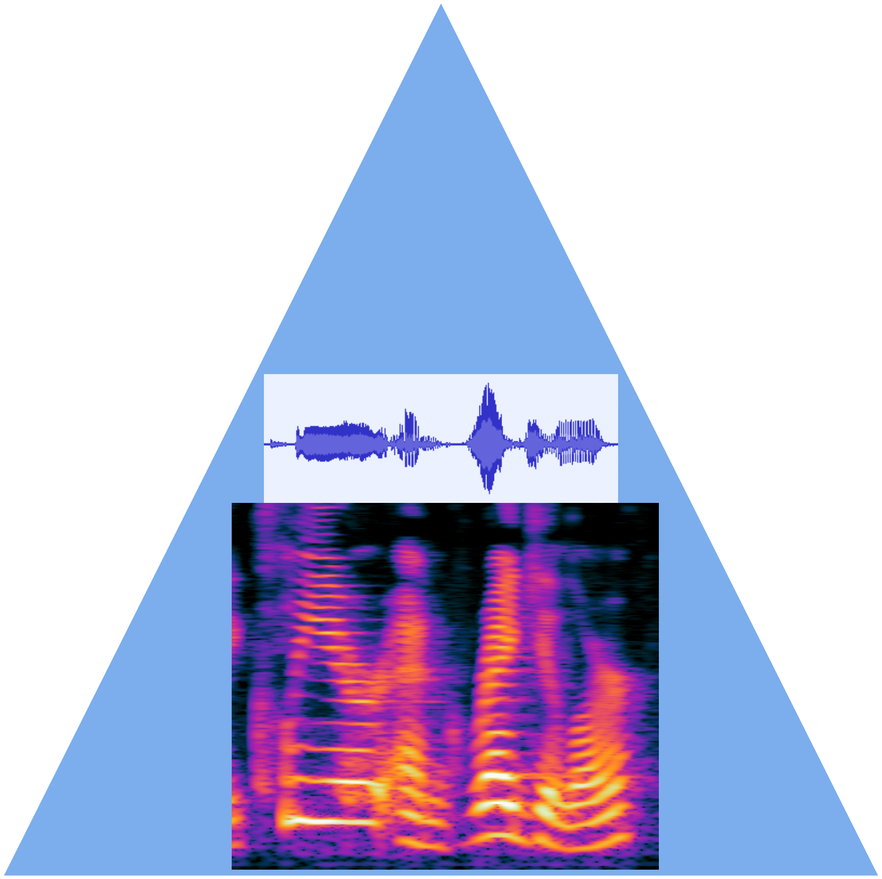
Lecture 7: Speech perception

Zhizheng Wu

Agenda

- Recap
- Speech chain
- Auditory system

Content



Timbre

Prosody

Source-filter model

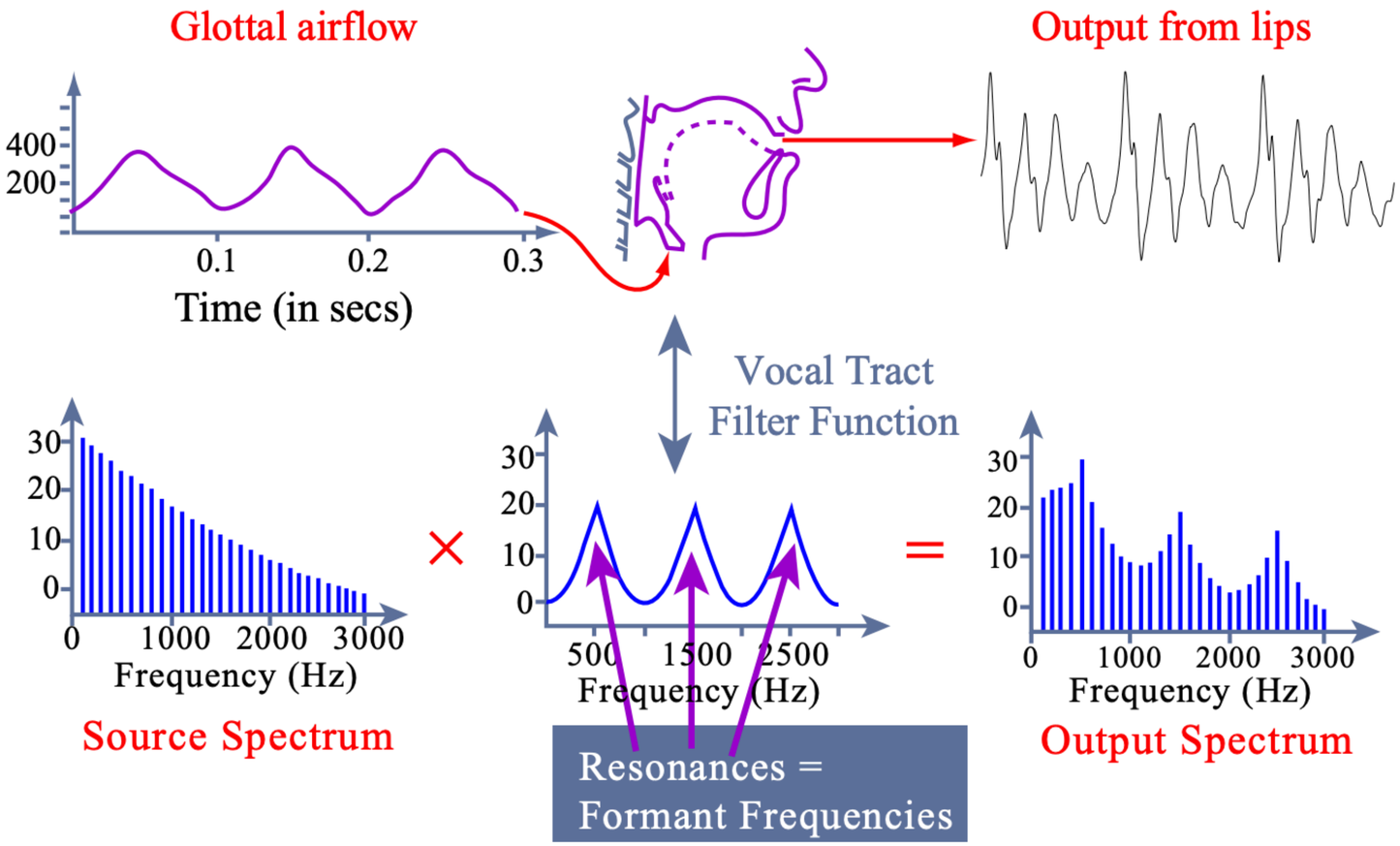
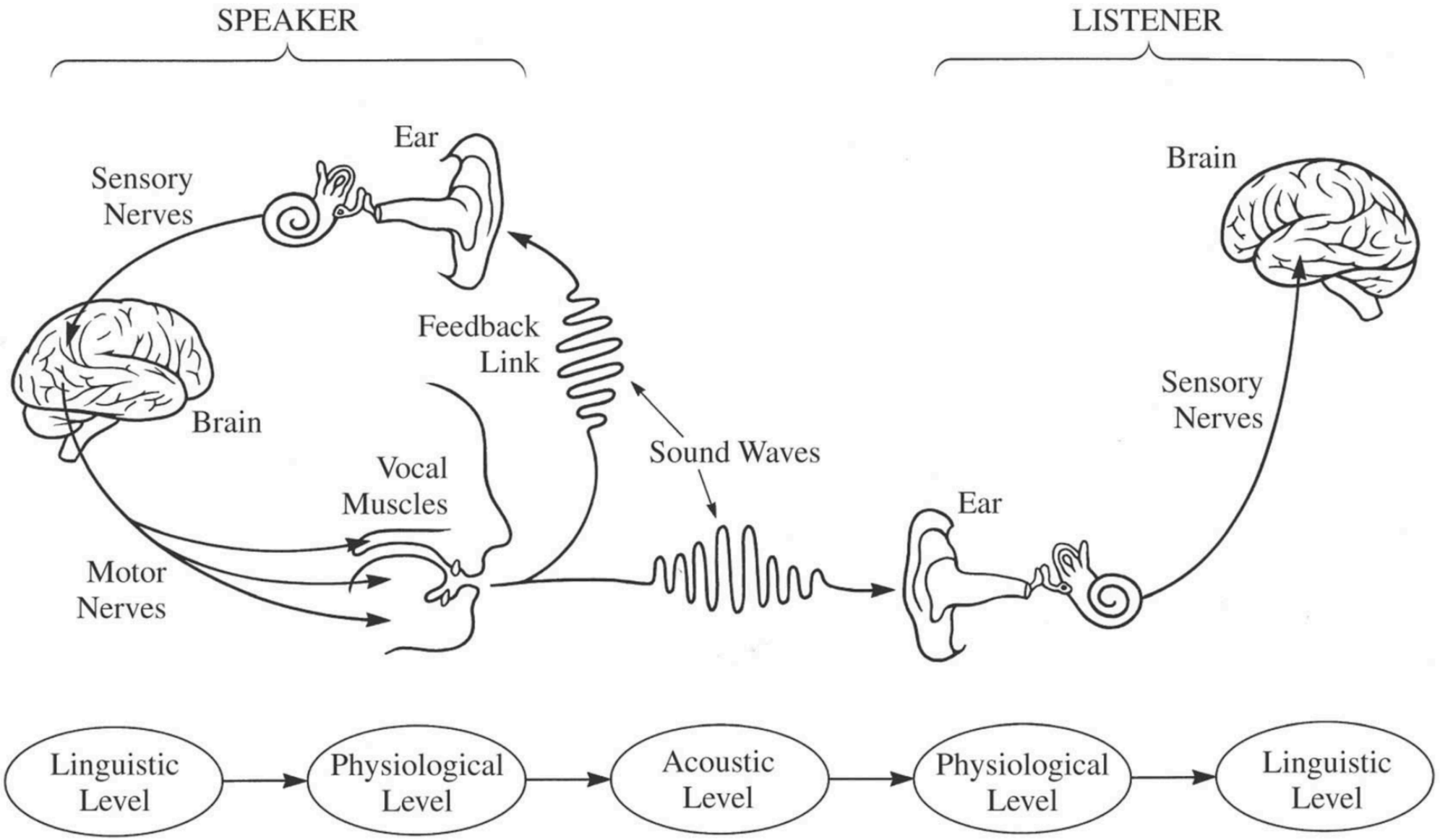


