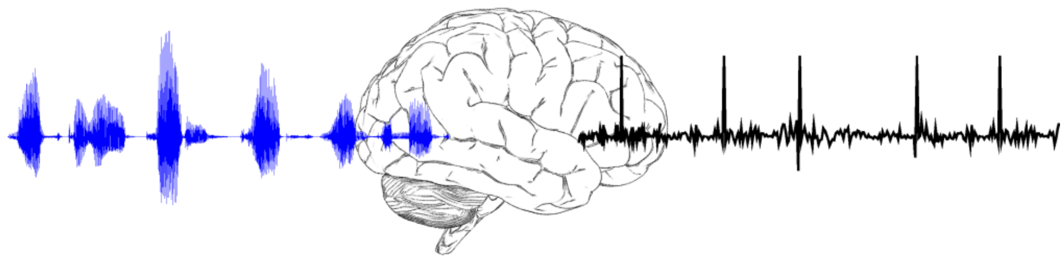


Human Auditory Perception

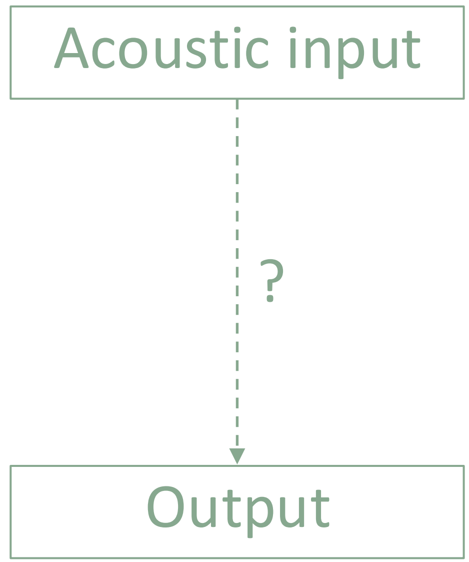
Moïra-Phoebé Huet

Dpt. of Electrical & Computer Engineering
Johns Hopkins University
mphuet@jhu.edu

Speech is just fluctuations in air pressure…

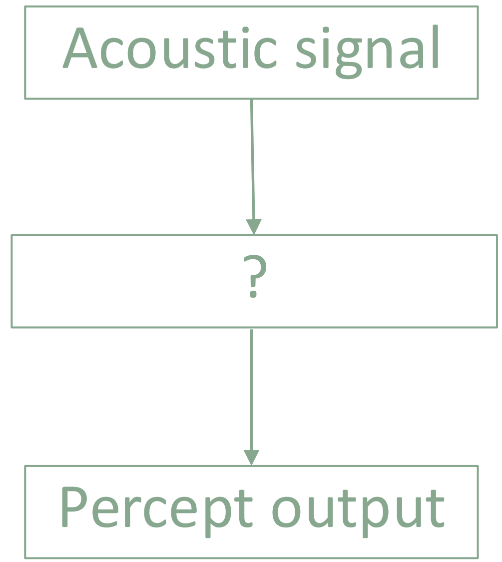
- None of these appear explicitly in the waveform
- Yet we recover all of them
- How?

From sound to perception



- What did you hear?

The missing piece



- Same input
- Different percepts
- Perception is an active interpretation process

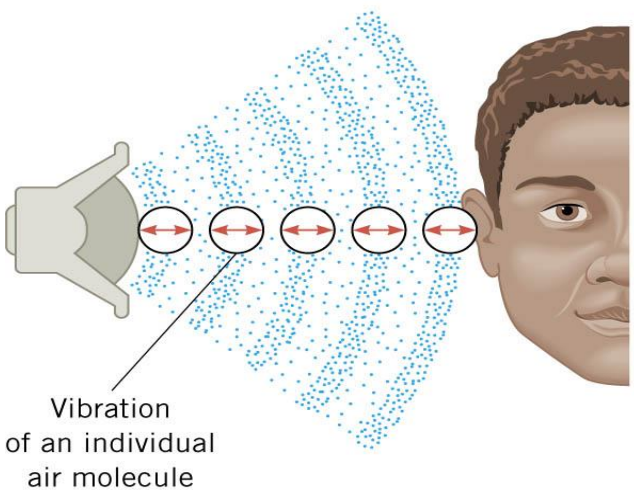
Roadmap

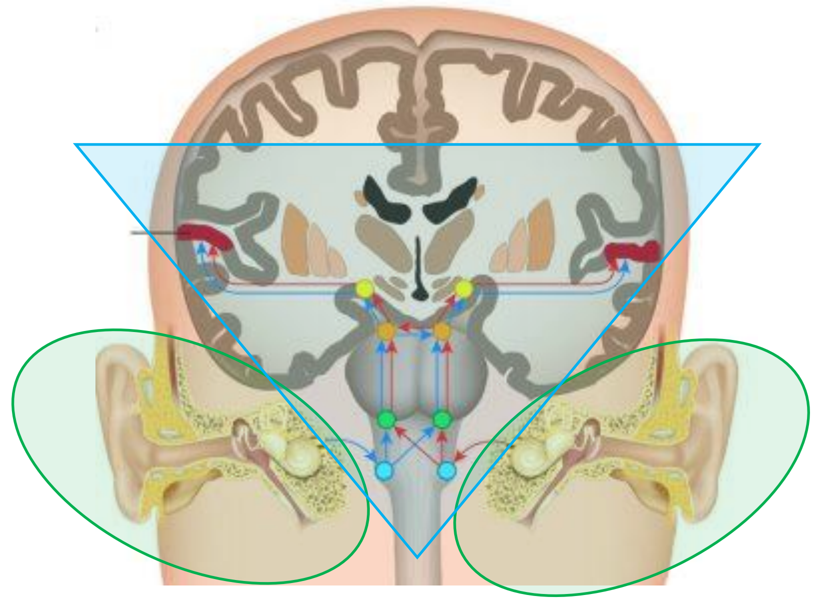
1. Biological feature extraction
2. Neural feature representation
3. Perceptual dimension & construction
4. Perception in the real world

1. Biological feature extraction

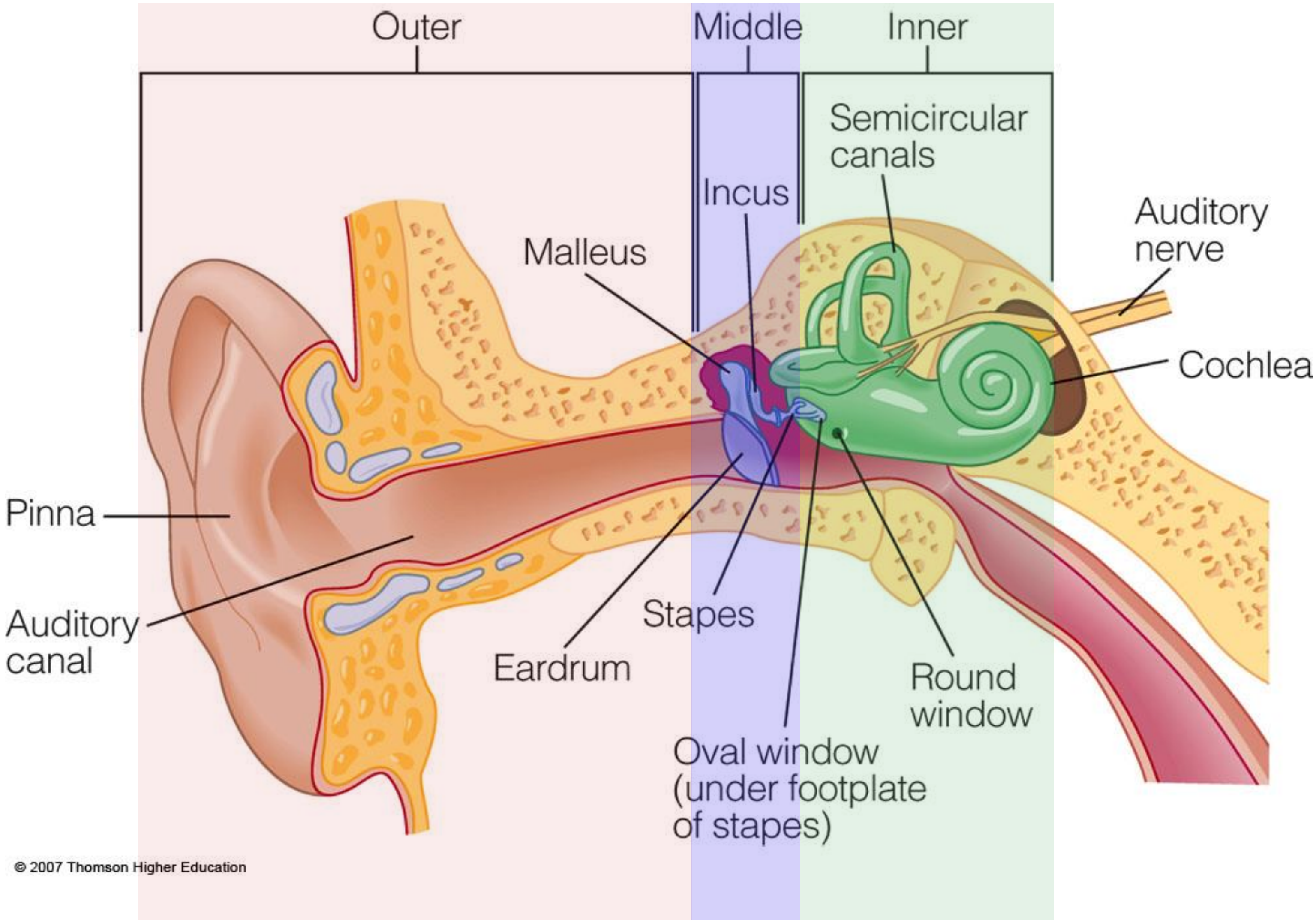
The auditory system transform sound

- The ear is not a passive microphone; it transforms pressure fluctuations into neural activity.
- Two major components in the auditory system
  - Peripheral Auditory System (the ear; green)
  - Central Auditory System (the brain; blue)



