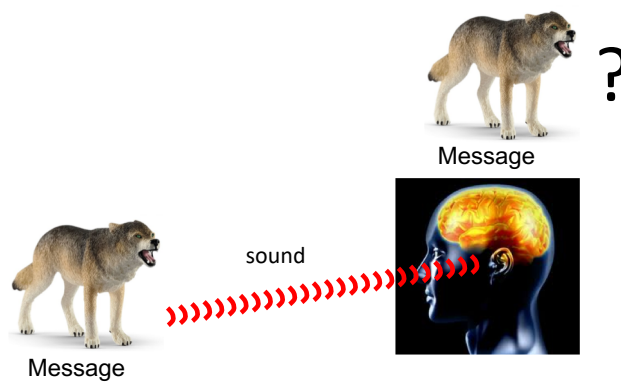
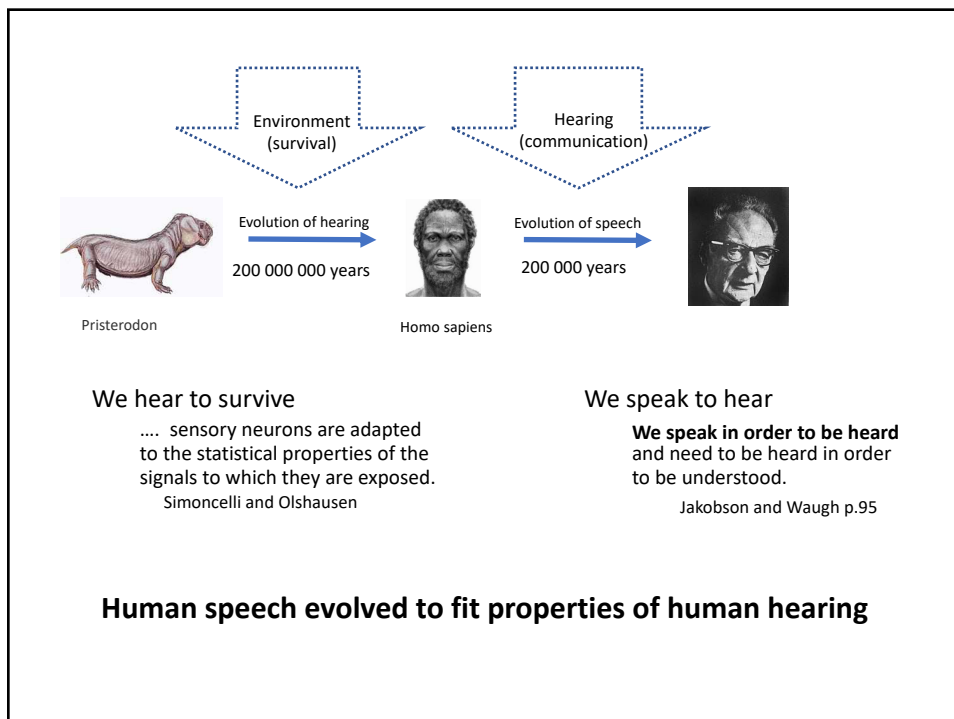
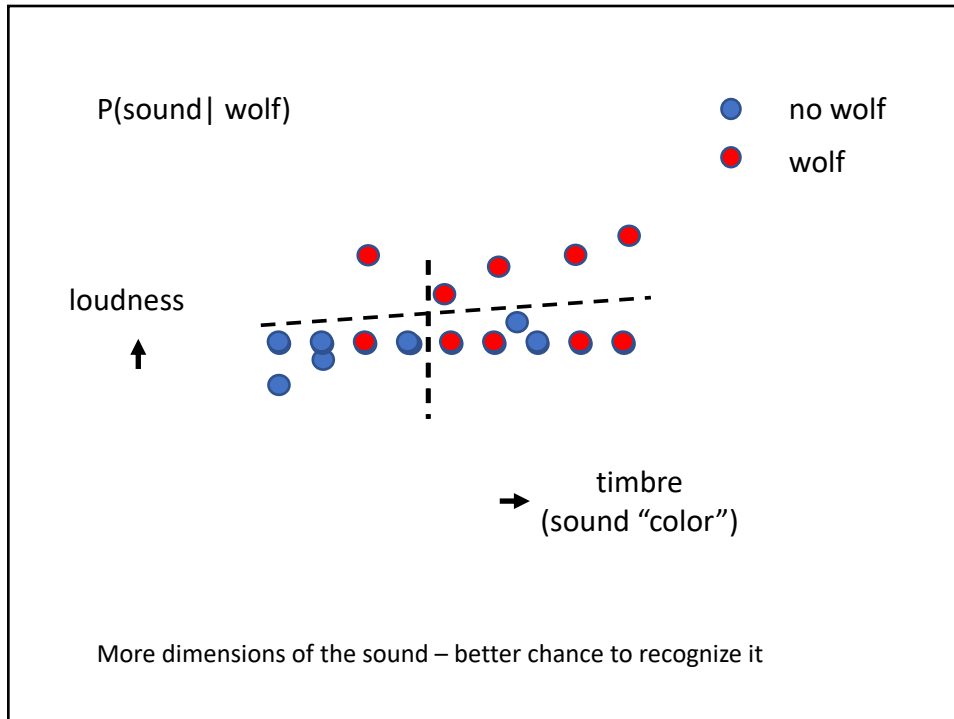