Image credit: MIT OCW

Speech perception

- Understand how we hear sounds and how we perceive speech
- Better design and implementation of robust and efficient systems for analyzing and representing speech
- Try to understand speech perception by looking at the physiological models of hearing

Speech chain

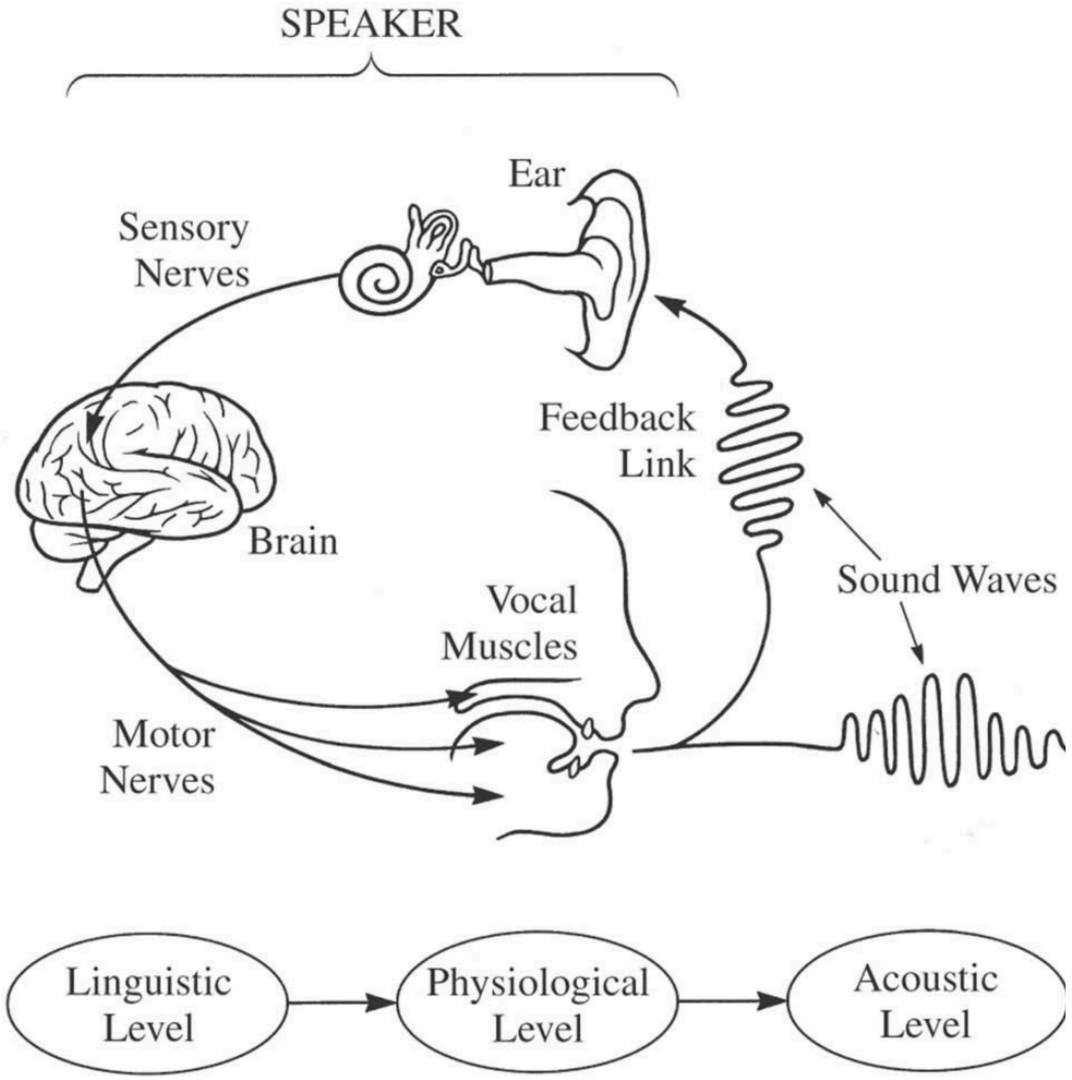


Speech chain

- The Speech Chain comprises the processes of:
  - speech production
  - auditory feedback to the speaker
  - speech transmission (through air or over an electronic wire)
  - communication system (to the listener)
  - speech perception and understanding by the listener.