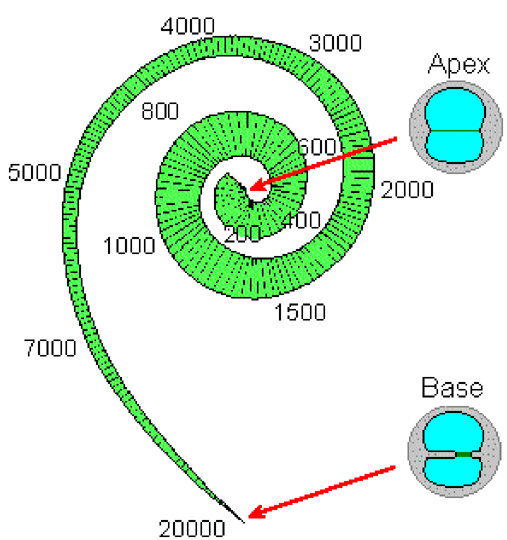


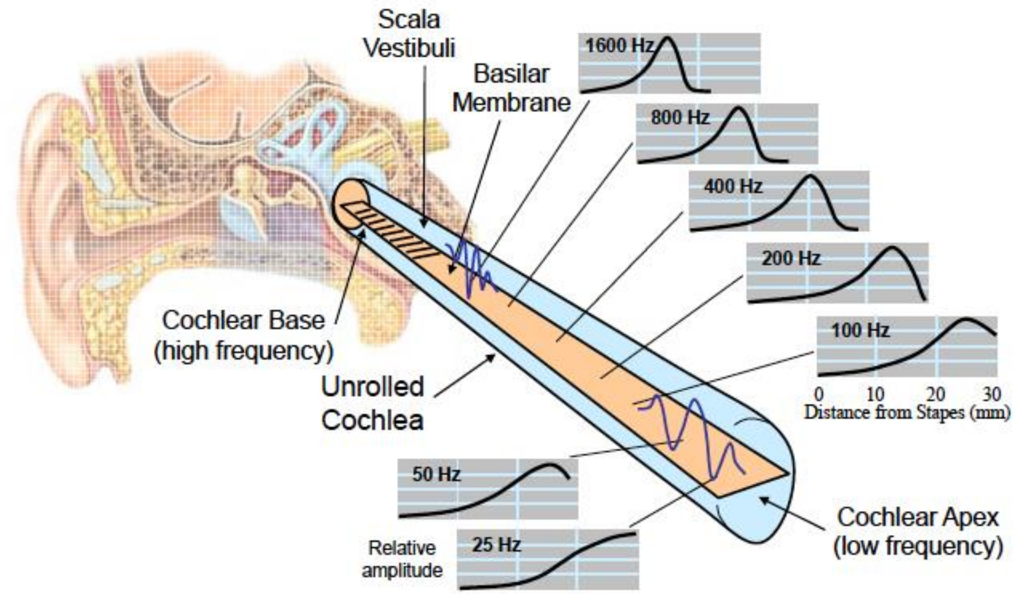
The ear



Outer
Middle
Inner
Semicircular canals
Incus
Malleus
Auditory nerve
Cochlea
Pinna
Auditory canal
Stapes
Eardrum
Round window
Oval window (under footplate of stapes)

Cochlea: a biology frequency analyzer

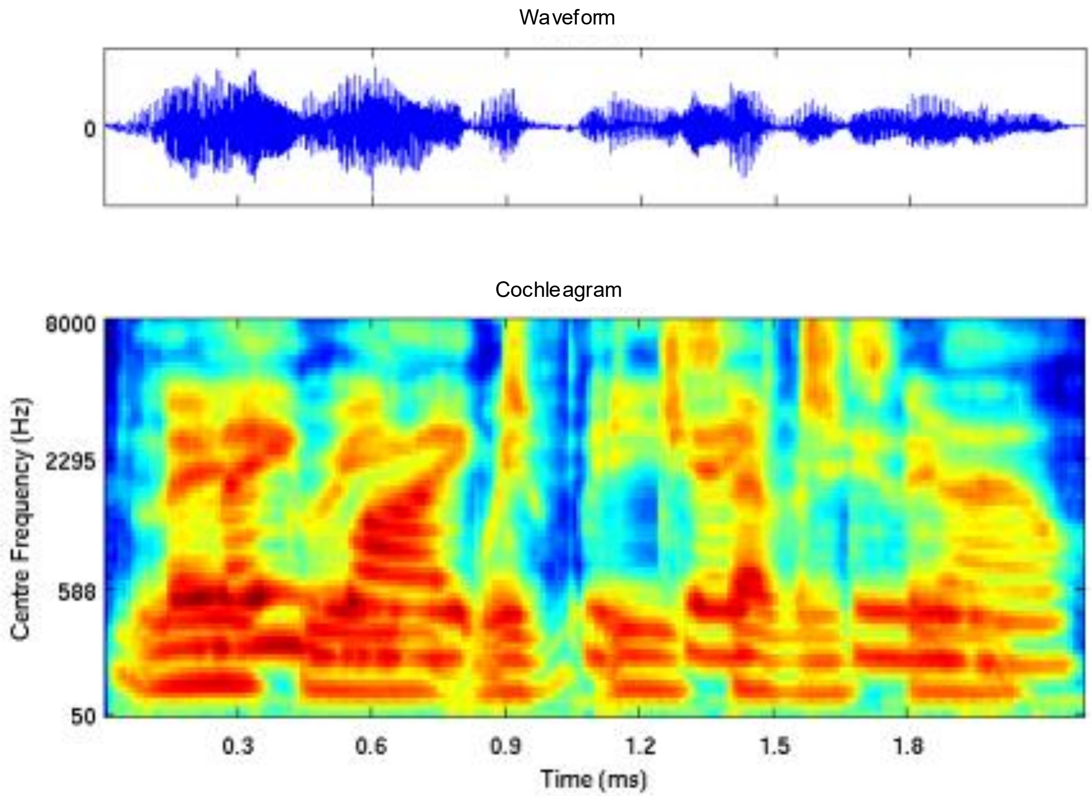




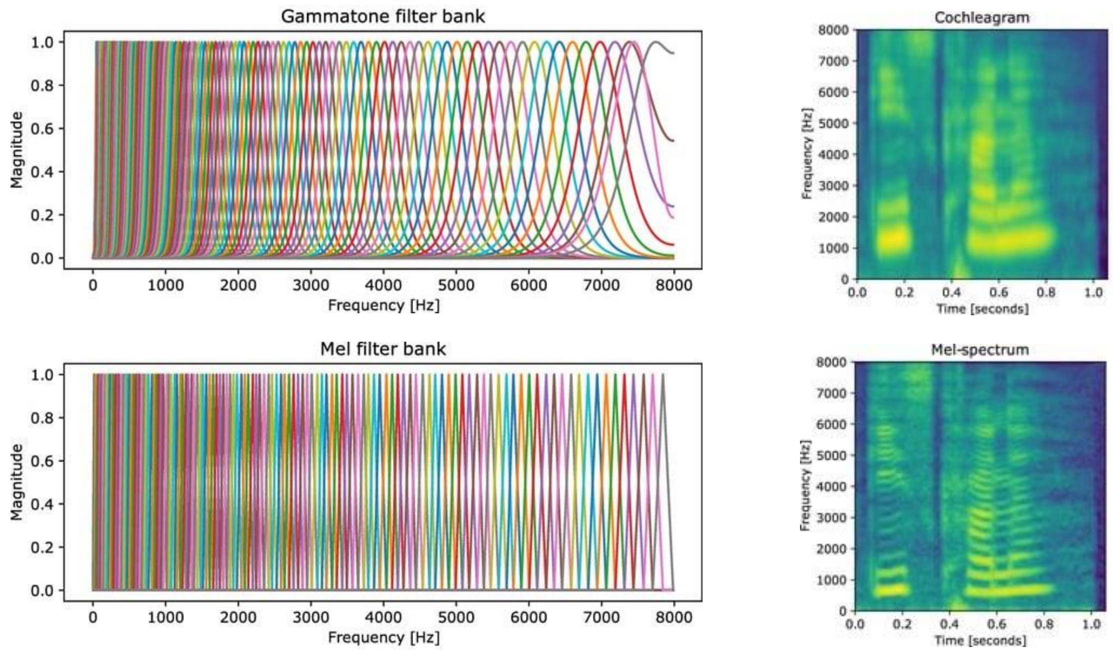
- Different frequencies activate different locations along the basilar membrane.
- Tonotopy : mapping between frequency into space

From cochlea to cochleagram

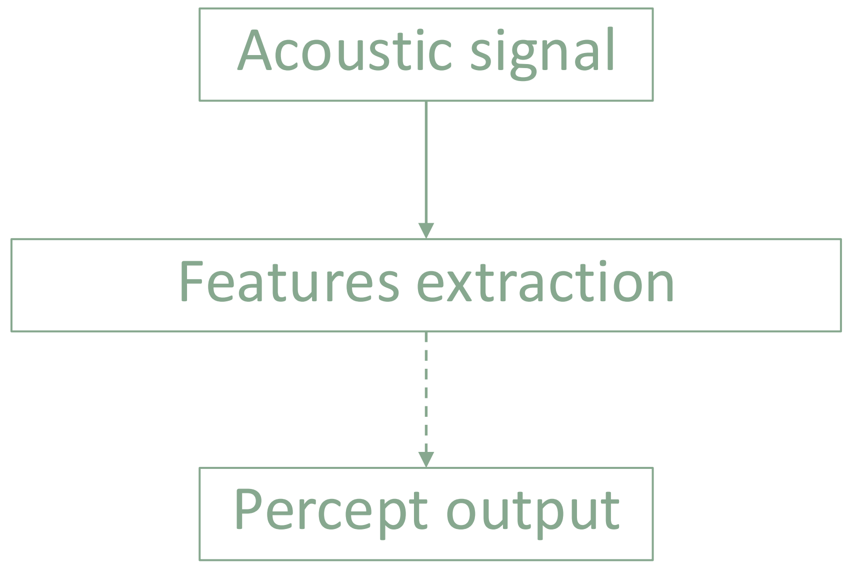
- The activity of all cochlear filters over time forms a cochleagram.



