

The logo of The University of Texas at Dallas, featuring a circular seal with the letters 'UTD' in the center, the text 'THE UNIVERSITY OF TEXAS AT DALLAS' around the top, and 'EST. 1969' at the bottom. Two stars are positioned on either side of the 'EST. 1969' text.

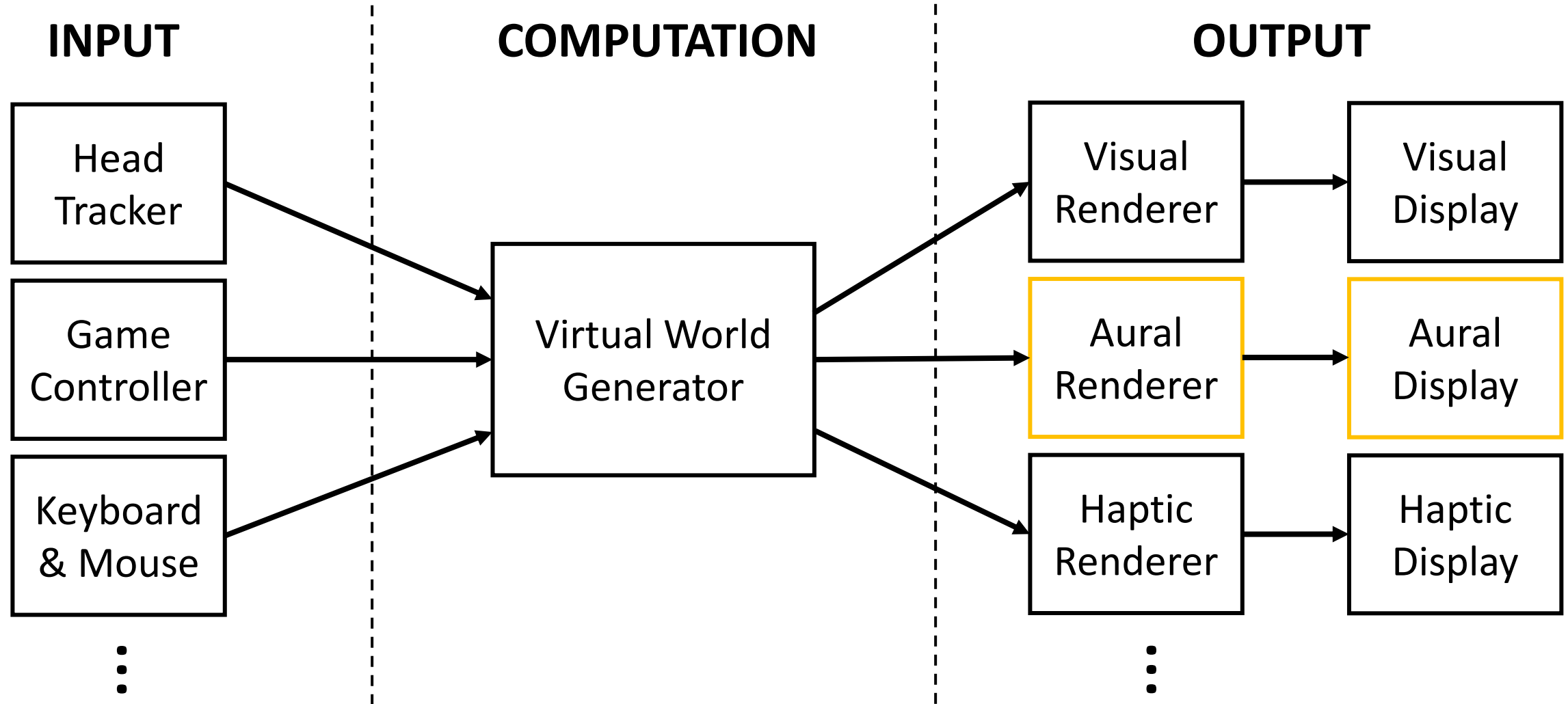
Audio: Physics, Physiology and Perception

CS 6334 Virtual Reality

Professor Yu Xiang

The University of Texas at Dallas

Review of VR Systems



The Physics of Sound

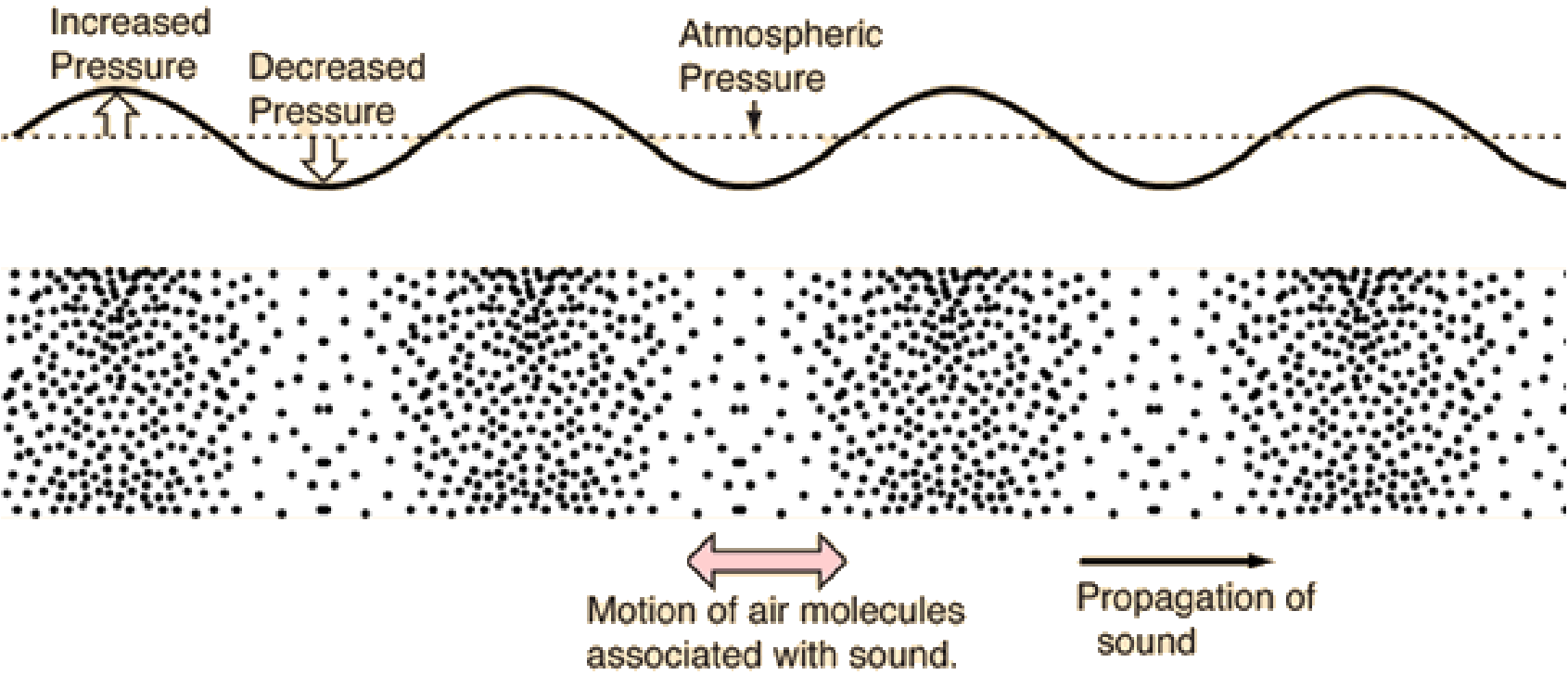


Figure by Dale Pond

The Physics of Sound

- Sound corresponds to vibration in a medium
 - Air, water, solids
 - No sound in vacuum
- Pressure variation
 - Compression extreme to a decompressed, rarefaction extreme
- Sound waves are typically produced by vibrating solid materials
 - Striking a large bell
 - Air flow: flute
 - Humans: lungs to force air through the vocal cords

