The Perception of Sound

EECS 4462 - Digital Audio

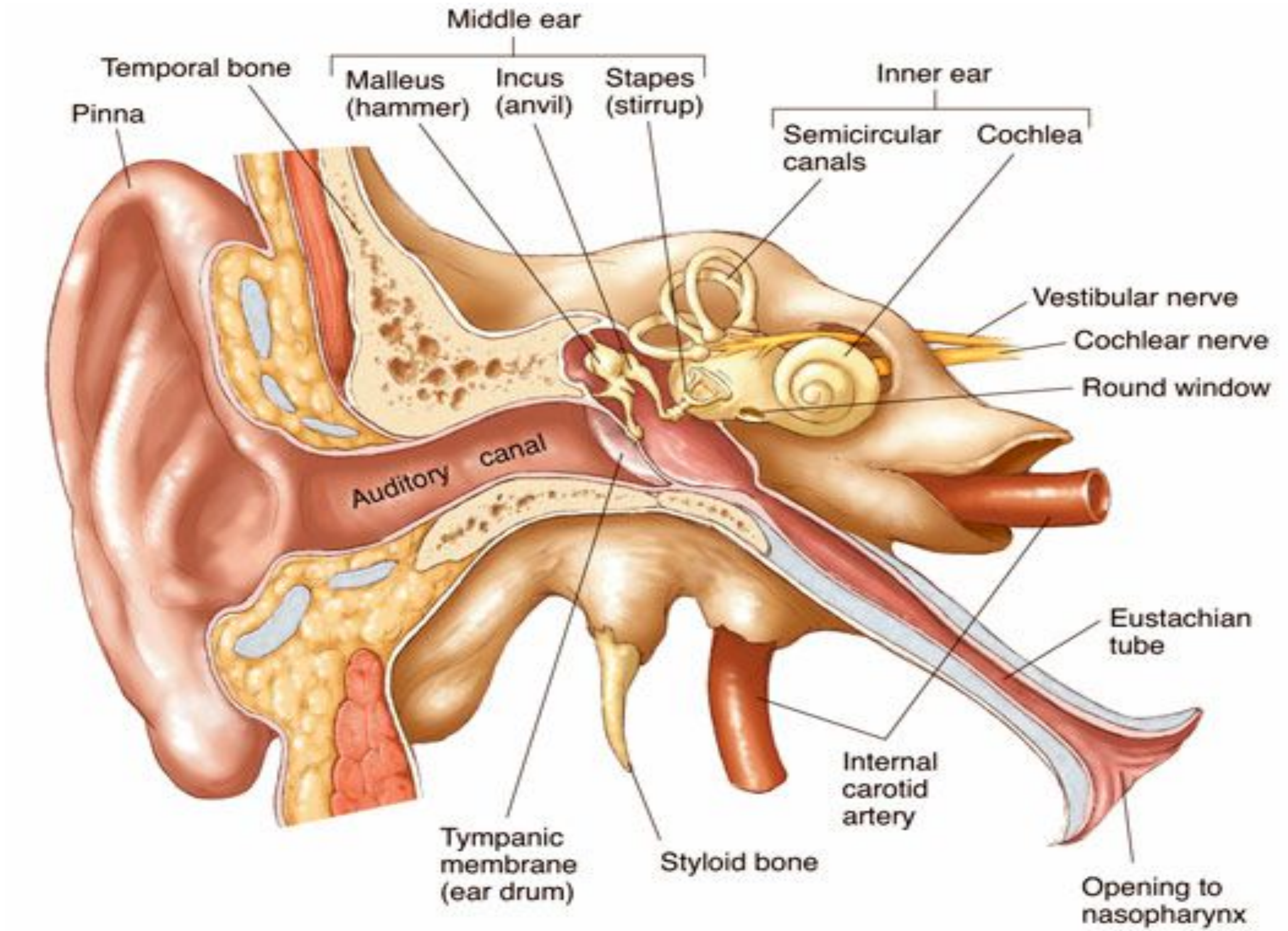
September 17, 2018

YORK UNIVERSITÉ UNIVERSITY

Let’s start with a video

- Auditory Transduction

The human ear



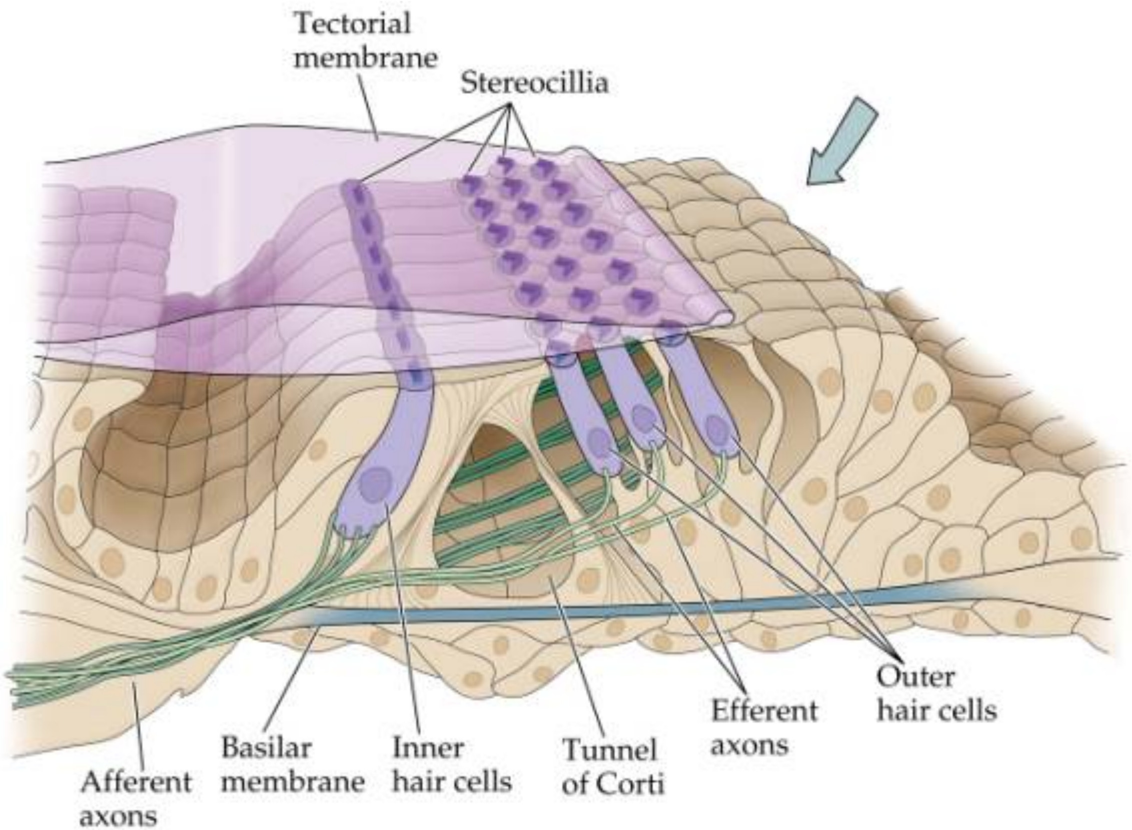
https://biologyboom.com/describe-the-structure-of-human-ear-give-a-summary-of-process-of-hearing-and-equilibrium-in-man/

Parts of the ear

- The Pinna: Gathers incoming sound and focuses it to the auditory canal
- The Auditory Canal: Transmits the sound to the eardrum
- The Middle Ear: Takes sound waves traveling through air and matches the impedance of the cochlear fluid which is 3000 times greater than that of air.
- The Inner Ear: Contains the cochlea and the Organ of Corti
  - Different areas of the cochlea resonate at different frequencies
  - Hair cells in the Organ of Corti respond to these vibrations sending electrical signals to the brain

The Organ of Corti

- The tectorial membrane vibrates the fluid surrounding the hair cells.