From Cochlea to MFCCs

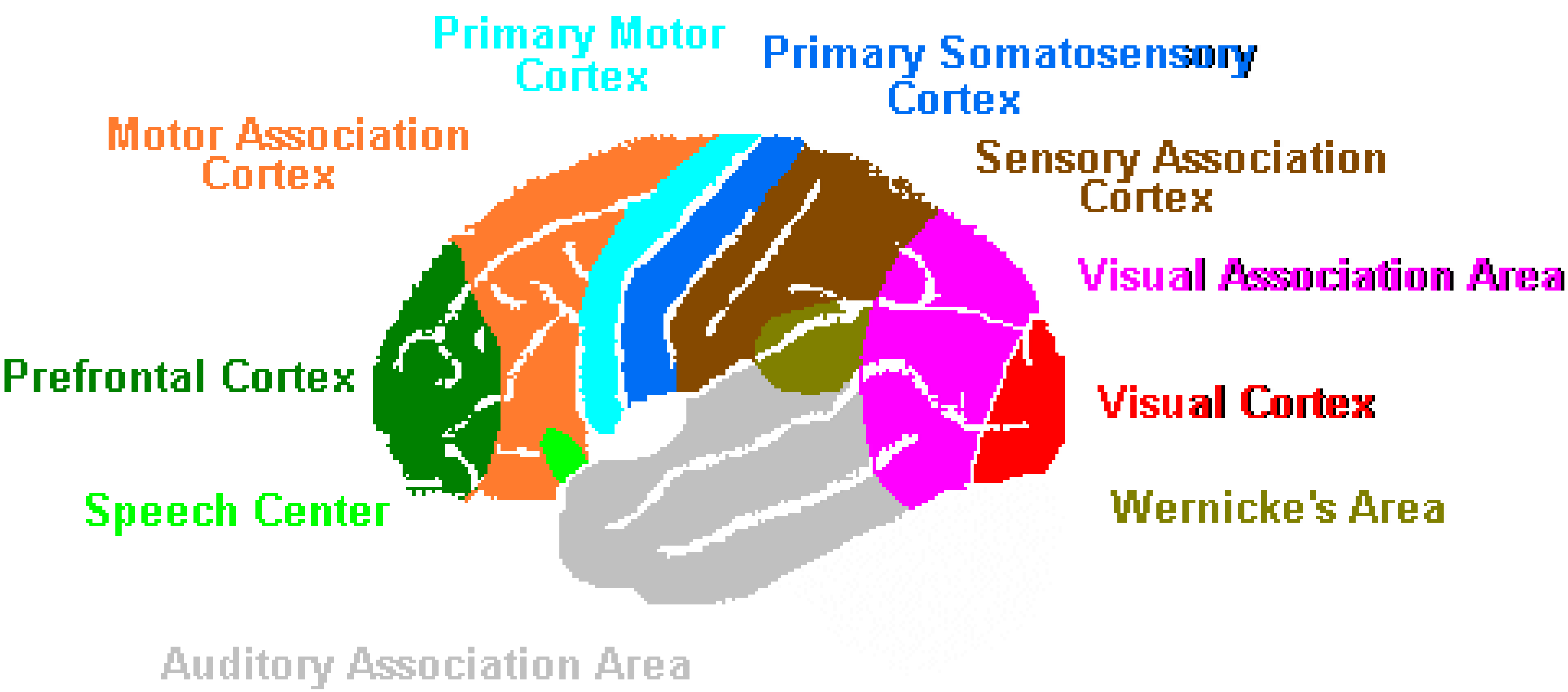


- Shared principles:
  - Frequency decomposition
  - Nonlinear frequency spacing
  - Coarser resolution at high frequencies
- Engineering approximation, not a biological model


Acoustic signal
Features extraction
Percept output

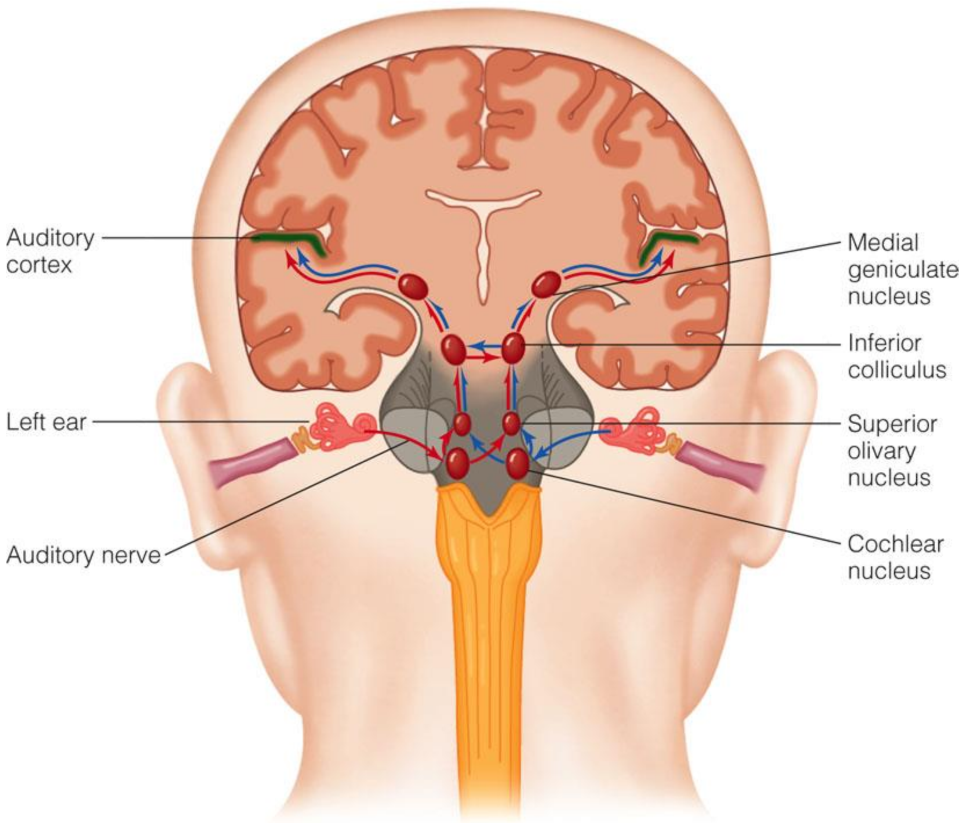
2. Neural feature representation

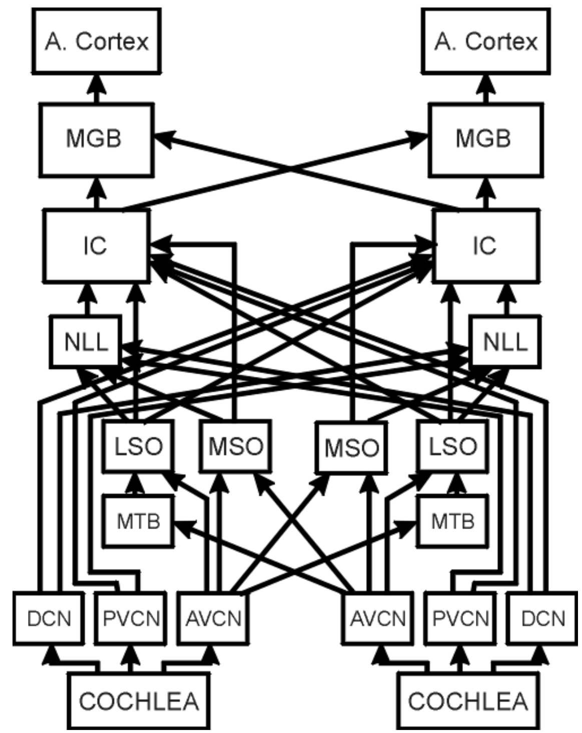
The Brain


Primary Motor Cortex
Primary Somatosensory Cortex
Motor Association Cortex
Sensory Association Cortex
Visual Association Area
Prefrontal Cortex
Visual Cortex
Speech Center
Wernicke's Area
Auditory Association Area

Auditory Pathway

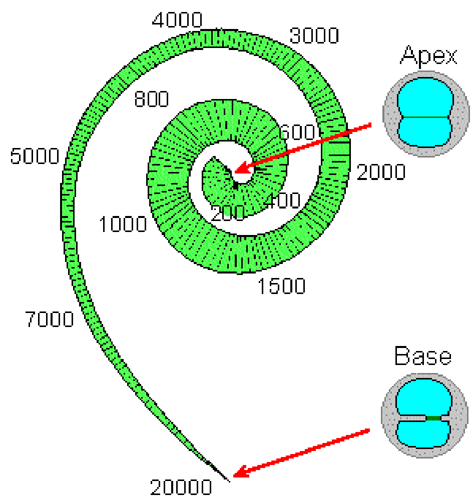
- The representation of sound does not stop at the cochlea.
- Information is progressively transformed at successive stages of the auditory pathway.



