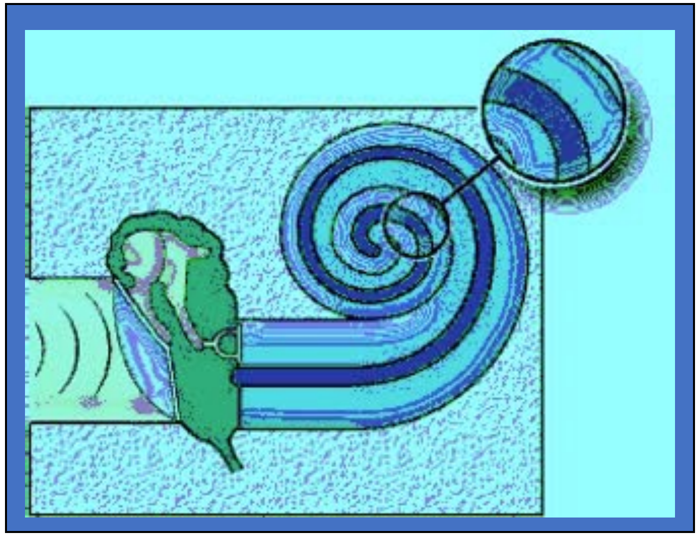


1. the cochlear duct
2. the scala vestibuli
3. the scala tympani
4. the spiral ganglion
5. auditory nerve fibers

The red arrow is from the oval window,
the blue arrow points to the round window.

The Basilar Membrane

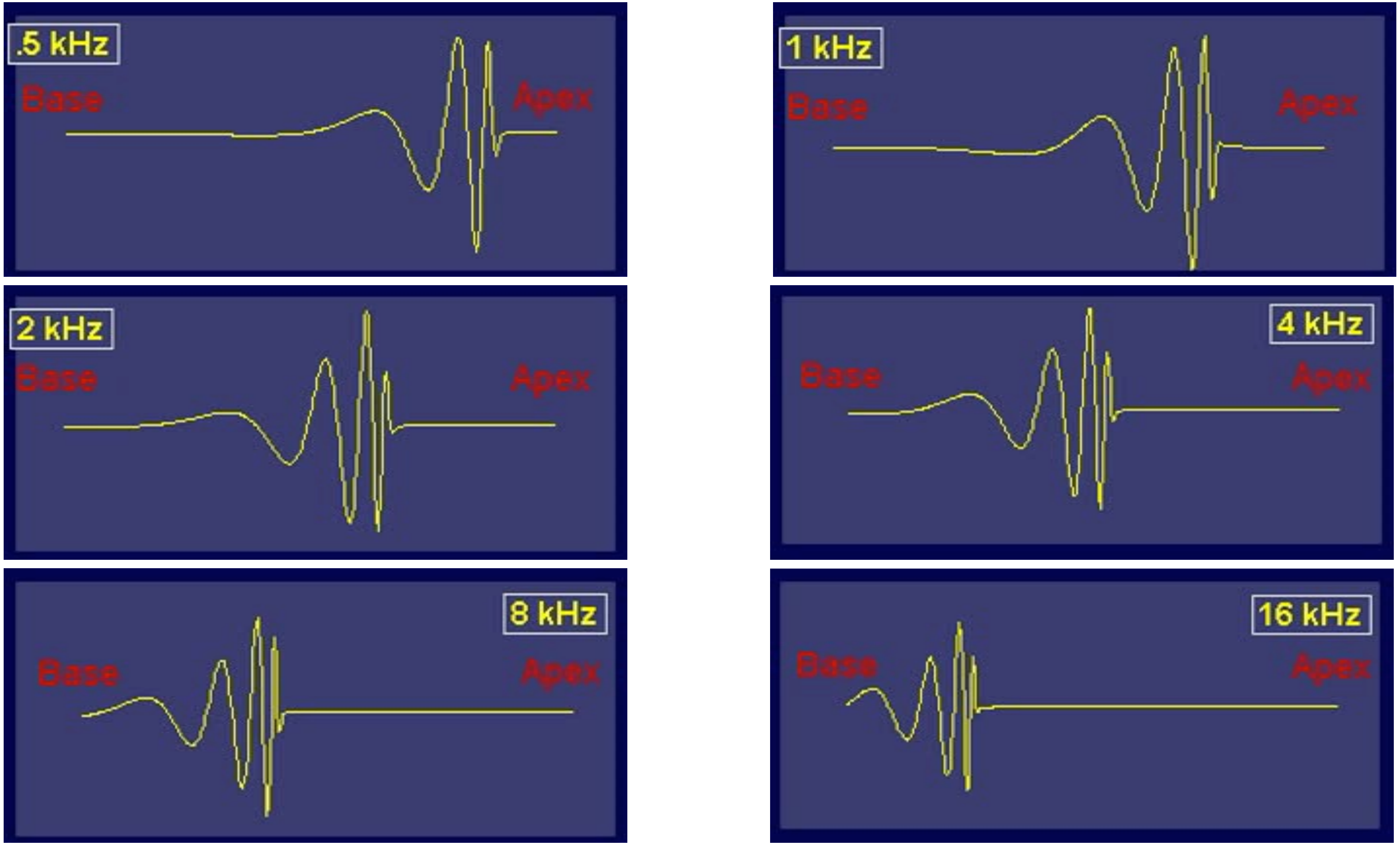
Depending on the frequency, the vibration has a maximum resonance at a different point along the basilar membrane.