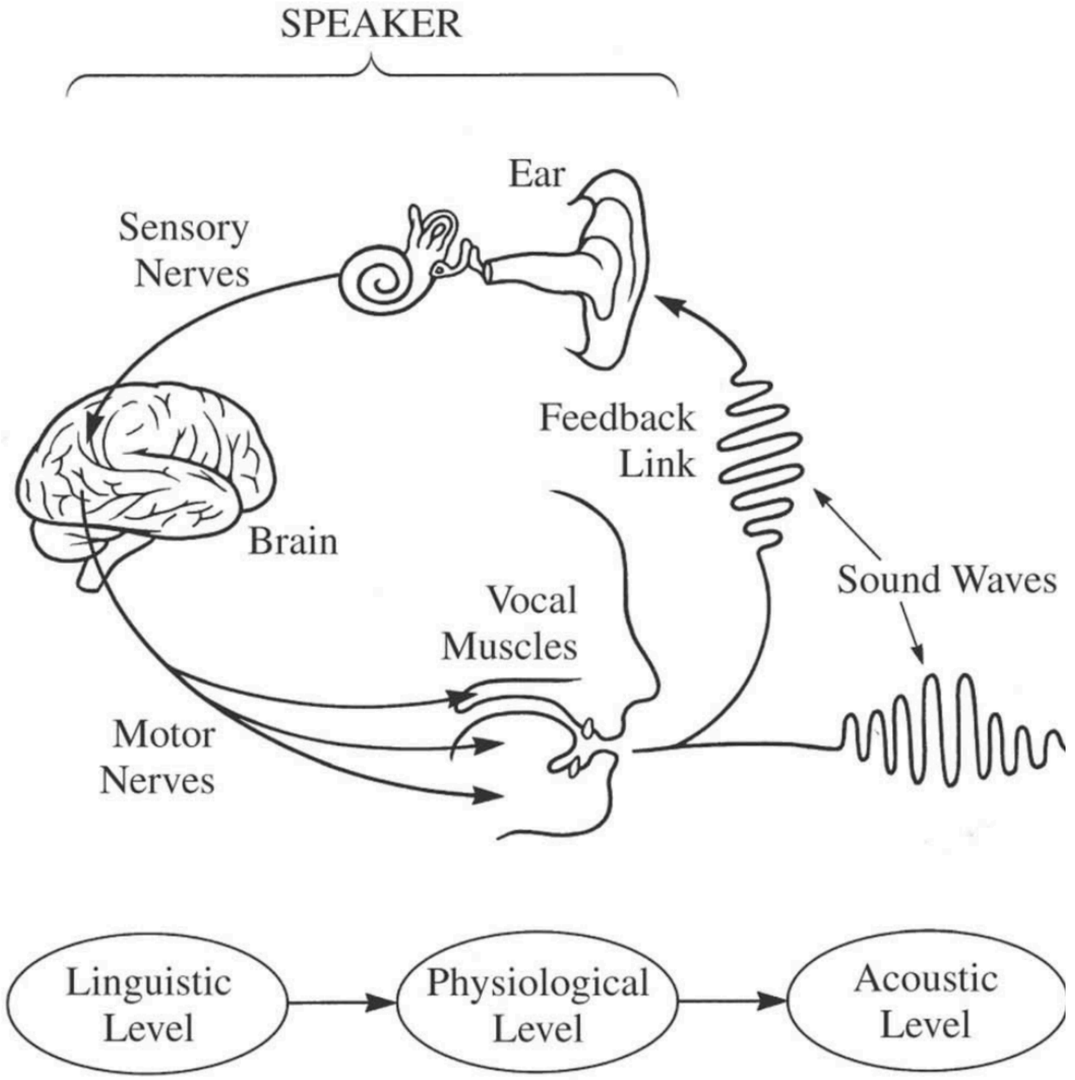
Speech chain： Linguistic level

The basic sounds of the communication are chosen to express some thought of idea



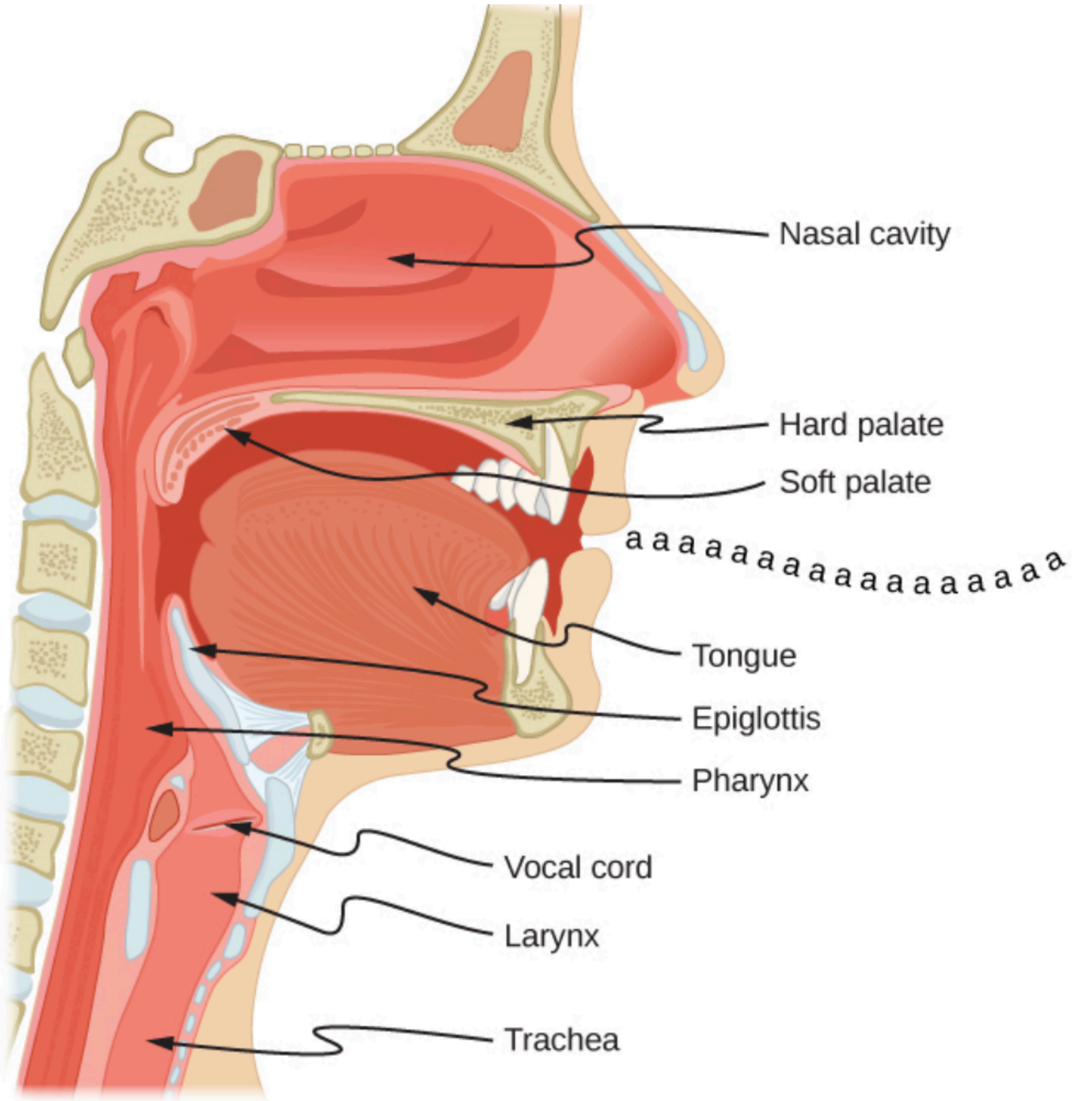
Speech chain： Physiological level

The vocal tract components produce the sounds associated with the linguistic units of the utterance



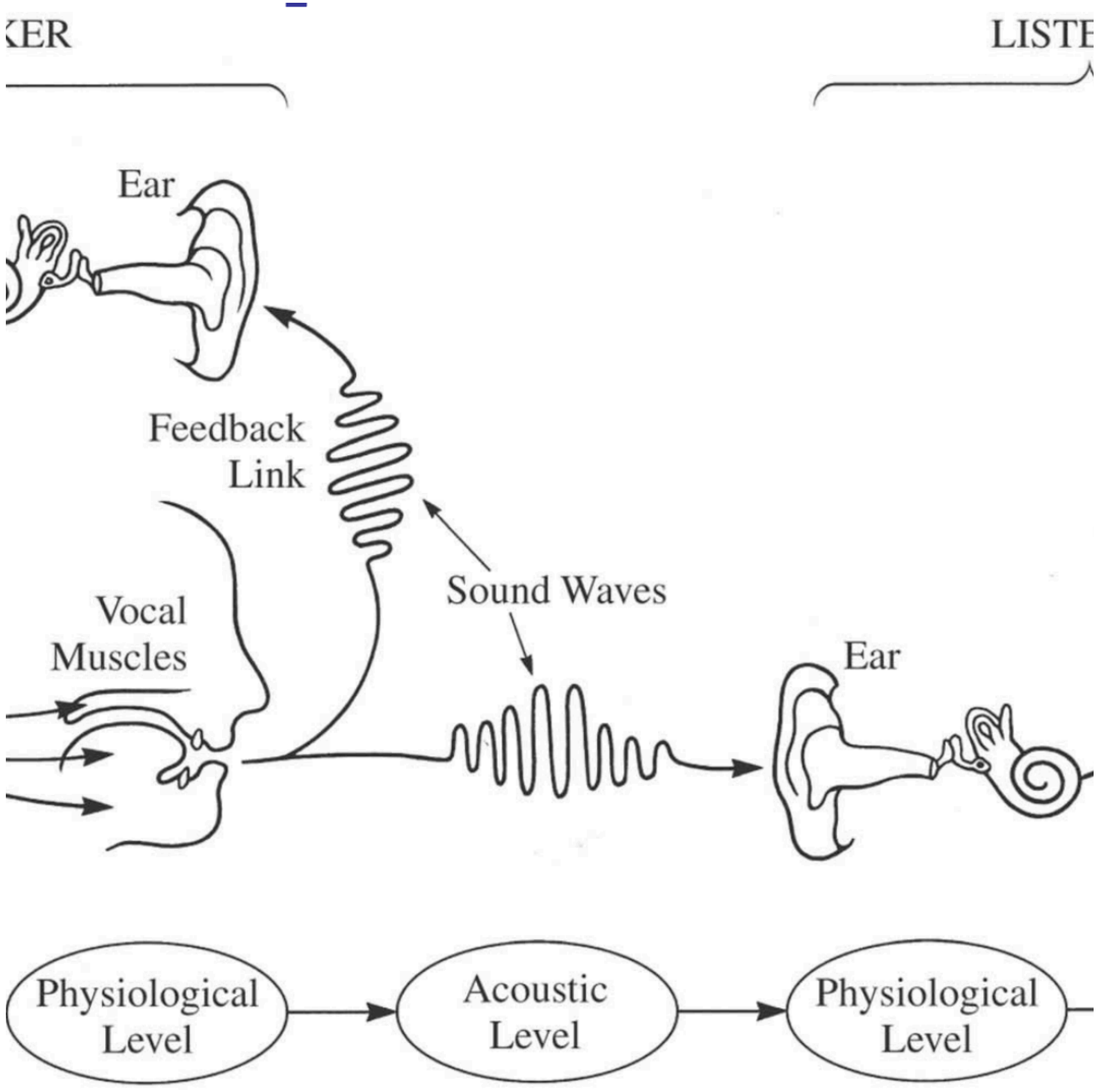
Review: Speech production

- Source-filter model
  - Source produces an initial sound
  - Vocal tract filter modifies it
- Source
  - An input of acoustic energy into the speech production system
- Vocal tract filter
  - Articulators: tongue, teeth, lips, velum etc