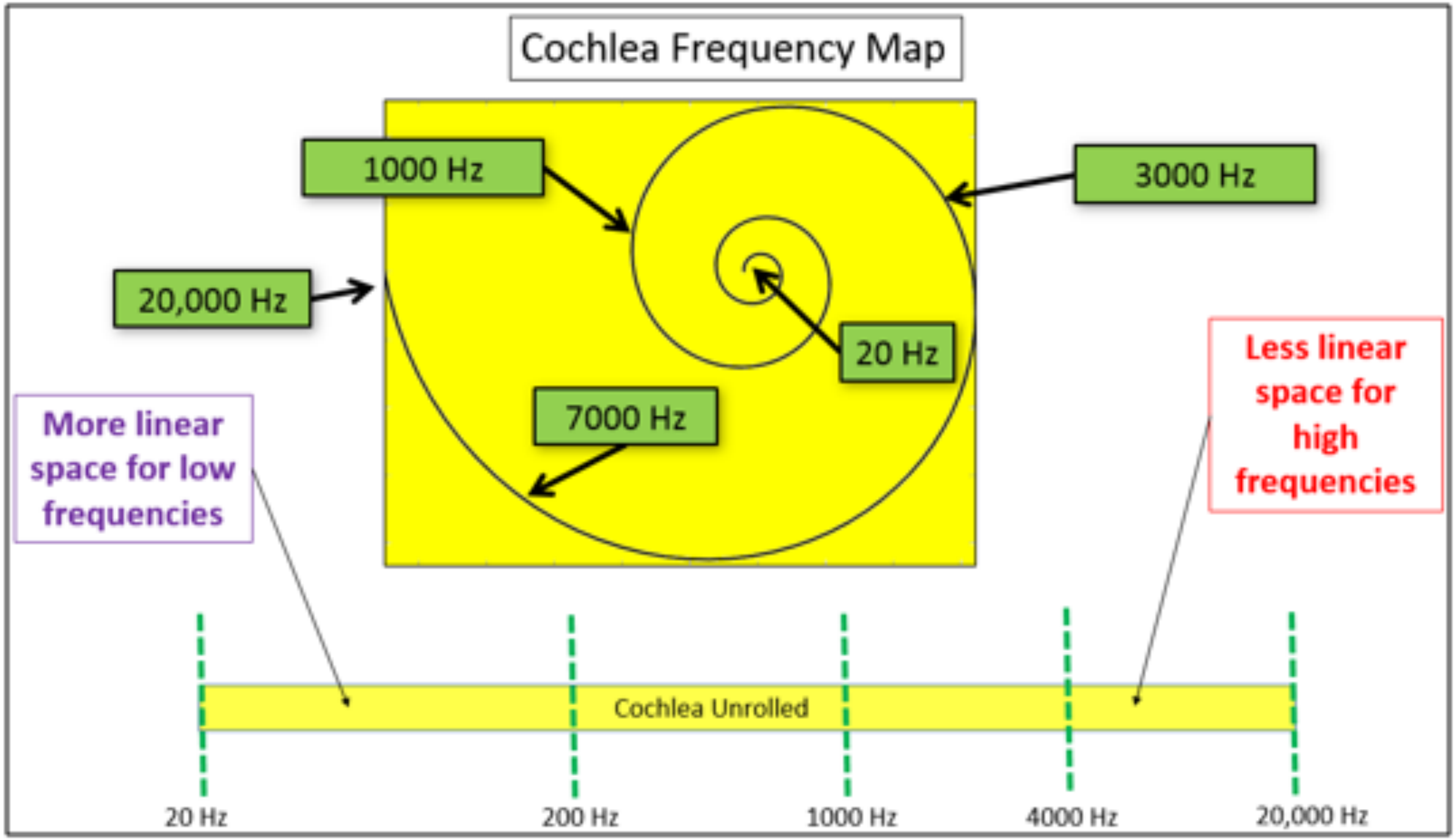
The place on the basilar membrane where maximum excitation of the hair cells takes place determines the perception of pitch.