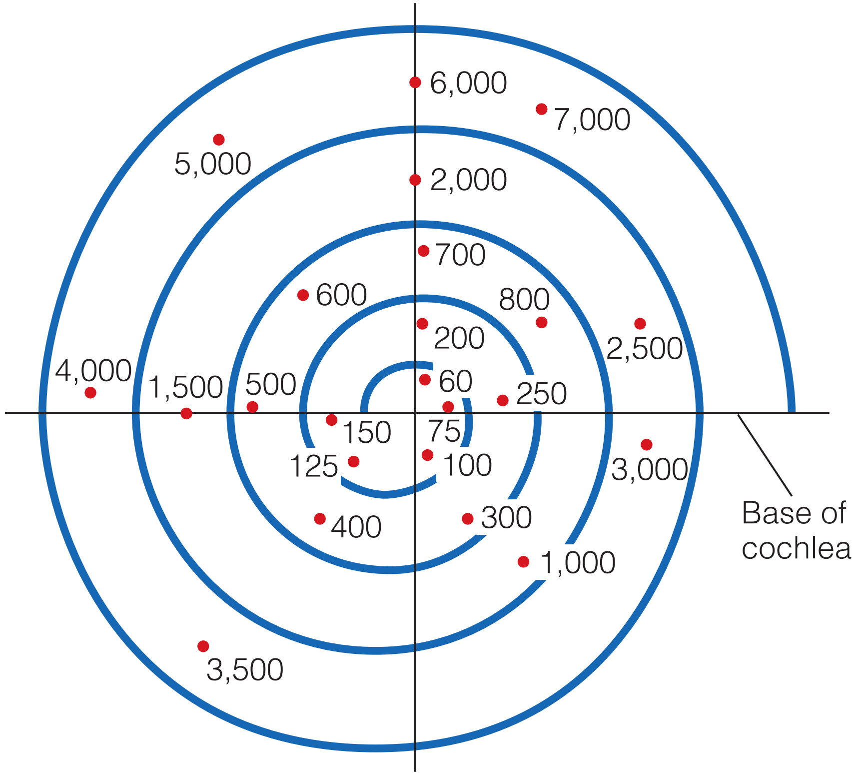
low frequency

high frequency

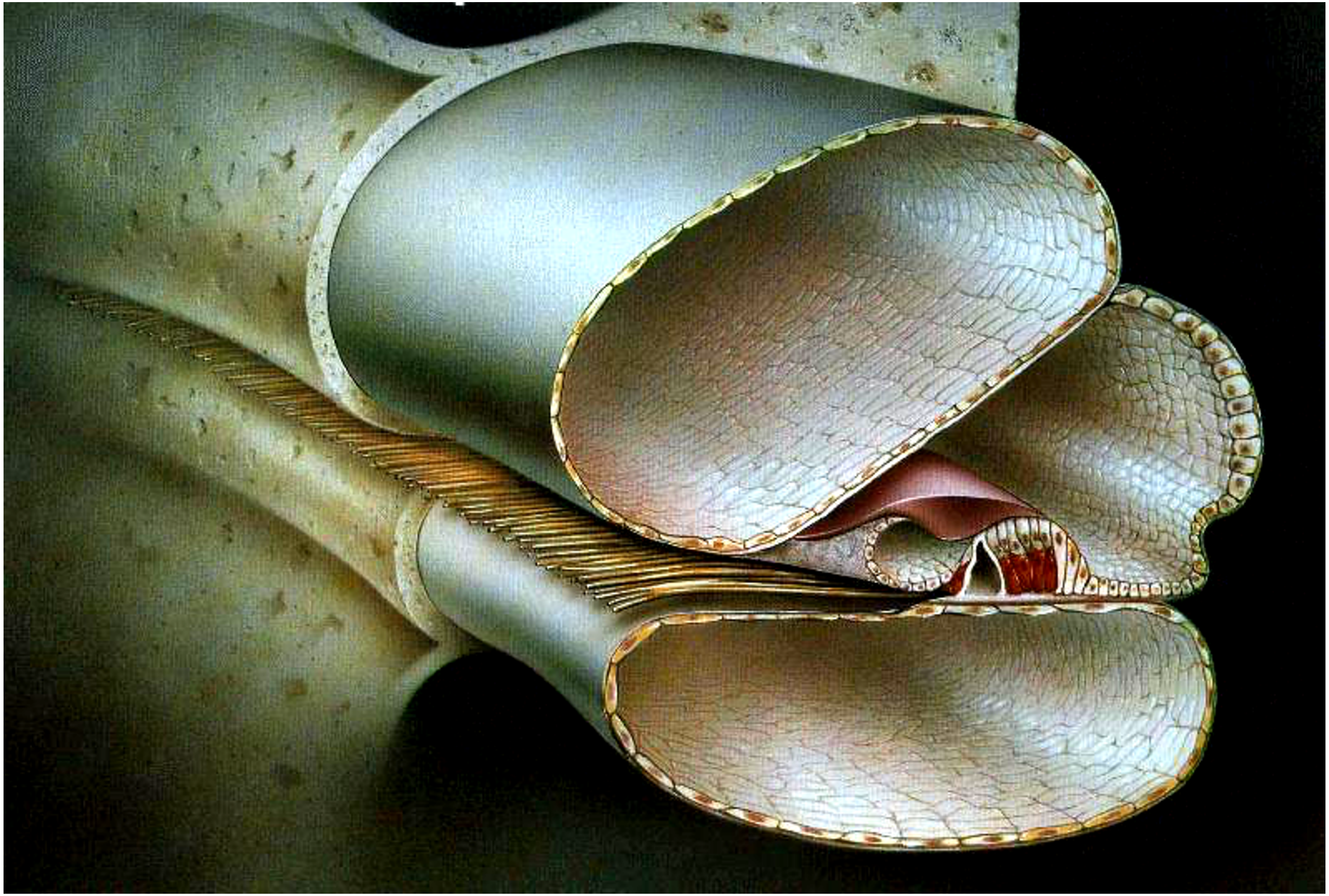
Traveling Wave on the Basilar Membrane



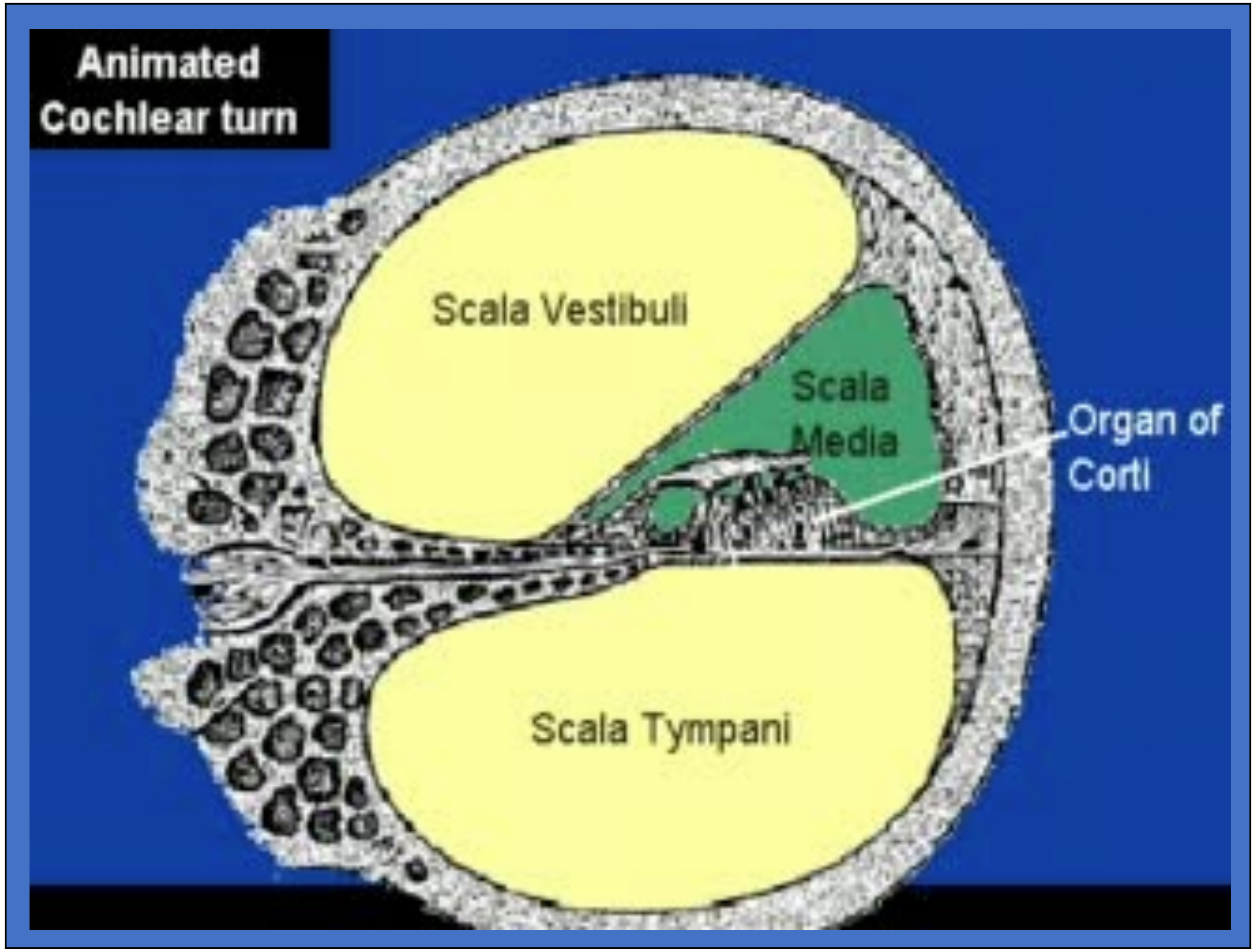
Tonotopic map of the guinea pig cochlea



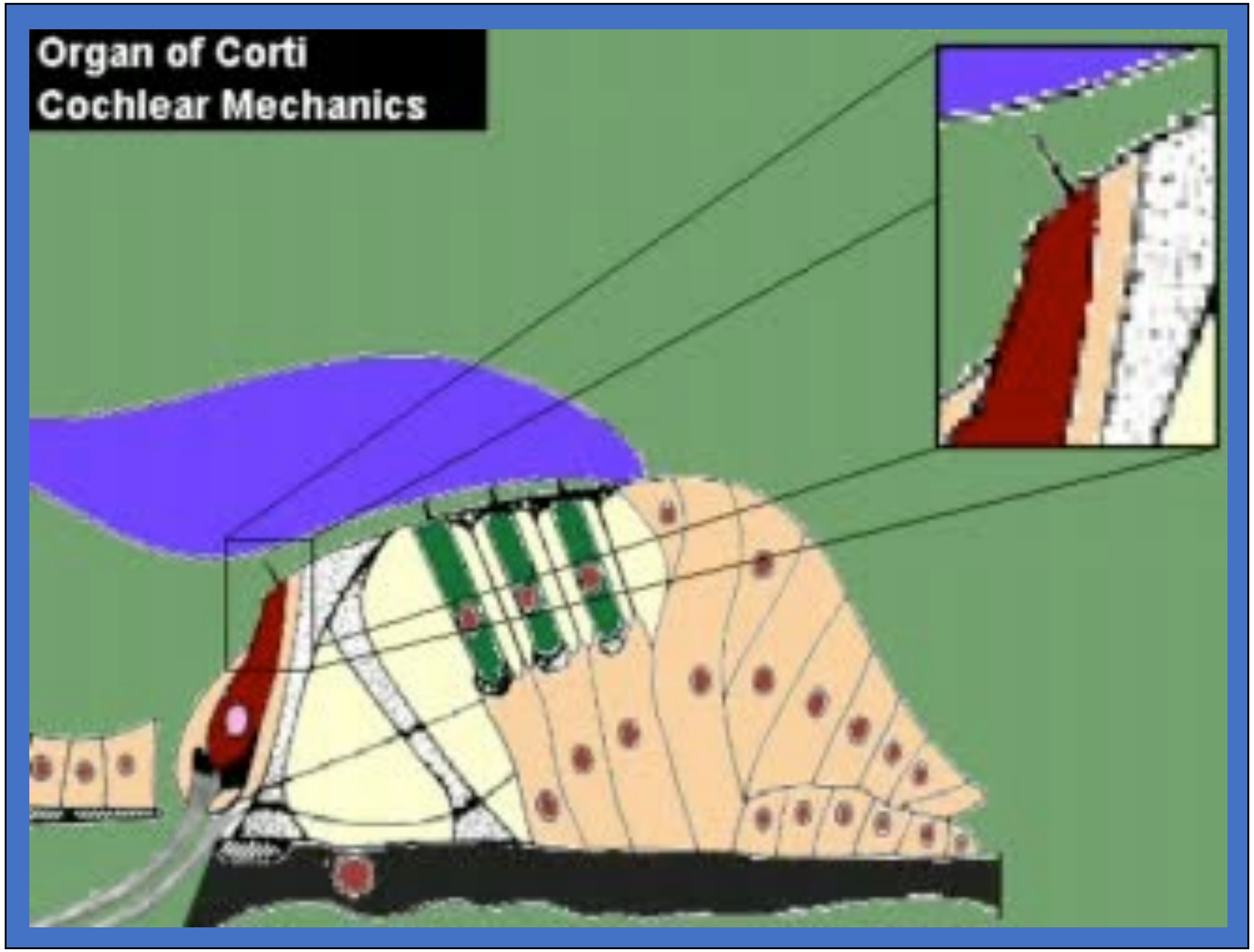
Cochlea cross-section



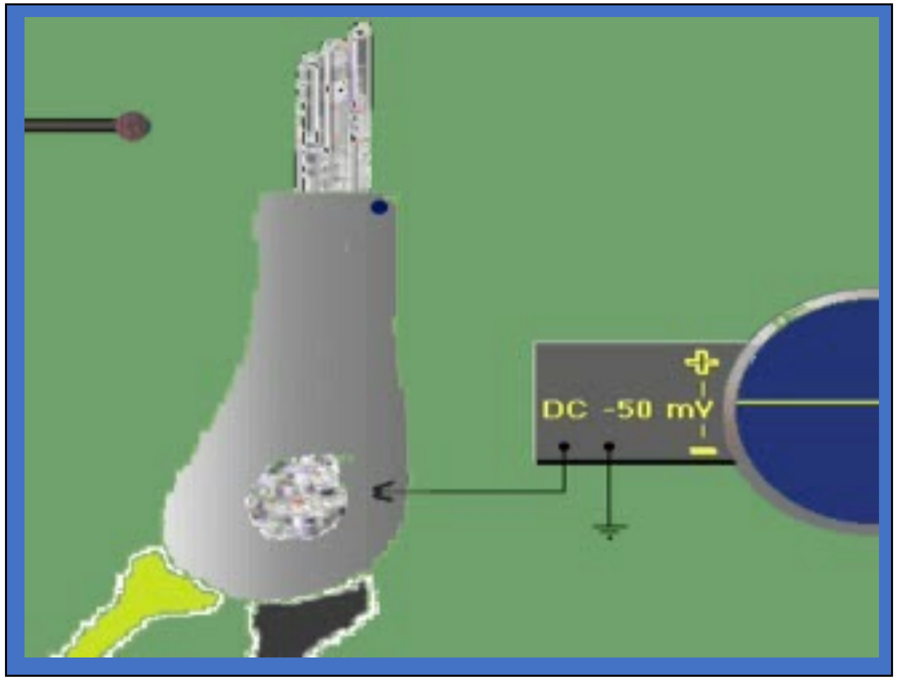
The Organ of Corti

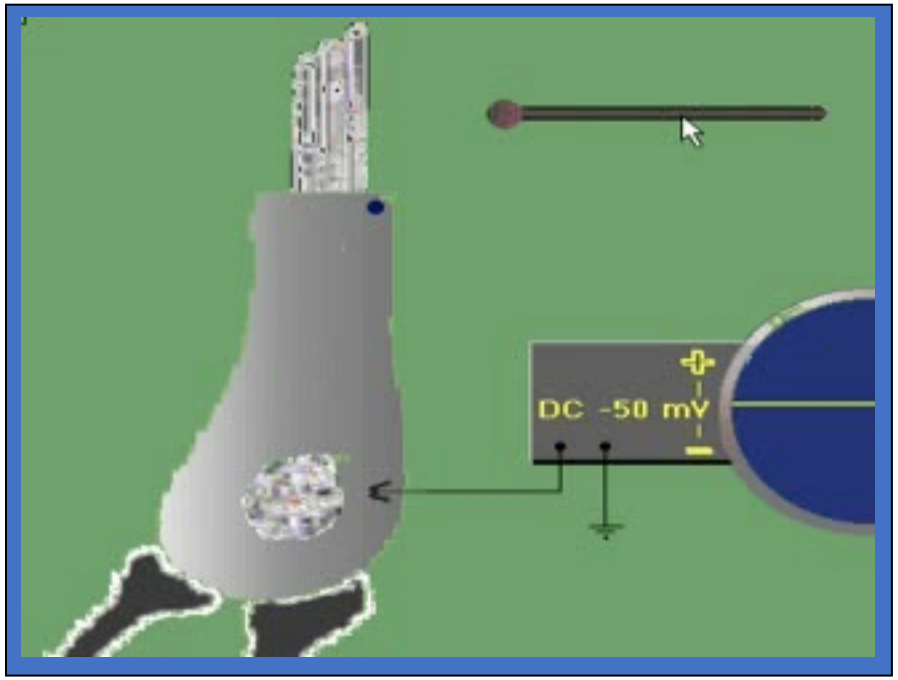


Cochlear Mechanics in Organ of Corti



The Inner Hair Cell





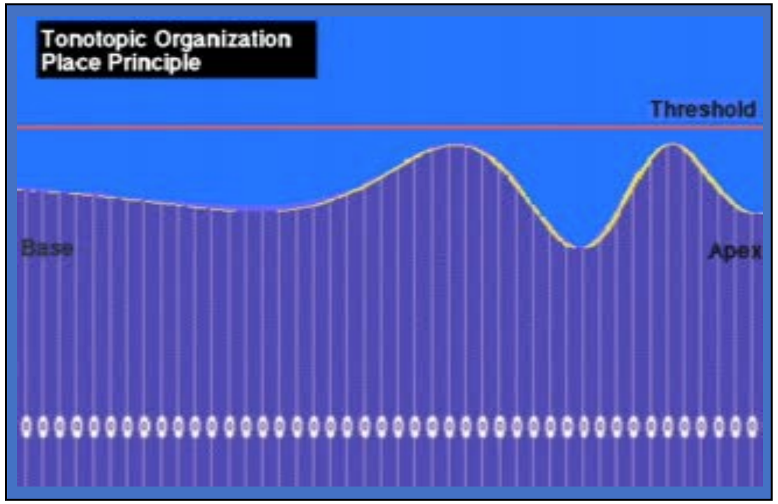
