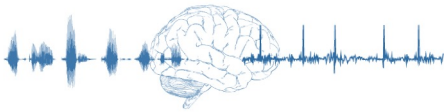


Primer on Auditory Processing

Mounya Elhilali
Department of Electrical & Computer Engineering
Johns Hopkins University
mounya@jhu.edu

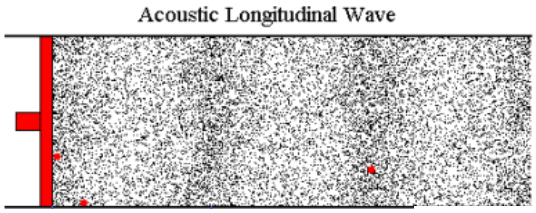
1

Speech as waves

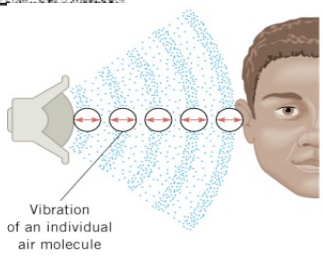
2

Sound is a wave

- Sound is a mechanical wave caused by a vibrating source
- The vibrating source that causes the matter around it to move



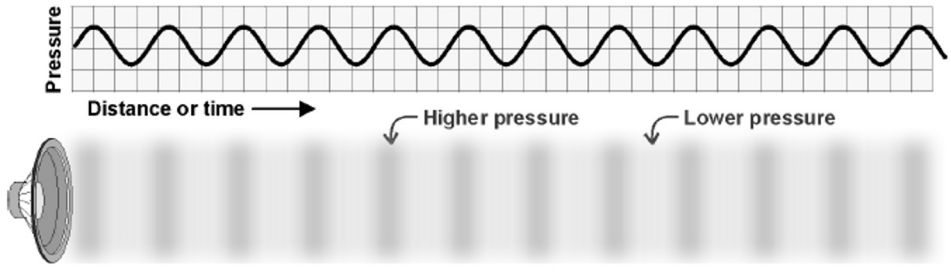
- No sound is produced in a vacuum
  - Matter (air, water, earth) must be present
- Individual air molecules do not move the wave. A given molecule vibrates forth about a fixed location.



3

Sound waves

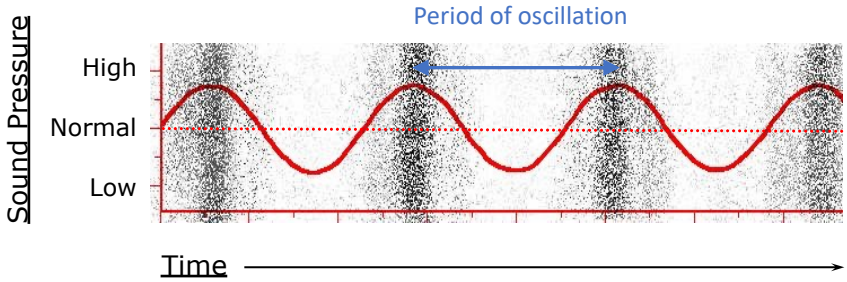
A sound wave is a wave of alternating high-pressure and low-pressure regions of air.



- Vibrating object compresses the air around it (high pressure)
- Pushes air away leaving an area of low pressure (low pressure)
- then compresses again creating a periodic pattern

4

Sound waves

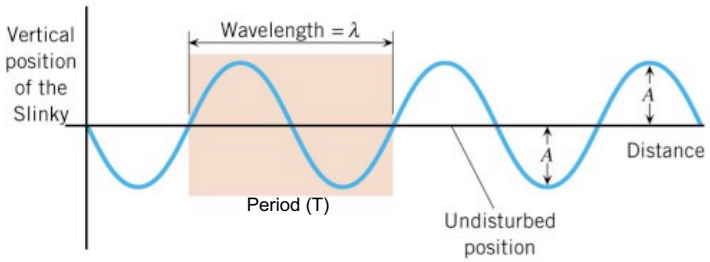


- Motion air particles do not travel, they oscillate around a point in space
- The rate of oscillation is called frequency (f)
  - ✓ denoted in cycles per second (cps) or hertz (Hz).

5

5

Physical Dimensions of Sound



<u>Amplitude</u>

- Height of a cycle

Related to measure/perception of loudness

<u>Frequency</u> (F)

- Cycles per second

Related to measure/perception of pitch/spectrum

<u>Wavelength</u> (λ)

- Distance traveled by one cycle

Affected by medium (how sound travels)

6

6

Amplitude ↔ Loudness

- Sound Pressure Level (SPL) is a relative measure of sound intensity

$$L = 10log_{10}\frac{I}{I_0}$$

Intensity of target sound (*rate of energy flow over an area in* W/m^2)