http://flipper.diff.org/apprulesitems/items/info/6238

Critical bandwidths of the ear

Critical Band (Bark)	Center Frequency (Hz)	Bandwidth (Hz)
1	50	100
2	150	100
3	250	100
4	350	100
5	450	110
6	570	120
7	700	140
8	840	150
9	1000	160
10	1170	190
11	1370	210
12	1600	240
13	1850	280
14	2150	320
15	2500	380
16	2900	450
17	3400	550
18	4000	700
19	4800	900
20	5800	1100
21	7000	1300
22	8500	1800
23	10500	2500
24	13500	3500

Critical bandwidths of the ear



https://community.plm.automation.siemens.com/t5/Testing-Knowledge-Base/Critical-Bands-in-Human-Hearing/ta-p/416798

Critical bandwidths of the ear

- Sounds occupying a critical bandwidth will drown out any other sound that is quieter in the same bandwidth
- The ear can distinguish frequencies better at the low end
- However, musical notes are very close together at the low end
  - C_1 = 33Hz E_1 = 41Hz G_1 = 49 Hz
- For this reason, it is not pleasing to have several bass notes played together
- Piano demo!

YORK U UNIVERSITÉ UNIVERSITY

Measuring loudness

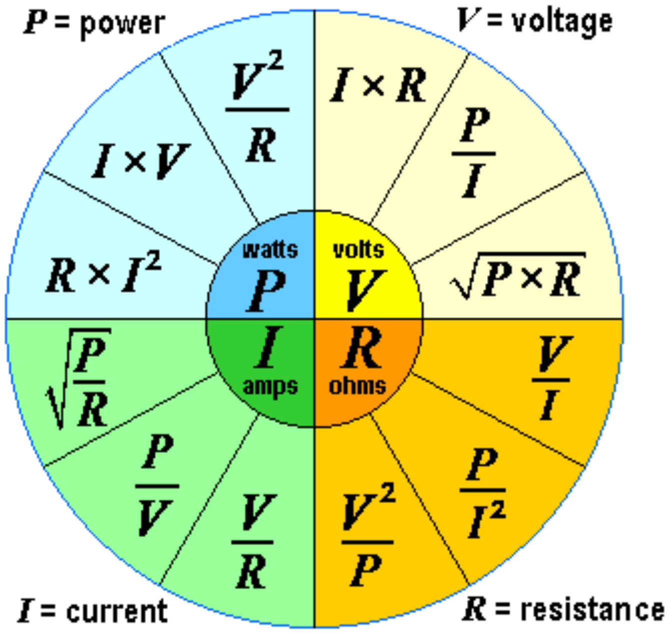
- Loudness relates to perception, so it can be quite subjective
- We can objectively measure air pressure p
- Sound intensity I is proportionally related to the square of sound pressure