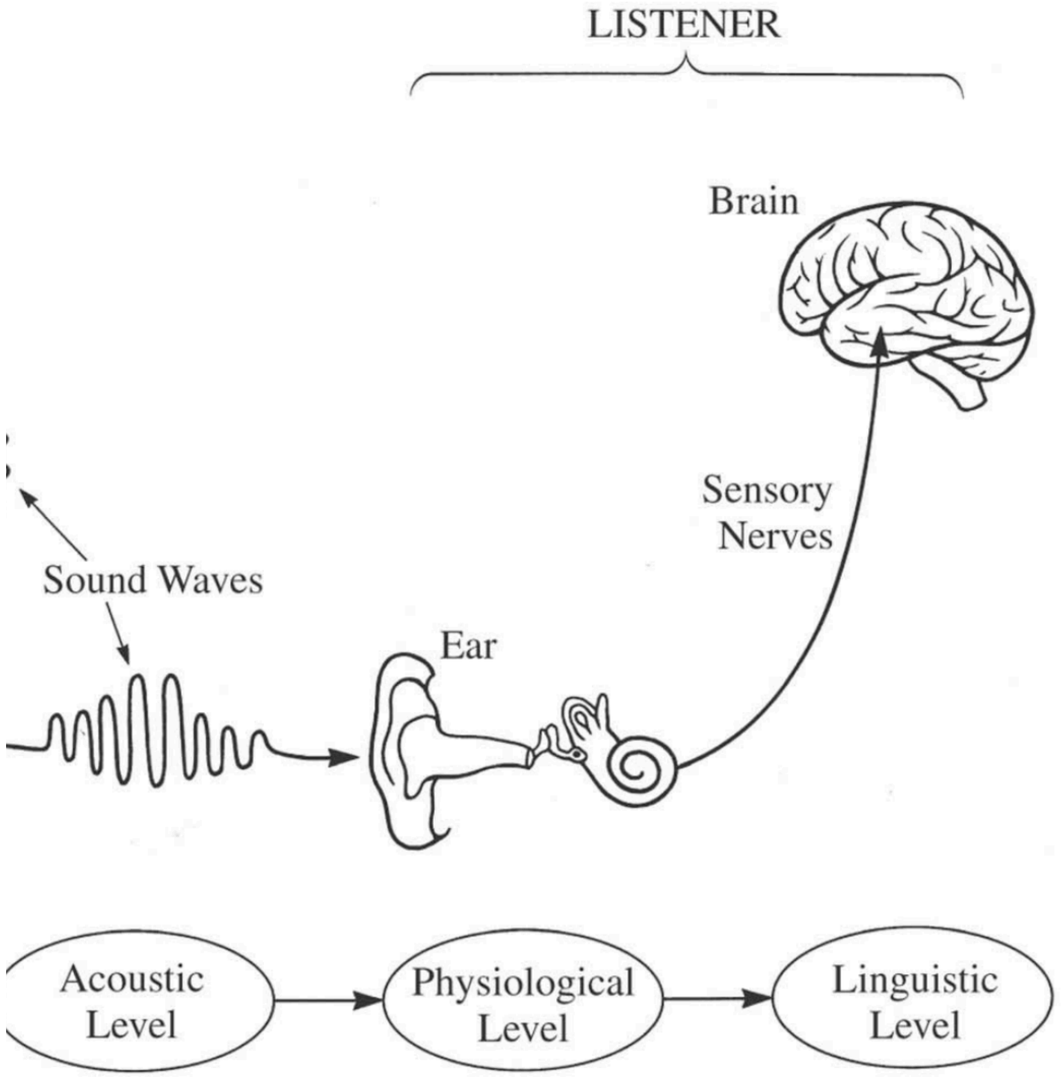
Speech chain： Acoustic level

sound is released from the lips and nostrils and transmitted to both the speaker (sound feedback) and to the listener