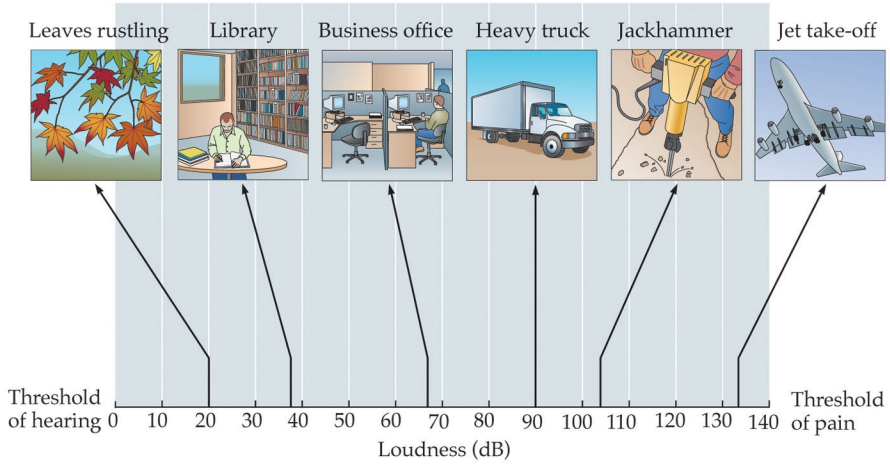
Reference Intensity:
Human absolute hearing threshold
$I_0 = 1 \times 10^{-12}\ (W/m^2)$
(related to $P_0 = 20\ \mu Pa$)

- SPL is technically a unitless measure; but uses unit of Decibels or dB or dB-SPL
- Decibels provide a **relative** measure of sound intensity.

7

7

Sound pressure level of everyday sounds

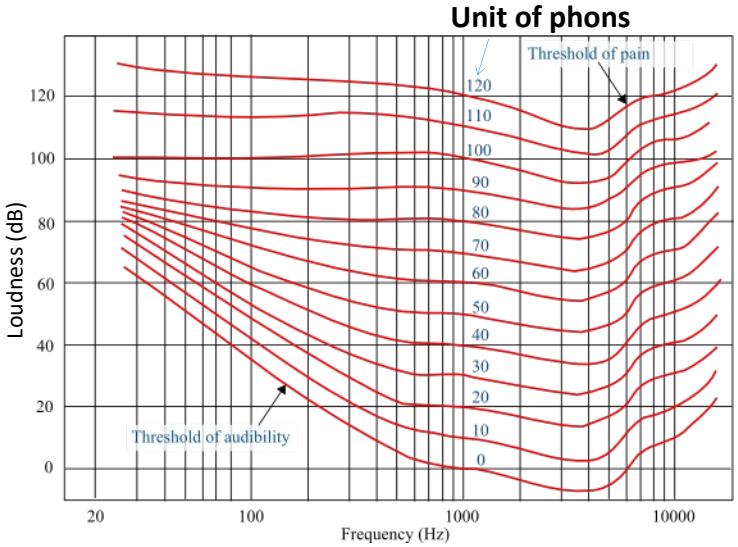


Note: Listening to loud music will gradually damage your hearing!

8

8

Physical sound pressure level vs. perceived loudness



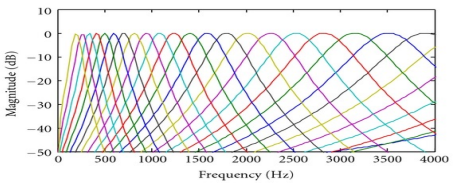
Each contour represents equally-perceived tones

9

9

Critical band

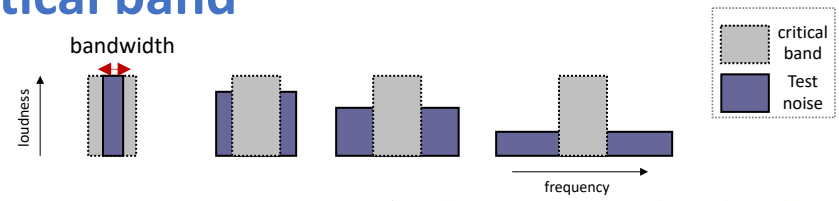
- A critical band is a frequency region over which energy is integrated
  - Idea is:
    - If two frequencies are close enough to each (within a critical band), they activate *same* region in the ear; so the sound is not perceived as loud.
    - If they are far apart (larger than a critical band), a new region in the ear is activated making the sound seem louder



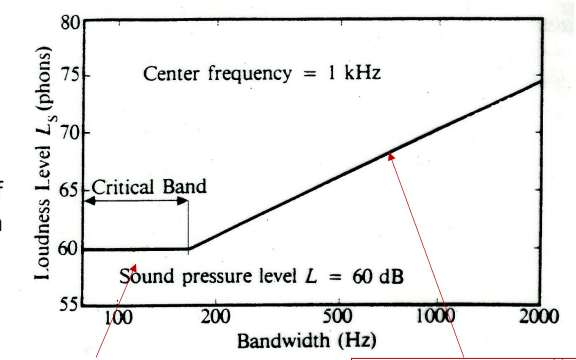
10

10

Critical band



- A reference noise band compared to test noise band with increasing bandwidth (constant power).
- When the bandwidth of the test noise exceeds a certain level (**critical band**), the loudness begins to increase.