The Physics of Sound

- Sound pressure level is reported in decibels (dB)

$$N_{db} = 20 * \log_{10}(p_e/p_r)$$



Pressure level of the
peak compression

Reference pressure level
 2×10^{-7} newtons / square meter

Breathing: 10 dB

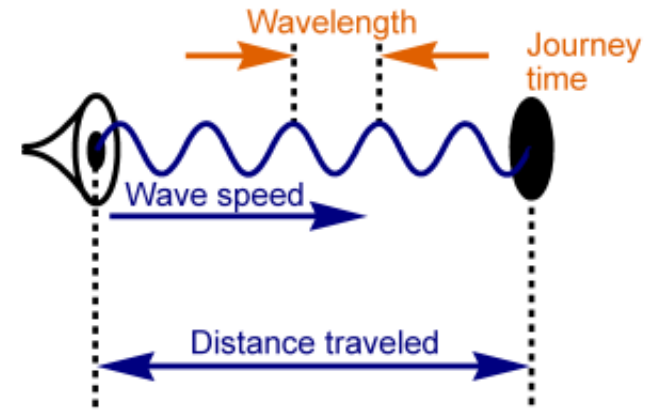
Conversion in restaurants: 60 dB

Motorcycle at ft: 90 dB

Jet take-off at 25 meters: 150 dB

The Physics of Sound

- Sound ray
 - Perpendicular to the sound propagation wavefront
- Attenuation
 - Sound intensity decreases by a constant factor (fixed percentage) for every unit distance from the source (exponential decay)
- Propagation speed
 - 343.2 meters per second through air at 20° C (68 ° F)
 - Light is about 874,000 times faster



The Physics of Sound

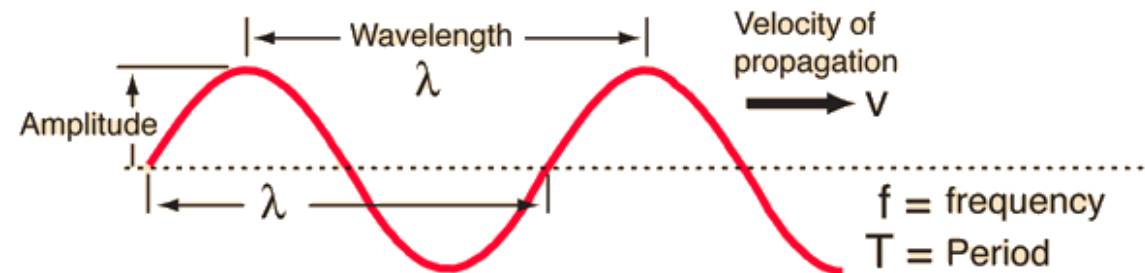
- Frequency

- The number of compressions per second (called pitch)
- 20 Hz to 20,000 Hz for human hearing
- Ultrasound: above 20,000 Hz
- Infrasound: below 20 Hz

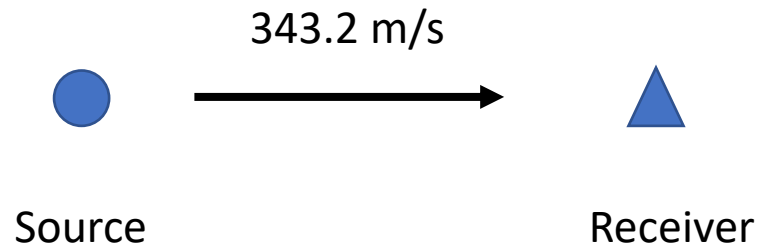
$$f = \frac{v}{\lambda}$$

- Wavelength

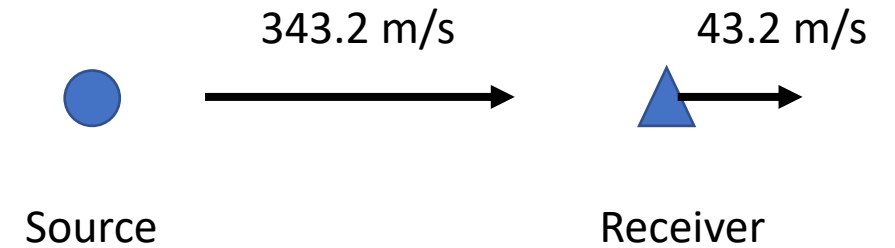
- At 20 Hz: $\lambda = 343.2/20 = 17.1\text{m}$
- At 20,000 Hz: $\lambda = 17.1\text{mm}$



Doppler Effect



$$f = \frac{v}{\lambda} \quad 343.2 \text{ m/s}$$



$$f = \frac{v}{\lambda} \quad 300 \text{ m/s}$$

Doppler Effect

- The received frequency shifts due to the relative motion between the source and the receiver

$$f_r = \left(\frac{s + v_r}{s + v_s} \right) f_s$$

$$s = 343.2$$

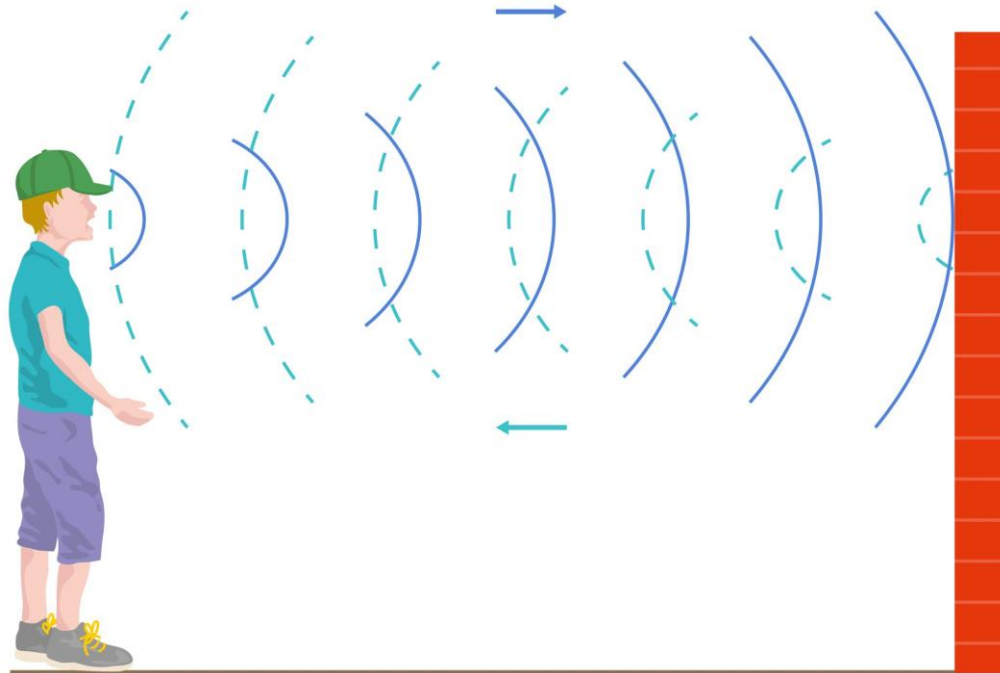
$$v_r = -43.2$$

$$v_s = 0$$

- s is the propagation speed in the medium
- v_r is the velocity of the receiver
- v_s is the velocity of the source

