

Lecture 21: Auditory Physiology

Mark Hasegawa-Johnson

ECE 537: Speech Processing

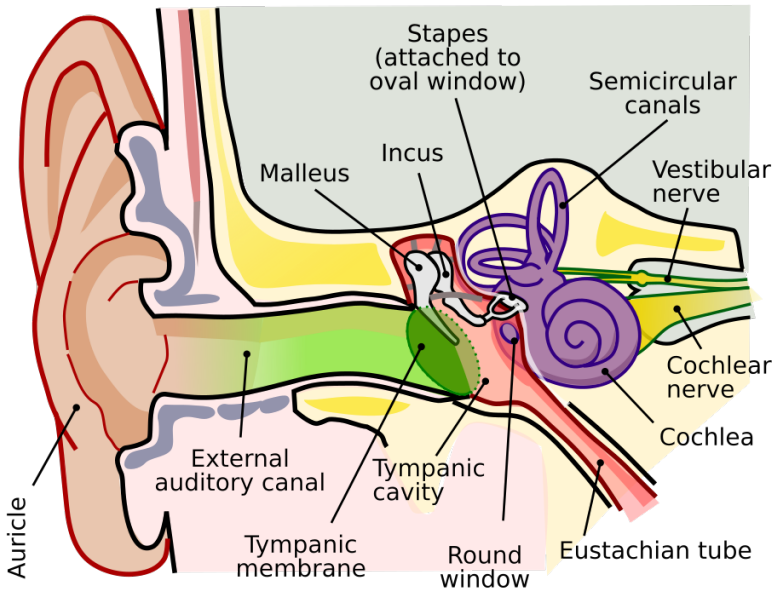
Outline

1 Outer Ear

2 Middle Ear

3 Inner Ear

4 Summary



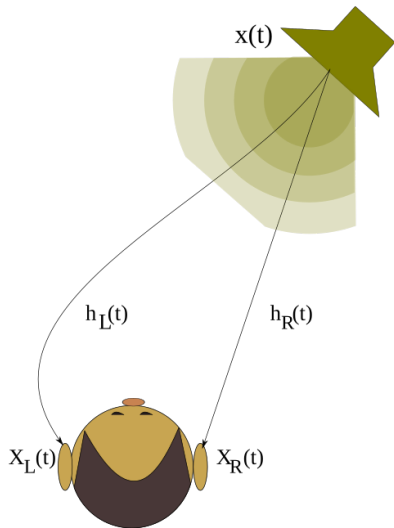
Some functions of the auricle and ear canal

- Auricle: Direction-dependent filtering
- Ear canal: Pre-emphasizes sounds up to $\sim 3\text{kHz}$

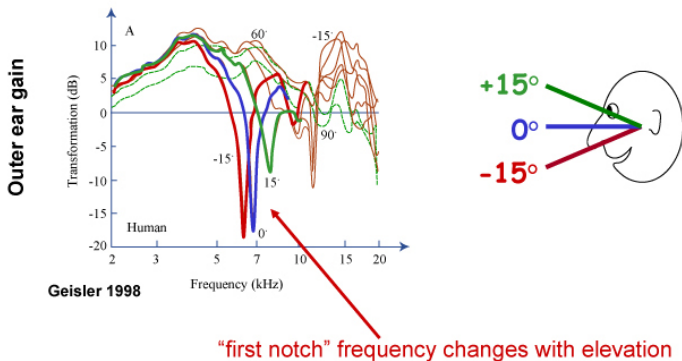
The **head-related transfer functions** (HRTFs) are

$$H(\Omega, r, \theta, \phi) = \frac{X_L(\Omega)}{X(\Omega, r, \theta, \phi)}$$

as functions of the source distance, r , azimuth θ , and elevation ϕ .