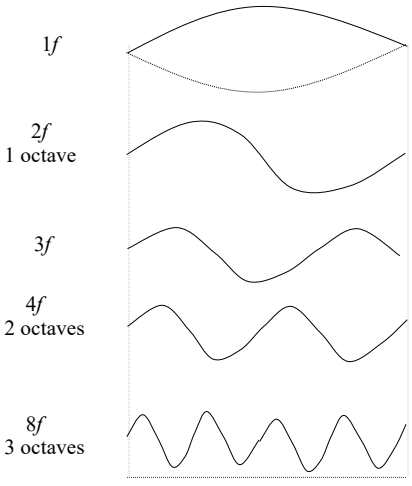


Loudness does not change

Loudness increases with BW

11

Frequency ↔ Pitch

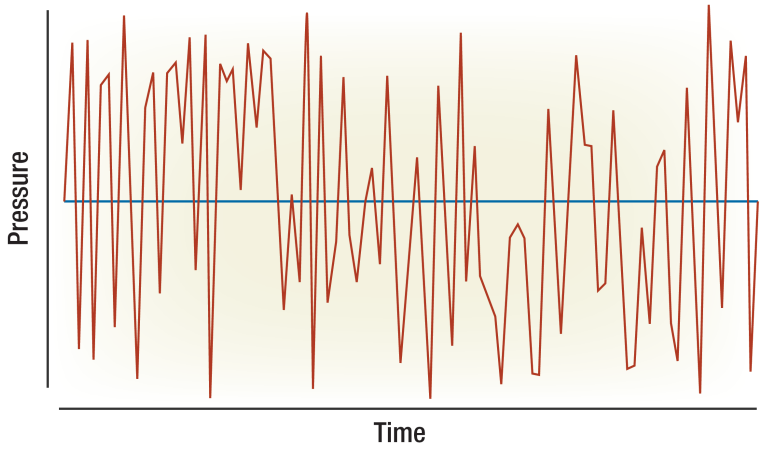


- At first approximation, the pitch of a simple periodic signal is determined by its frequency.
- Most oscillators (guitar string, vocal chords) naturally oscillate at a fundamental frequency (F_0) as well as its integer multiples (called harmonics/partials/overtones).
- The pitch of a *complex* period signal is often determined by its *fundamental* frequency (F_0)

12

12

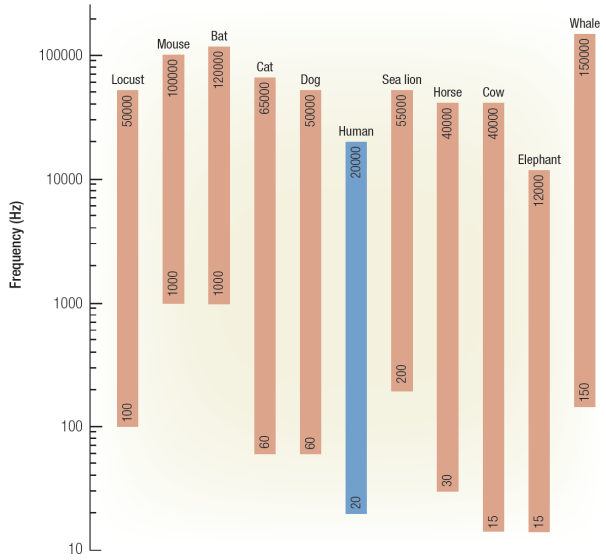
Most sounds are complex (not simple tones)



13

13

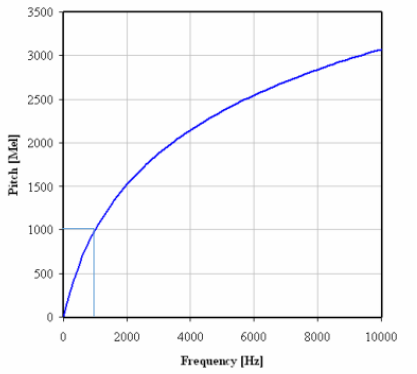
Species-Specific Frequency Range



14

a Pitch scale

- Perceptual scale of pitch: mel scale
  - *How far in frequency do we have to be in order to feel a tone as doubled in pitch?*



➡ It's a relative scale, based on pitch comparisons

$$m = 2595 log_{10}\left(1 + \frac{f}{700}\right)$$

✓ Mel-scaling is used in signal processing to build filters that approximate human pitch perception (MFCC)

15

15

Masking

- Hearing phenomenon
  - When the perception of one sound is affected by presence of *another* sound
  - one sound being *masked* by another
- Term masking is used to describe effects of noise and interference in sound perception
- We experience masking everyday

16

16

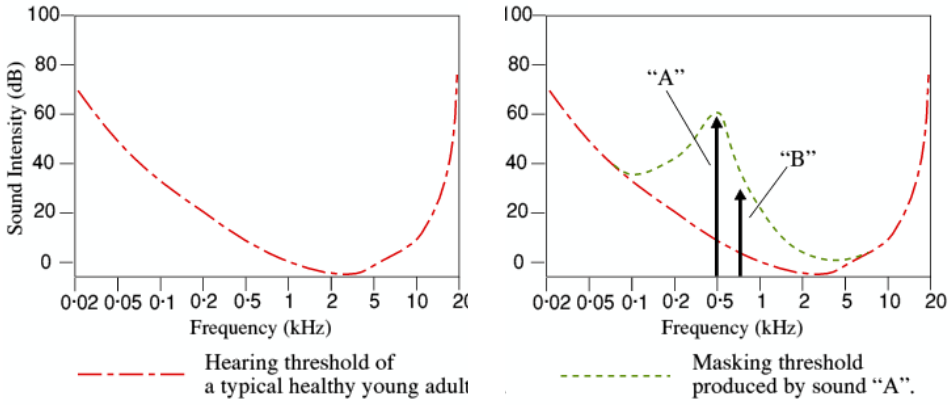
Masking

17

17

Frequency Masking

(simultaneous masking)