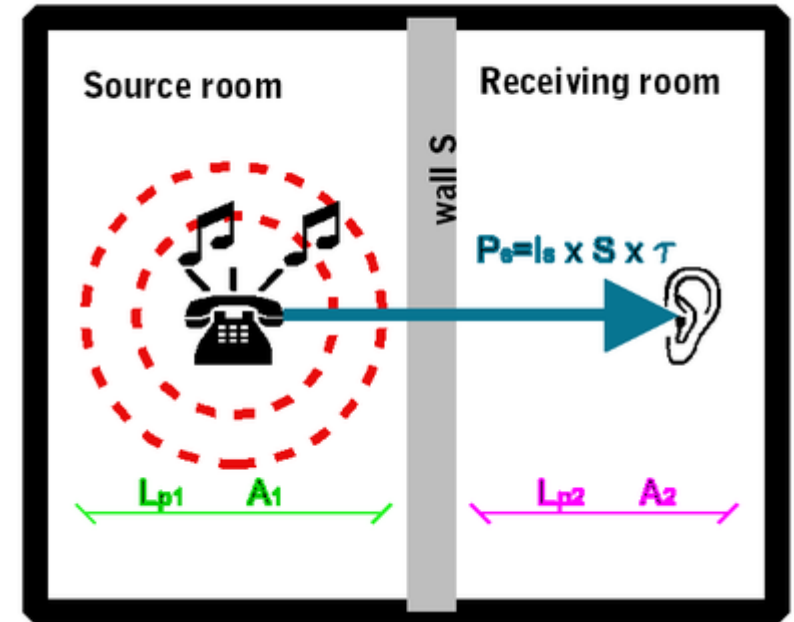
Siren seems to change pitch as a police car passed by

Reflection and Transmission



Reflection

Airborne sound transmission



Transmission

Diffraction

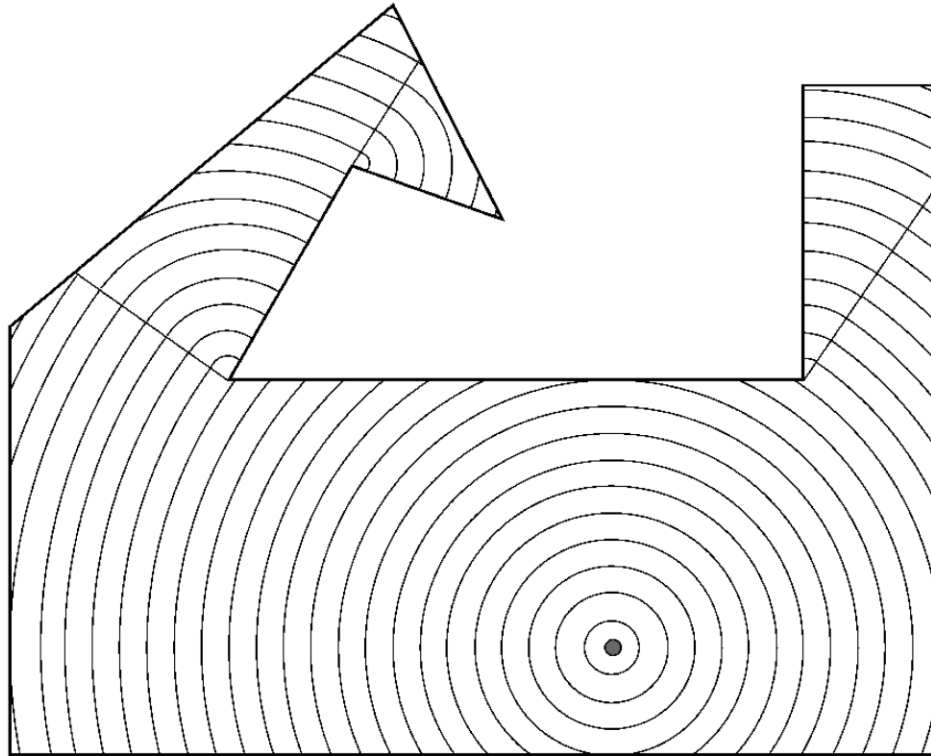
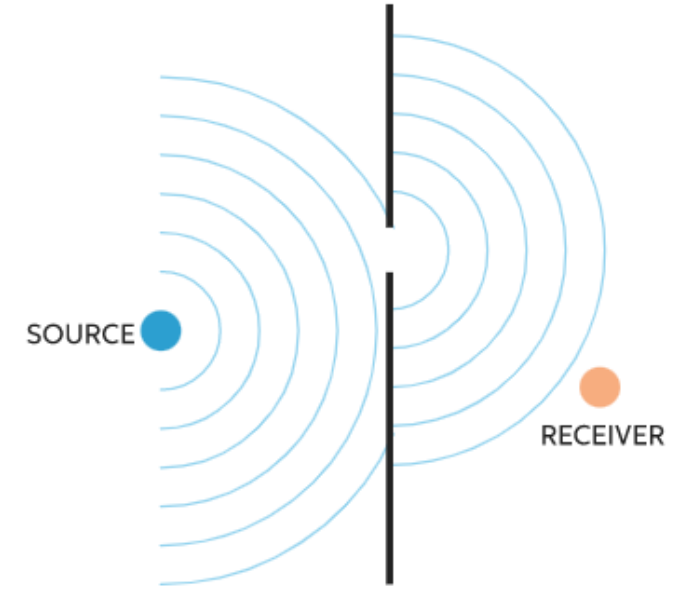


Figure 11.2: Waves can even bend around corners, due to *diffraction*. A top-down view of a room is shown. At each of the three interior corners, the propagating wavefront expands around it.