Speech

Hynek Hermansky
Electrical and Computer Engineering
Hackerman 324Fp

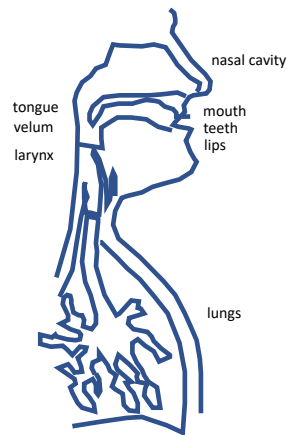


$$P(\text{wolf}|\text{sound}) \approx P(\text{sound}|\text{wolf}) \times P(\text{wolf})$$



Human vocal tract

means for generation many different sounds (many dimensions)



breathing
eating
biting

speaking?

Message

Message

word

$$P(\text{wolf} | \text{word}) \approx P(\text{word} | \text{wolf}) \times P(\text{wolf})$$

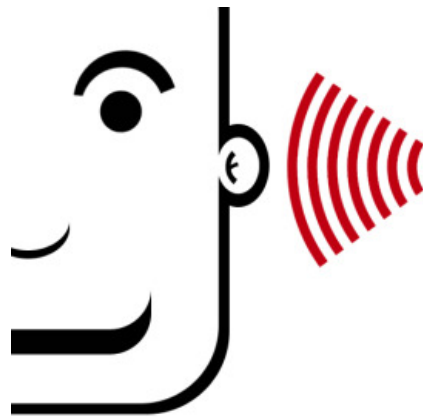
When more than one signal (e.g., audio and visual)