“Head Related Transfer Function” (HRTF)



CC-BY-SA 4.0, https://commons.wikimedia.org/wiki/File:MIT_HST.722J_-_9.044J_Brain_Mechanisms_for_Hearing_and_Speech,_Fall_2005_-_Head_Related_Transfer_Function_-_4809307652.jpg

The Ear Canal: Emphasis up to 3kHz

- The ear canal is a 3cm quarter-wave resonator, i.e., its first resonance is

$$\frac{c}{4l} \approx 3000\text{Hz}$$

- At frequencies $f < 3000$, its frequency response is

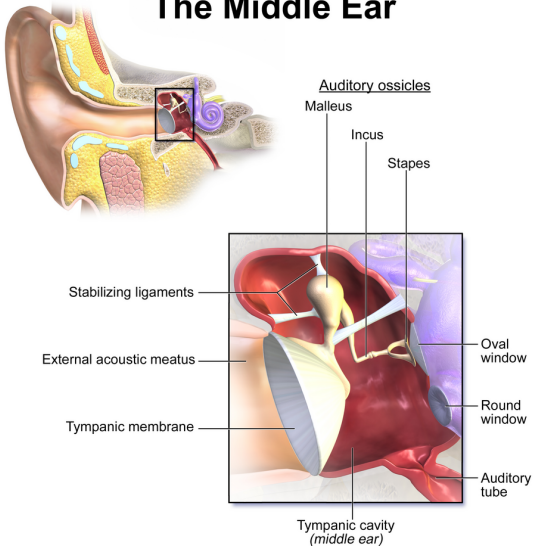
$$|H(\omega)| = \frac{1}{\sqrt{|(1 - p_1 e^{-j\omega})(1 - p_1^* e^{-j\omega})|^2}} \approx \frac{F_s/4\pi}{|3000 - f|}$$

- The voice source drops off at $1/f$ at low frequencies ($1/f^2$ at high frequencies); the ear canal compensates with a factor of $1/|3000 - f|$ so that all frequencies in the range $0 \leq f \leq 3000$ are audible.

Outline

- 1 Outer Ear
- 2 Middle Ear
- 3 Inner Ear
- 4 Summary

The Middle Ear



https://commons.wikimedia.org/wiki/File:Blausen_0330_EarAnatomy_MiddleEar.png

Middle Ear: Impedance Matching

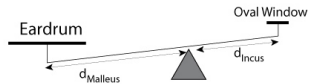
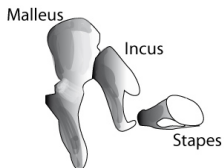
- The impedance of the air-filled ear canal is

$$\frac{P_+(\Omega)}{U_+(\Omega)} = \frac{\rho c}{A} \approx \frac{1.22 \left[\frac{\text{kg}}{\text{m}^3} \right] \times 350 \frac{\text{m}}{\text{s}}}{10^{-4} \text{m}^2} \approx 0.043$$

- Impedance of the water-filled cochlea is

$$\frac{P_+(\Omega)}{U_+(\Omega)} = \frac{\rho c}{A} \approx \frac{1000 \left[\frac{\text{kg}}{\text{m}^3} \right] \times 1000 \frac{\text{m}}{\text{s}}}{10^{-6} \text{m}^2} \approx 1$$

- Factor of 25: Compensate by reducing flow by $5\times$, increasing pressure by $5\times$.