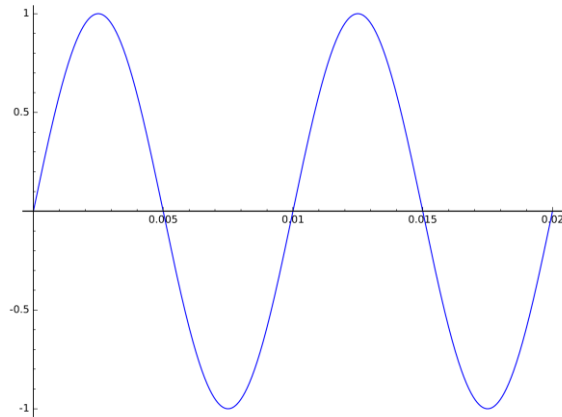


More diffraction occurs for longer wavelengths

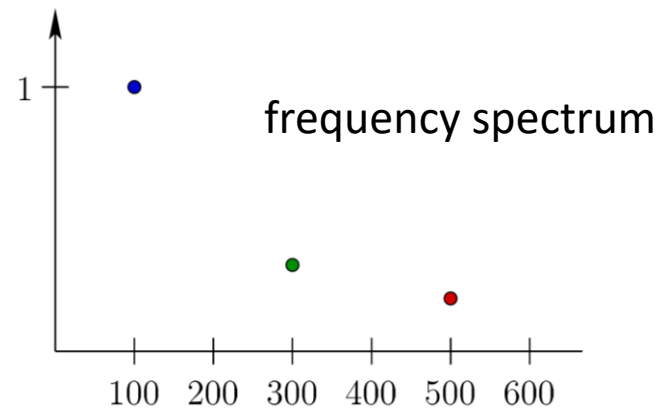
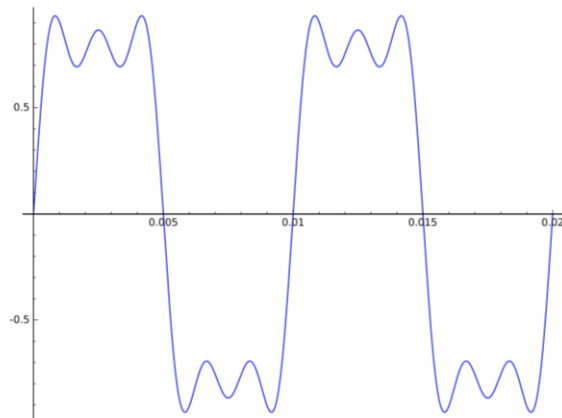
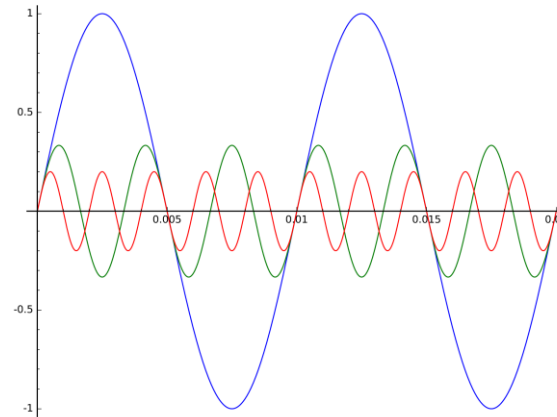


Spectral Decomposition

A pure tone (sinusoid)