$$P(o | x_1, x_2) = \frac{P(x_1 | o) P(x_2 | o) P(o)}{P(x_1) P(x_2)}$$

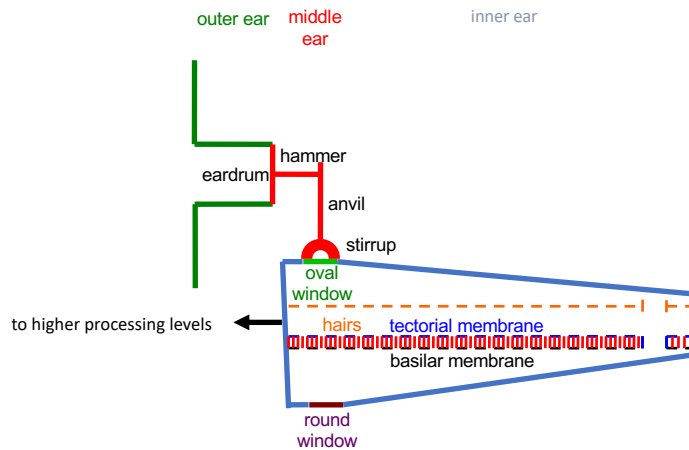
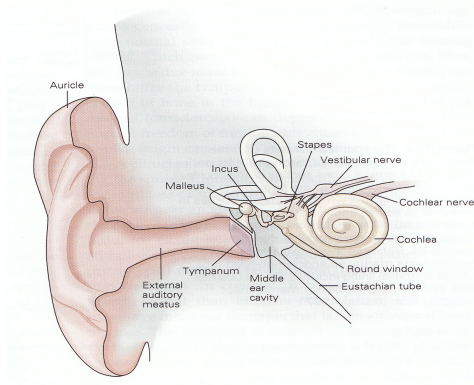