depolarization

hyperpolarization

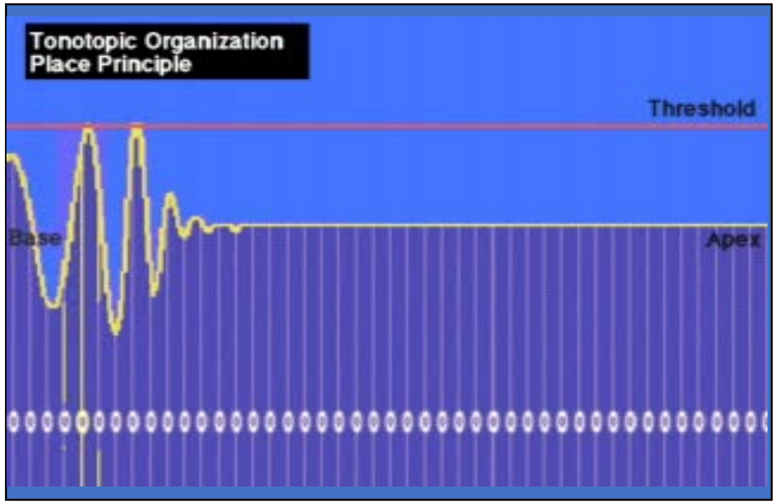
Three Theories of Frequency Coding

- ✔ Place Coding Theory
- ✔ Phase locking Coding Theory
- ✔ Volley Theory

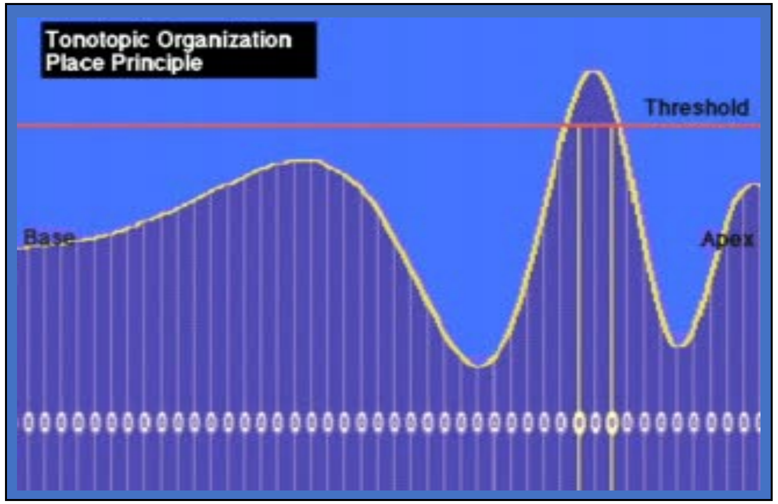
Place Coding Theory



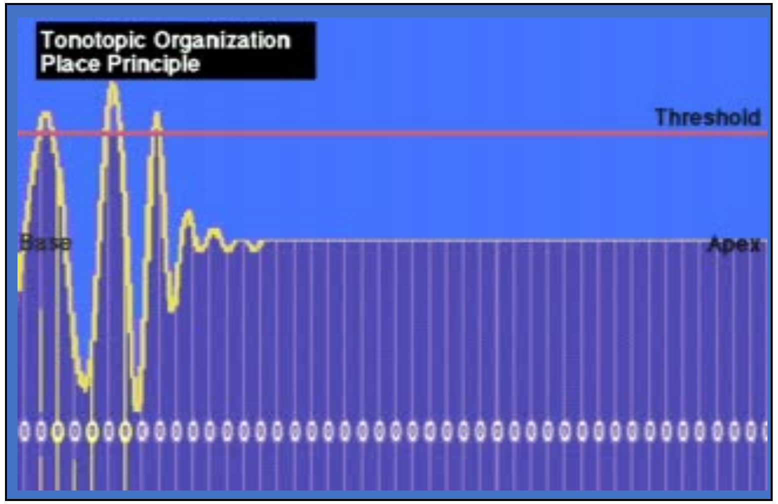
softer sound with 1kHz



softer sound with 8kHz

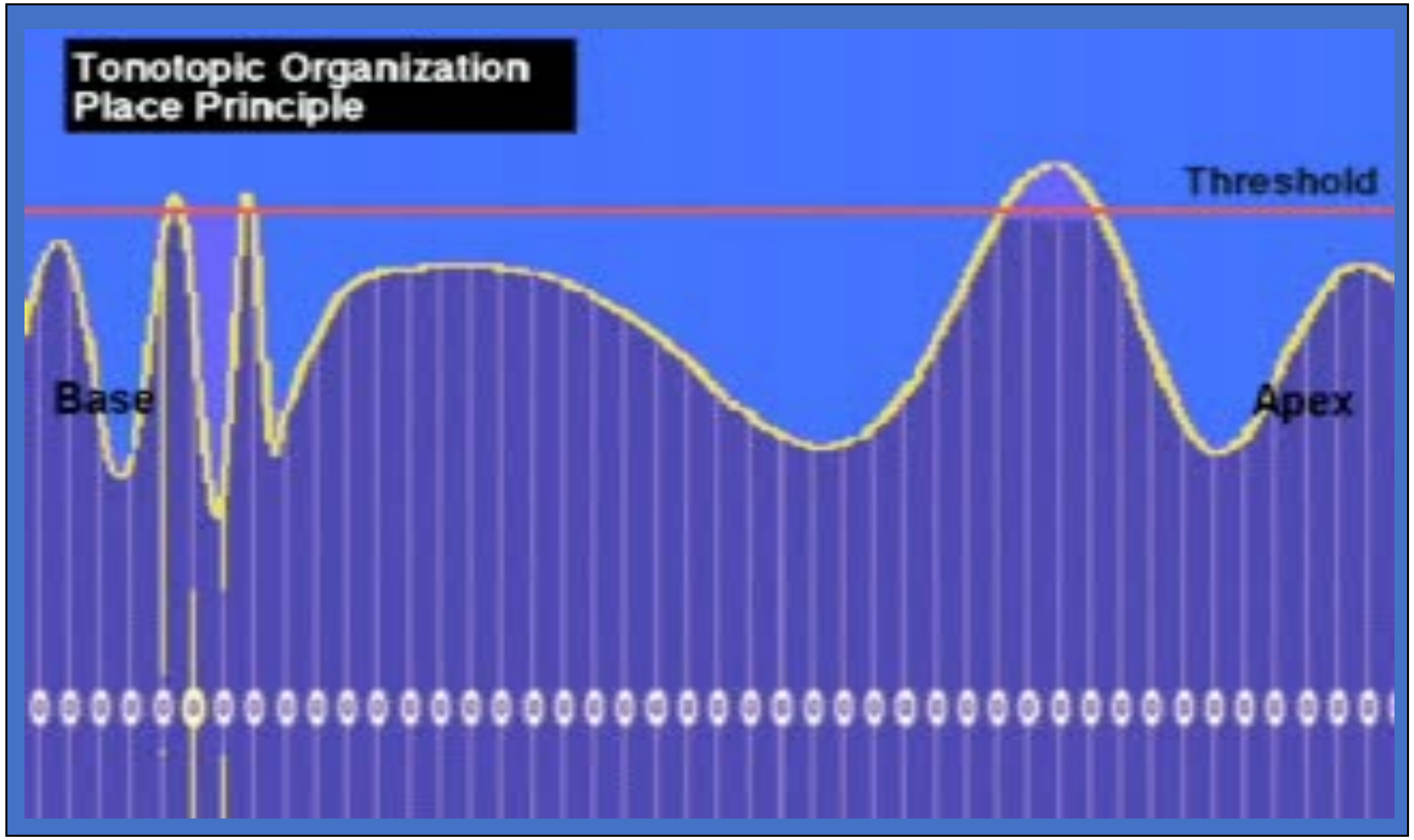


louder sound with 1kHz



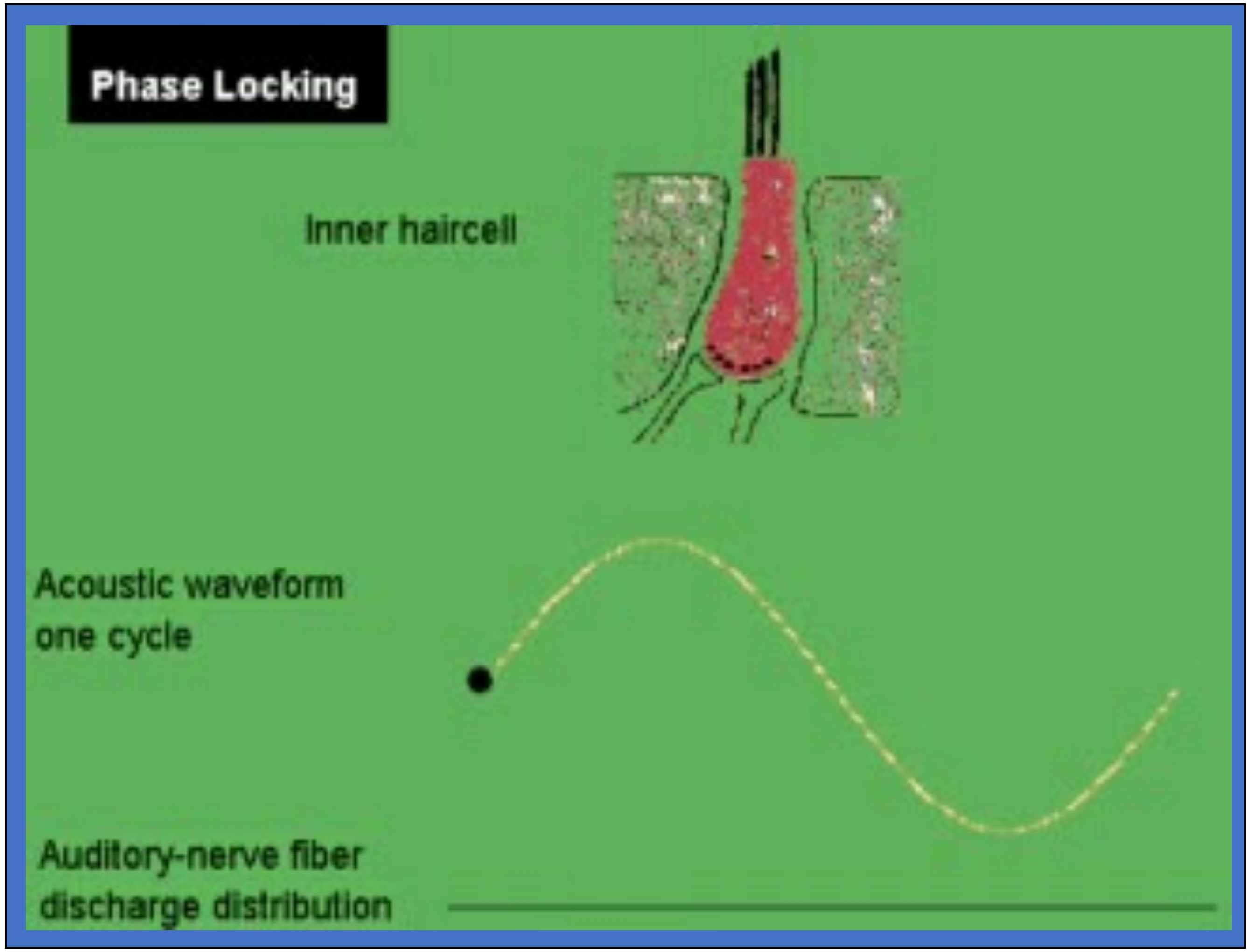
louder sound with 8kHz

Place Coding Theory (cont.)