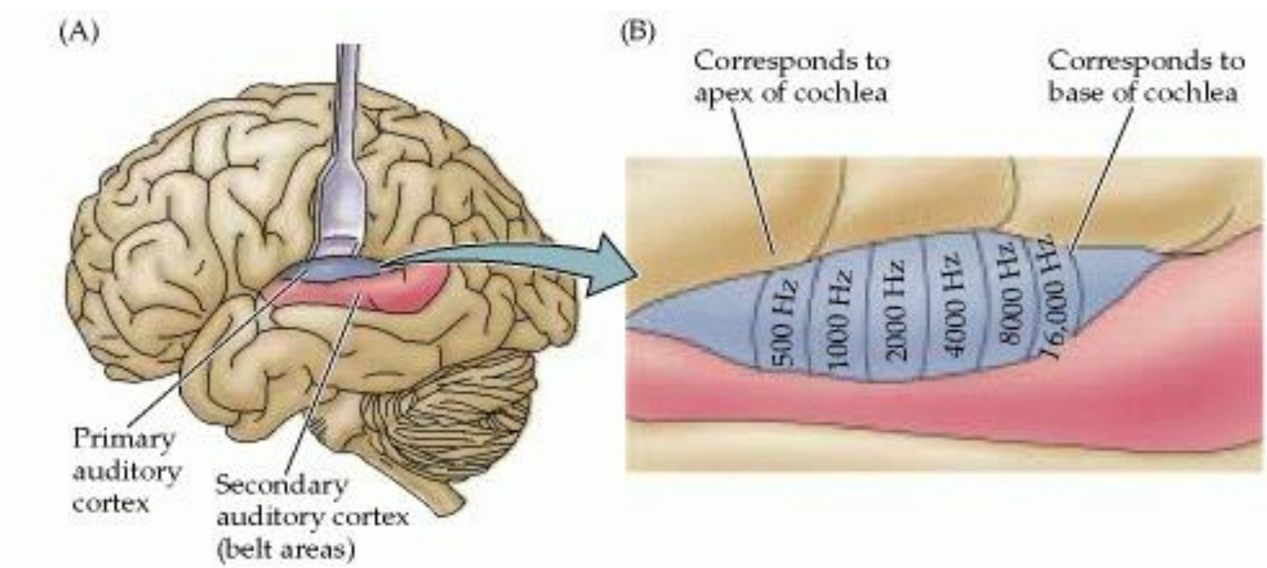


Tonotopy is preserved throughout the auditory system

- Adjacent cells in auditory cortex form a frequency-map, similar to the one observed in the cochlea.



Cochlea

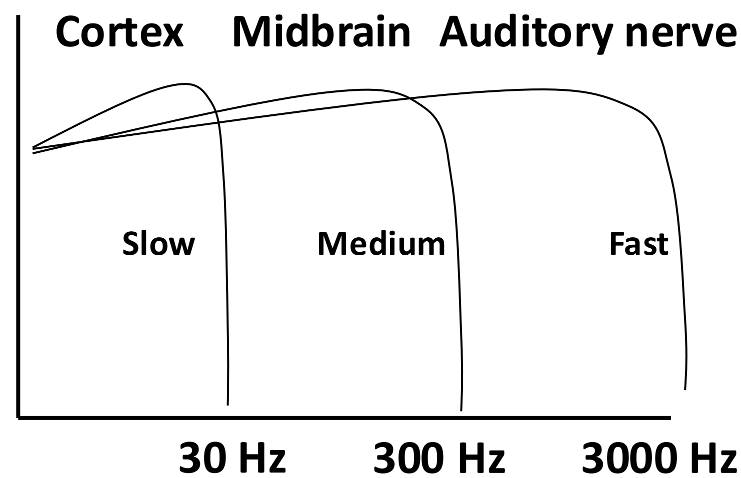


A1

From Temporal Precision to Feature Selectivity

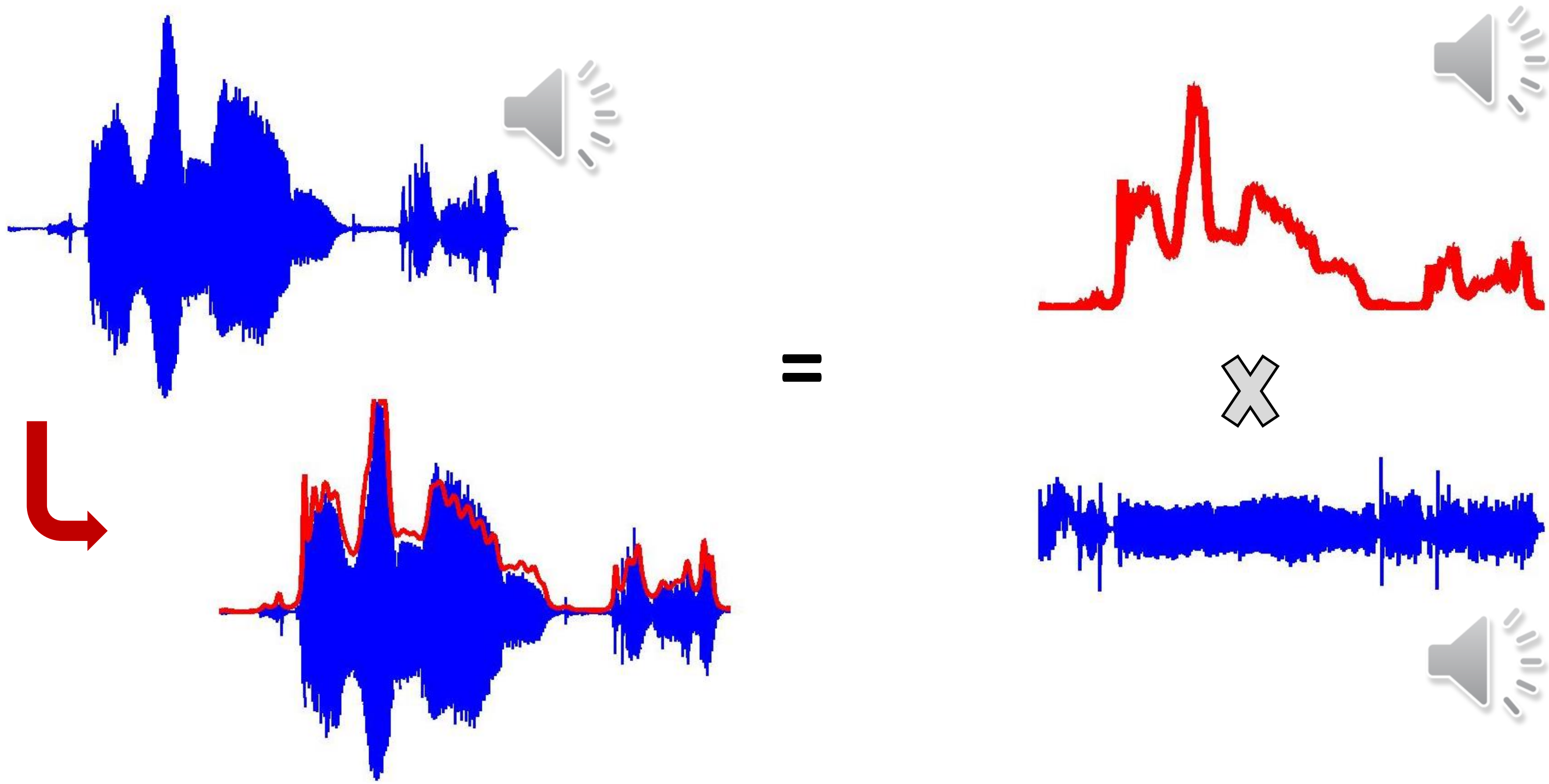
- Auditory nerve neurons can follow very rapid temporal fluctuations
- Cortical neurons are primarily sensitive to slower modulations
- Neural coding gradually shifts from precise temporal synchrony to feature selectivity