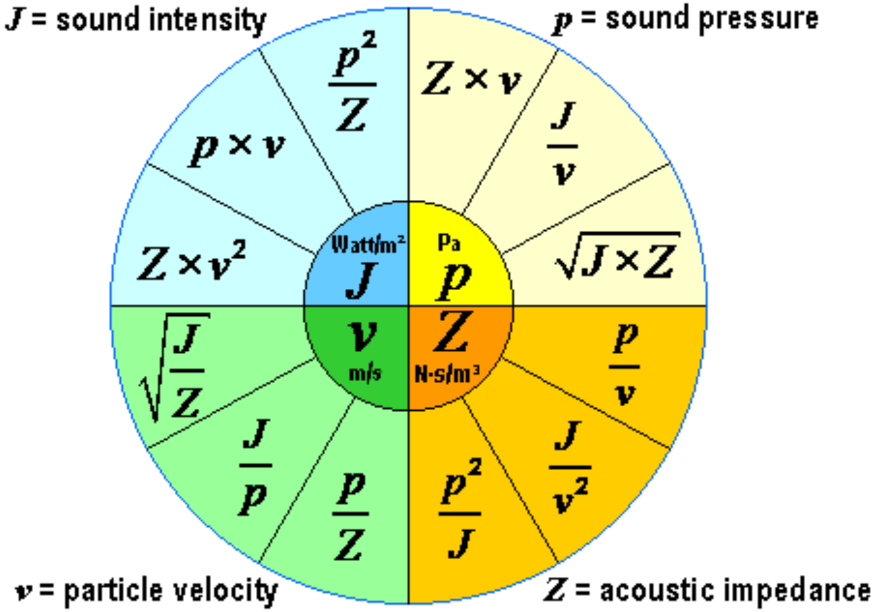
$$I = p^2 / Z_0$$

- Z_0 is the acoustic impedance of the air

YORK UNIVERSITÉ UNIVERSITY

Analogy to electrical circuits





http://www.sengpielaudio.com/FormulaWheel-Acoustics.htm

Measuring loudness

- We express the magnitude of acoustic signals as a ratio to a fixed reference point
- These ratios can get very large, so we use the logarithm of the ratio
- Our ears also perceive loudness in a fashion that is roughly logarithmic related to pressure, rather than linear

dB

- A sound wave with an intensity of I_1 is said to be $\log(I_1/I_0)$ Bels loud
- I_0 is some reference point
- Instead of Bels, we usually use deciBels (dB)
- 1 Bel = 10 deciBels
- A dB is a unitless measure
- When combined with a reference point, it can be used as a pseudo unit

dB SPL

- The quietest sound that can be perceived by a human has a sound pressure of $p_0 = 20\mu Pa$
- This is the “threshold of hearing”
- We measure the magnitude of acoustic signals with respect to this reference point
- An acoustic signal with double the sound pressure of the threshold of hearing is $10\log 2^2 \cong 6dB$ louder
- This gives us the dB SPL scale
  - SPL = sound pressure level

dB SPL (examples)

Source of sound	Distance	dB SPL
Jet engine	1m	150
Threshold of pain	At ear	130-140
Jack hammer	1m	100
Car	10m	60-80
Normal Conversation	1m	40-60
Light leaf rustling	Ambient	10
Threshold of hearing	At ear	0