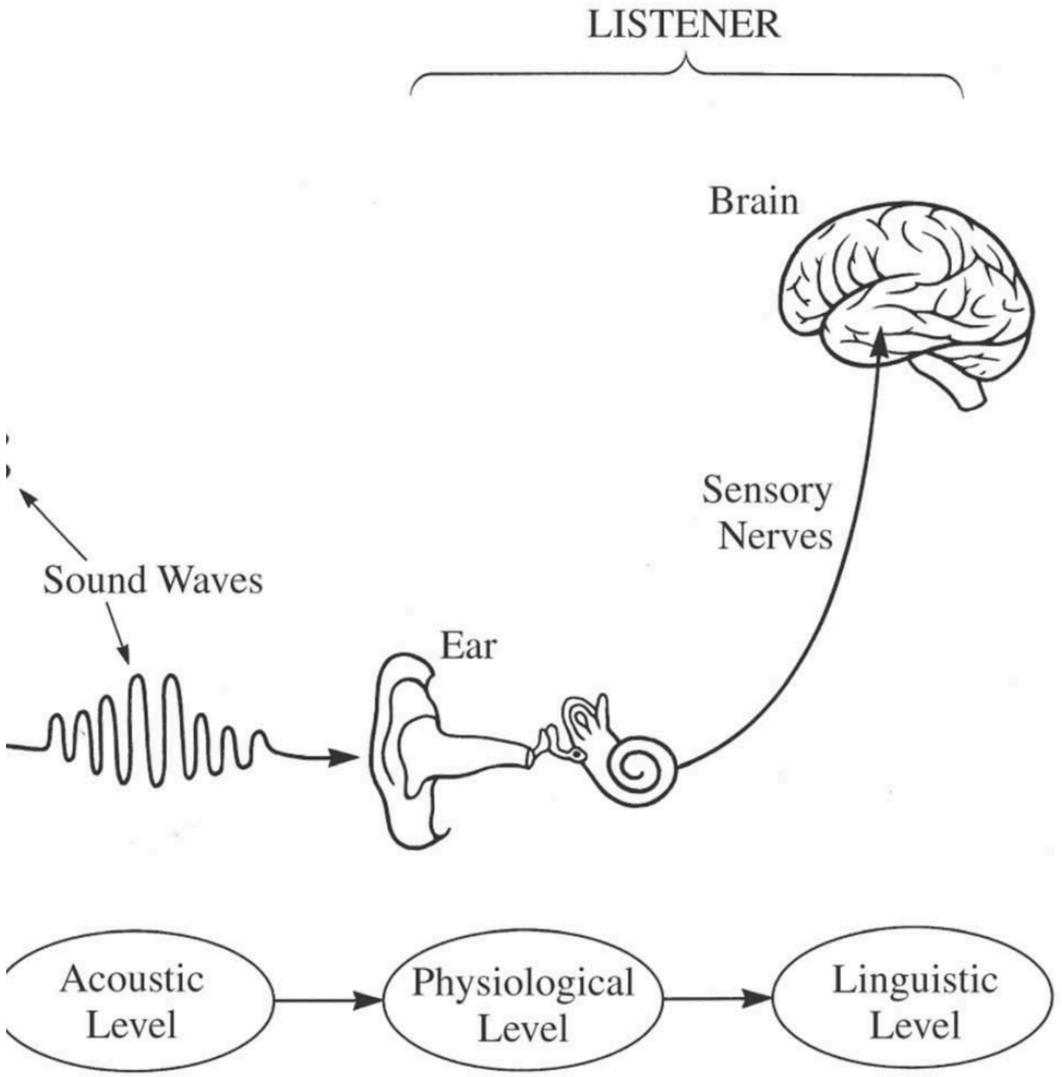
Speech chain： Physiological level

The sound is analyzed by the ear and the auditory nerves

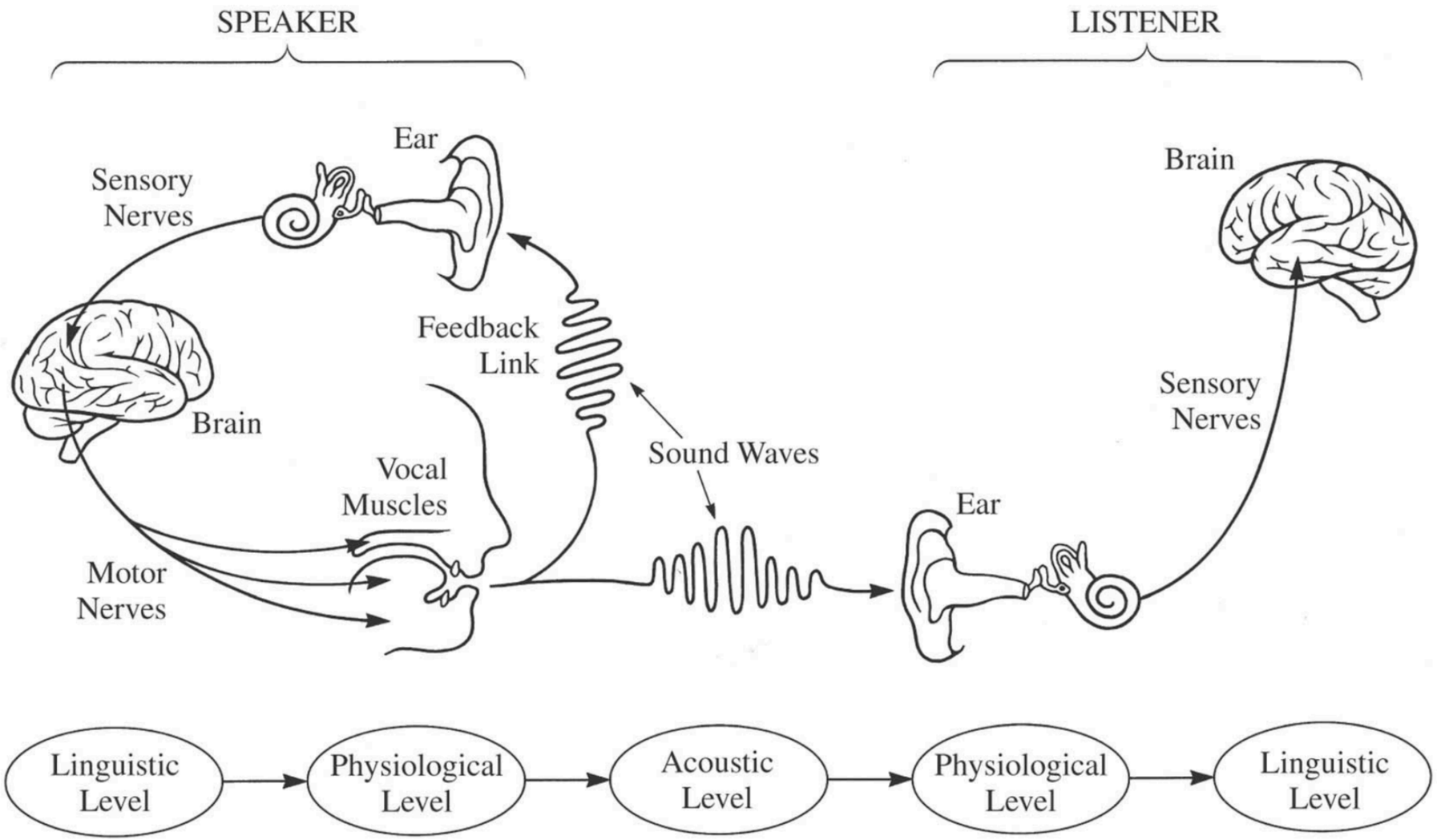


Speech chain： Linguistic level

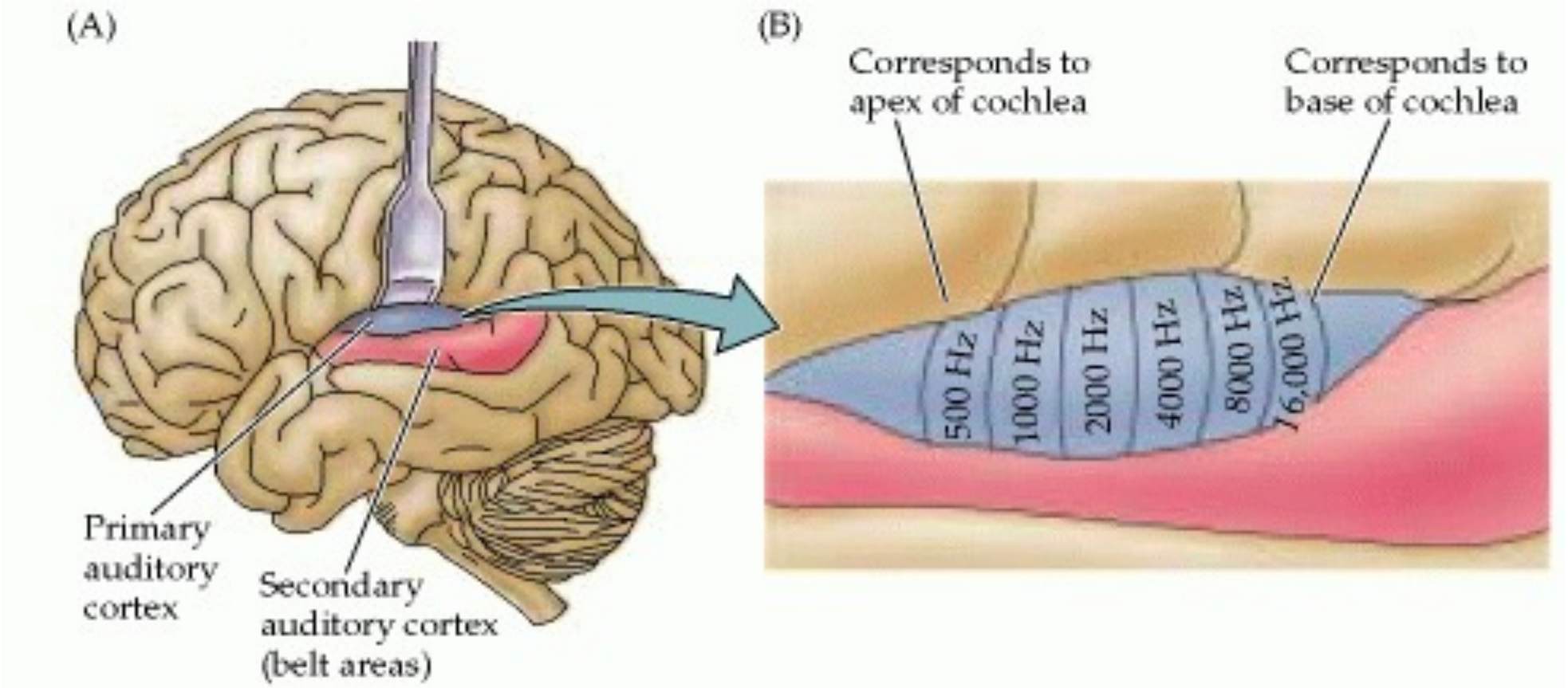
The speech is perceived as a sequence of linguistic units and understood in terms of the ideas being communicated