Range of Temporal modulations

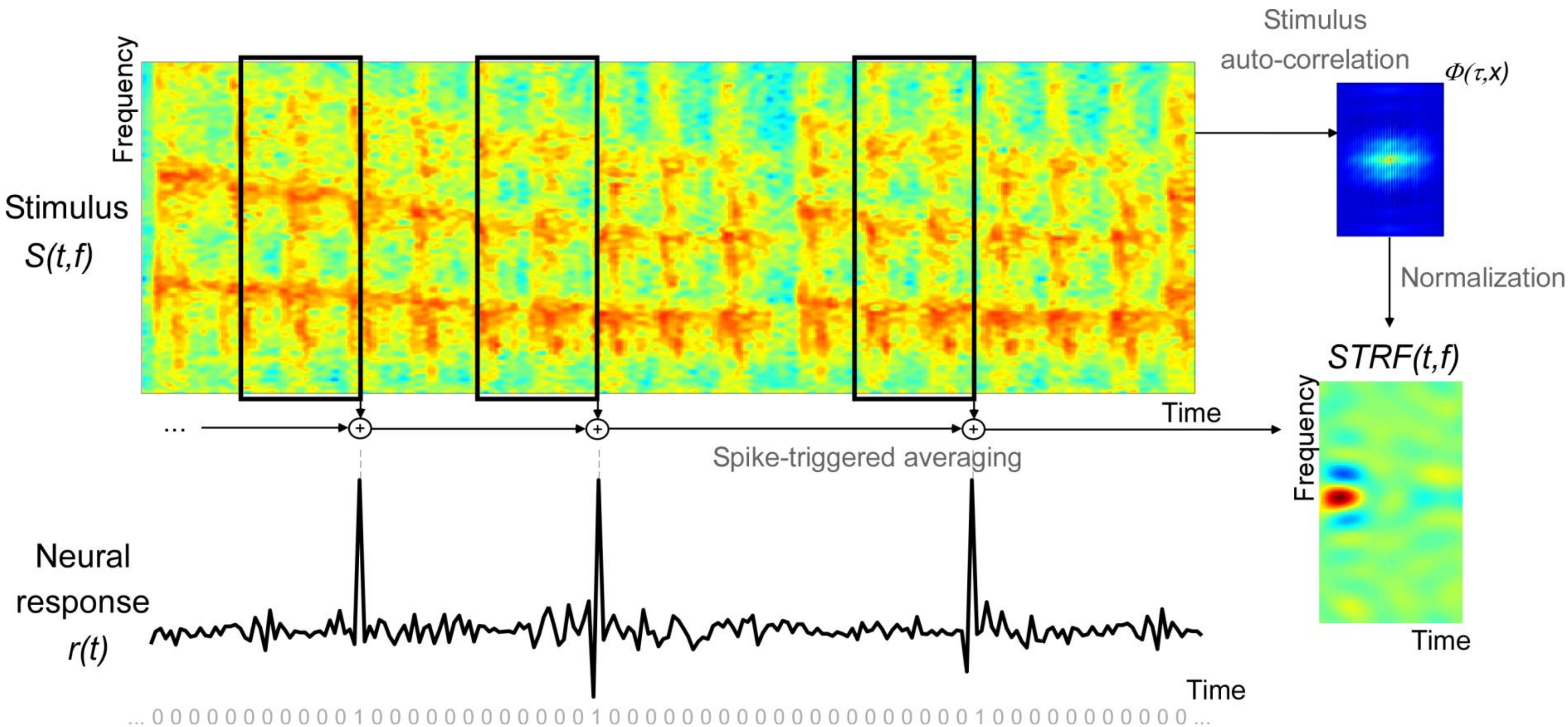


Speech carries information at multiple levels

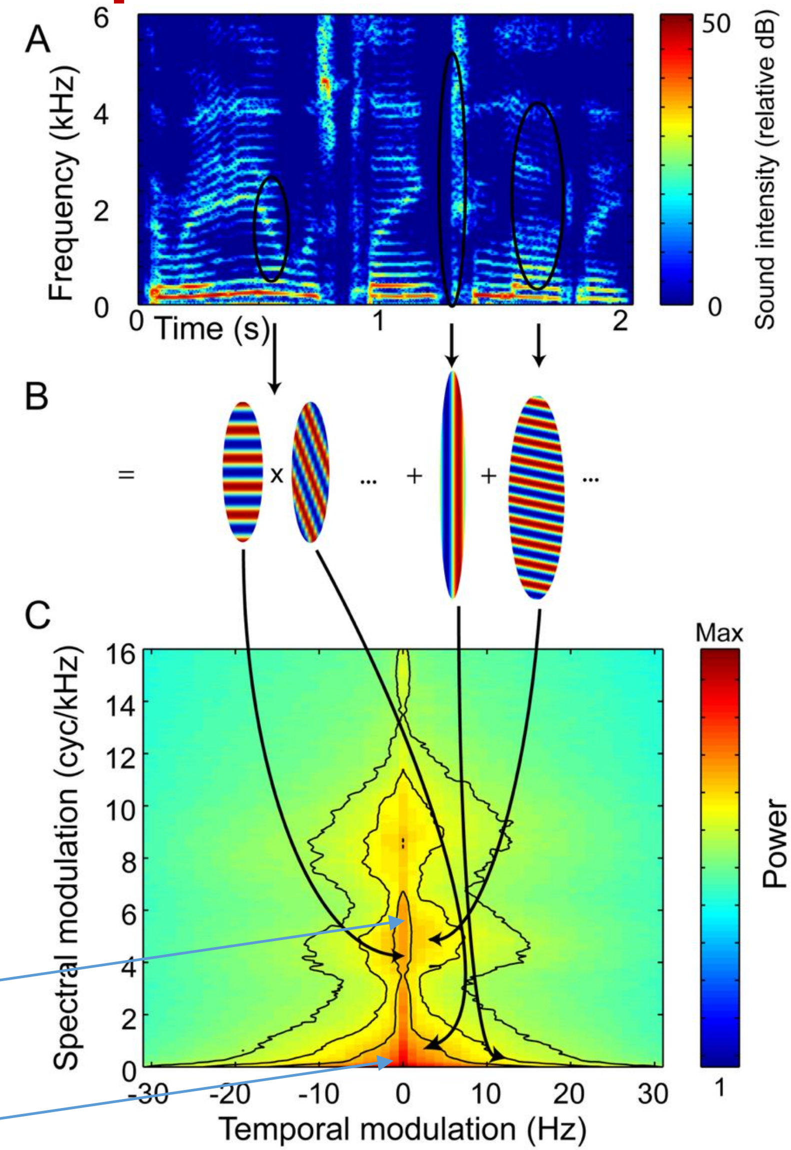
- Any speech signal can be separated into two signals.



Spectro-Temporal Receptive Field (STRF)