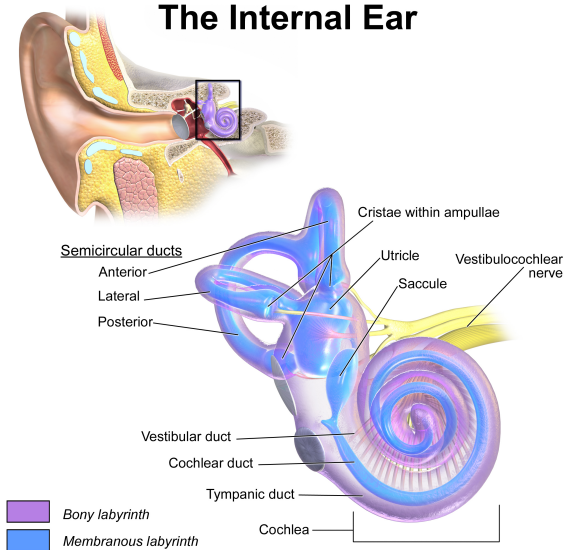


https://commons.wikimedia.org/wiki/File:MiddleEar_Amplification.jpg

Outline

- 1 Outer Ear
- 2 Middle Ear
- 3 Inner Ear**
- 4 Summary

The Internal Ear

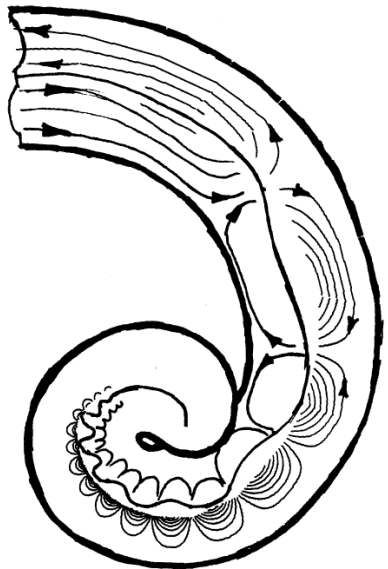


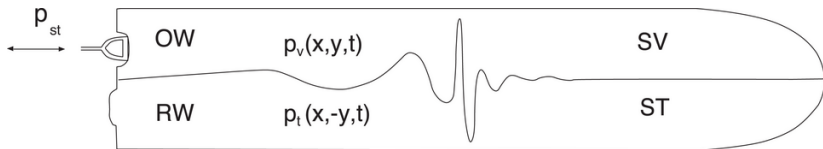
https://commons.wikimedia.org/wiki/File:Blausen_0329_EarAnatomy_InternalEar.png

Vestibular duct, Basilar membrane, Tympanic duct

- Sound propagates down the length of the vestibular duct
- The vestibular duct and tympanic duct are separated by the floppy basilar membrane
- Each point on the basilar membrane is tuned to a different frequency
- Sounds at that frequency are sucked out of the vestibular duct by vibration of the basilar membrane

When a frequency component (f) in the traveling wave reaches the x at which the basilar membrane resonant frequency is $f(x) = f$, the energy at that frequency is sucked out of the vestibular duct



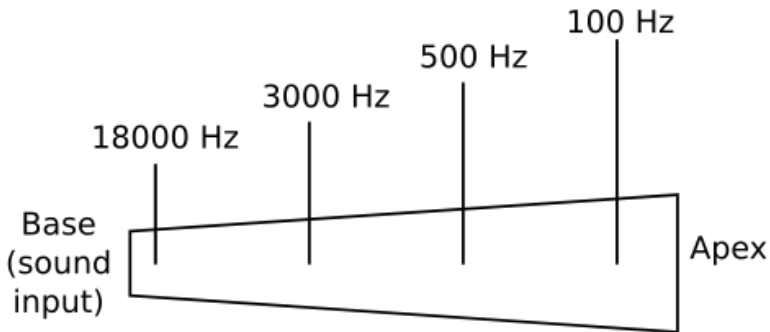


https://commons.wikimedia.org/wiki/File:Schematic_uncoiled_cochlea.png

- The energy sucked out of the vestibular duct causes large vibration of the basilar membrane.
- The response of each x on the basilar membrane is a bandpass filter, tuned to a center frequency $f(x)$
- We call these bandpass filters the “auditory filters”