Speech chain

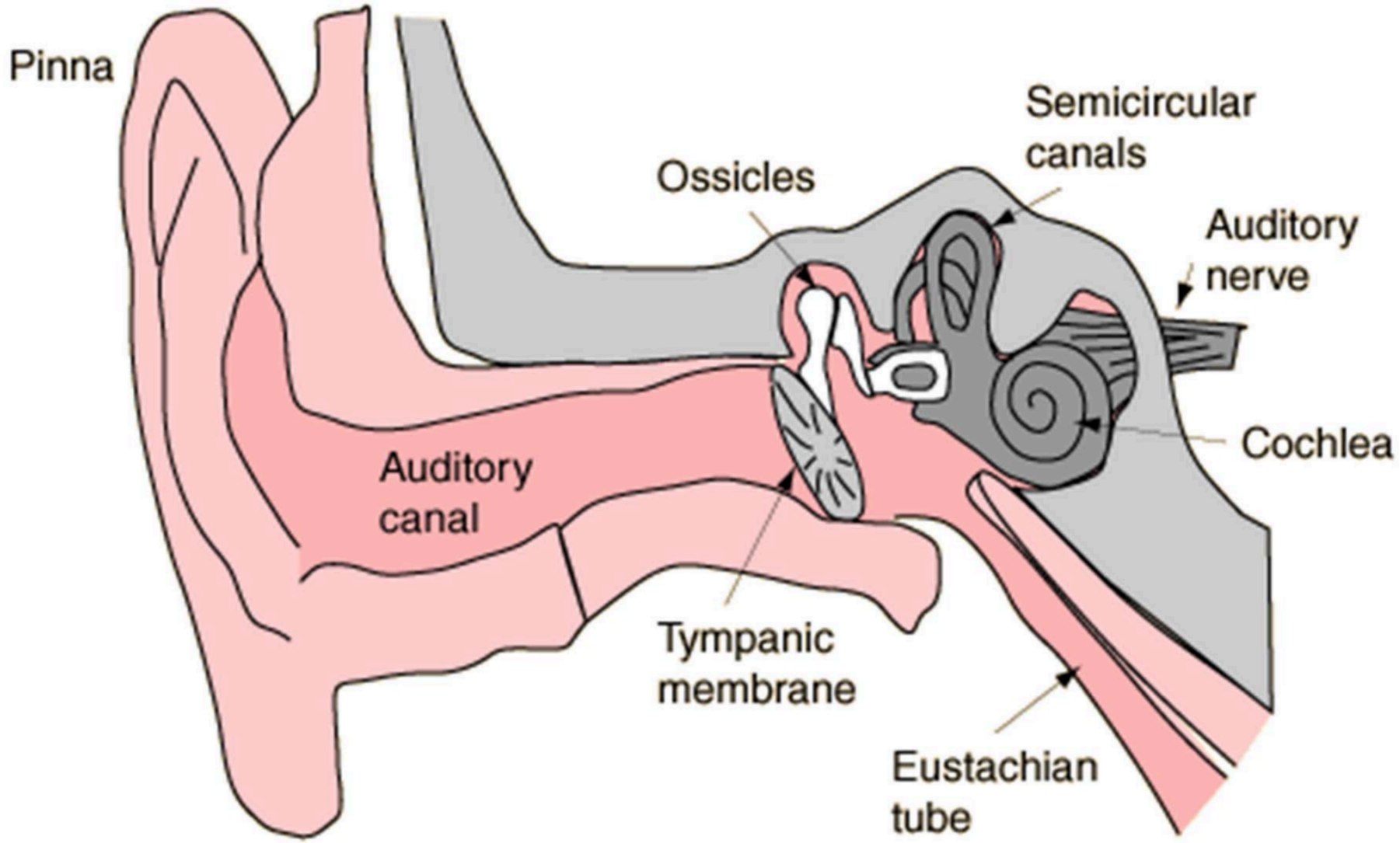


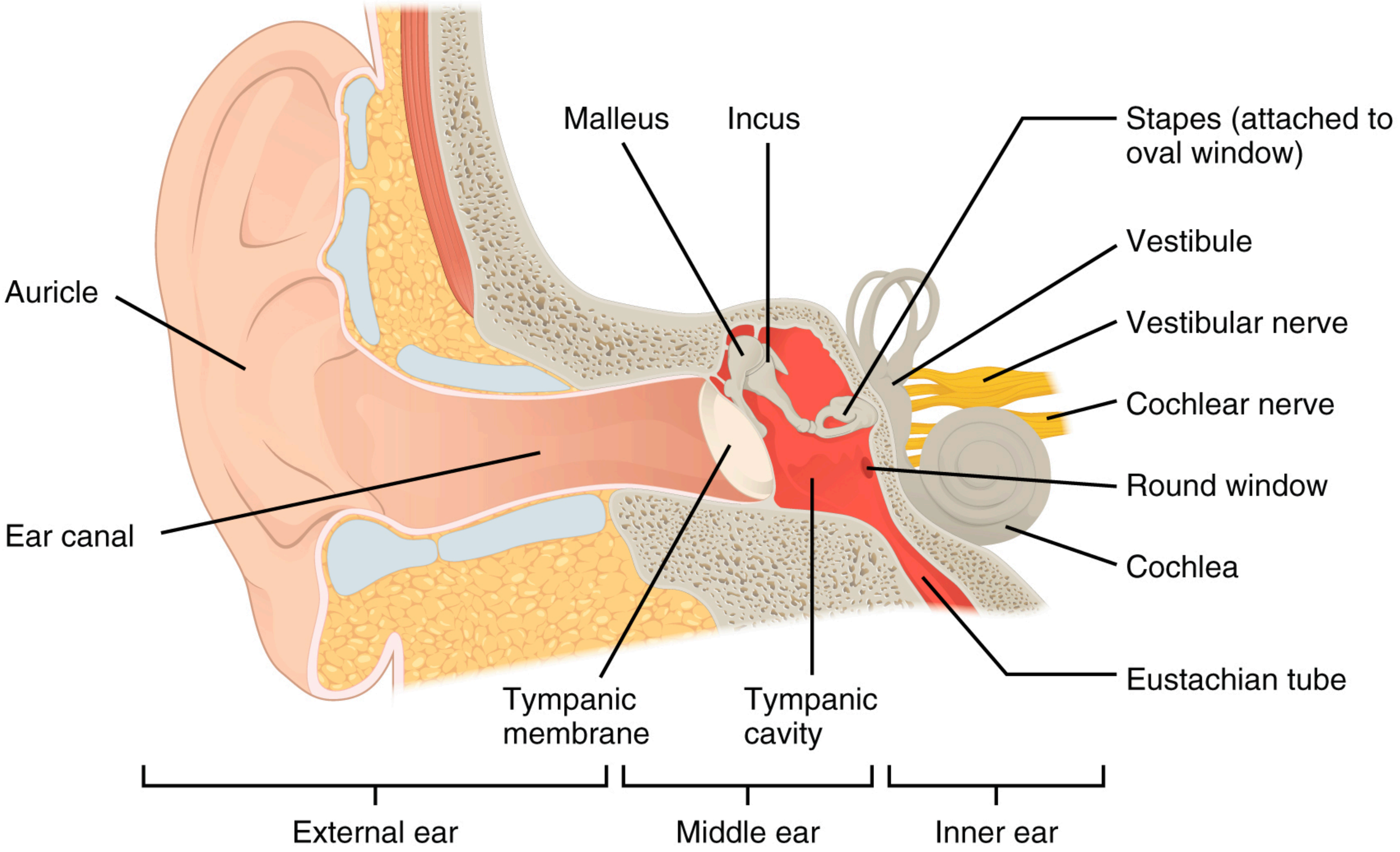
Auditory cortex



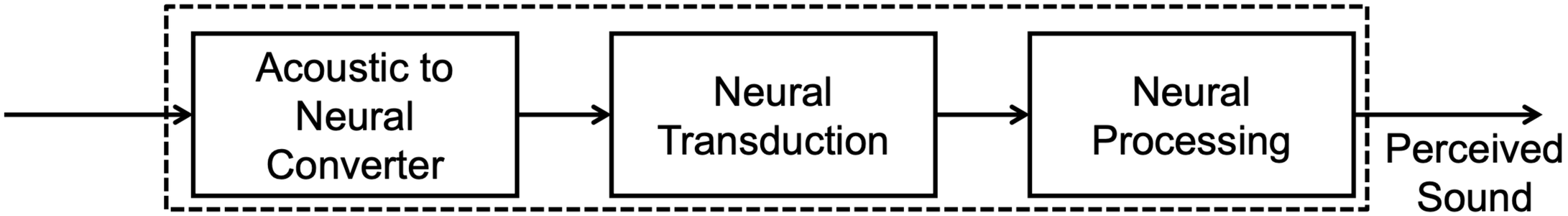
https://www.youtube.com/watch?v=WeQluIdıhnQ

Ear and Hearing

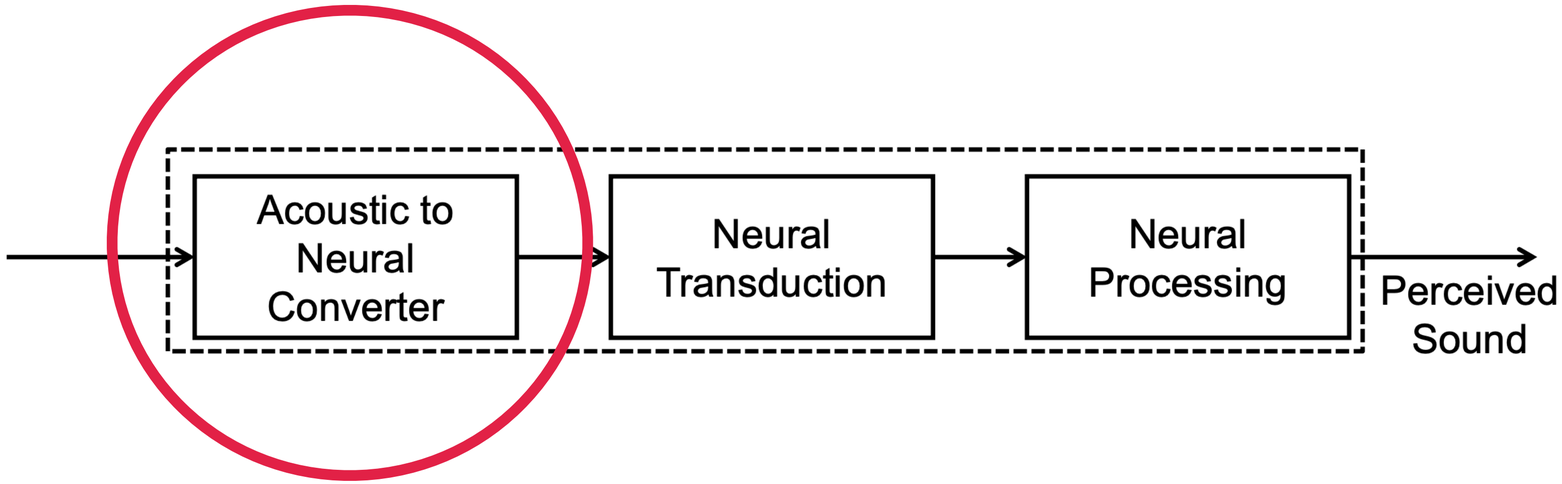



Malleus
Incus
Stapes (attached to oval window)
Vestibule
Vestibular nerve
Cochlear nerve
Round window
Cochlea
Eustachian tube
Auricle
Ear canal
Tympanic membrane
Tympanic cavity
External ear
Middle ear
Inner ear