https://en.wikipedia.org/wiki/Sound_pressure

YORK UNIVERSITÉ UNIVERSITY

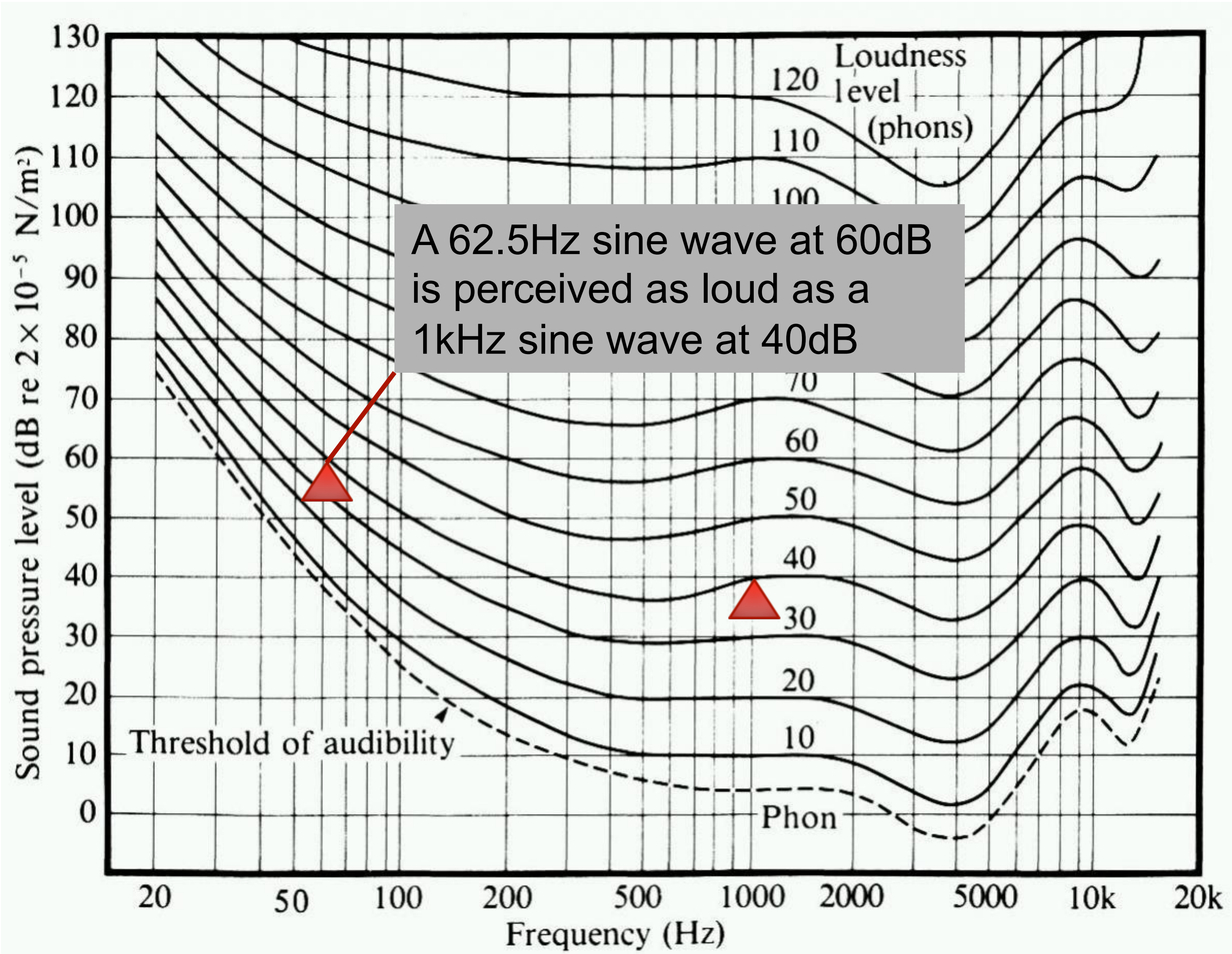
dB FS

- In digital audio, the loudest level possible (to avoid issues similar to overflow) is used as the reference point
- As a result, the loudest signal in digital audio is 0 dB FS
  - FS = Full Scale
- A signal that reaches 50% of the maximum level has a level of −6 dB FS

Sensitivity of the human ear

- The human ear can distinguish a difference when sound intensity changes by as little as 1dB
- This is called the Just Noticeable Difference (JND)
- An increase of about 10dB is perceived as twice as loud
- Also, our sensitivity varies a lot based on frequency
- We need much higher sound pressure to perceive low frequencies
  - Low frequencies diffract over the pinna due to their large wavelength