18

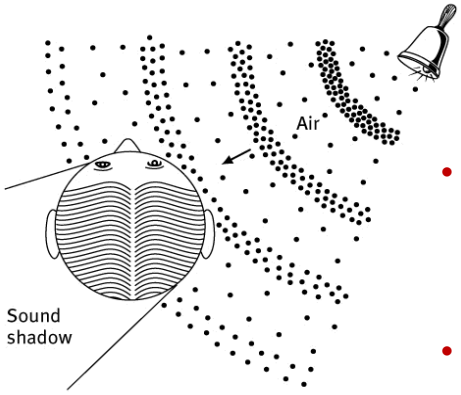
18

Sound localization

19

19

Sound Localization



- The direct path from the acoustic source to the two ears will generally be different.
- The signal needs to travel further to more distant ear
- More distant ear partially occluded by the head

20

20

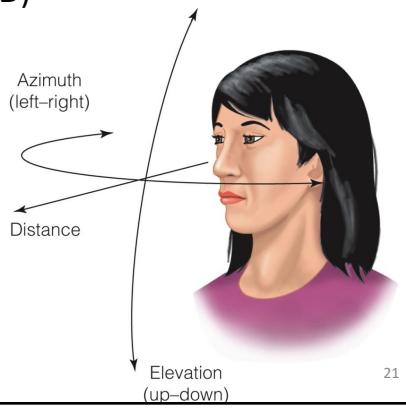
Auditory Localization cues

between 2 ears

1. inter-aural timing differences (**ITD**)
2. inter-aural level differences (**ILD**)
3. monaural cues (pinnae)

…

4. head movements



21

21

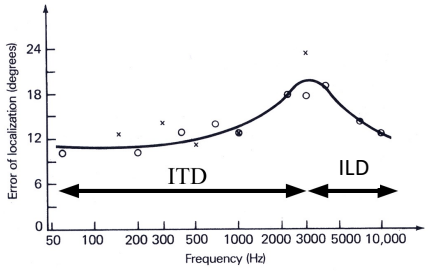
Interaural cues

Interaural level differences (ILDs)

- Threshold ILD ≈ 1 dB
  => Effective for high frequencies

Interaural time differences (ITDs)

- Threshold ITD ≈ 10-20 µs (*~ 0.7 cm)*
  ⇒ Effective for low frequencies



22

22

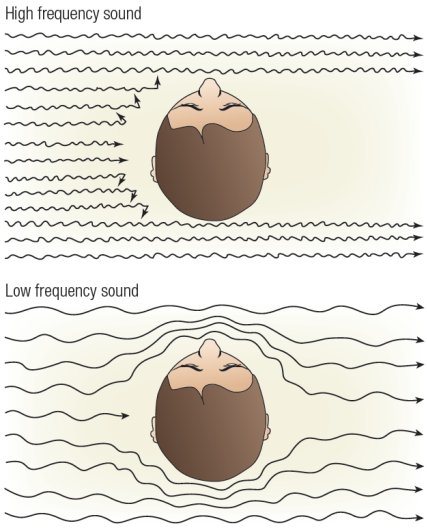
Interaural cues

- **ILD**
  - Head-size dependent: larger heads create bigger ILDs for the same frequency
  - Very-frequency dependent – larger effect at higher frequencies
- **ITD**
  - Head-size dependent: larger heads create bigger range of ITDs
  - Less-frequency dependent – works over large freq range
  - Requires extraordinarily exquisite temporal mechanisms (10 – 20 µs sensitivity)

23

23

Inter-aural cues – Head shadowing



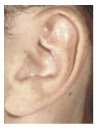
High-frequency sound waves are "blocked" by the human head and cast a "shadow" at the far ear **(Strong ILD cue)**

Low-frequency sound waves wrap easily around the head and cast little or no sound shadow **(Weak ILD Cue)**

24

Monaural cues

- Rely only on 1 ear -> monaural
  - Effective because of filtering properties of outer ear
  - Mostly pinnae



- Pinnae acts as directional filter
  - ✓ It amplifies sounds above and below differently
  - ✓ It acts mostly on high frequencies (above 5KHz)
  - ✓ Shoulder reflection causes changes in signal in 2-3KHz

➡ Monaural localization is not as accurate as binaural localization

25

25

Head-related transfer function

- A **Head-related transfer function** (**HRTF**) is a function that characterizes how a particular ear receives sound from a point in space
- It allows an audio system to simulate effects of sound in 3D space by mixing audio tracks with right filtering and delay parameters

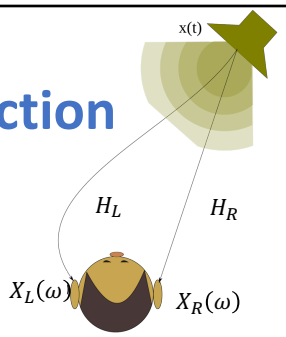
26

26

Head-related transfer function

- The Head Related Transfer Function :

$$X_L(\omega) = H_L(\omega,\theta,\phi)X(\omega)$$
$$X_R(\omega) = H_R(\omega,\theta,\phi)X(\omega)$$



- Where:
  - ω represents frequency and θ, ϕ represent elevation and azimuth respectively
  - $H_L(\omega,\theta,\phi)$ and $H_R(\omega,\theta,\phi)$ are the left and right HRTFs
  - $X_L(\omega)$ and $X_R(\omega)$ are the Fourier Transforms of the signals received by the Left and the Right ears
  - $\mathrm{X}(\omega)$ is the Fourier Transform of the source signal $\mathrm{x}(\mathrm{t})$