The Auditory System

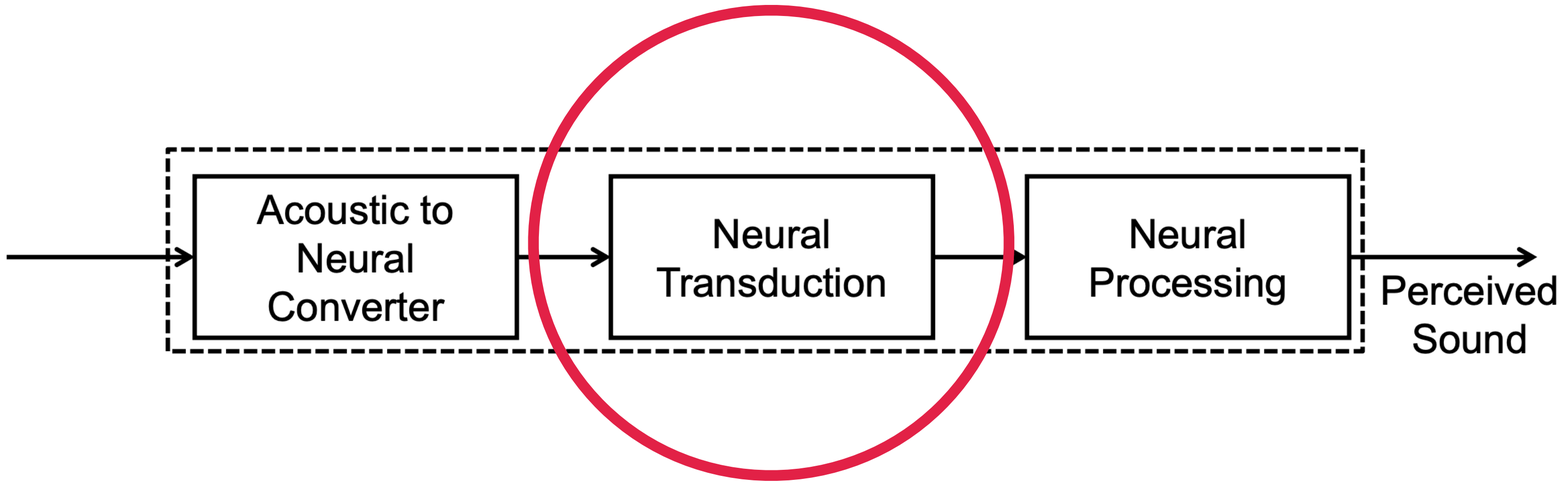


The Auditory System



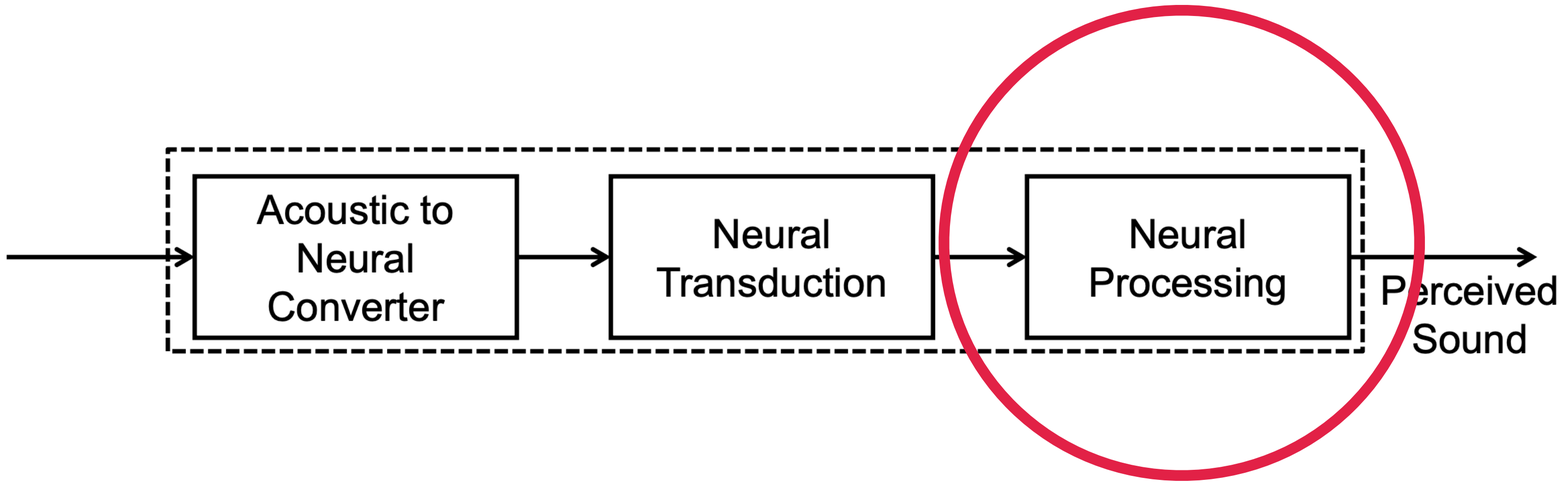
- An acoustic signal first converted to a neural representation by processing in the ear

The Auditory System



- The neural transduction takes place between the output of the inner ear and the neural pathways to the brain

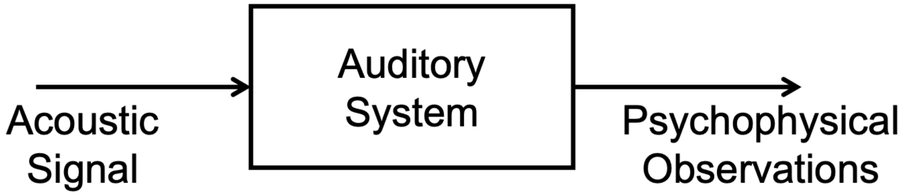
The Auditory System



- The nerve firing signals along the auditory nerve are processed by the brain to create the perceived sound corresponding to the spoken utterance

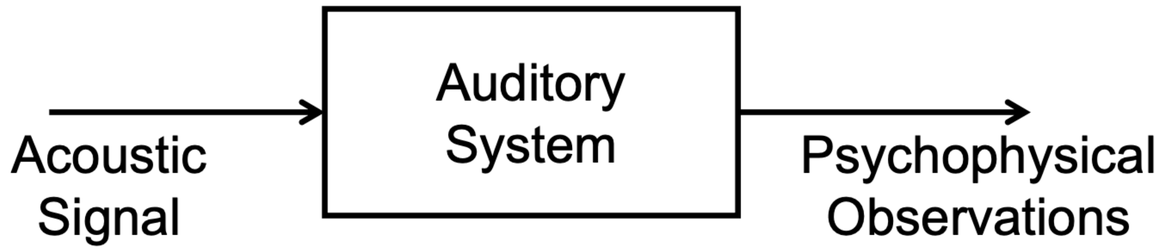
The Black Box Model of the Auditory System

- A “black box” behavioral model of hearing and perception
  - Assumption: an ***acoustic signal*** enters the auditory system ***causing behavior*** that we record as ***psychophysical observations***



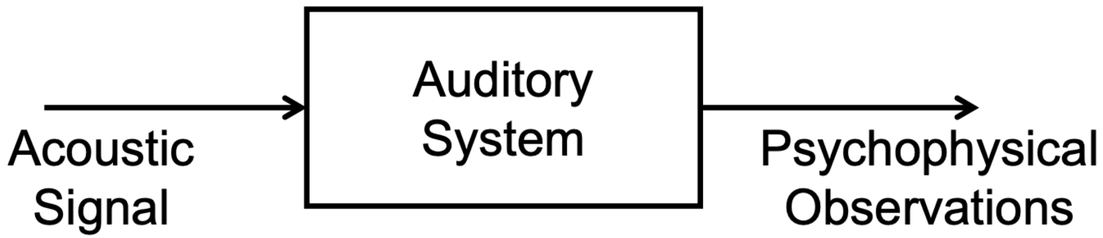
The Black Box Model of the Auditory System

- A “black box” behavioral model of hearing and perception
  - Psychophysical methods and sound perception experiments determine how the brain processes signals with
    - different loudness levels
    - different spectral characteristics
    - different temporal properties



The Black Box Model of the Auditory System

- A “black box” behavioral model of hearing and perception
  - Characteristics of the physical sound are varied in a systematic manner and the psychophysical observations of the human listener are recorded and correlated with the physical attributes of the incoming sound
  - We then determine how various attributes of sound (or speech) are processed by the auditory system



The Black Box Model Examples

Physical Attribute	Psychophysical Observation
Intensity	Loudness
Frequency	Pitch

- Experiments with the “black box” model show:
  - correspondences between sound intensity and loundess, and between frequency and pitch are complicated and far from linear
  - attempts to extrapolate from psychophysical measurements to the processes of speech perception and language understanding are, at best, highly susceptible to misunderstanding of exactly what is going on in the brain

Why do human have two ears?

- The brain needs input from both ears in order to separate sounds efficiently

- Sound localization
  - Spatially locate sound sources in 3-dimensional sound fields, based on two-ear processing, loudness differences at the two ears, delay to each ear
- Sound cancellation
  - Focus attention on a ‘selected’ sound source in an array of sound sources – ‘cocktail party effect’, Binaural Masking Level Differences (BMLDs)

Some facts about human hearing

- ***Masking*** is the phenomenon whereby one loud sound makes another softer sound inaudible
  - Masking is most effective for frequencies around the masker frequency
  - Masking is used to hide quantizer noise by methods of spectral shaping (similar grossly to Dolby noise reduction methods)

Sound Pressure Levels (dB)

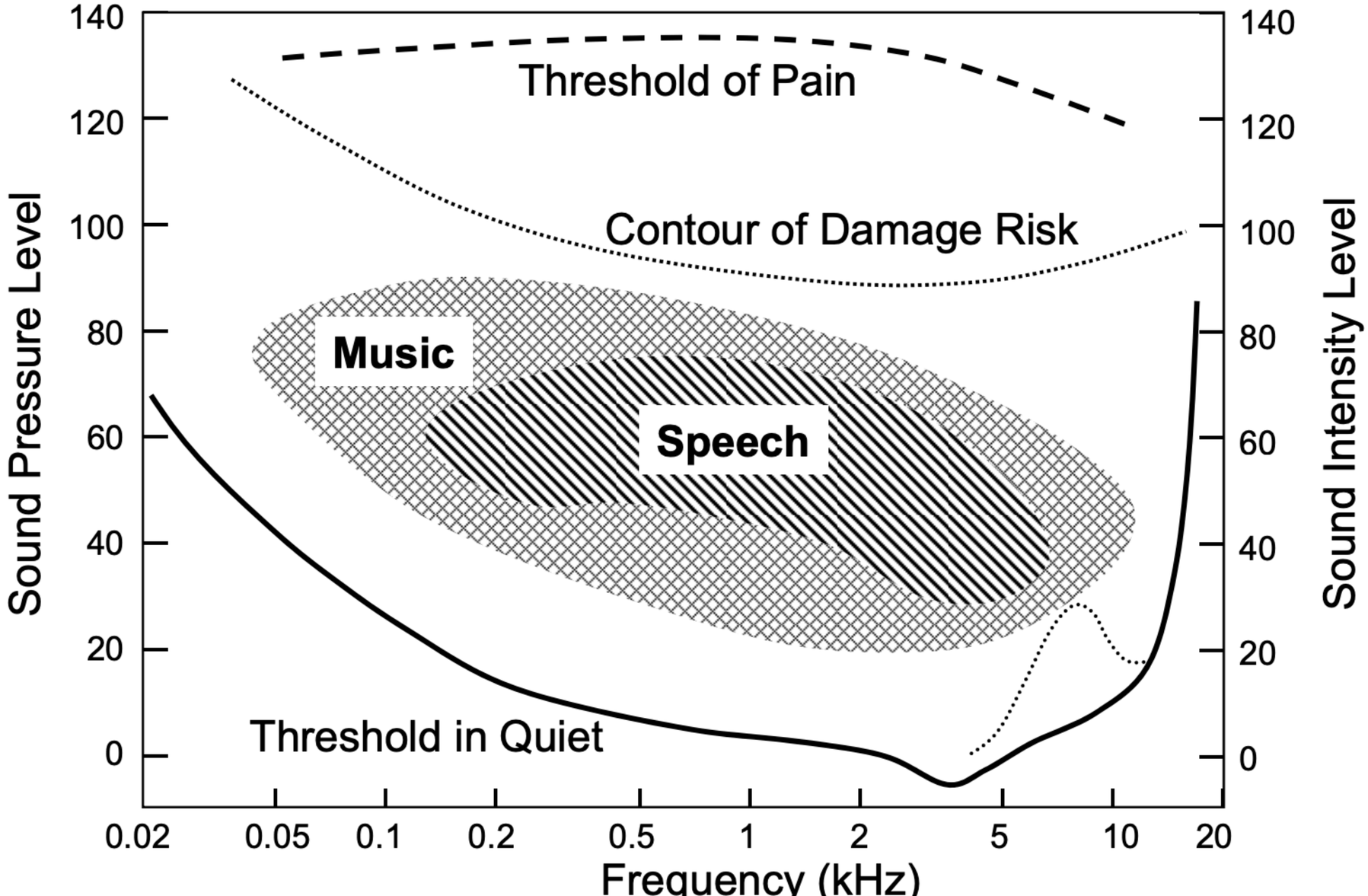
- A logarithmic measure of the effective pressure of a sound relative to a reference value