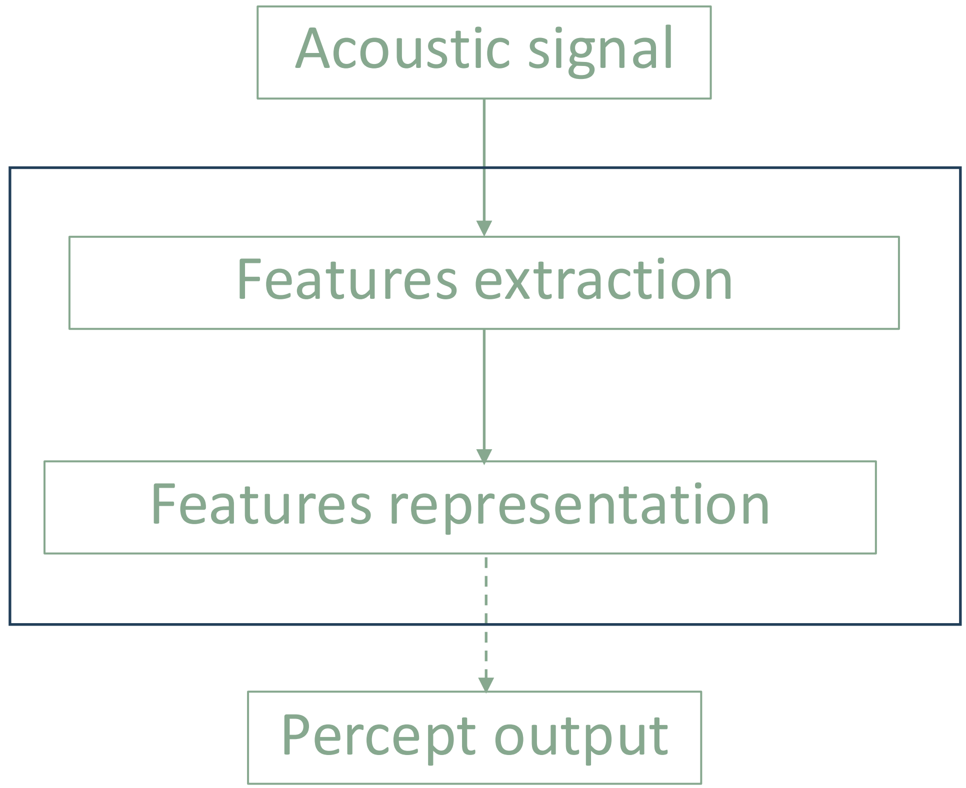


Modulation profiles



voice

Intelligibility


Acoustic signal
Features extraction
Features representation
Percept output

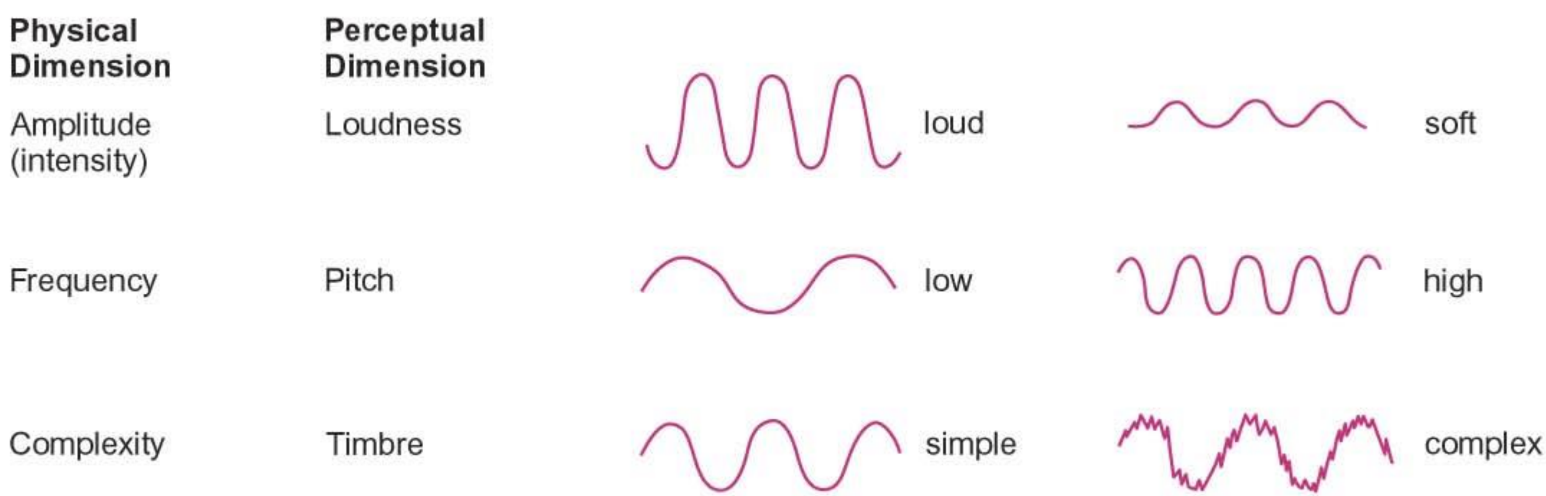
3. Perceptual dimension & construction

Auditory Perception

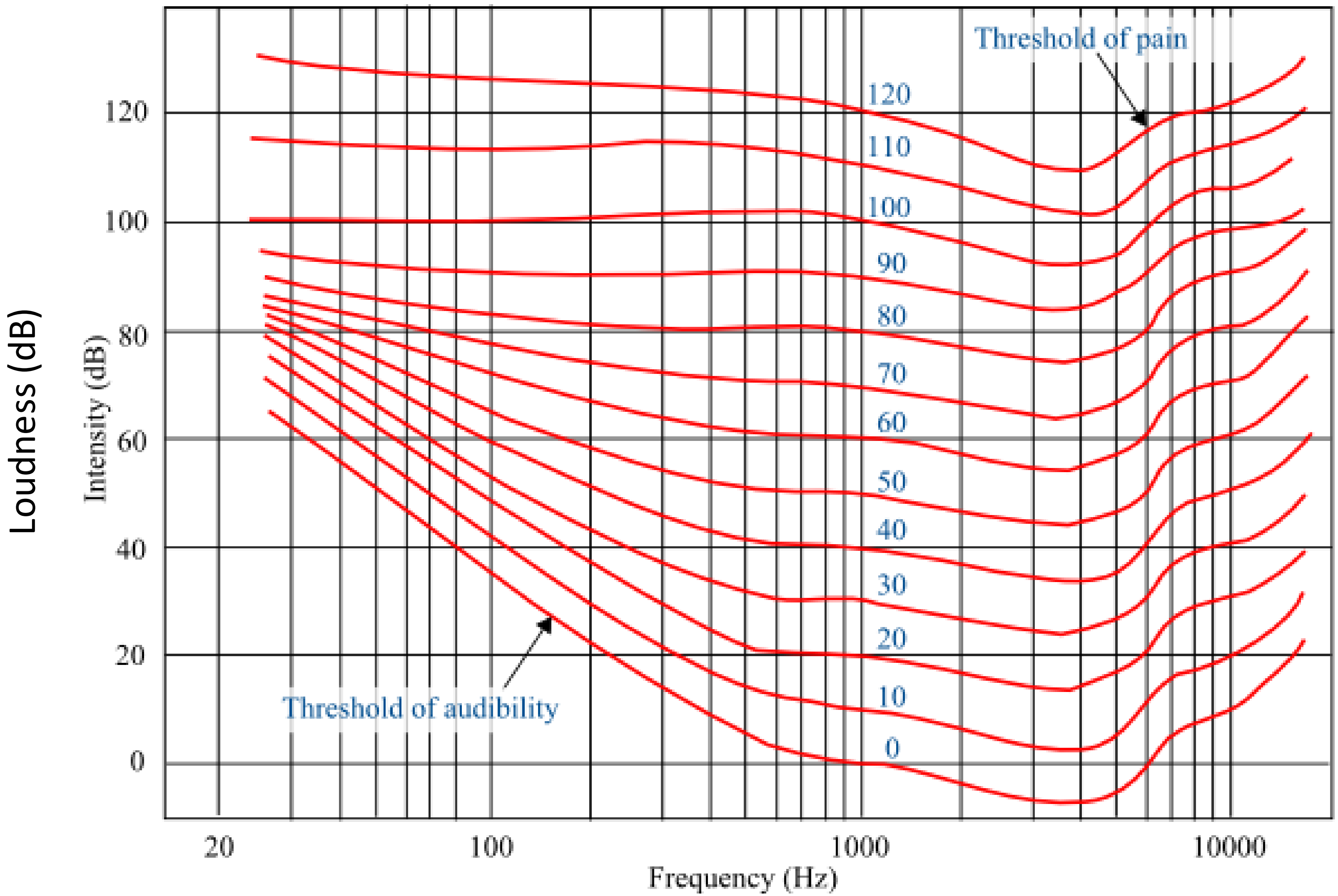
Auditory Perception

<u>Physical Properties of Sound</u>
Intensity
Frequency
Complexity
(frequency content & time)

<u>Perceptual Dimensions</u>
Loudness
Pitch
Timbre



Loudness depends on more than intensity



Each contour represents equally-perceived tones

Pitch perception is nonlinear

The ear is more sensitive to F_0 differences in the low frequencies than the higher frequencies. This means that:

300 vs. 350 $\neq$ 3000 vs. 3050

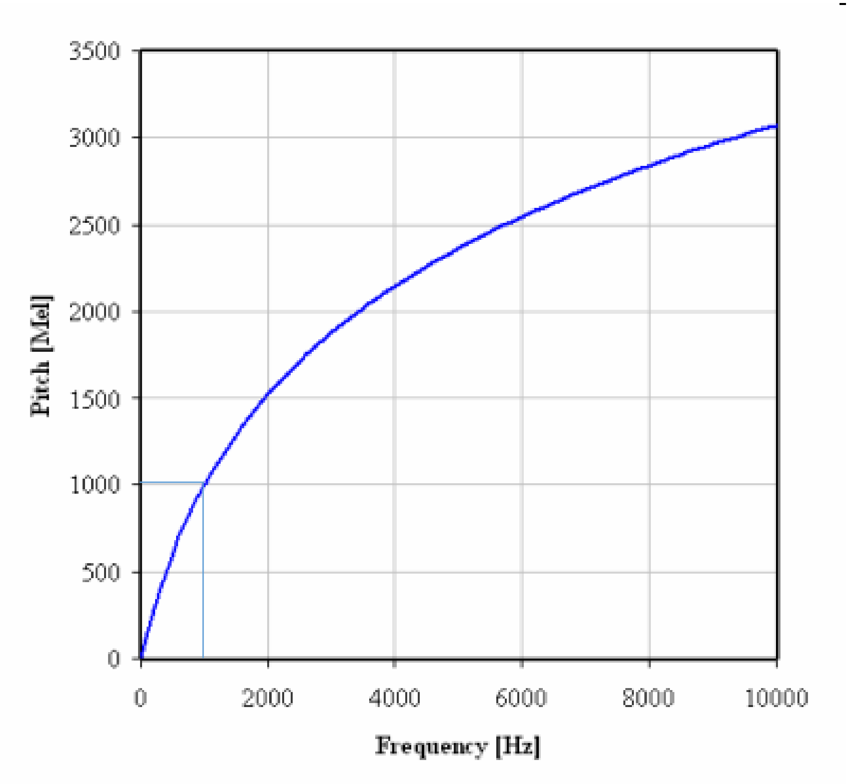
That is, the difference in ***perceived pitch*** (not F_0) between 300 and 350 Hz is **NOT** the same as the difference in pitch between 3000 and 3050 Hz, even though the ***physical*** differences in F_0 are the same.

300-350: **3000-3050:**

Pitch Perception – a Pitch scale

- Perceptual scale of pitch: mel scale

➡ 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)

Perceptual dimensions: Timbre

- Timbre, also known as sound quality or tone color, is oddly defined in terms of what it is not.

- Definition of timbre

  the attribute that allows us to distinguish between sounds having the same perceptual duration, loudness, and pitch, such as two different musical instruments playing exactly the same note

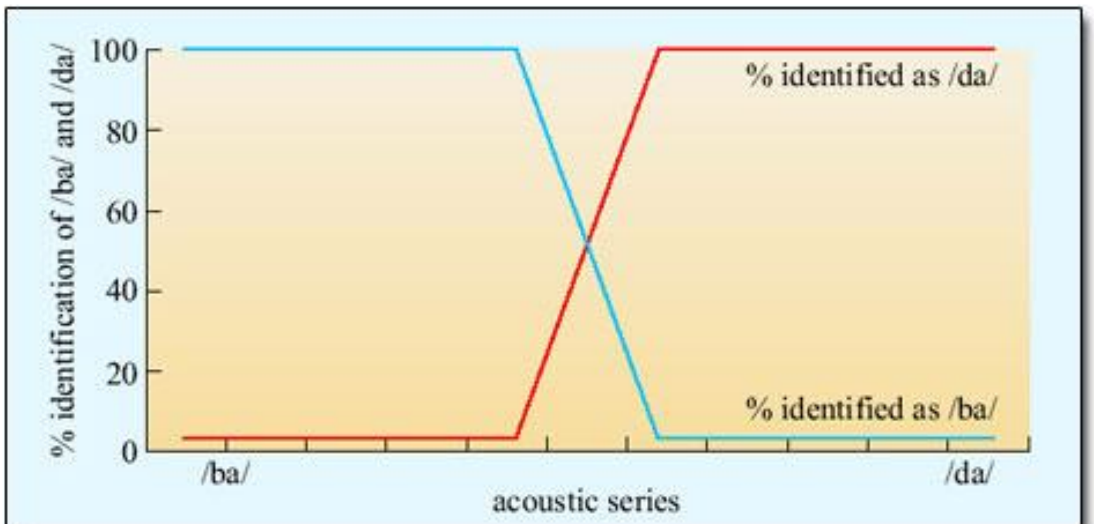
- Example of different sounds with unique sound quality (timbre)

  - ✓ a clarinet, a saxophone, and a piano all play a middle C at the same loudness and same duration.