McGurk effect
acoustic /ba/ and visual /ga/ yields /da/ or /tha/

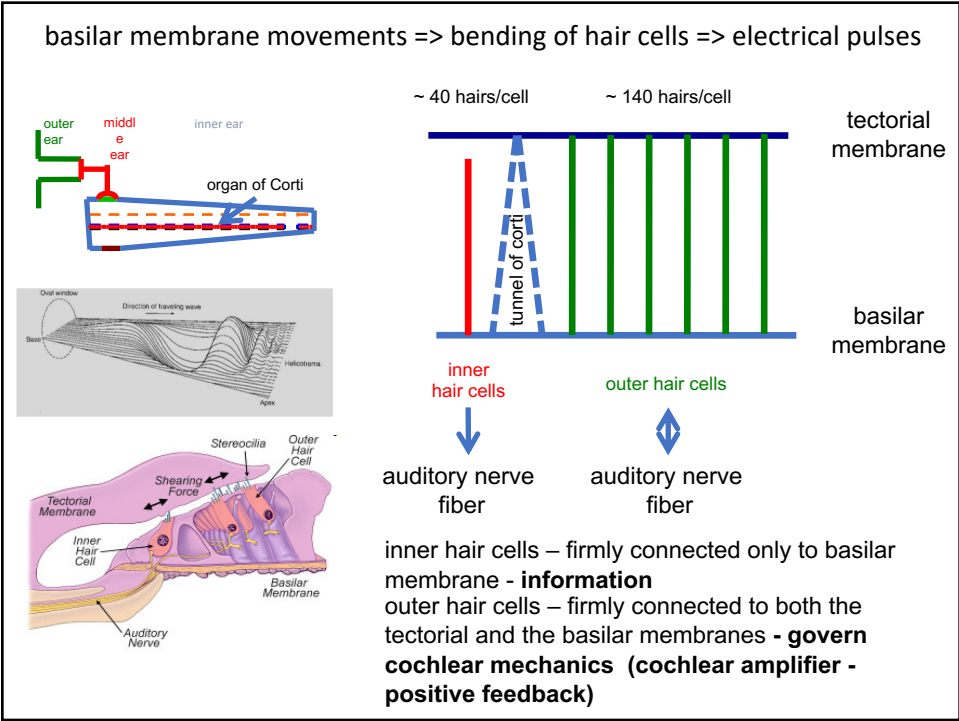
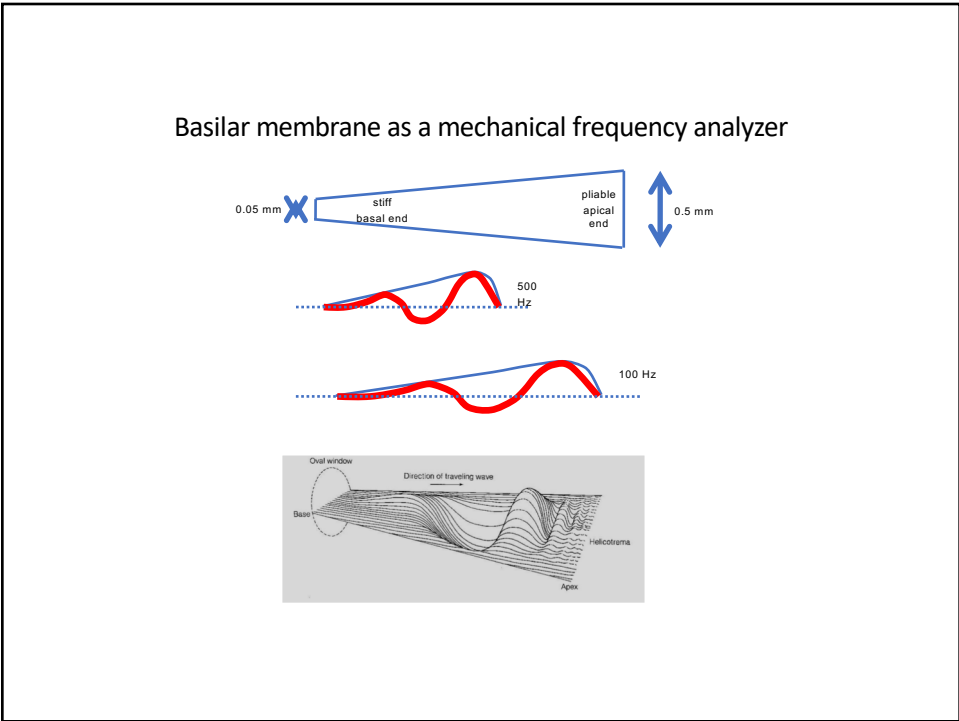