$$20 \log_{10} \left(\frac{p}{p_0} \right) \text{ dB},$$

p is the root mean square sound pressure
p_0 is a reference sound pressure

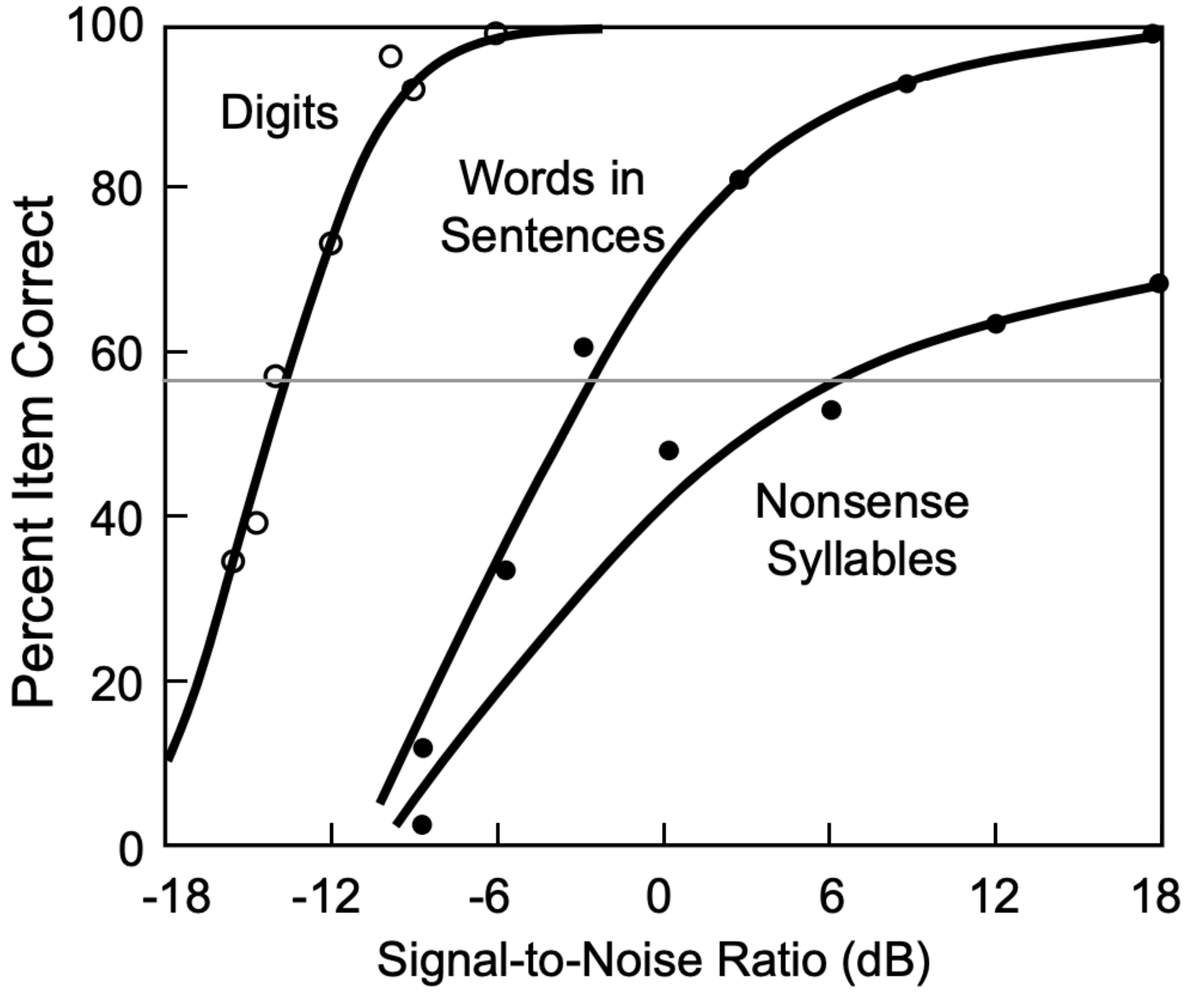
Commonly used p_0: the threshold of human hearing
(roughly the sound of a mosquito flying 3 m away)

Range of Human Hearing

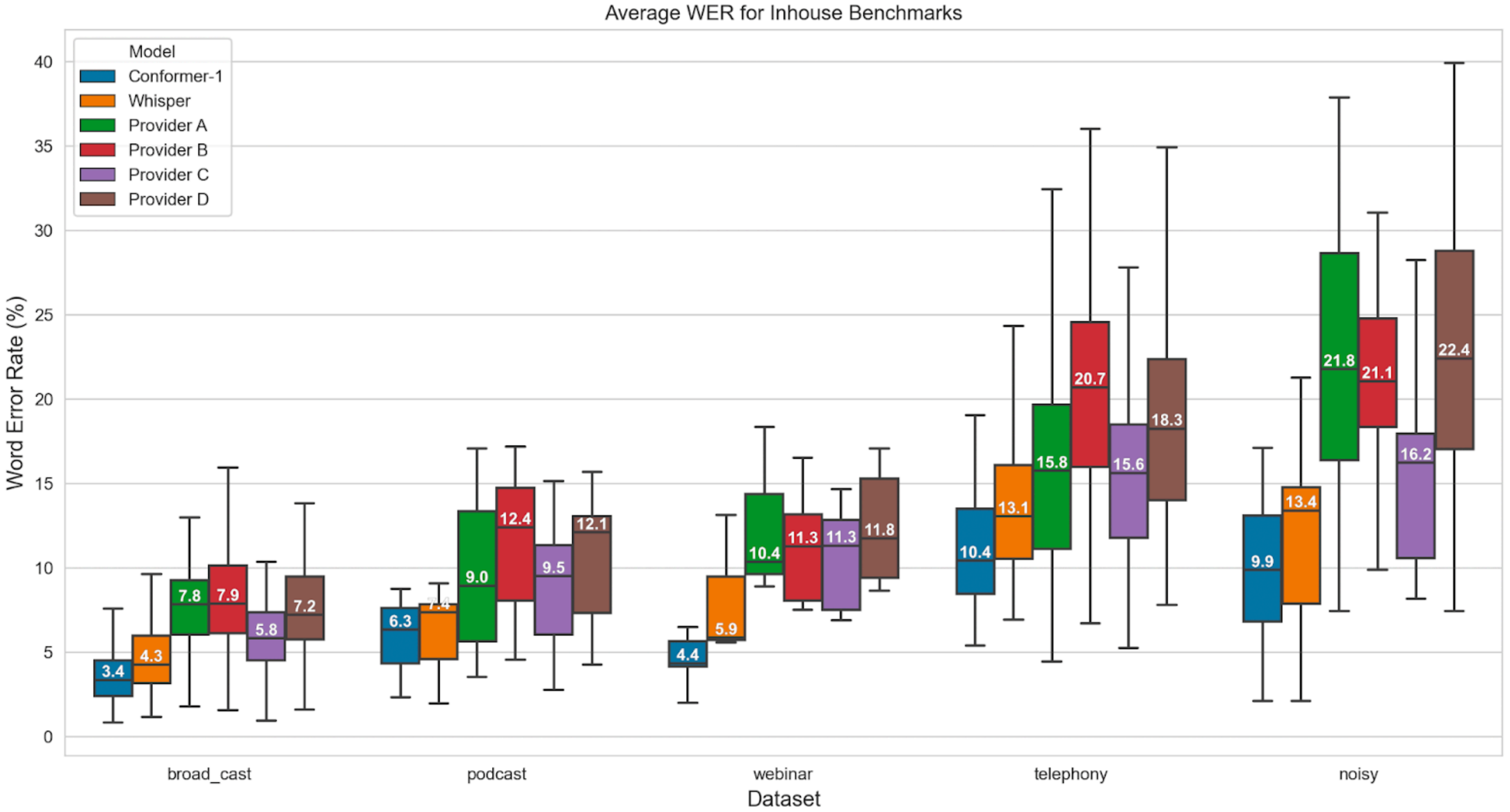


Speech perception

- Depending on multiple factors
  - Perception of individual sounds (based on distinctive features)
  - Predictability of the message
- The importance of linguistic and contextual structure cannot be overestimated



Word Intelligibility



Summary

- Speech chain
  - Linguistic level, Physiological level, Acoustic level, Physiological level, Linguistic level
- Auditory System
- Black box model of the auditory system
- Sound pressure