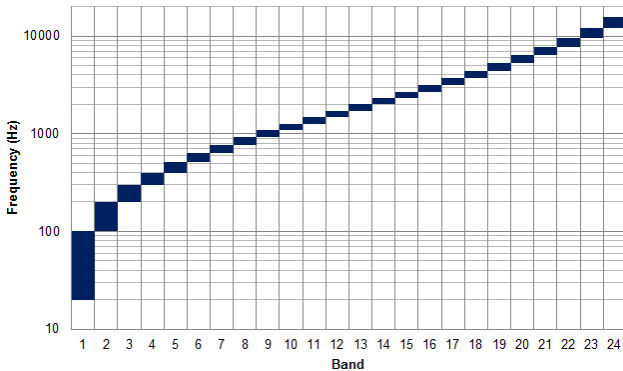
Tuning of the basilar membrane

The basilar membrane is thin and tight at its base, thick and floppy at its apex, so $f(x)$ is a decreasing function of x :

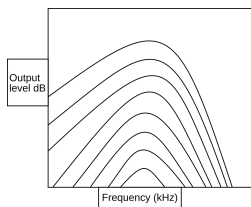


https://commons.wikimedia.org/wiki/File:Basilarmembrane2_topath.svg

- The basilar membrane is 24-32mm long
- Here is an estimate of the frequencies mapped to each millimeter, where $x = 1$ is the apex and $x = 24$ is the base
- The scale is roughly $f \propto x$ for $f < 500$, and $f \propto \ln(x)$ for $f > 500$.



Auditory Filters: Symmetric at Low Intensity, Skewed Left at High Intensity

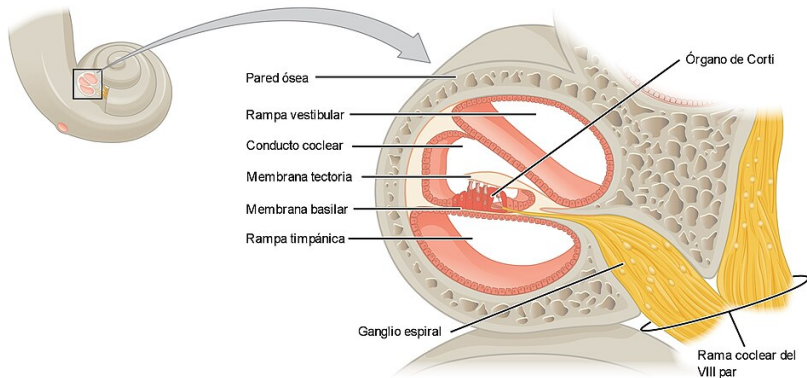


Public Domain,

<https://commons.wikimedia.org/wiki/File:OutputlevelMoore.svg>

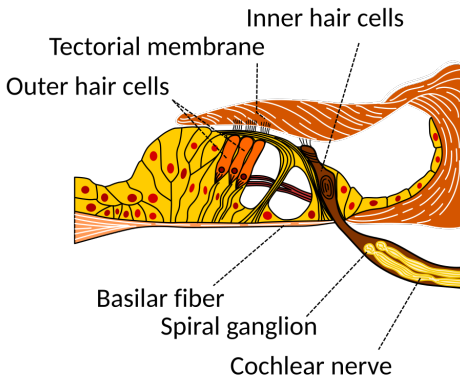
- The auditory filter at frequency f_1 also reacts to loud tones at frequency f_2 if $f_2 < f_1$ (f_2 is more apical than f_1), because the tone at f_2 still has lots of energy when it passes over the part of the basilar membrane tuned to f_1 .
- The filter at f_1 doesn't respond as well to frequency f_2 if $f_2 > f_1$ (f_2 is more basal than f_1), because the tone at f_2 has been absorbed by the basilar membrane before it ever reaches the part of the basilar membrane tuned to f_1 .