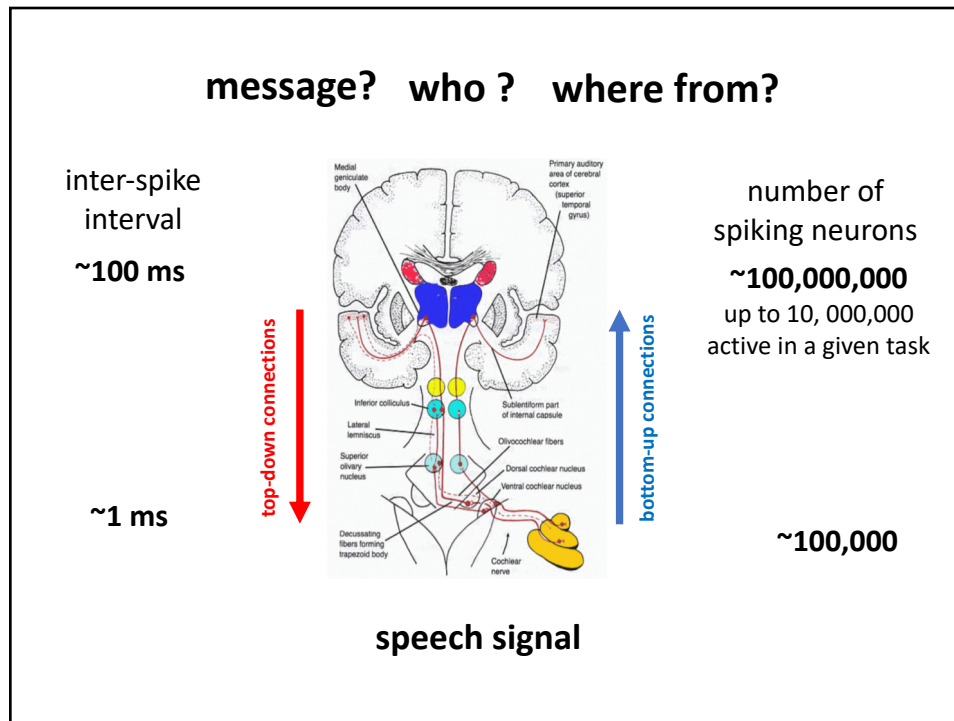
HEARING



Physiology of Hearing





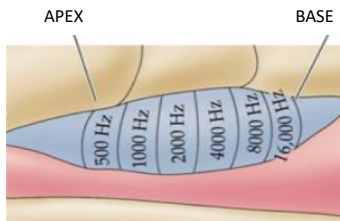


- massive increase in number of neurons from lower processing levels to cortex
- decrease in average spiking rates from periphery to cortex
- spikes in cortex are sparse (< 5% of cortical neurons active at any moment)

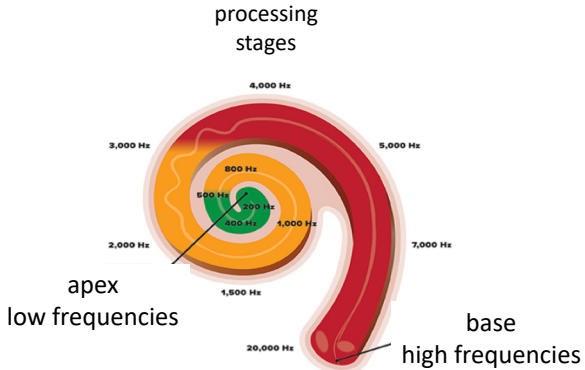
Hromadka et al PLOS Biology 2008

TONOTOPY

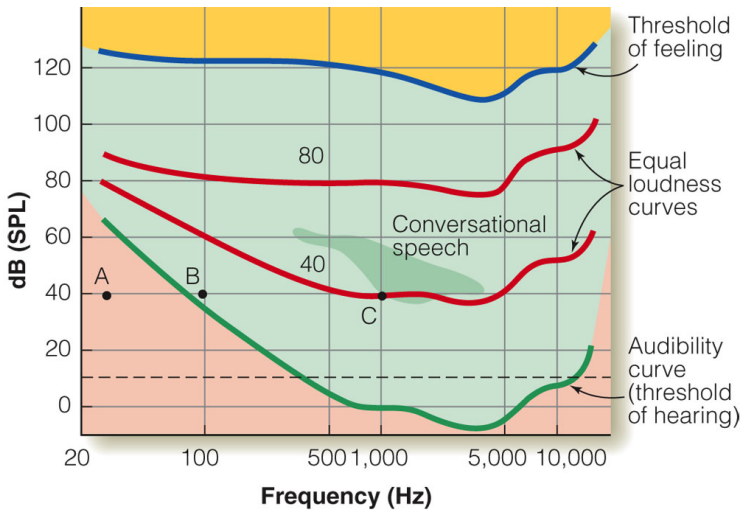
different frequencies excite different parts of the cortex



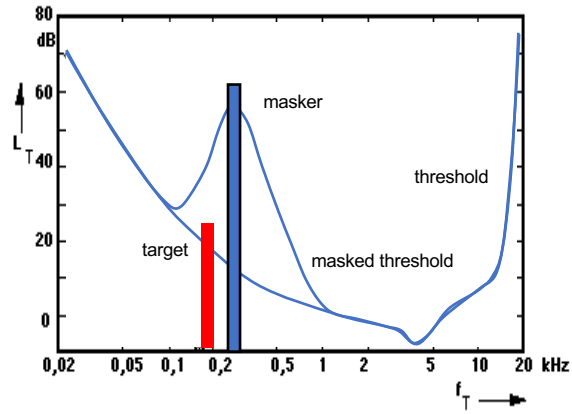
different frequencies excite different parts of the cochlea



Sensitivity of hearing



Simultaneous masking



Frequency selectivity of hearing (Critical bands of hearing)