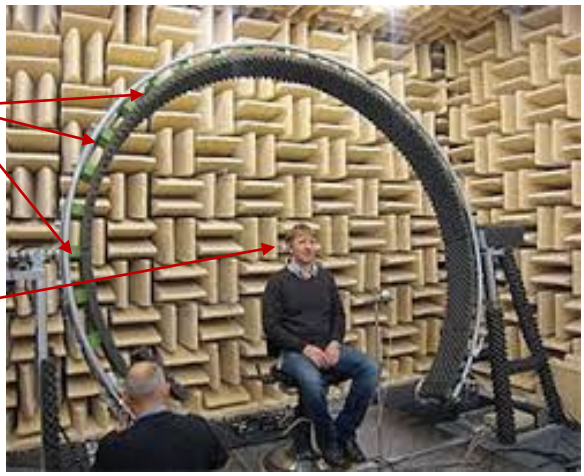
27

27

HRTF measurement

Vary speaker location

Record signal received



28

28

HRTFs

- HRTFs are widely used in gaming and virtual reality simulations to offer an immersive, realistic 3Dexperience without relying on visual cues

29

29

AV integration – McGurk effect

- "Compromise" between conflicting sound and visual cues in speech understanding
  - Compare auditory perception with eyes open vs. eyes closed

30

30

How do we perceive sounds?

31

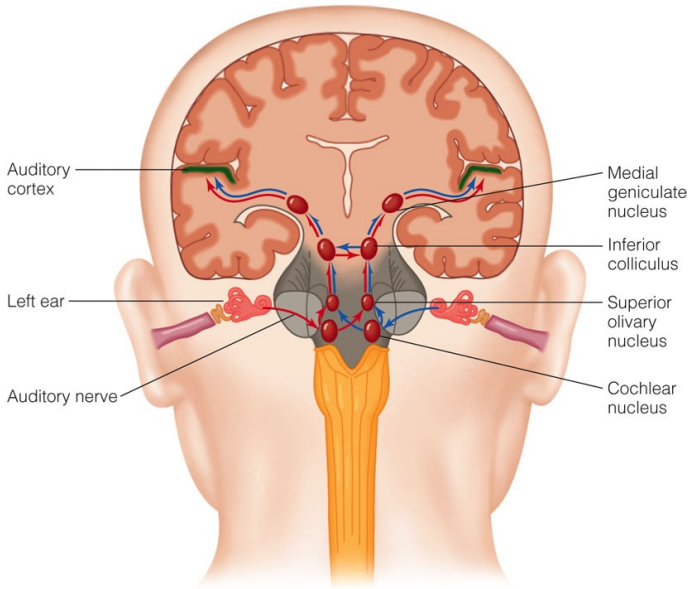
The auditory system

- Two major components in the auditory system
  - The peripheral auditory organs (the ear)
    - Converts sounds pressure into mechanical vibration patterns, which then are transformed into neural firings
  - The auditory nervous system (the brain)
    - Extracts perceptual information in various stages

32

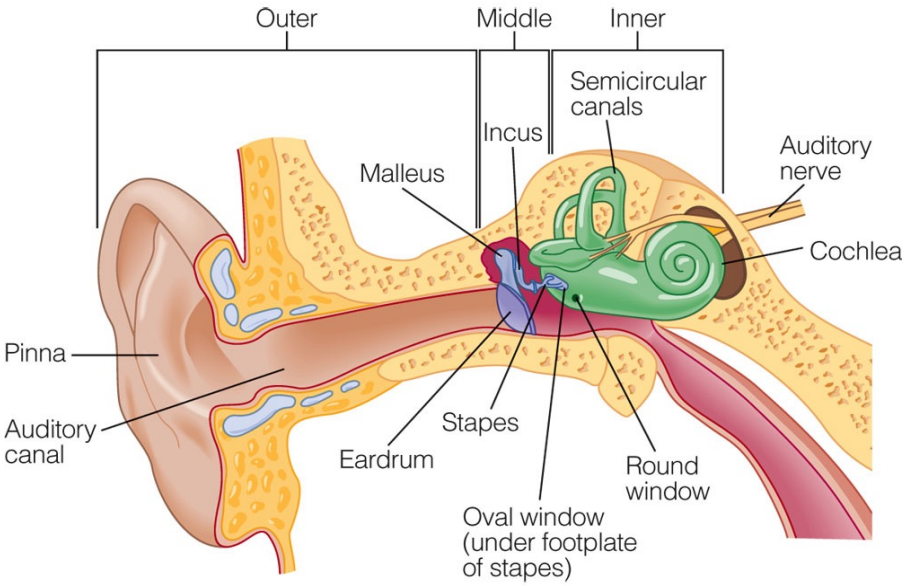
32

Auditory Pathway



33

33

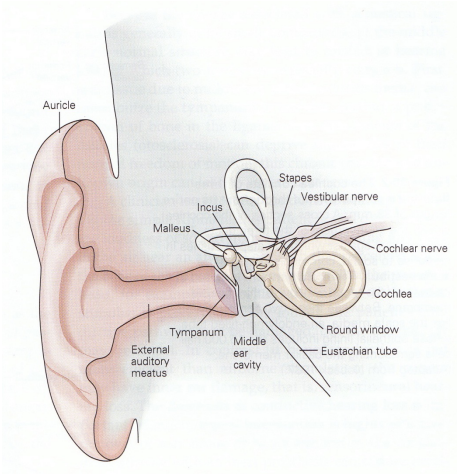


34

34

The ear

- The ear is the organ of hearing
- It changes sound pressure waves from the outside world into a signal of nerve impulses sent to the brain.
- It consists of 3 components:
  - Outer ear
  - Middle ear
  - Inner ear