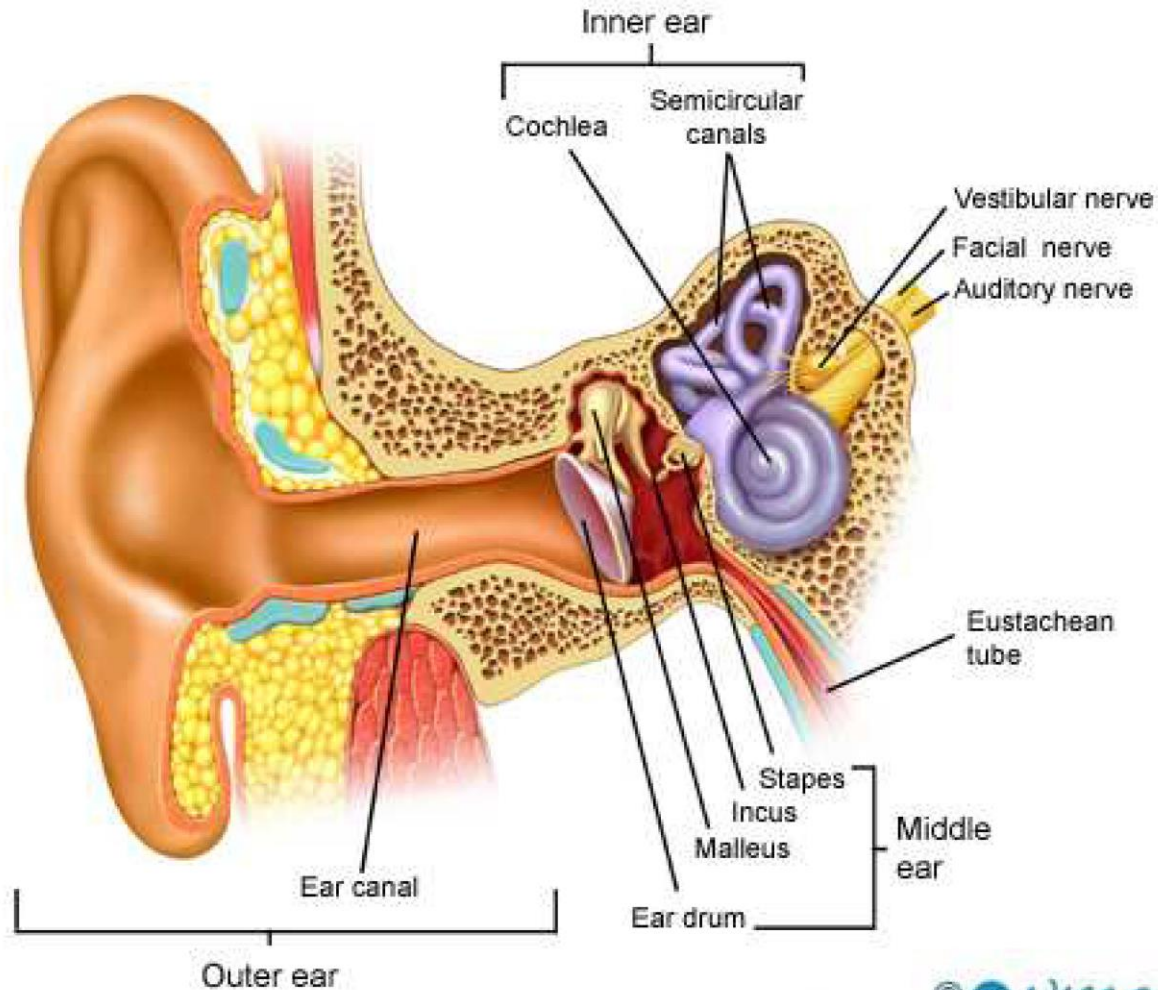
Three pure tones



Fourier analysis: any periodic function can be decomposed into sinusoids

Directly adding the three pure tones

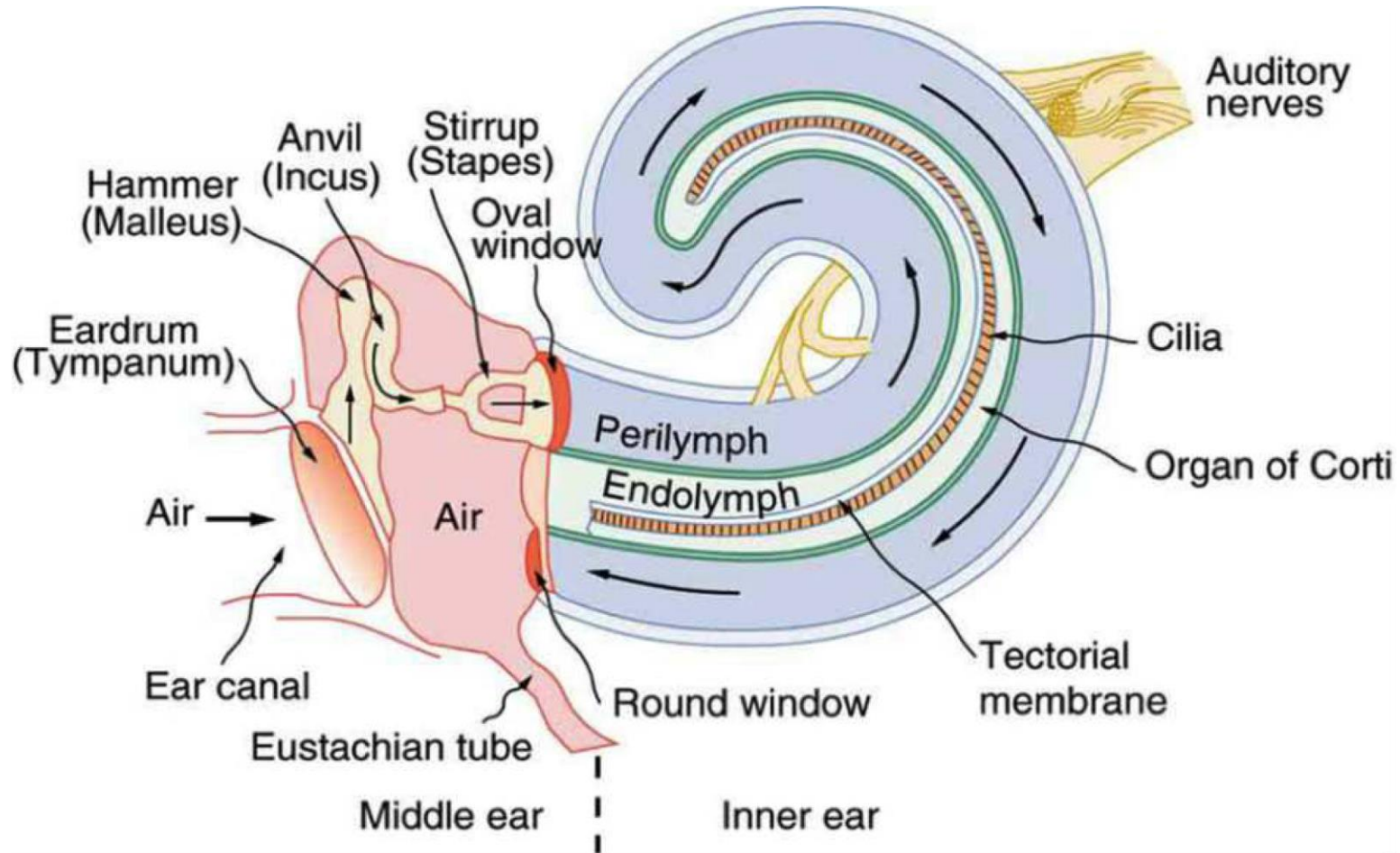
The Human Ear



© VMC
Virtual Medical Centre

- Pinna, amplifying sounds in 1500 Hz and 7500 Hz range
- Ear drum, 55mm², one pixel camera
- Middle ear (3 bones), converting vibrating air molecules in the outer ear into vibrating liquid in the inner ear
- Inner ear, the vestibular organs and the cochlea