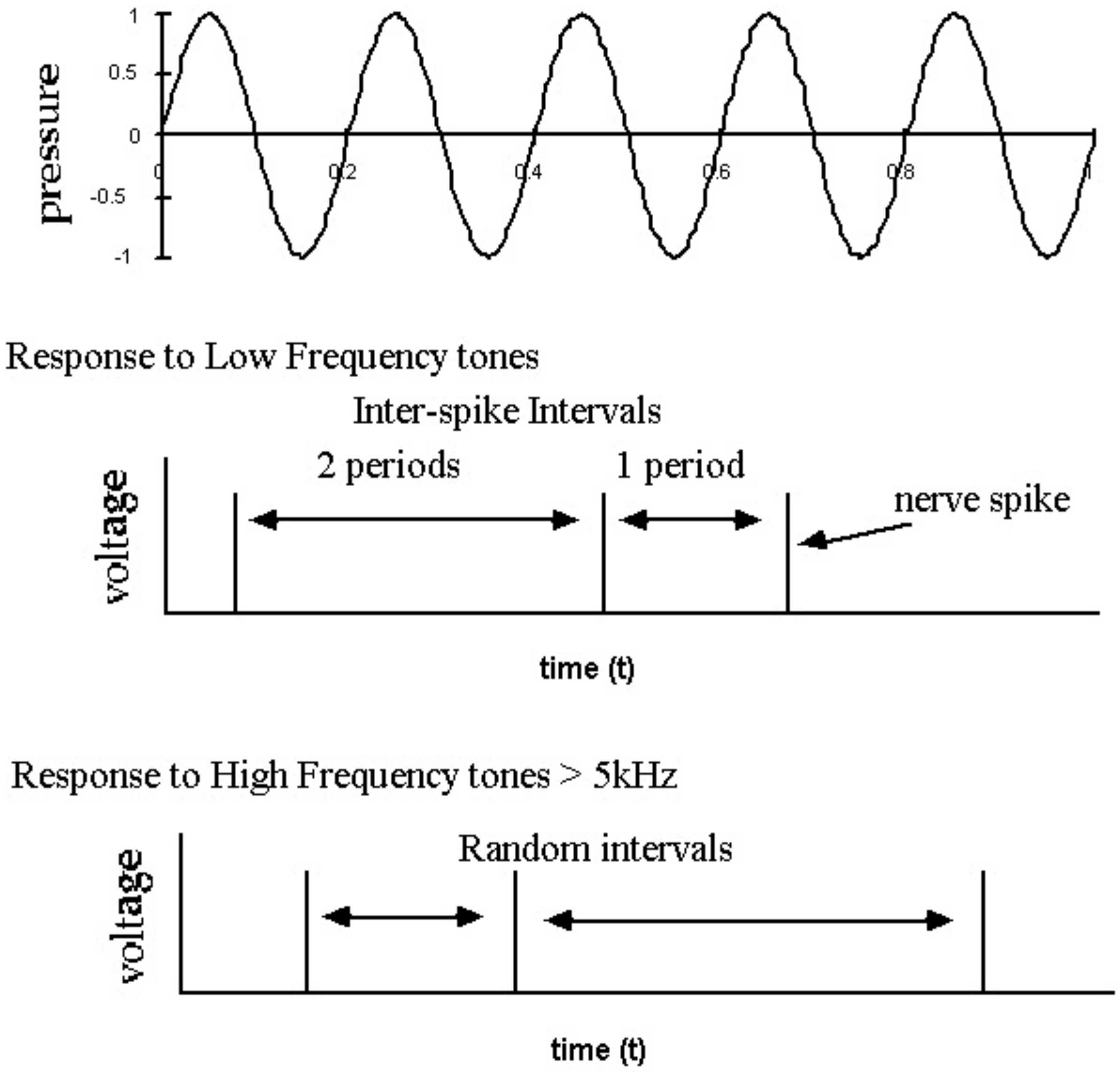


louder complex sound with 1kHz and 8kHz

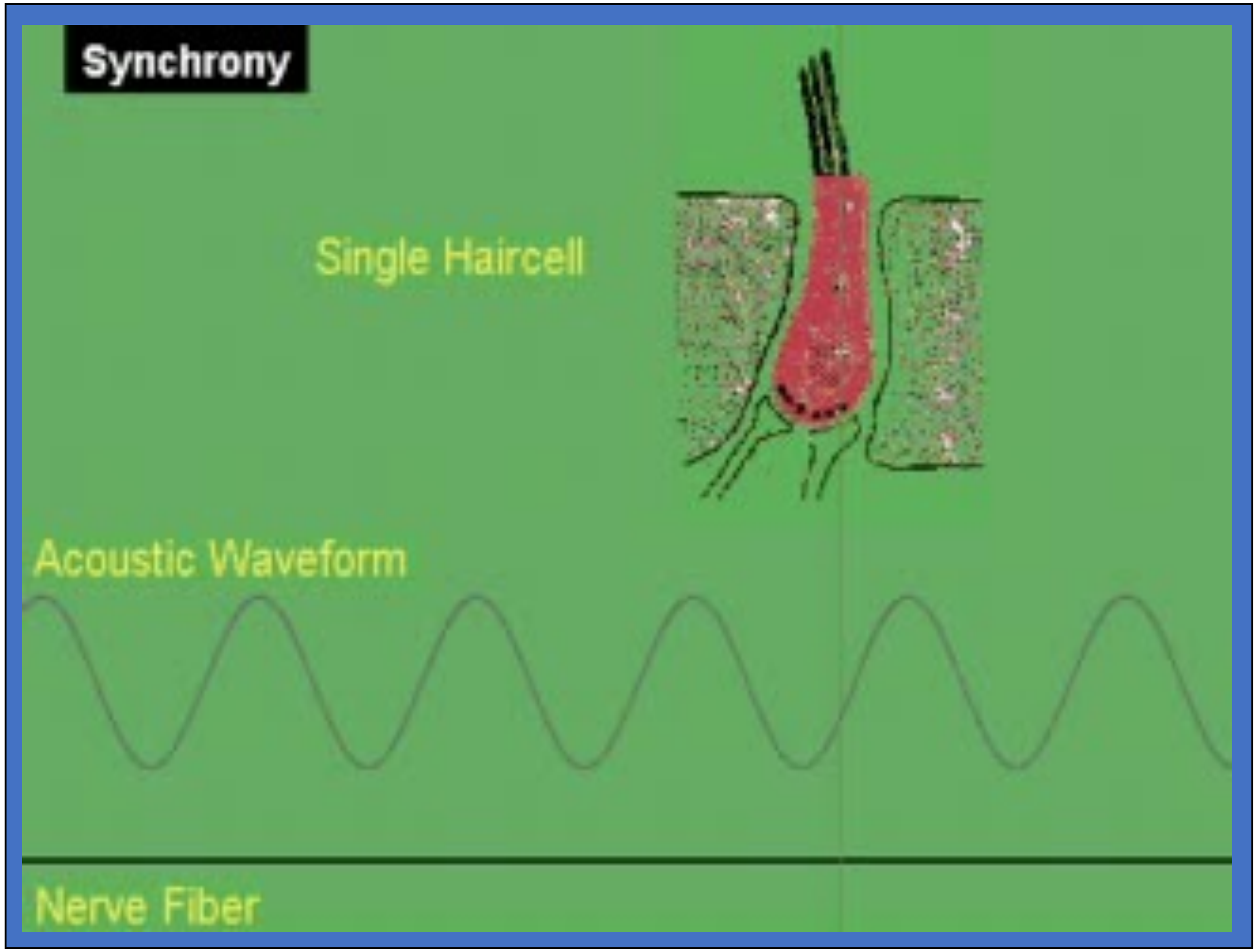
Phase Locking Coding Theory



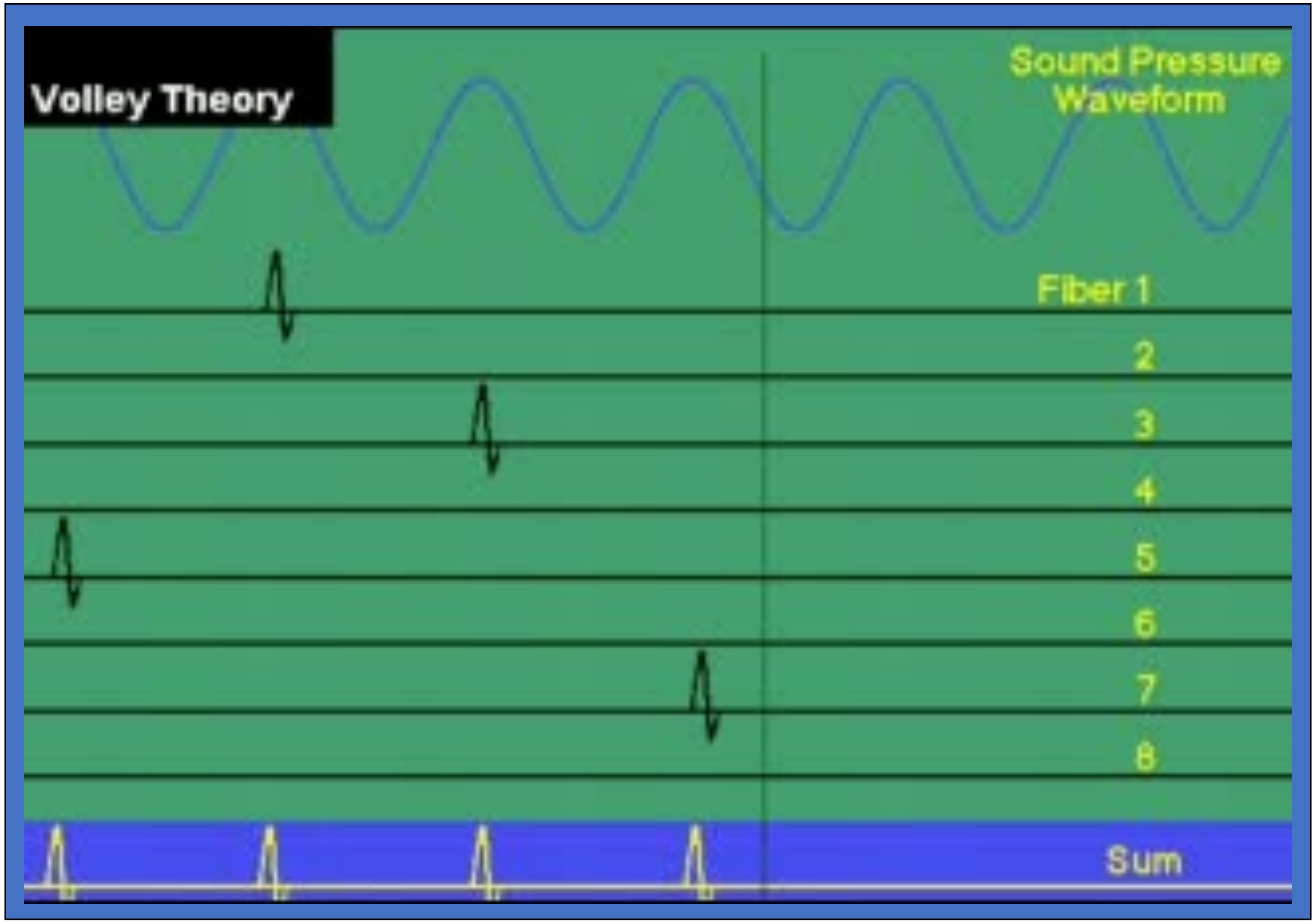
Phase-locking



Definition of Synchrony Firing



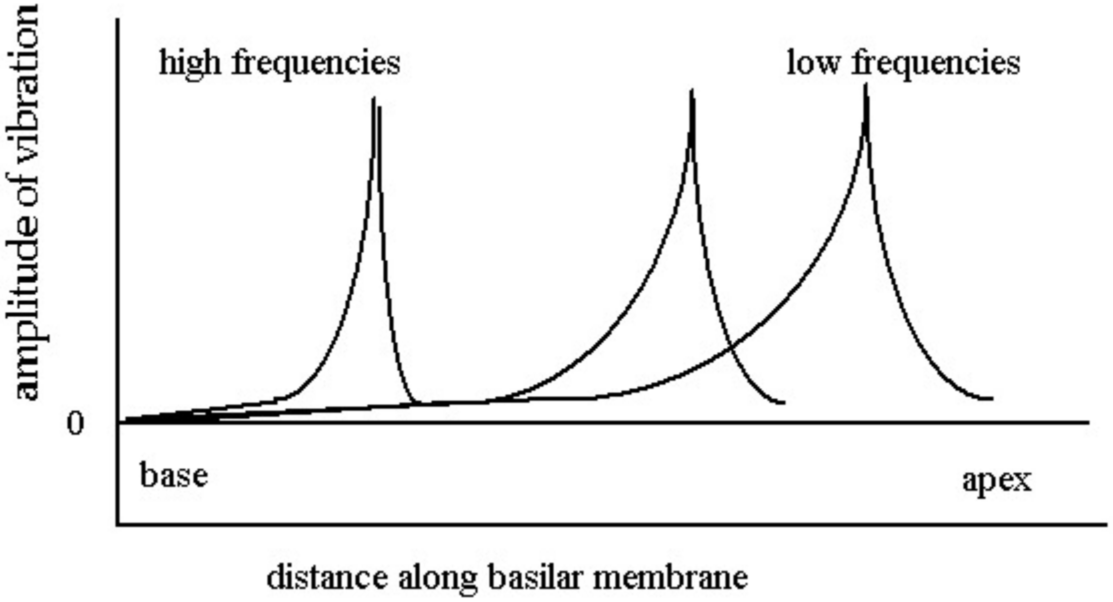
Volley Theory



Three Theories of Frequency Coding

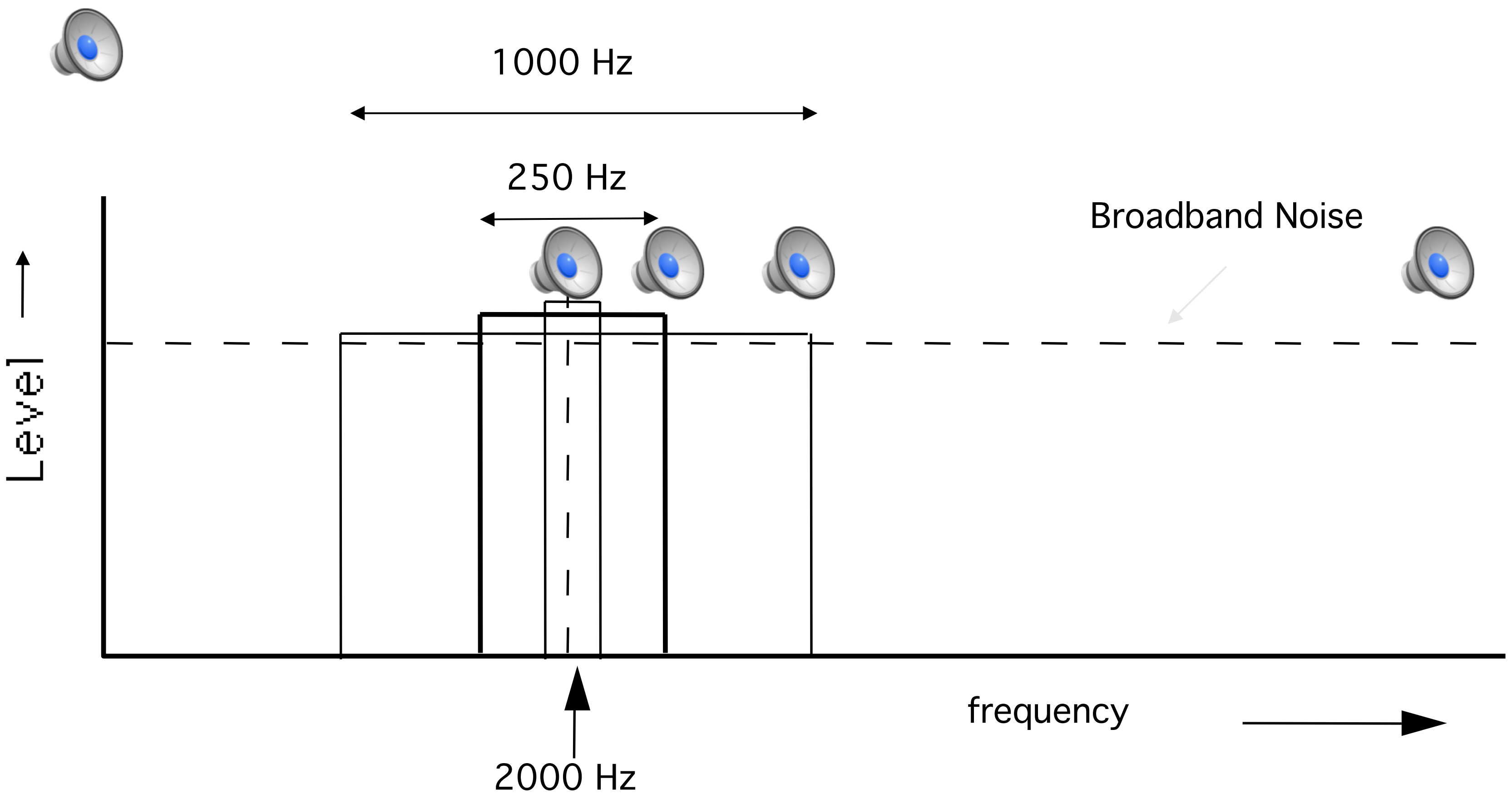
- ✔ Place Coding Theory
- ✔ Phase locking Coding Theory
- ✔ Volley Theory