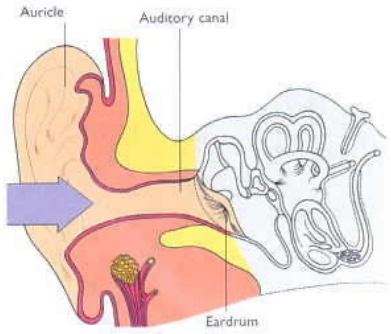


35

35

Organ of hearing outer ear

- The external ear plays the role of an acoustic antenna,
- It diffracts and focuses sound waves (pinna), while the ear canal acts as a resonator => amplifies sounds in 2-5 kHz range
- The end of the canal has an eardrum which vibrates with sound

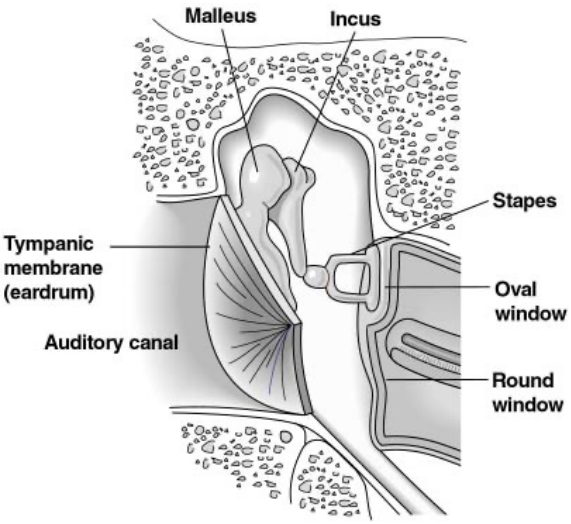


36

36

Organ of hearing middle ear

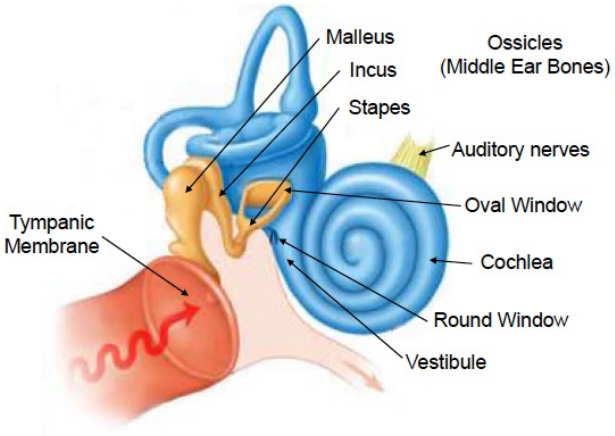
- Eardrum (or tympanic membrane) vibrations cause mechanical motion of the small bones of the middle ear (malleus, incus & stapes) [3 smallest bones in the human body]
- The middle ear acts as an impedance adapter to adjust energy difference between air environment and fluid environment



37

Organ of hearing inner ear

- Cochlea translates physical vibrations into electrical signals for the brain to process
- Cochlea acts a frequency analyzer of sound signals

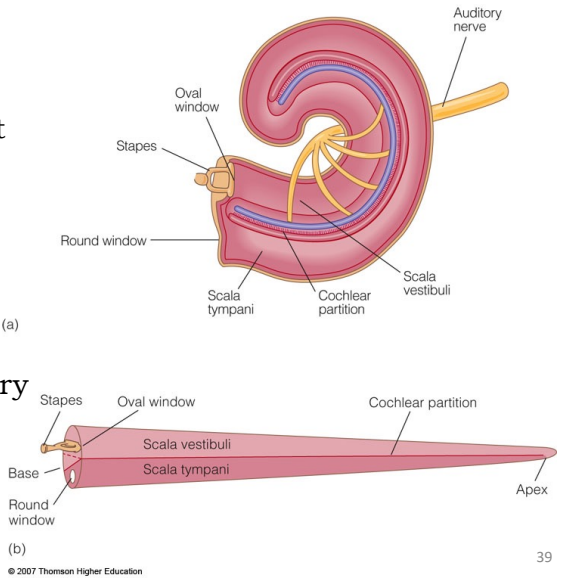


38

38

The Cochlea

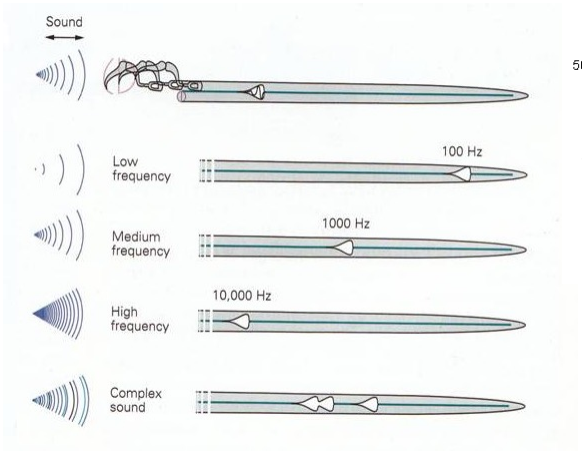
- The cochlea is the inner ear organ that converts sound waves into neural signals.
- The neural signals are passed to the brain via the auditory nerve.