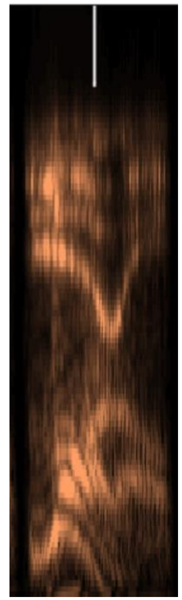
  - ✓ two vowels (e.g., /a/ and /i/) spoken at the same loudness and same pitch differ from one another in timbre

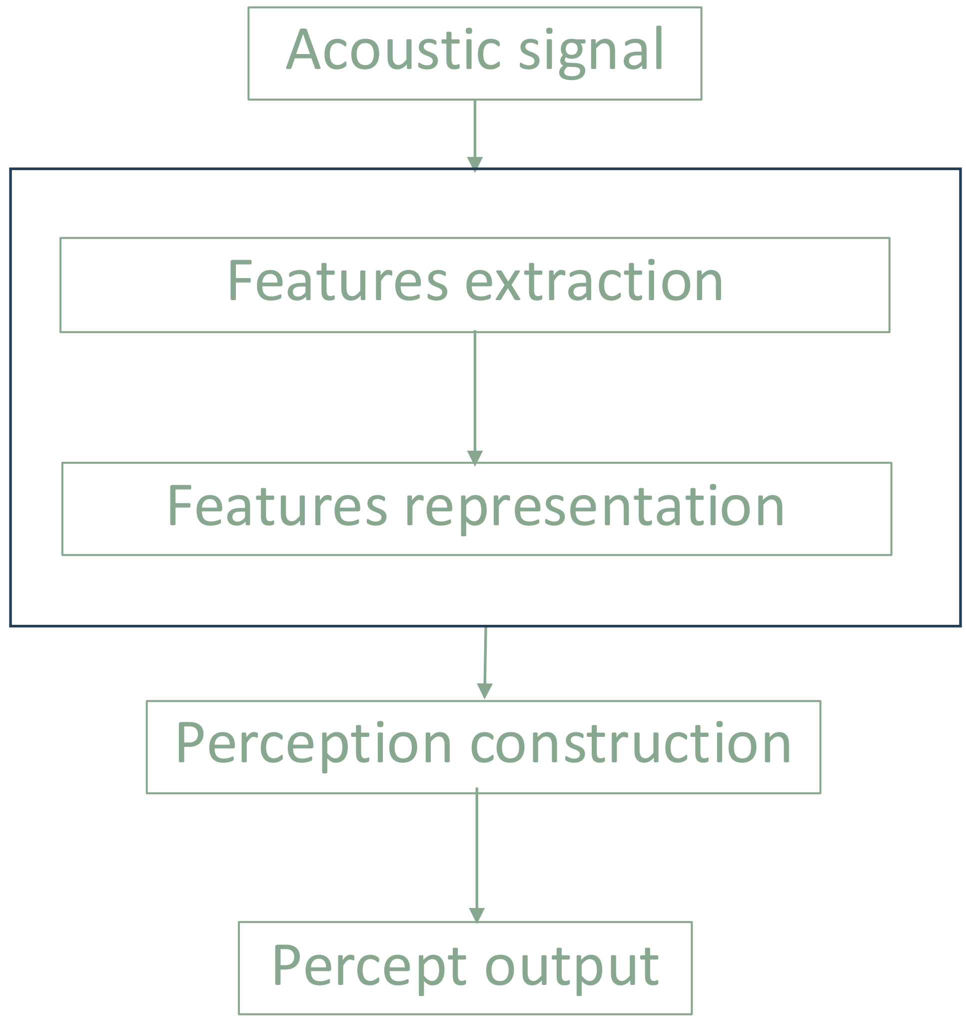
- Physical attributes that relate to timbre are still a topic of research

Categorical Perception



Revisiting Laurel/Yanny




Acoustic signal
Features extraction
Features representation
Perception construction
Percept output

4. Perception in the real world


8oz 12oz 16oz
2.50 3.50 4.50
ITALIAN SODA 3.00
MINERAL WATER 1.75
sparkling
SWEET&PASTRIES
FRESH BAKED DAILY

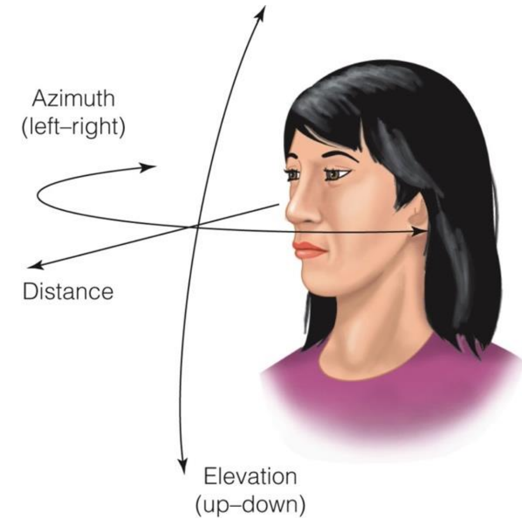
Auditory scene analysis

- Acoustic mixture → auditory objects
- **Segmentation**
  - Decompose the acoustic scene into elementary components
- **Object formation**
  - Extract acoustic features and candidate auditory objects
- **Grouping**
  - Combine components that likely originate from the same source
- **Result**: separate auditory streams

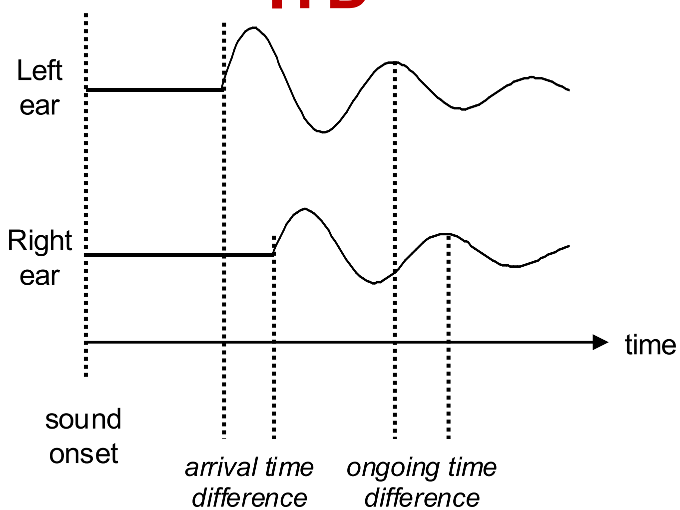
Separate sources (binaural cues)

between 2 ears

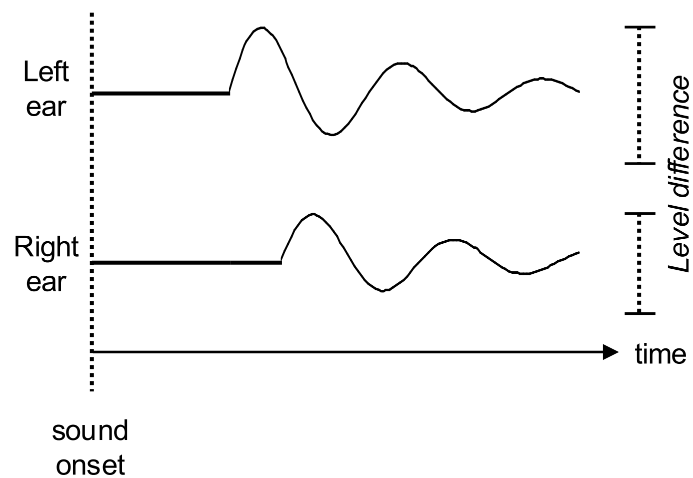
1. inter-aural timing differences (**ITD**)
2. inter-aural level differences (**ILD**); Also called IID: interaural intensity difference