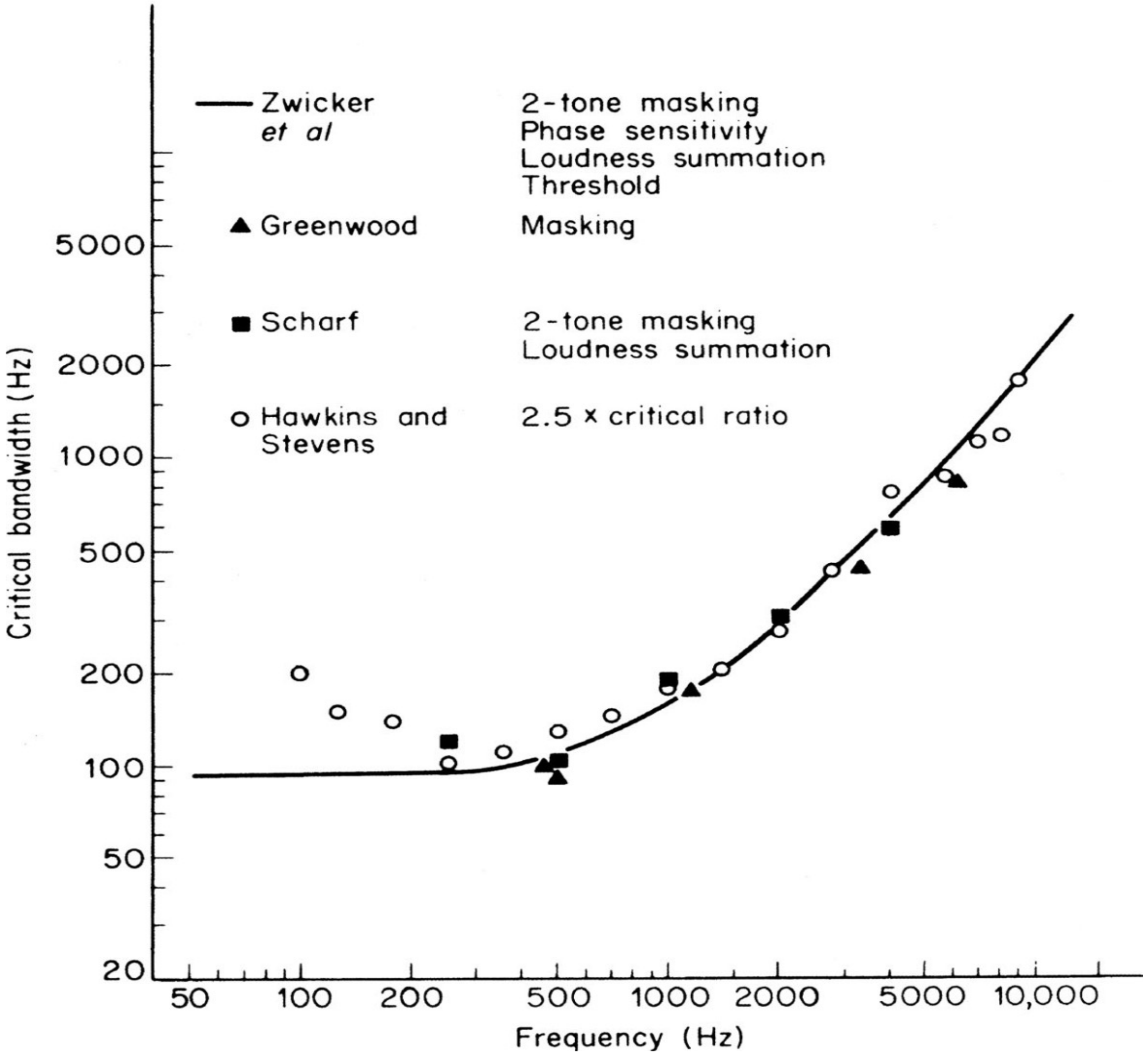
Excitation patterns (envelope of excitation)



Basilar membrane excitation pattern is like a spectrum

Measurement of auditory bandwidth with band-limited noise





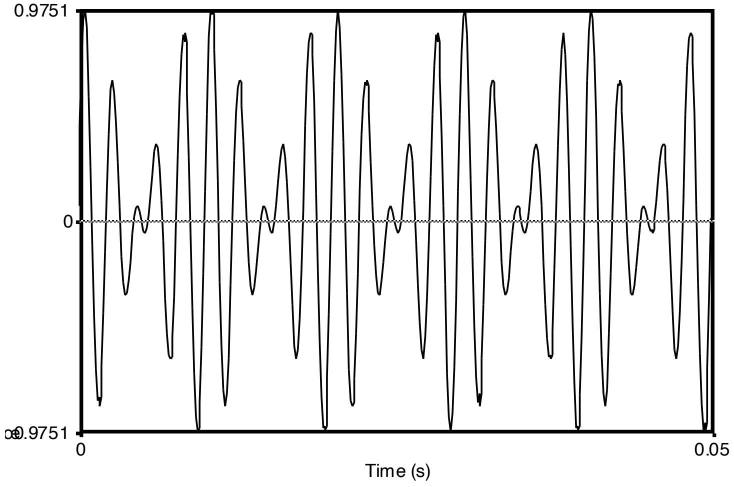
The width of critical bands as a function of center frequency

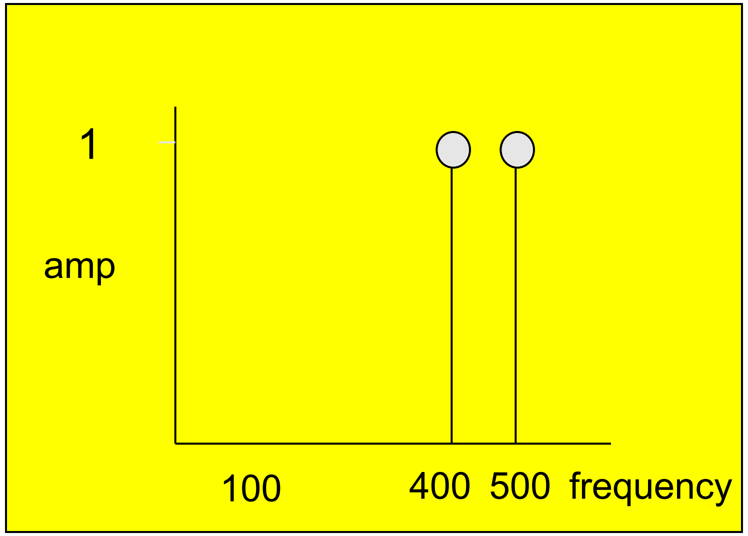
Beats

Repetition rate is the difference in frequency between the two sine-wave components