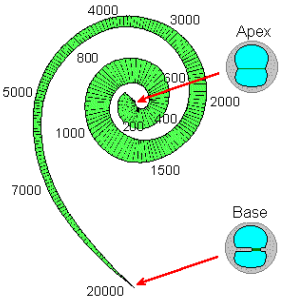


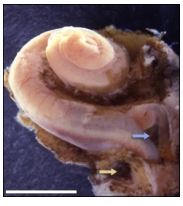
39

39

Cochlea as frequency analyzer



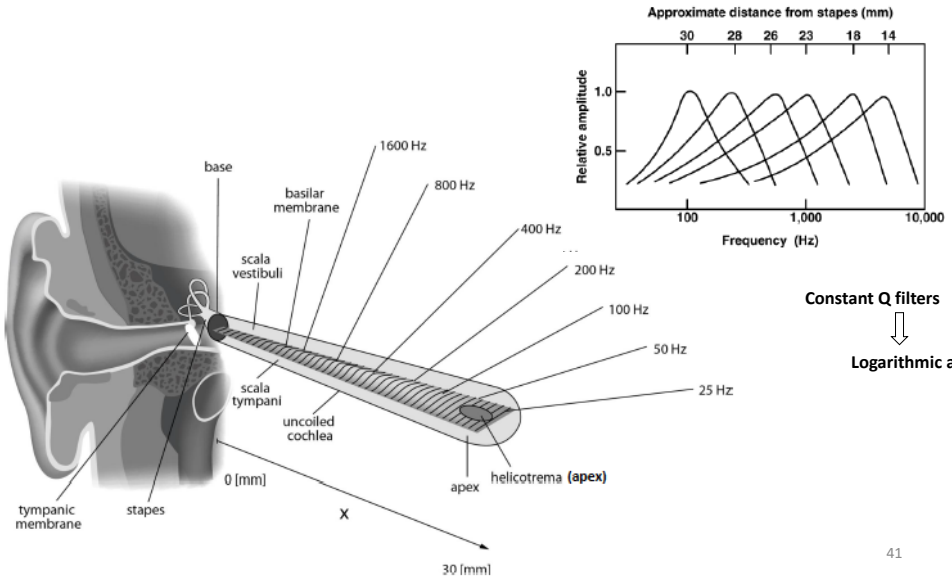




40

40

Modeling the cochlea ~ Bank of filters



Constant Q filters

⇩

Logarithmic axis

41

41

How the ear works (review)

42

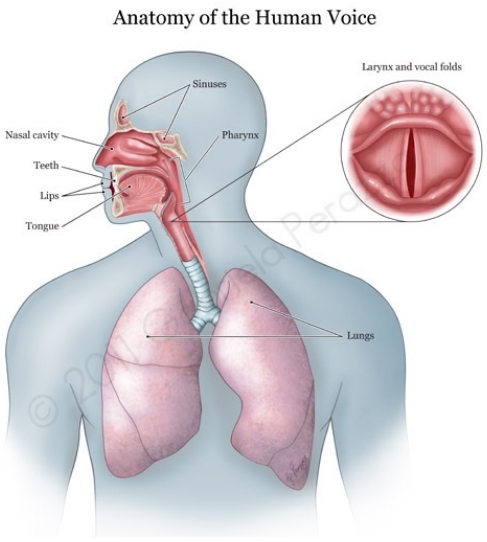
42

How is speech produced?

43

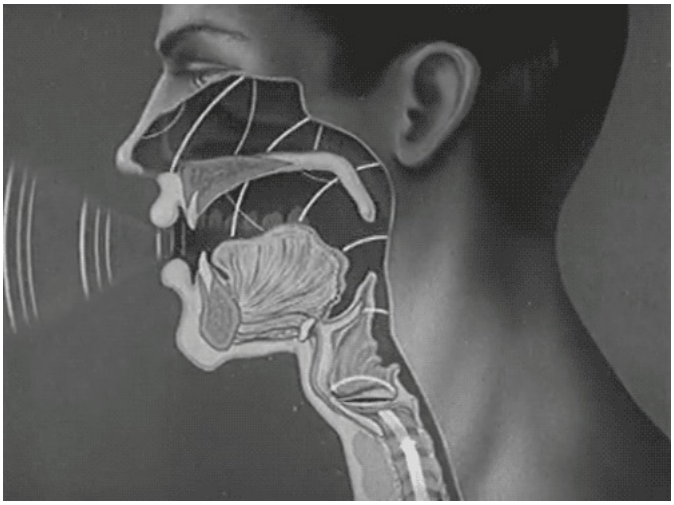
Anatomy & Physiology of speech organs

- Speech production apparatus starts from the lungs to the lips & nose.