The Cochlea



Eardrum vibration



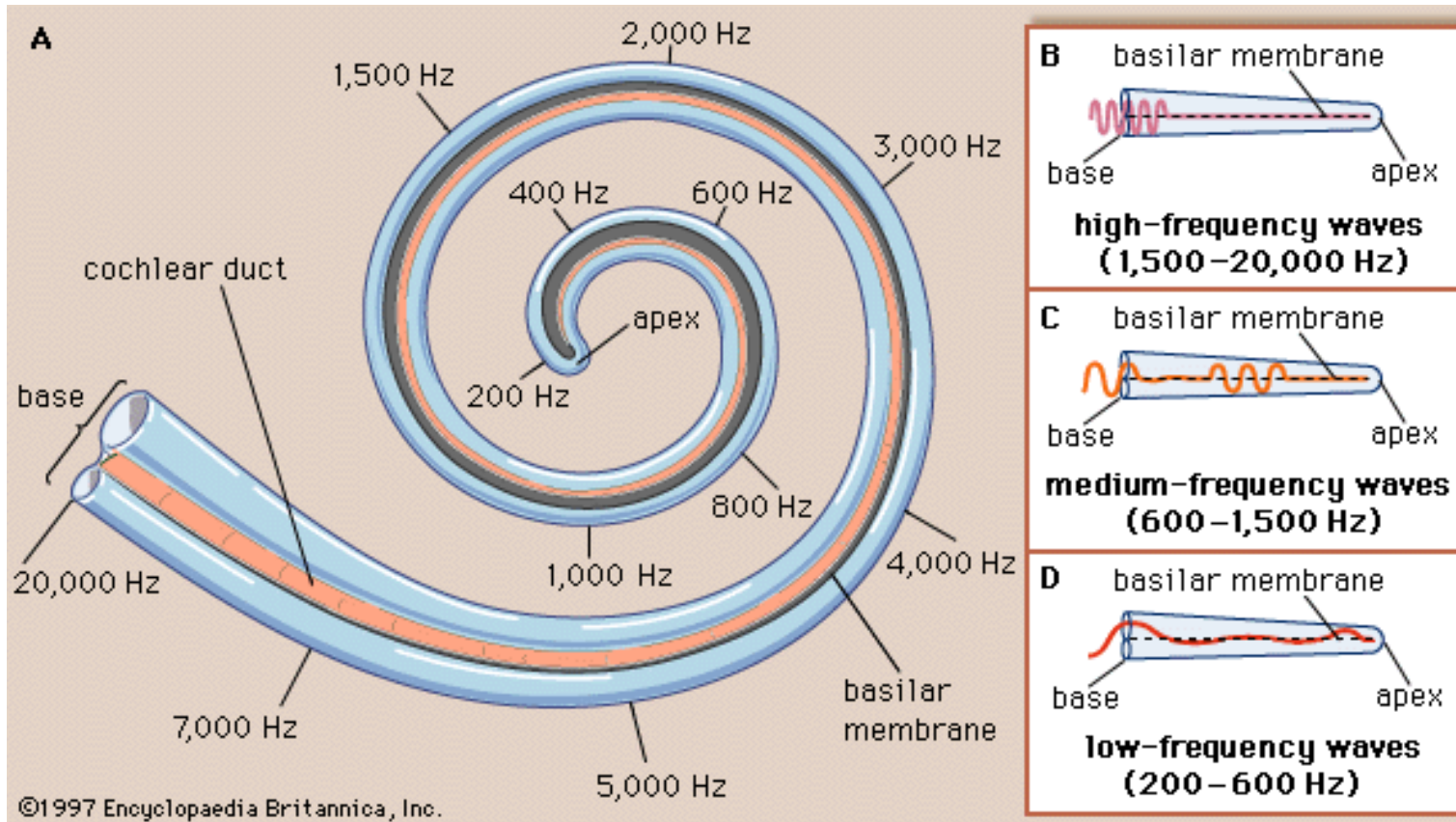
Oval window



Perilymph

Round window

The Cochlea

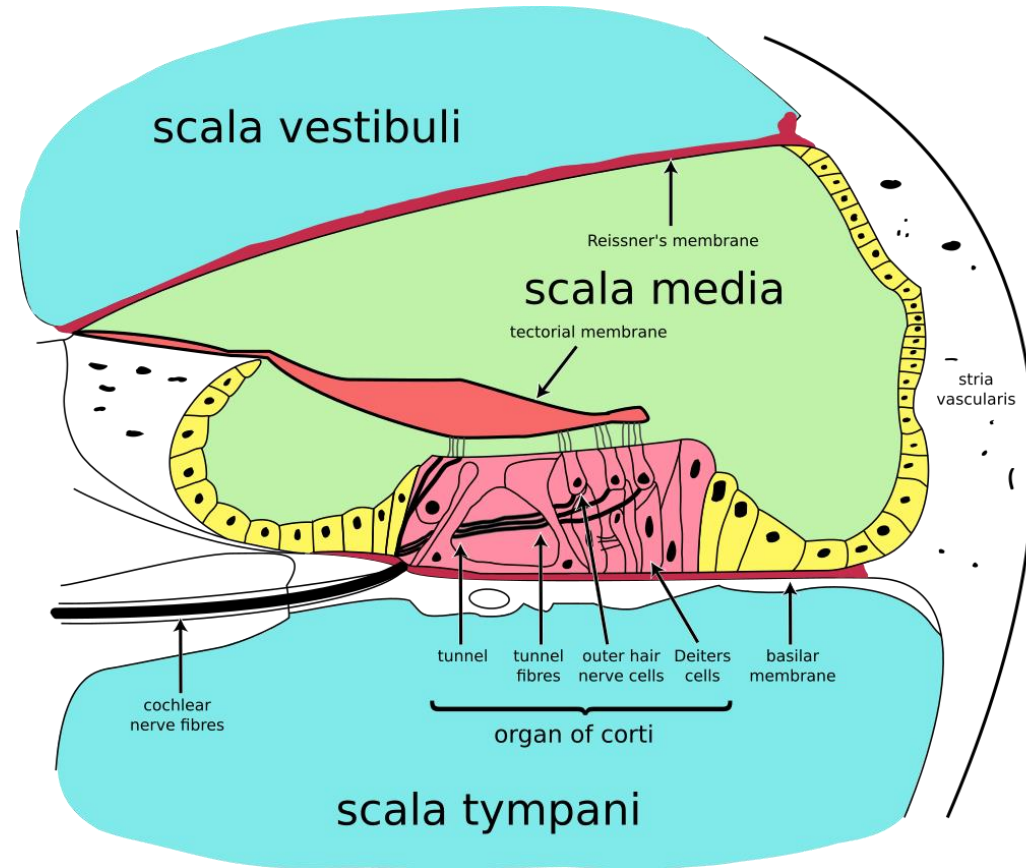


Basilar membrane

- Thin and stiff near the base
- Soft and floppy near the apex

Spectral decomposition

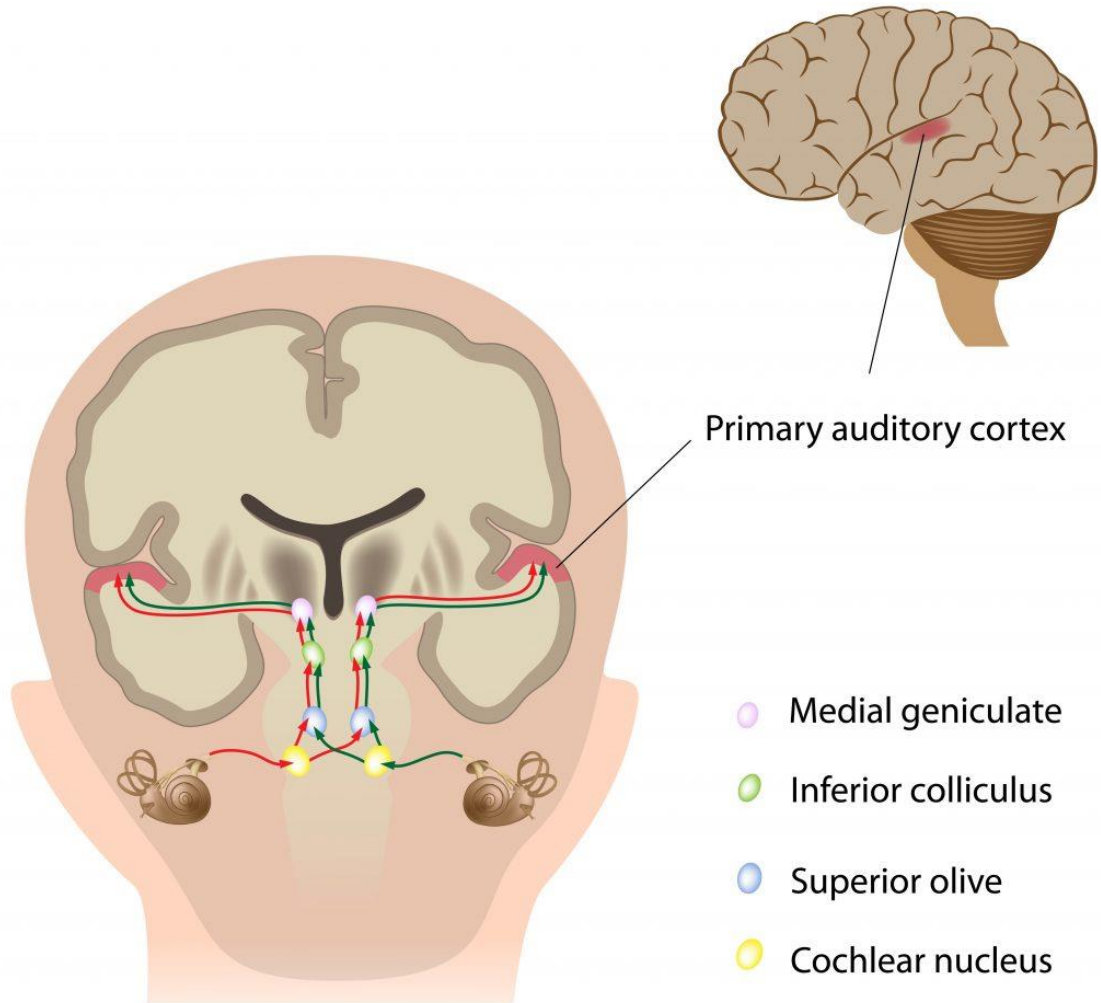
Mechanoreceptors



A cross section of the organ of Corti. The basilar and tectorial membranes move relative to each other, causing the hairs in the mechanoreceptors to bend.