Equal Loudness Curves



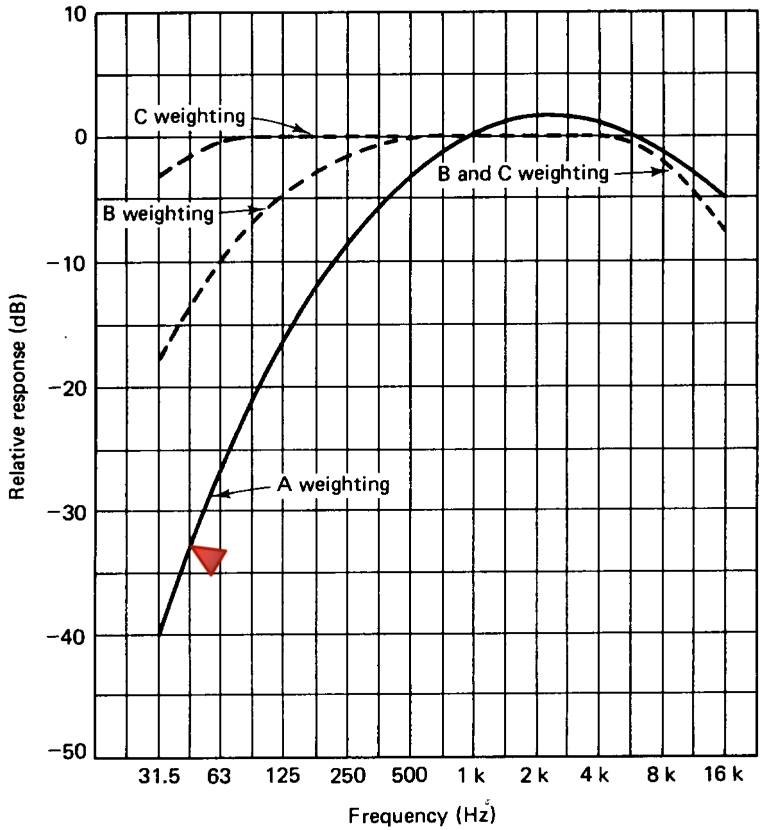
https://www.researchgate.net/figure/Equal-loudness-curves-after-DW-Robinson-and-RS-Dadson-1956_fig2_228682953

The Phon scale of loudness

- A sound has a loudness of **X** phons if it is equally as loud as a sine wave of **X** dB SPL at 1KHz
- A 62.5 Hz sine wave at 60dB has a loudness of 40 phons
- When measuring loudness, different *weighting networks* can be used to compensate for this behaviour

Weighting network	dB SPL
A	20-55
B	55-85
C	85-140

Weighting network A



When measuring a signal at 50 Hz, subtract 33 dB from the measurement to account for the fact that humans cannot perceive low frequencies well

If the original measurement was 70dB, the adjusted measurement will be 37 dBA

http://personal.cityu.edu.hk/~bsapplec/single.htm

Binaural localization

- Because we have two ears, we can localize sound
- Our brain uses three different cues to do this:

1. Interaural intensity difference.
2. Interaural arrival time difference.
3. Effects of the pinnae.

Interaural intensity difference

- The ear that is in direct earshot of the sound will hear the sound louder than the ear that is in the acoustic shadow of the head.
- Works mostly with high frequencies

Interaural arrival time difference

- Sound arrives at the ear in line of sight first and bends around the head to enter the other ear a fraction of a second later
- Our brain is able to pick up this difference to determine the direction the sound is coming from
- The difference is at the level of 10-20 μsec!

Effects of the pinnae

- The pinnae help with localizing sound in terms of front vs back
- High frequencies, due to their small wavelength, cannot diffract around the pinnae as well as low frequencies
- Can also work for localizing sound in terms of up vs down, but the effect is not as strong

Our brain and reflections

- When reflections are present, our brain processes differently based on how far apart they are
- Fusion zone: When reflections are early enough (0 – 40ms), the brain fuses the original sound and the reflections.
- Echo zone: More than 40ms.
- Our brain uses information from reflections to determine
  1. The position of the sound source
  2. The size of the room we're in
  3. The material of the walls (based on the amount of reverberation present)

Threshold Shift

- Temporary Threshold Shift - Listening to music in the car. Threshold of hearing shifts to protect ears.
- Permanent Threshold Shift - Comes with age, need hearing aid.
- Ringing in ears after loud sounds is because hair cells are in a state of shock, are still moving
- Tinnitus: Hair cells are broken off or stuck in the ON position.

Auditory Scene Analysis

- Our brain can perform amazing analysis of complex audio scenes
- We can focus on a particular speaker in a noisy environment
- We can focus on one instrument in a 100-person orchestra
- Computers are not as great at this kind of source separation
- Active area of research involving deep learning

YORK UNIVERSITÉ UNIVERSITY

Auditory illusions

- https://www.youtube.com/watch?v=yKR2pGwavqE
- https://www.youtube.com/watch?v=0amvhGzeCnQ