44

Anatomy & Physiology of speech organs

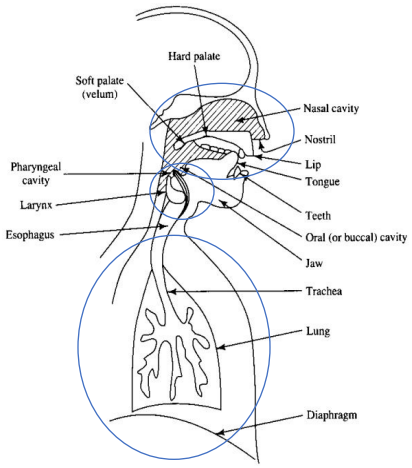


45

45

Speech Organs

- Speech organs:
  1. Lungs
  2. Larynx (vocal cords)
  3. Vocal tract

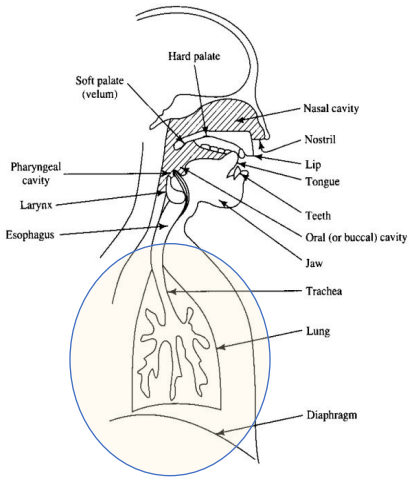


46

46

Speech Organs: Lungs

- Lungs play respiratory role: inhaling and exhaling air
- Inhalation results in expanding chest cavity to fill it with oxygen (lowering diaphragm which separates chest cavity from abdomen).
  - Lowers air pressure in the lungs
  - Causes air to rush through the vocal tract into lungs

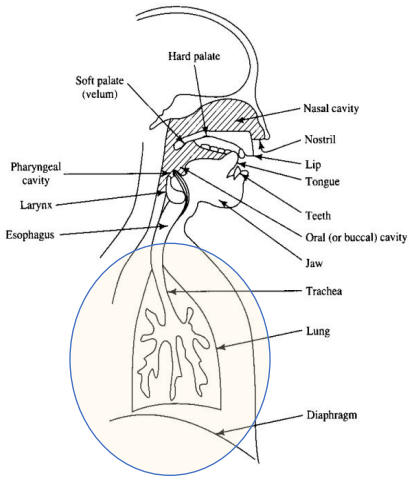


47

47

Speech Organs: Lungs

- Exhalation results in reducing volume of chest (contracting muscles of rib cage)
  - increases air pressure in the lungs
  - Causes air (carbon dioxide) to rush outside the lungs
  - Exhaling normally takes 60% of breathing cycle.

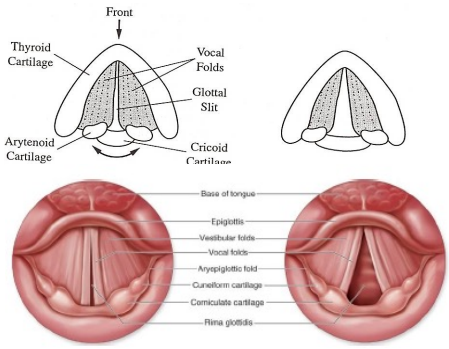


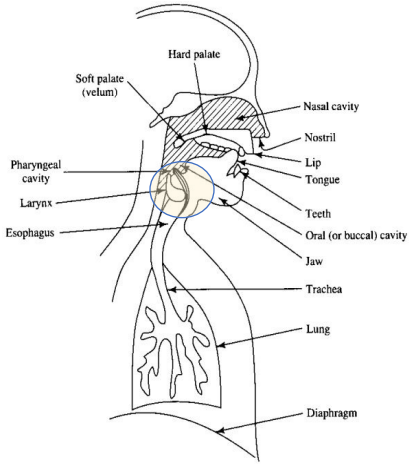
48

48

Speech Organs: Larynx

- Larynx: system of cartilages, muscles and ligaments
- Primary role is to control vocal cords (or vocal folds)

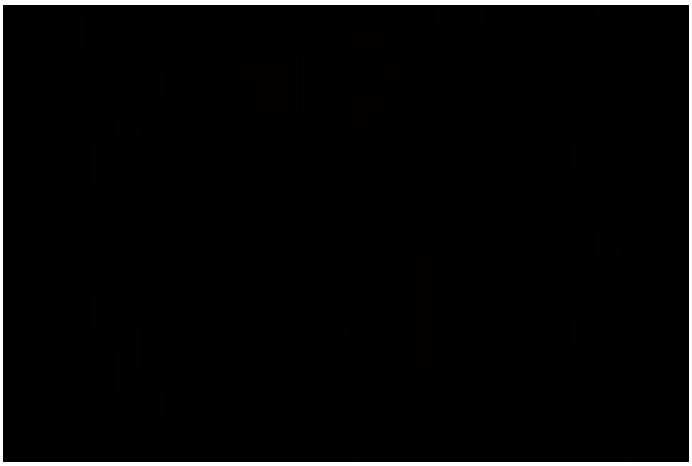




49

49

Vocal fold animation



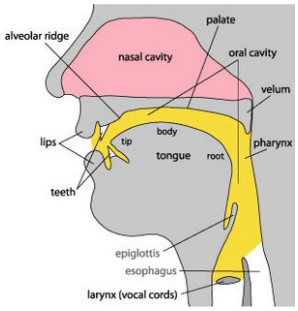
50

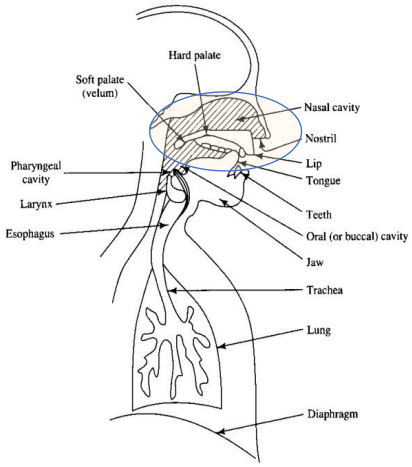
Vocal fold video

51

Speech Organs: Vocal tract

- Vocal tract is comprised of:
  - Oral cavity (from larynx to the lips)
  - Nasal cavity (coupled with oral tract through velum)





52

52