Microscopic hairs on the basilar membrane flex back and forth as the membrane vibrates



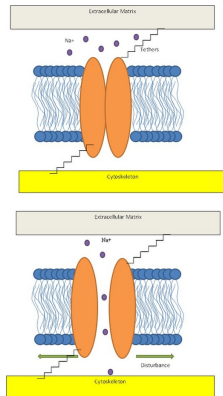
https://commons.wikimedia.org/wiki/File:1406_Cochlea_esp.jpg

Microscopic hairs on the basilar membrane flex back and forth as the membrane vibrates



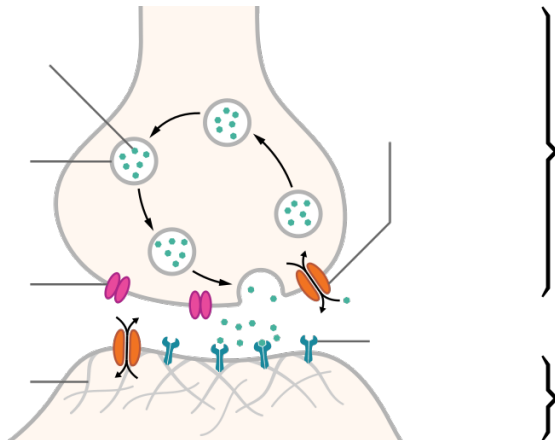
https://commons.wikimedia.org/wiki/File:Organ_of_corti.svg

- Like neurons, resting voltage inside the hair cell is -50mV .
- Bending of the hair cell opens ion gates in its membrane, depolarizing it.
- Depolarization causes it to release neurotransmitter, activating a nearby neuron.

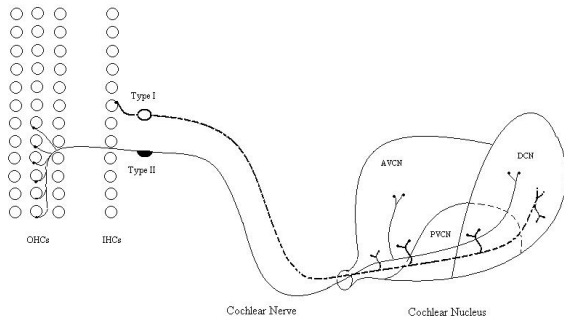


https://commons.wikimedia.org/wiki/File:Mammalian_Hair_Cell_Model.jpg

Hair cell releases neurotransmitter, activating a nearby neuron

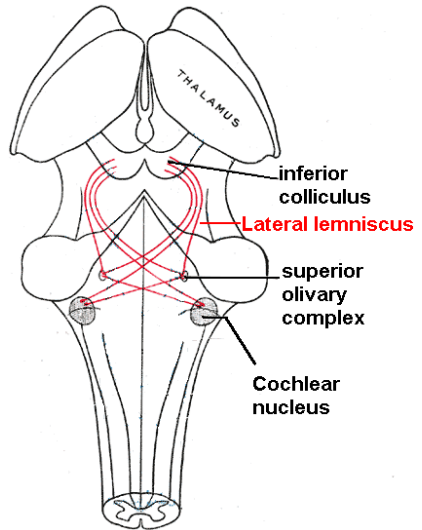


https://commons.wikimedia.org/wiki/File:SynapseSchematic_lines.svg



- Inner hair cells (IHC) send perception (afferent) to posterior ventral cochlear nucleus (PVCN)
- Anterior ventral cochlear nucleus (AVCN) sends feedback (efferent) to outer hair cells (OHC)
- Both are tonotopic (neurons sorted by $f(x)$)

From brainstem's cochlear nerve, neurons travel to the inferior colliculus (with 0, 1, or 2 synapses on the way), from which connections are made to the auditory cortex. All neuron fields in brainstem and primary auditory cortex remain tonotopic (sorted by frequency).



https://commons.wikimedia.org/wiki/File:Lateral_lemniscus.PNG

Outline

- 1 Outer Ear
- 2 Middle Ear
- 3 Inner Ear
- 4 Summary

Summary

- Outer ear: Direction-dependent filtering (HRTF) and pre-emphasis up to 3kHz
- Middle ear: Impedance-matching between air and water
- Basilar membrane: mechanically tuned bank of bandpass filters, tuning is roughly linear below 500Hz, roughly logarithmic above
- Hair cell: mechanical-to-electric transducer that activates neurons when bent
- Brainstem: all processing is tonotopic (sorted by $f(x)$), with temporal processing done temporally