1/100th second

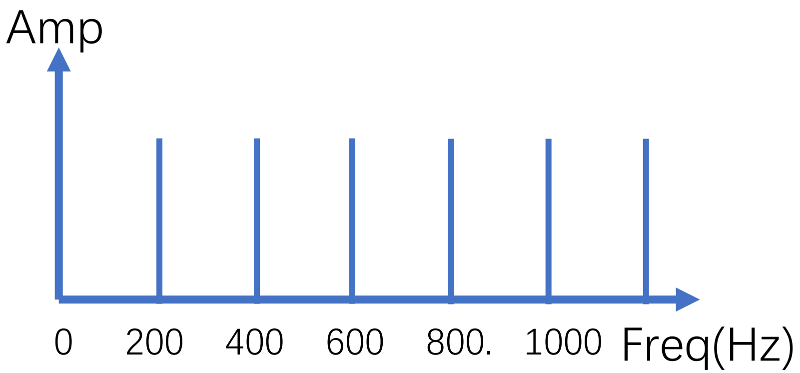
500 - 400 = 100 Hz

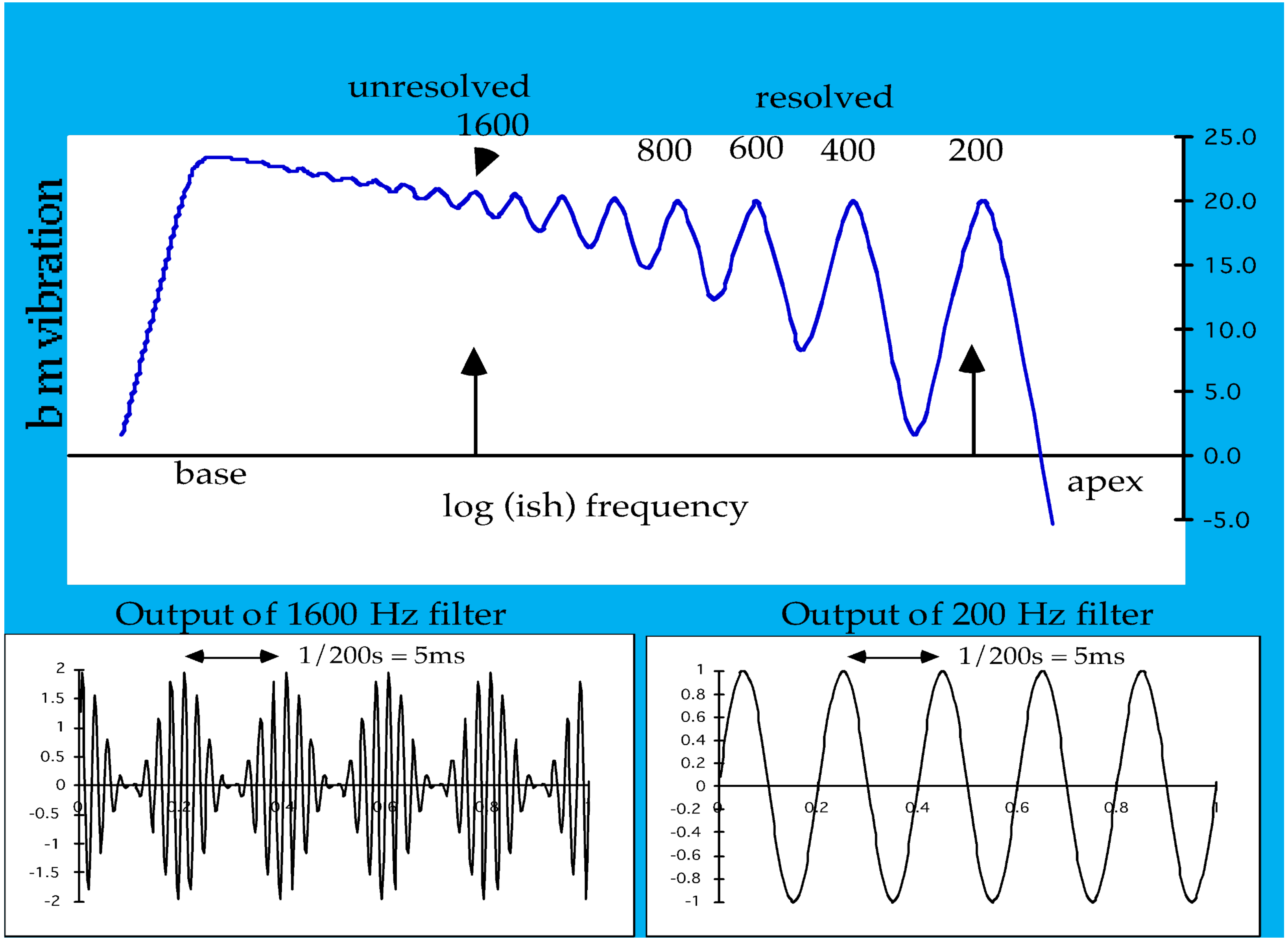




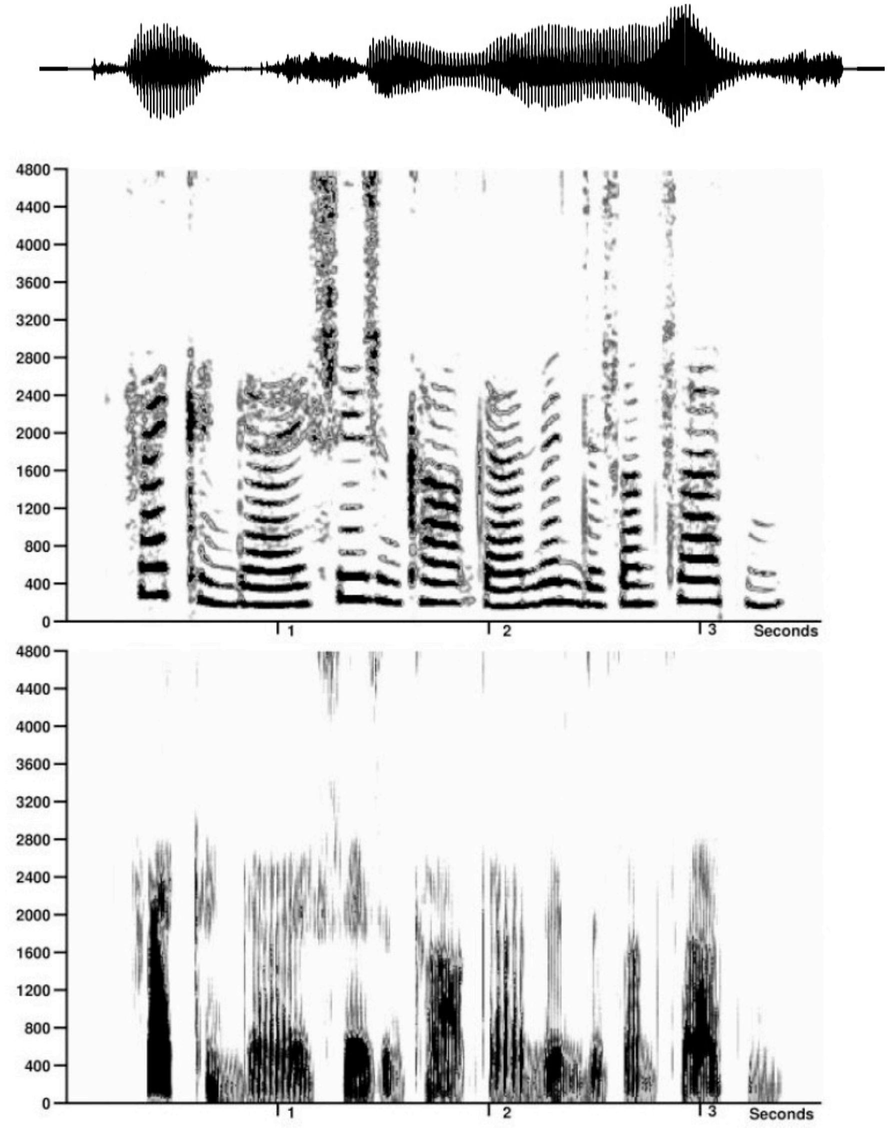
Excitation pattern of complex tone on BM

Sound stimuli:





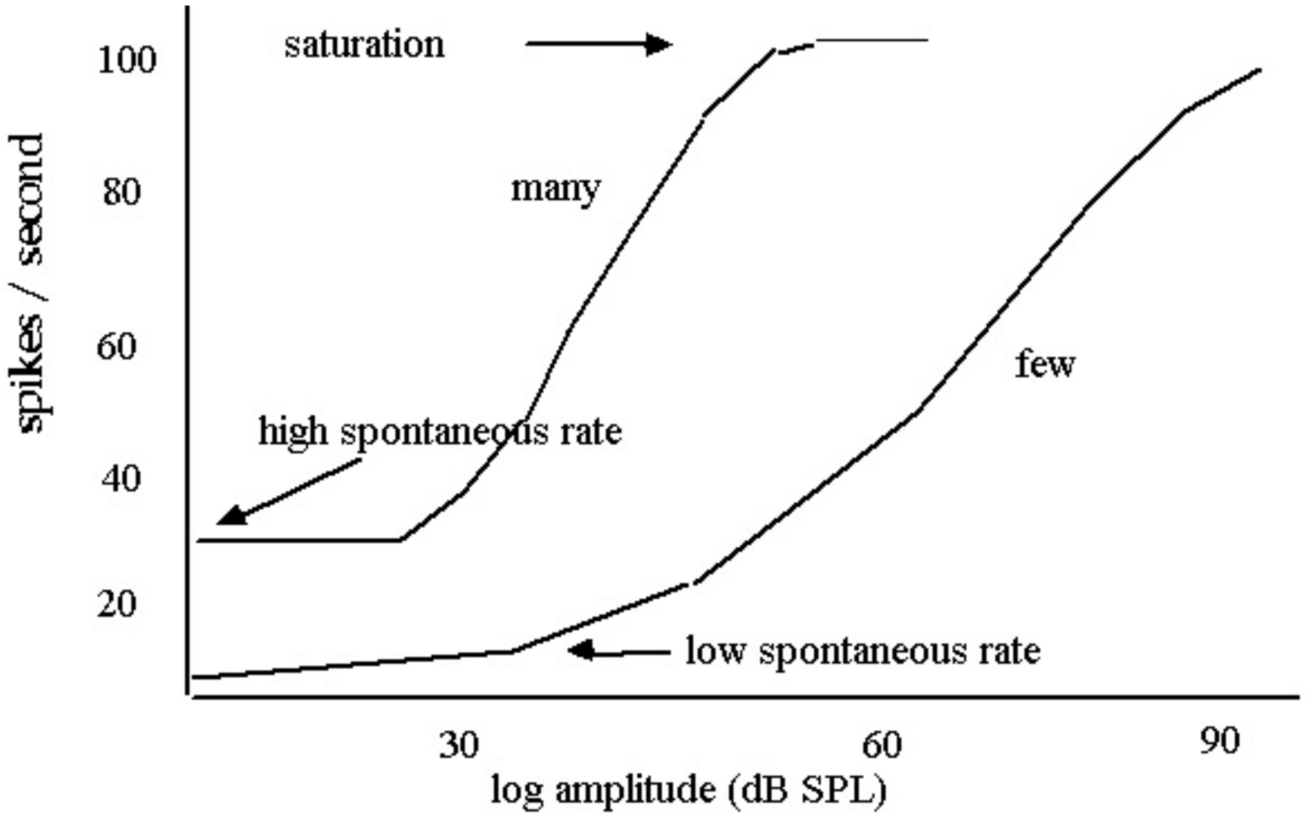
What auditory structure should be represented?



Conventional spectrograms of speech:
- upper time window: 20Hz
- lower time window: 200Hz

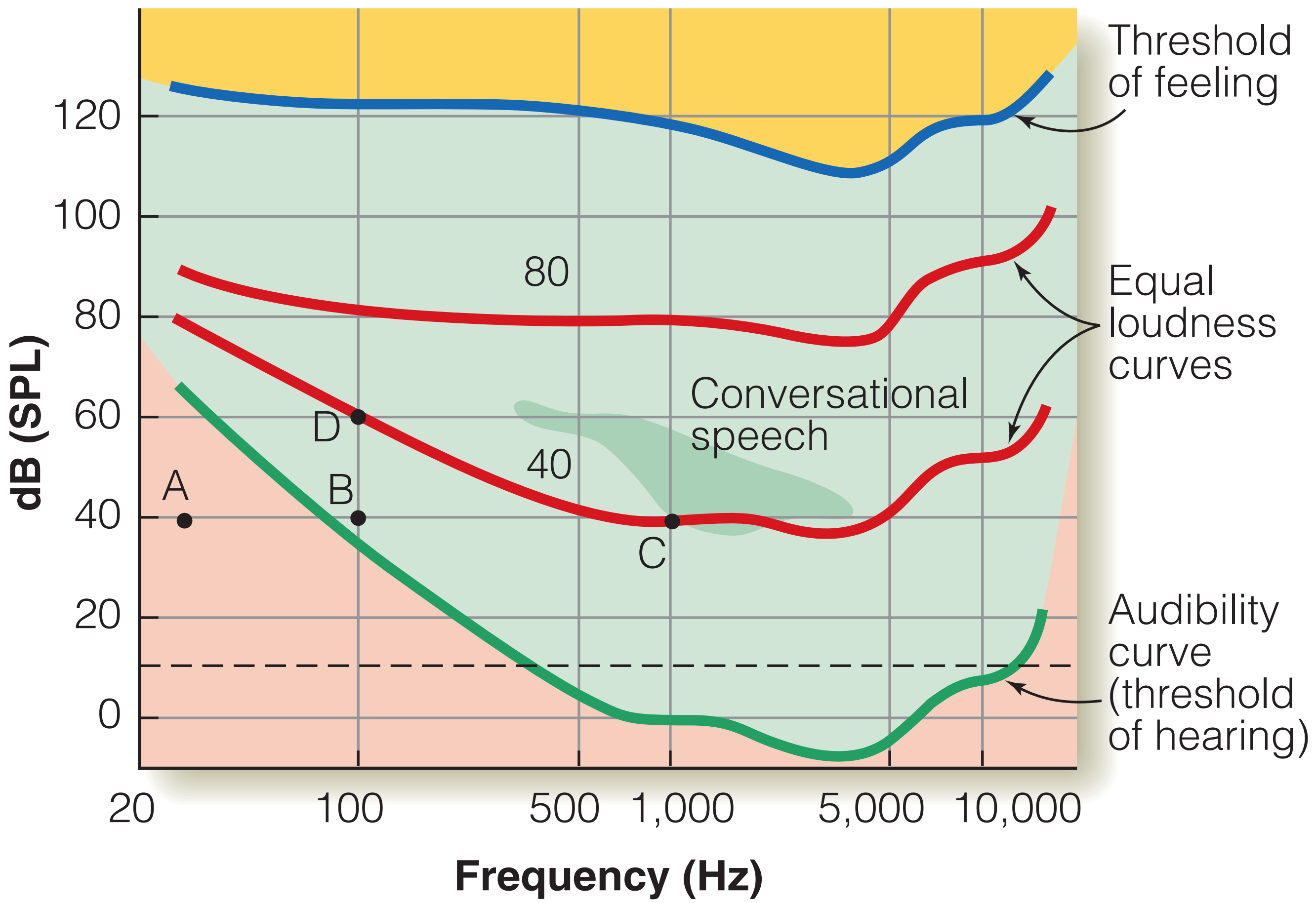
Auditory representation:
- low frequency: narrow bandpass filter
- high frequency: wide bandpass filter