ITD



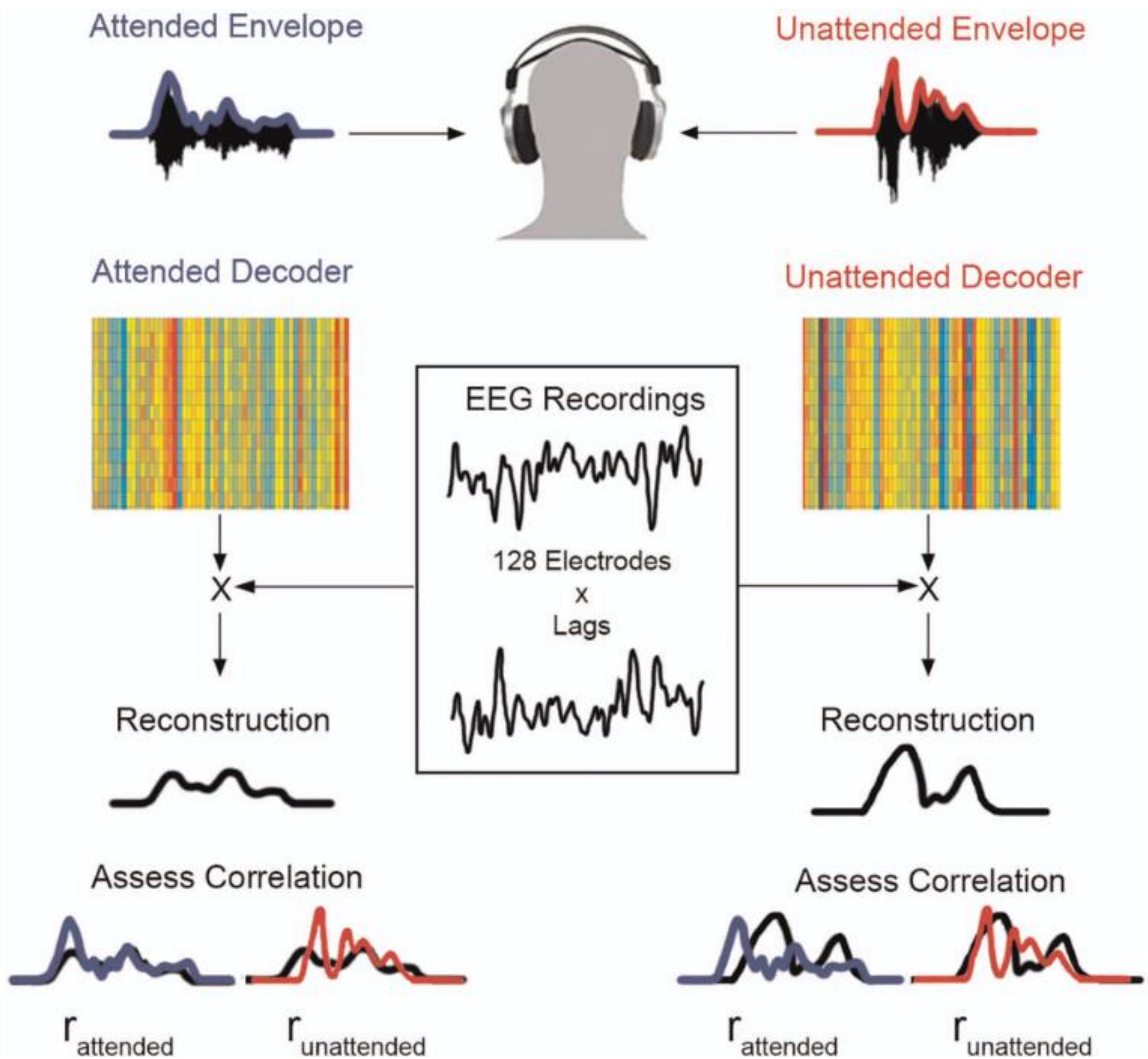
ILD



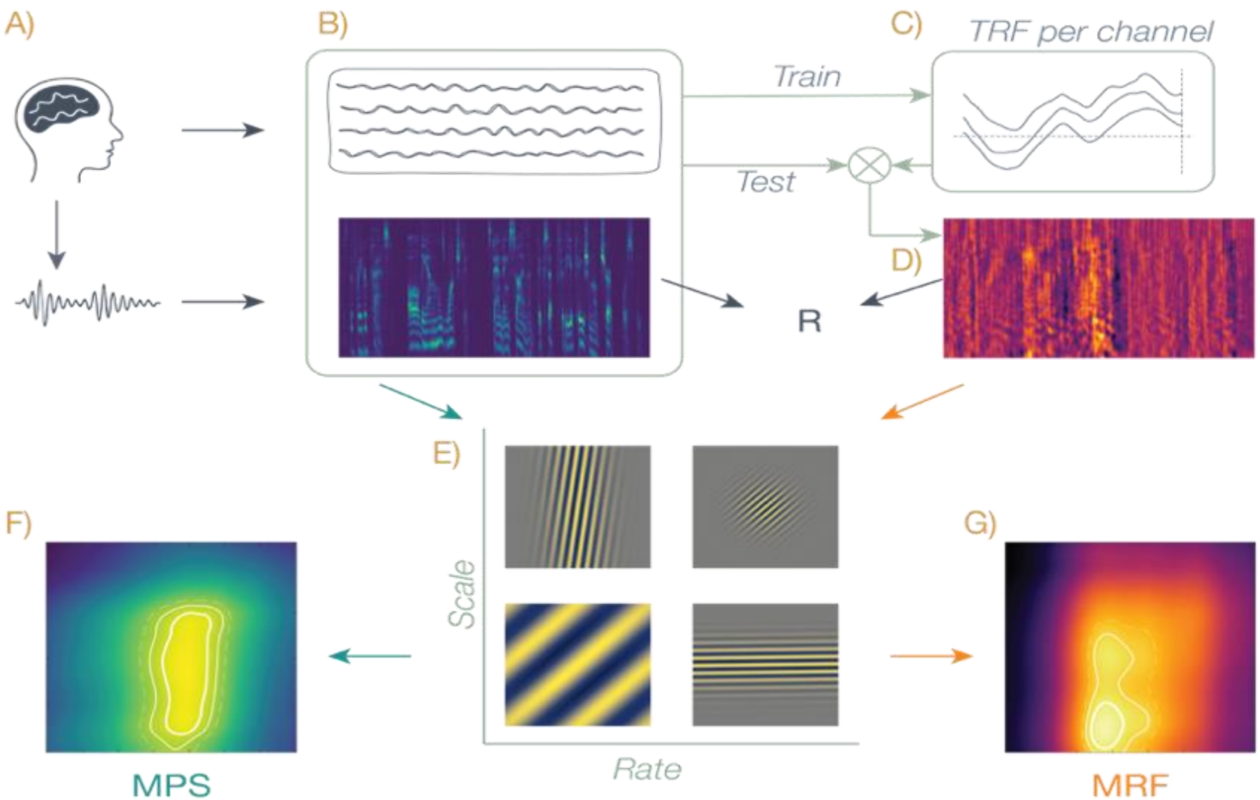
Selective Attention

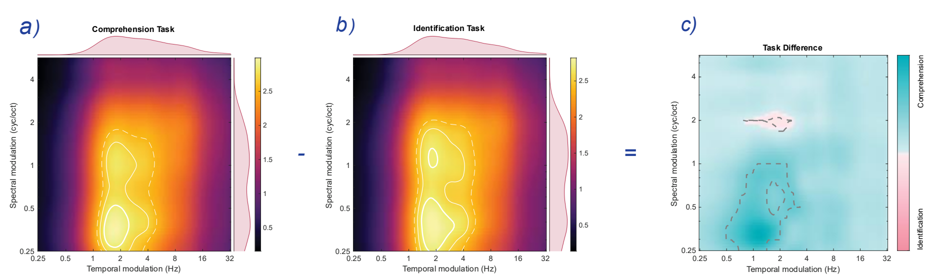
- Multiple streams may be present simultaneously
- Not all information can be processed equally
- Attention prioritizes relevant sounds

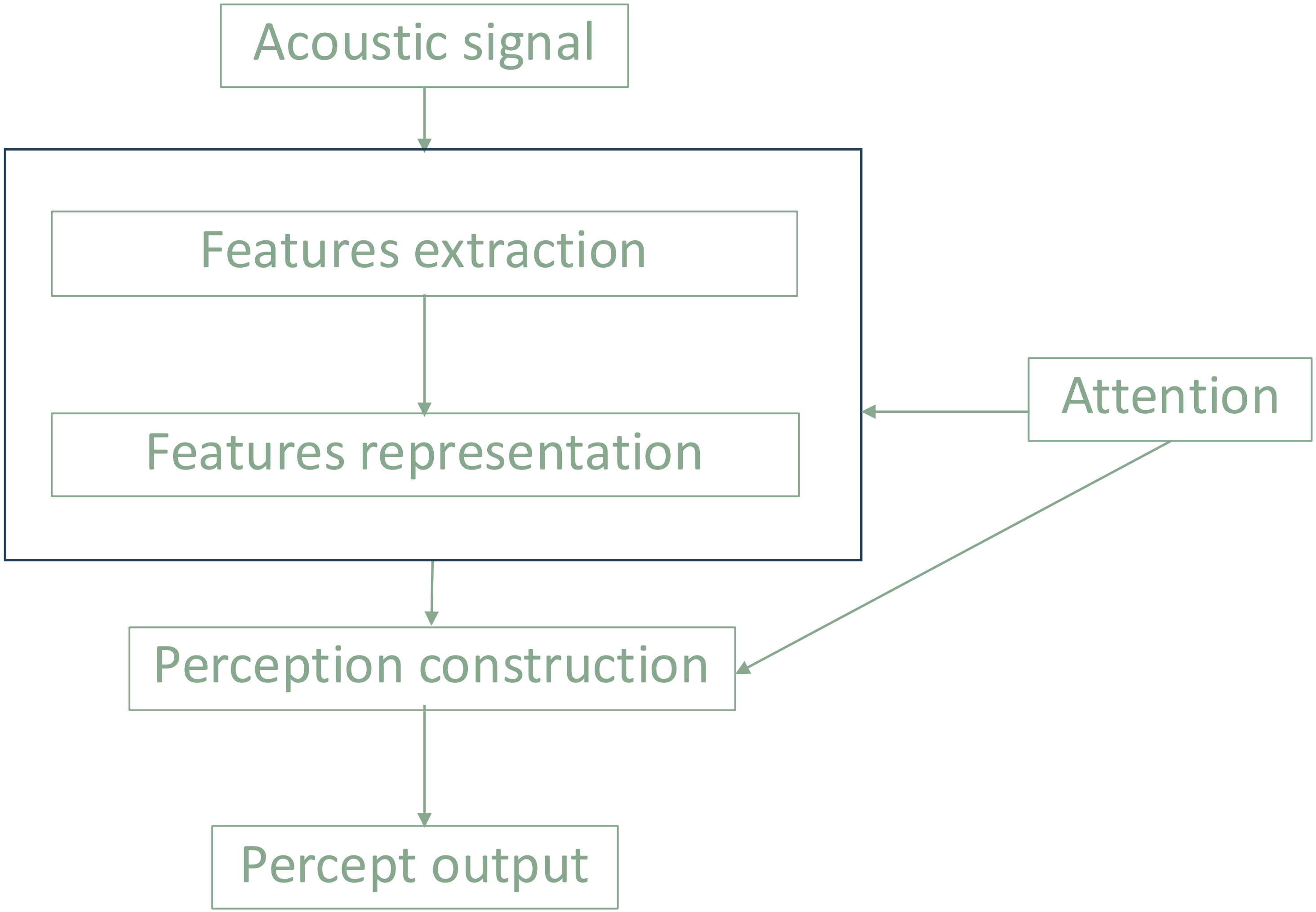
The brain tracks the attended speaker



Attention and modulation profiles






Acoustic signal
Features extraction
Features representation
Attention
Perception construction
Percept output

Key messages

- The auditory system does not passively record sound; it actively **transforms** it into perceptual representations.

- Perception emerges from **multiple stages of processing**, including feature extraction, feature representation, perception construction.