- Mechanoreceptors convert display of hairs into neural impulses
- Hairs are between the tectorial membrane and the basilar membrane
- Each ear contains around 20,000 mechanoreceptors (vs. 100 million photoreceptors in the eyes)

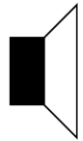
Auditory Pathways



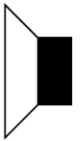
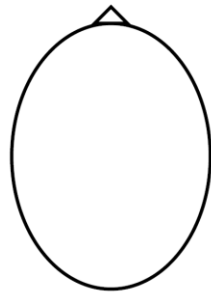
<https://teachmeanatomy.info/neuroanatomy/pathways/auditory-pathway/>

Auditory Perception

- Precedence effect
 - only one sound is perceived if two nearby identical sounds arrive at slightly different times

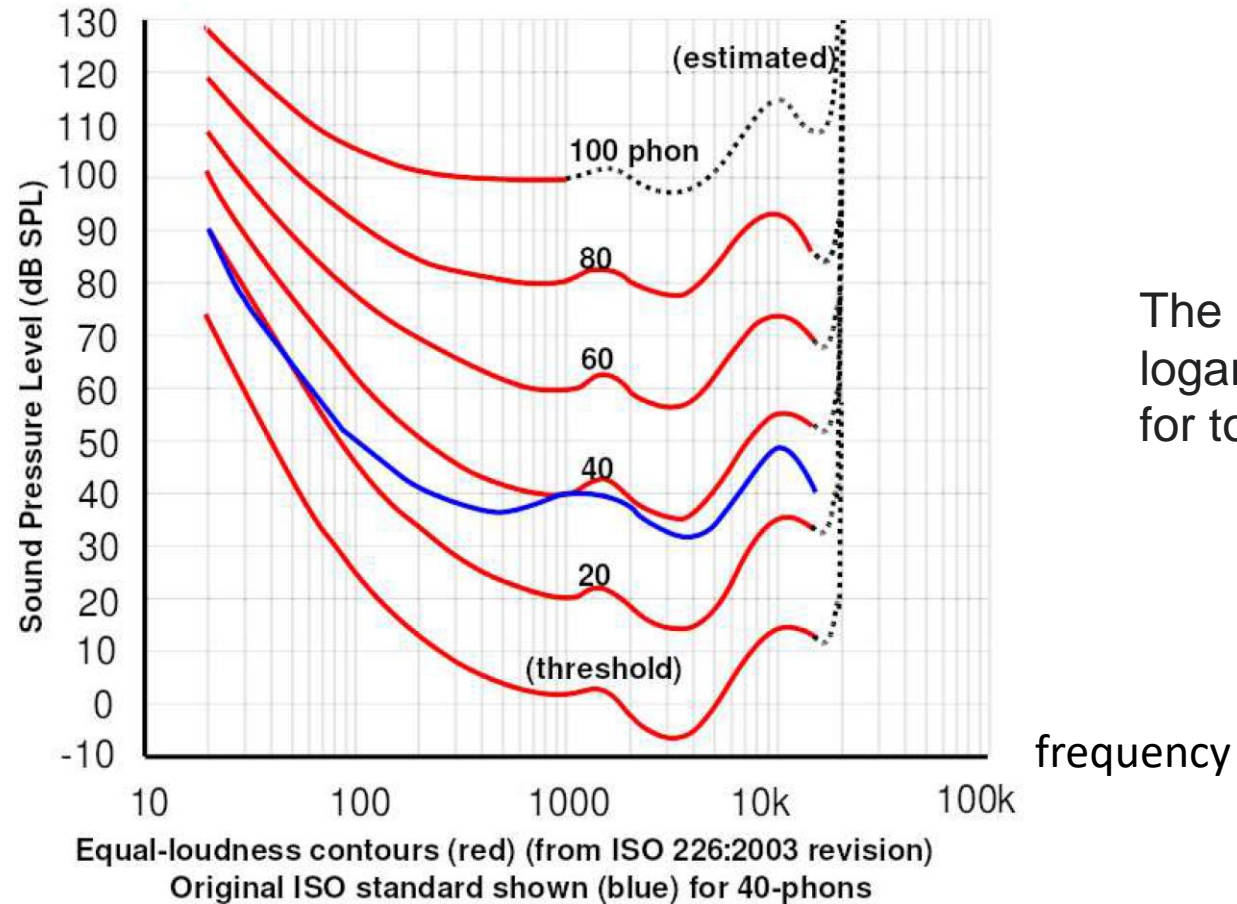


Left



Right

Loudness Perception

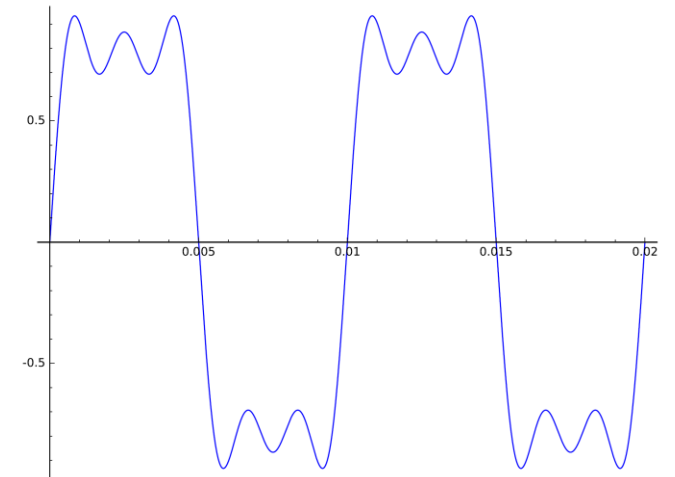


The **phon** is a logarithmic unit of loudness level for tones and complex sounds.