Intensity Coding



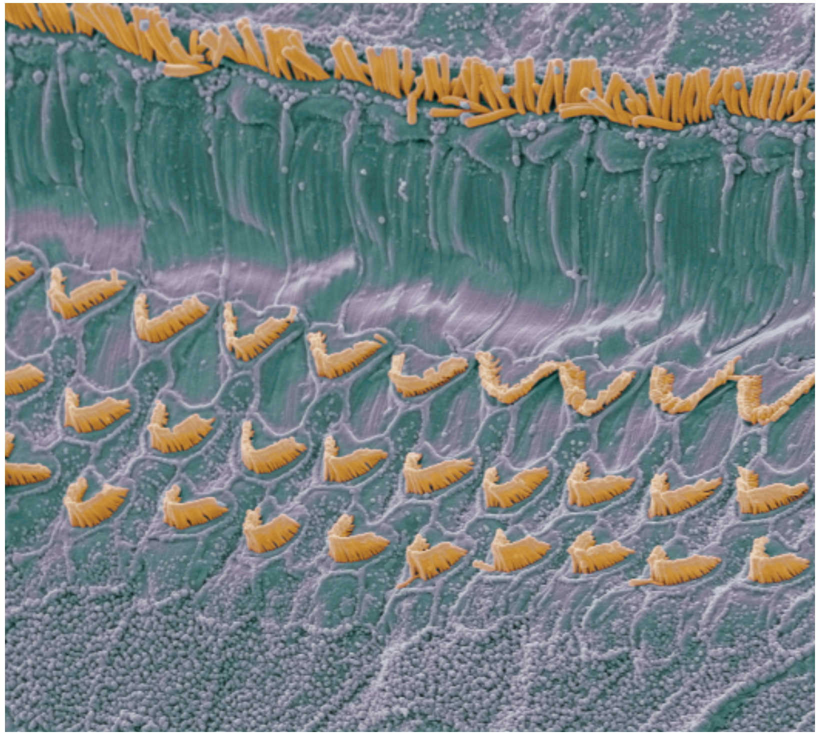
Auditory nerve rate-intensity functions

The audibility curve and the auditory response area



Inner vs Outer Hair Cells

Inner Hair Cells	Outer Hair Cells
Sensory	Motor
Afferent nerves	Efferent nerves
Single row	c.3 rows



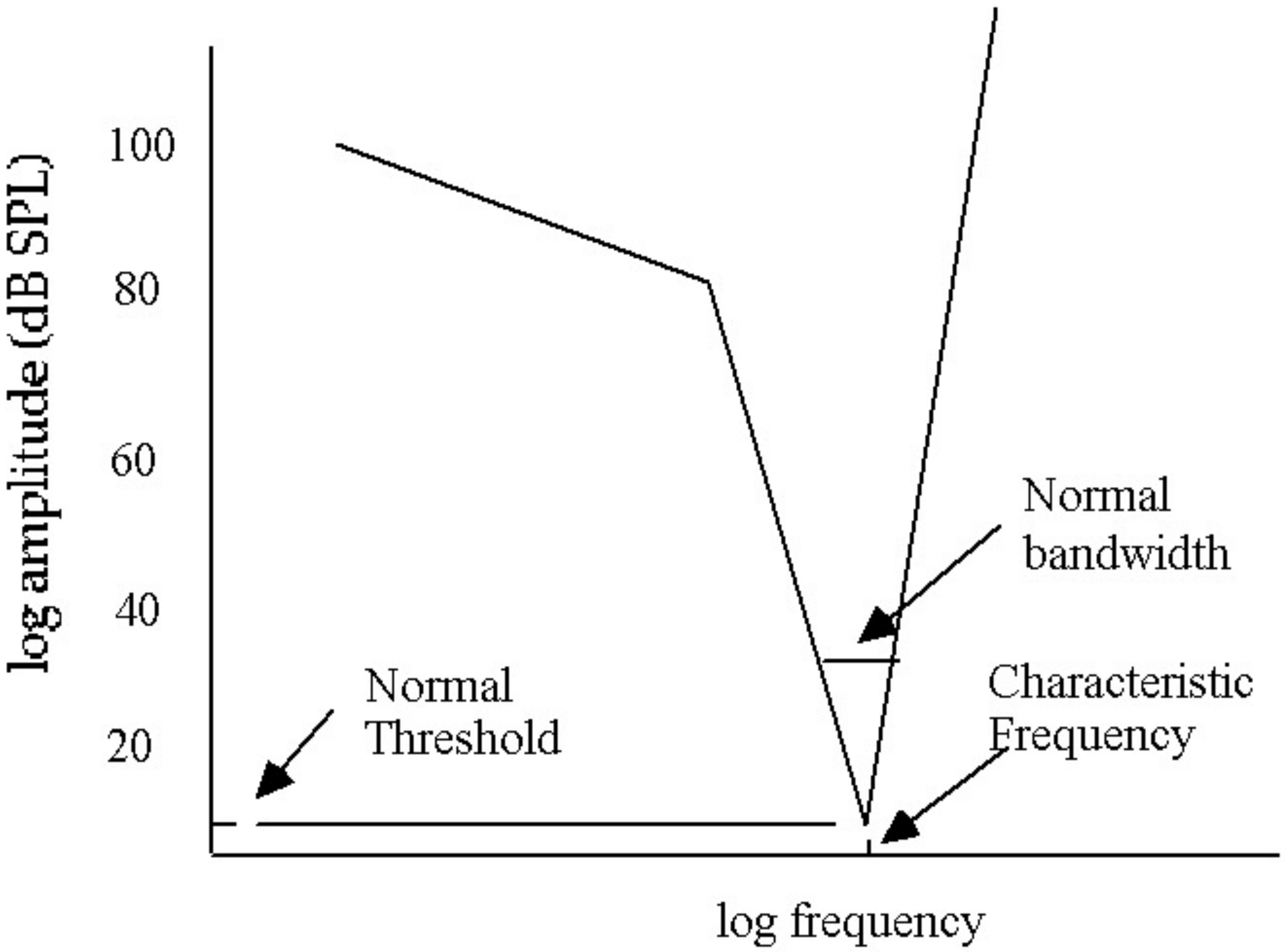
OHC activity

OHCs are relatively more active for quiet sounds than for loud sounds.

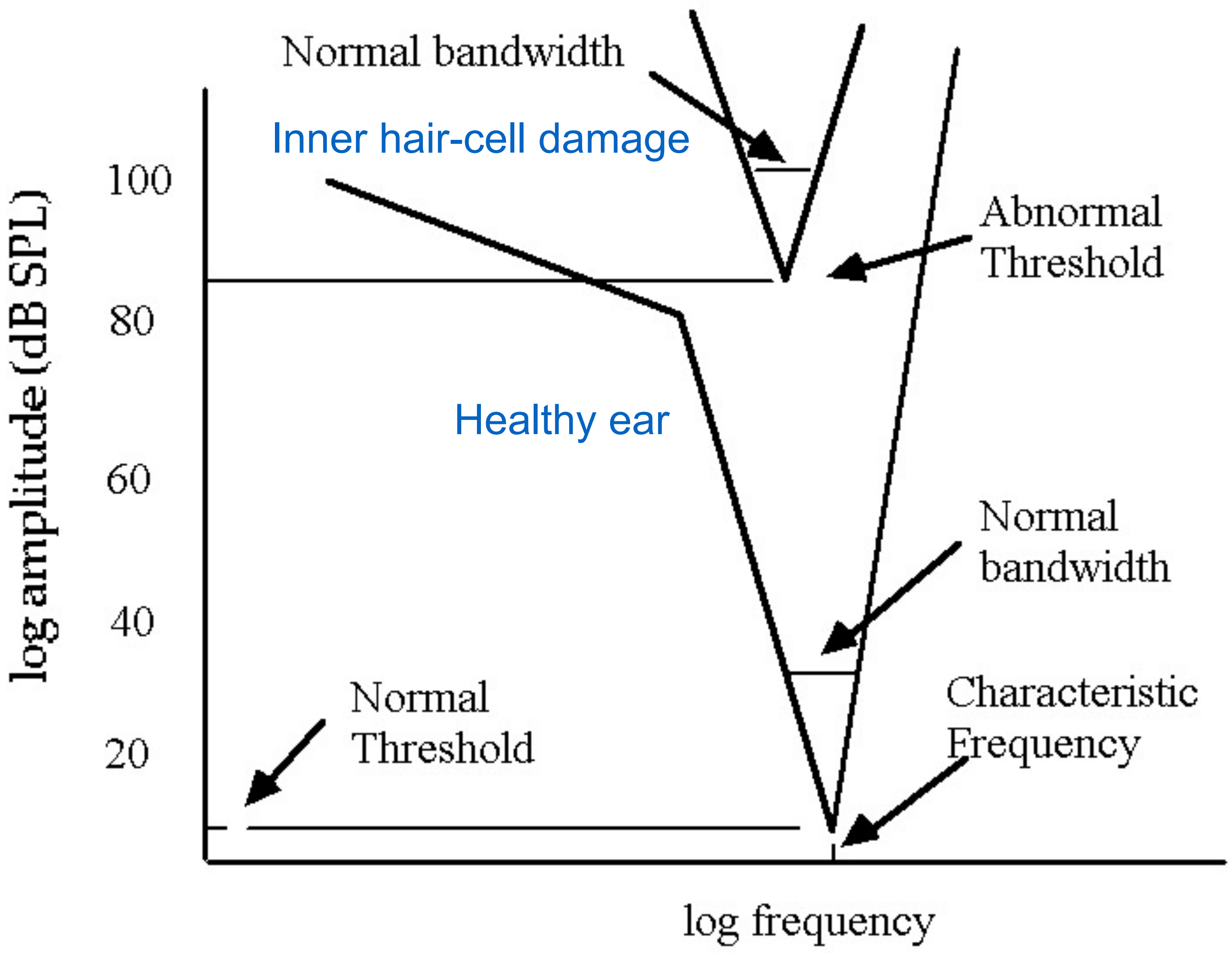
They only amplify sounds that have the characteristic frequency of their place.

- Increases sensitivity (lowers thresholds)
- Increases selectivity (reduces bandwidth of auditory filter)
- Gives ear a logarithmic (non-linear) amplitude response
- Produce Oto-acoustic emissions

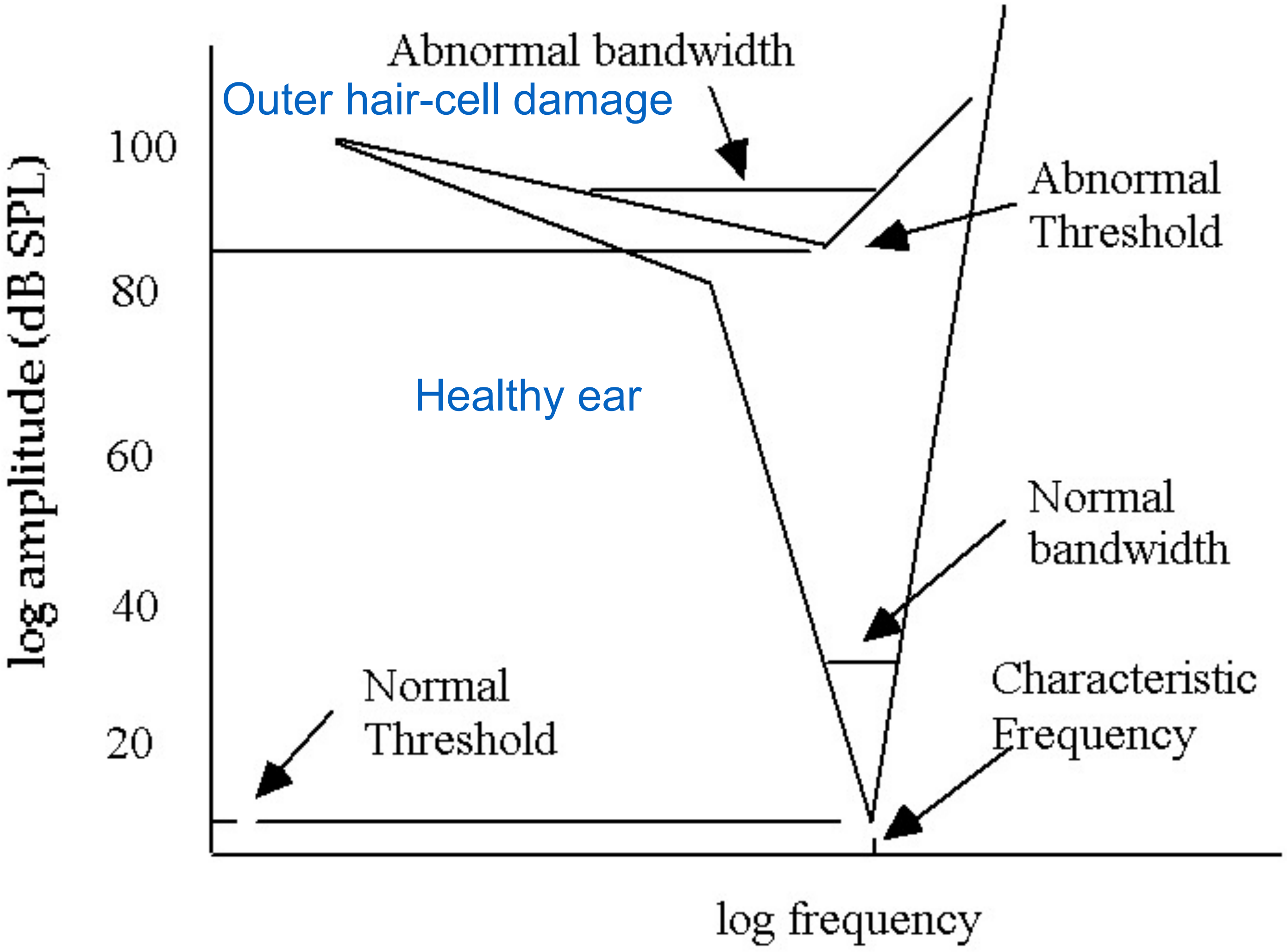
Frequency-threshold curves



Auditory tuning curves



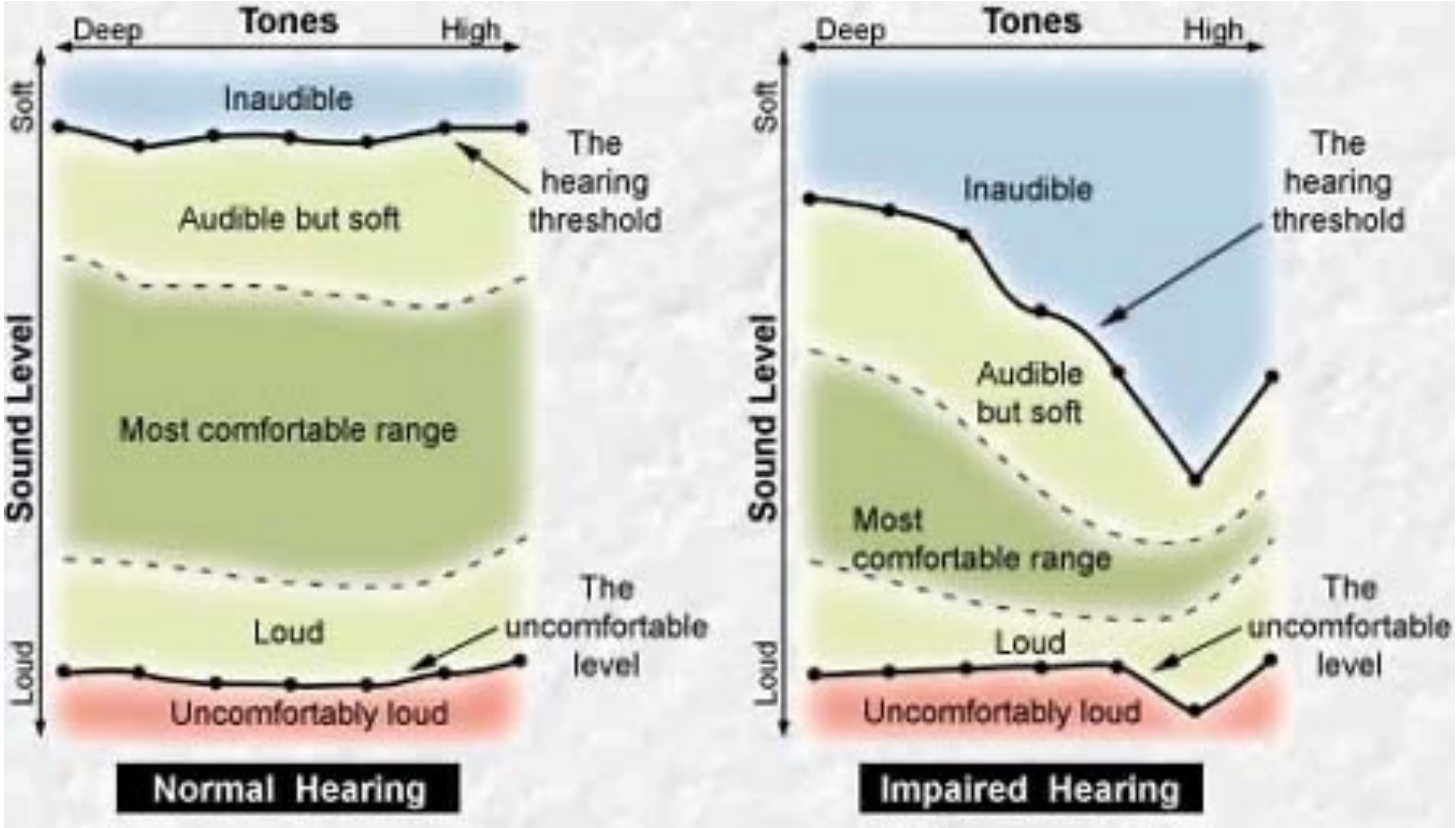
Outer-hair cell damage



Conductive vs. Sensori-neural deafness

	Conductive	Sensori-neural	Sensori-neural
Origin	Middle-ear	Cochlea (IHCs)	Cochlea (OHCs)
Thresholds	Raised	Raised	Raised
Filter bandwidths	Normal	Normal	**Increased**
Loudness growth	Normal	Normal	**Increased** (*Recruitment*

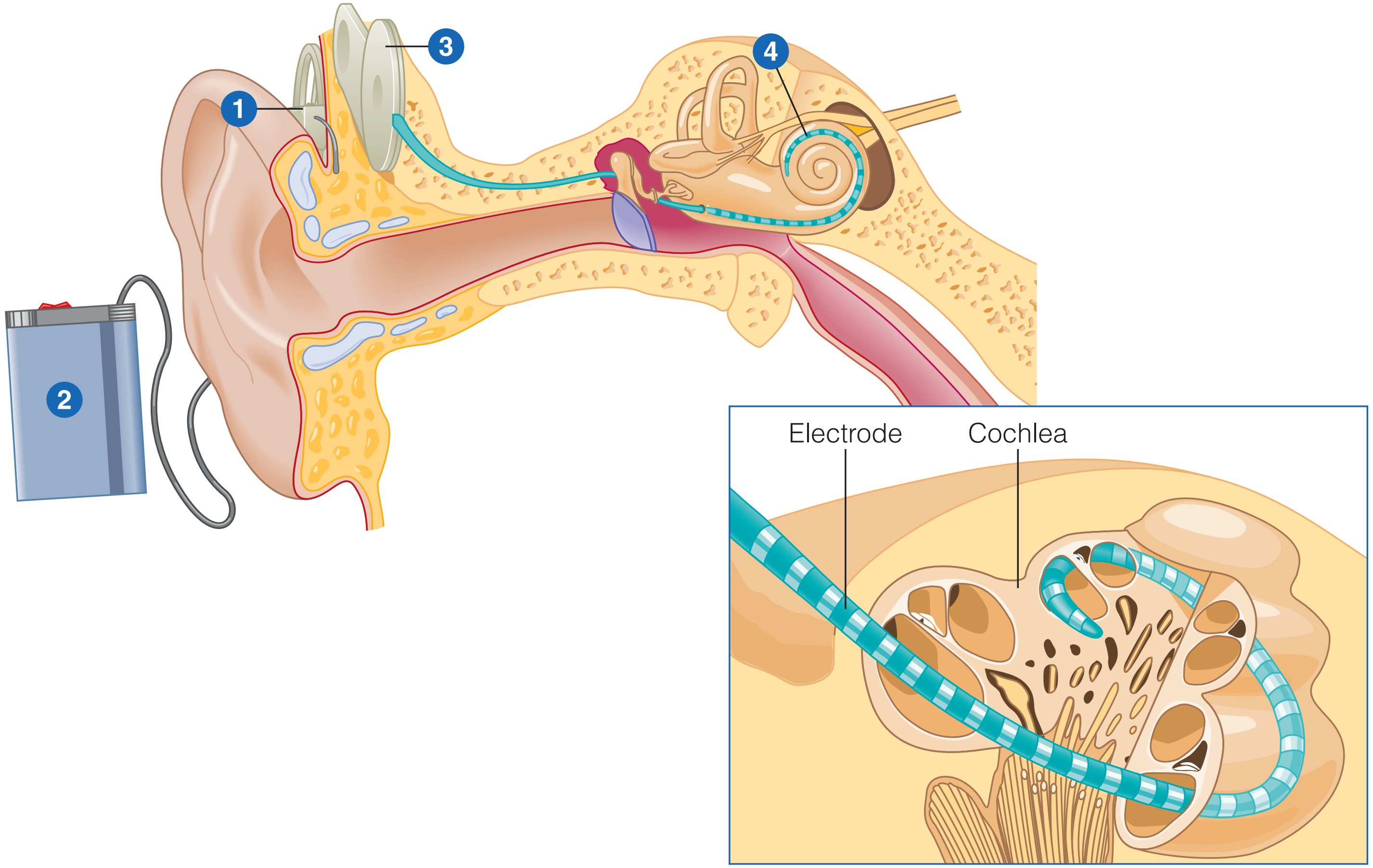
Normal vs. Impaired Dynamic Range



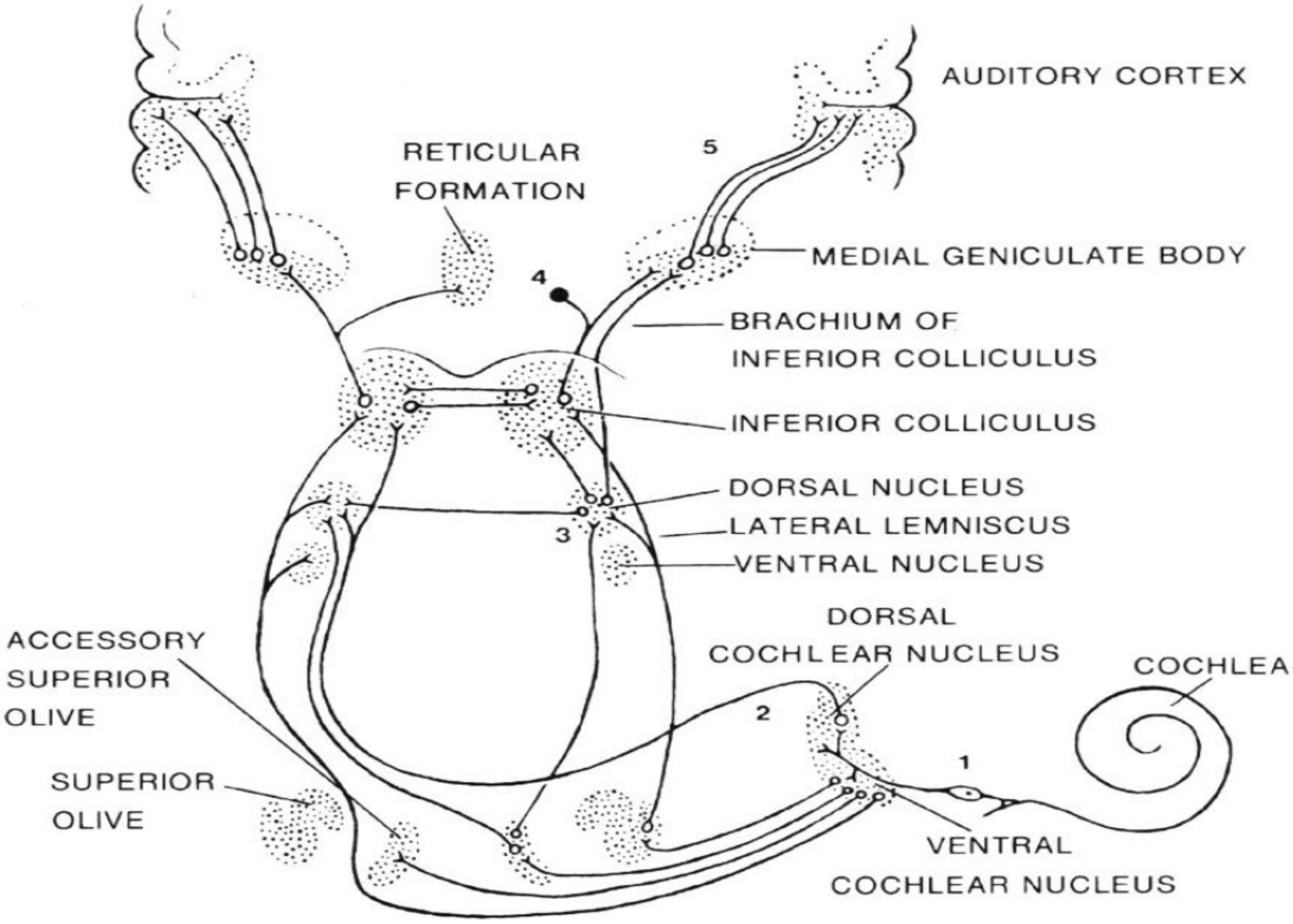
Normal auditory non-linearities

- Normal loudness growth (follows Weber’s Law, which is logarithmic, not linear)
- Combination tones
- Oto-acoustic emissions

Hearing aid and Cochlear implant device



Higher Levels in the Auditory System



Q&A