Contours of equal loudness perception as a function of frequency

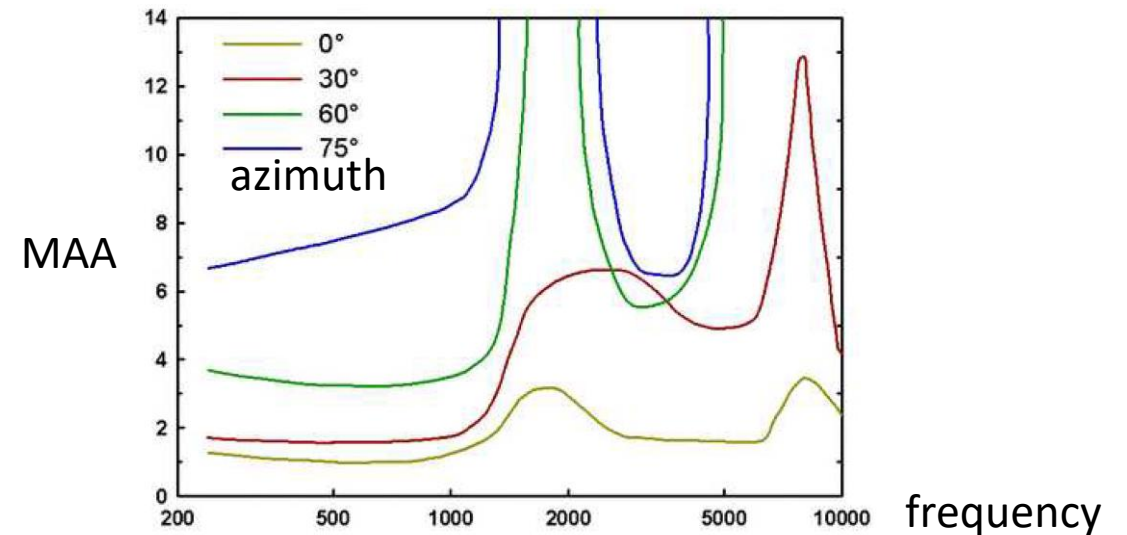
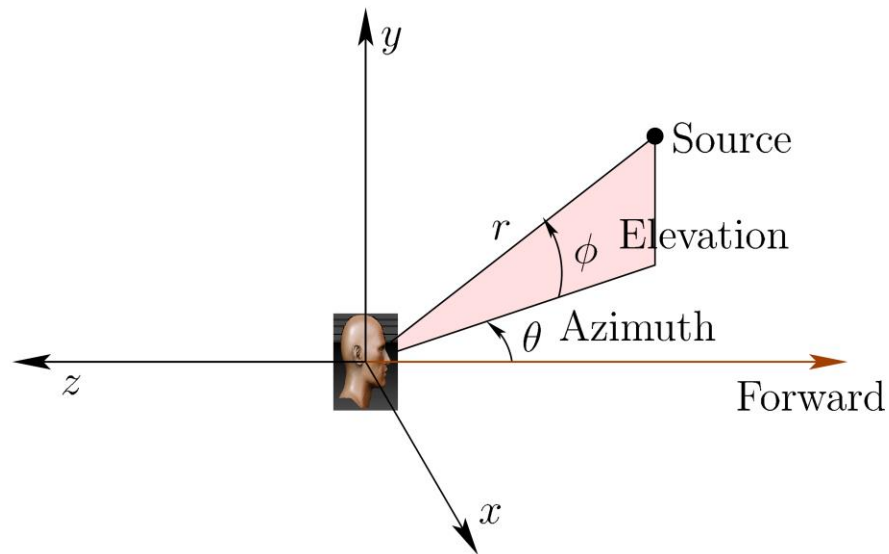
Pitch (Frequency) Perception

- Critical band masking
 - Block out waves that have frequencies outside of a particular range of interest
- Perception of differences in pitch
 - Just noticeable differences (JNDs)
 - $< 1000\text{Hz}$, JND 1Hz
 - $10,000\text{Hz}$, JND 100Hz
- People perceive the tone of the fundamental frequency even when it is removed, and only higher-order harmonics remain
 - Fundamental frequency: largest amplitude wave



Localization

- Estimating the location of a sound source by hearing it (crucial for VR)
- Minimum Audible Angle (MAA): minimum amount of angular variation that can be detected by a human listener

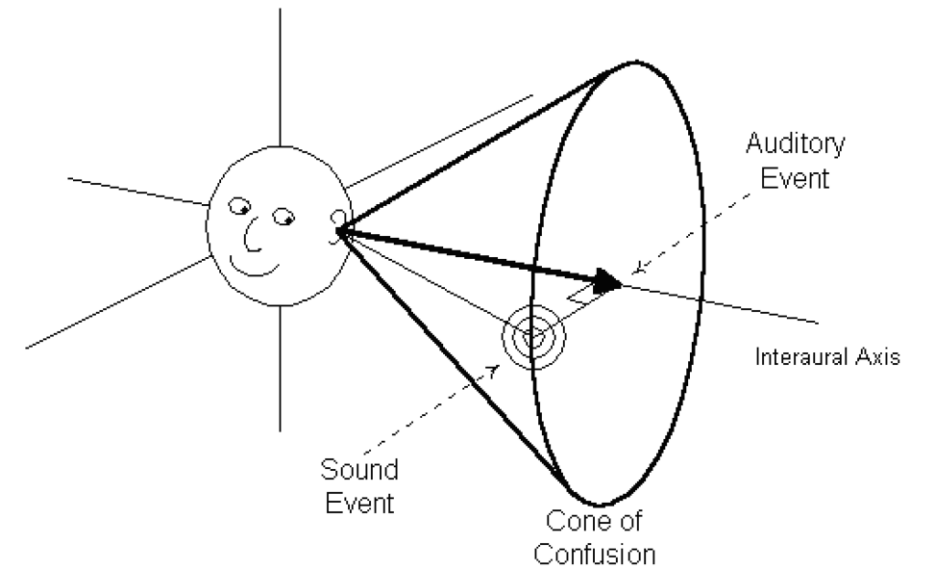
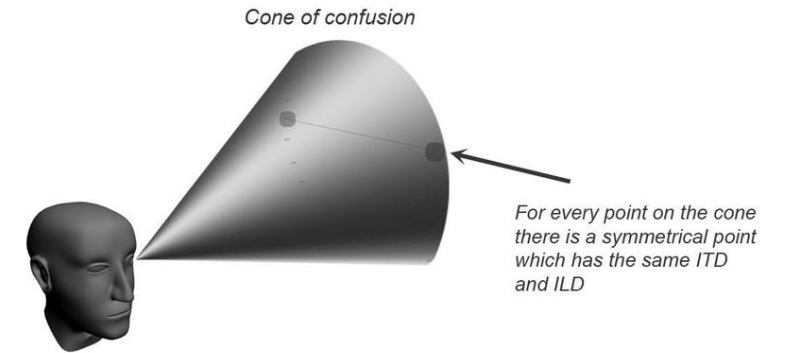


Monaural Cues for Localization

- The pinna is shaped asymmetrically so that incoming sound is distorted in a way that depends on the direction from which it arrives, especially the elevation.
- The amplitude of a sound decreases quadratically with distance.
- For distant sounds, a distortion of the frequency spectrum occurs because higher-frequency components attenuate more quickly than low-frequency components.
- The reverberations entering the ear as the sounds bounce around; this is especially strong in a room

Binaural Cues for Localization

- Interaural Level Difference (ILD), the difference in sound magnitude as heard by each ear
- Interaural Time Difference (ITD), the distance between the two ears is approximately 21.5cm, which results in different arrival times of the sound from a source



The cone of confusion is the set of locations where a point source might lie after using the ITD binaural cue.

Head Motion for Localization

- Auditory parallax, nearby audio sources change their azimuth and elevation faster than distant ones
- Integrating different cone of confusion for every head pose
- Doppler effect caused by the motion of a source relative to the receiver

Further Reading

- Section 11.1, 11.2 and 11.3, Virtual Reality, Steven LaValle