Outline

- Neural representation of speech sounds
  - Basic intro
  - Sounds and categories
  - How do we perceive sounds?
  - Is speech sounds special?
What is a phoneme?
It’s the basic linguistic unit of speech
Examples of phonemes?
How many phonemes in the word “cat”?
It comprises of 3 phonemes (k æ t)

k^h æ t
m æ t

{m æ t vs. k æ t} vs. {k æ t vs. k^h æ t}

To be more precise, a phoneme is the smallest contrastive speech unit that brings about a change of meaning.
- Phoneme is the basic contrastive unit of **speech**
- Language = Speech?
- How about sign languages?
Existence of signed languages tells us that speech is not a necessary modality for language: the meanings and complex structures of language can also be coded as manual gestures.

Nevertheless, no languages spoken by hearing communities are signed.

Thus in the presence of hearing, speech is the preferred modality for language.
How do we classify phonemes

- The standard classification of phonemes refers mainly to the *motor gestures* that produce them.
- For example, /x/ is Voiceless velar fricative in Arabic!
  - Place of articulation is: velar
  - Manner of articulation: *fricative
  - Voicing: voiceless (vocal folds don’t vibrate)

* Sounds that are produced by bringing the mouth into position to block the passage of the airstream, but not making complete closure so that air moving through the mouth generates audible friction.
How our brains perceive categories of sounds

- Sounds have distinctive features!
- Categorical perception helps us to distinguish between things!
- When items that range along a continuum are perceived as being either more or less similar to each other than they really are because of the way they are categorized!
### CONSONANTS (PULMONIC)

<table>
<thead>
<tr>
<th>Articulators</th>
<th>Voiced</th>
<th>Implosive</th>
<th>Ejective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilabial</td>
<td>V</td>
<td></td>
<td>E</td>
</tr>
<tr>
<td>Dental</td>
<td>V</td>
<td></td>
<td>E</td>
</tr>
<tr>
<td>Alveolar</td>
<td>V</td>
<td></td>
<td>E</td>
</tr>
<tr>
<td>Palatal</td>
<td>V</td>
<td></td>
<td>E</td>
</tr>
<tr>
<td>Posteralveolar</td>
<td>V</td>
<td></td>
<td>E</td>
</tr>
</tbody>
</table>

---

### CONSONANTS (NON-PULMONIC)

<table>
<thead>
<tr>
<th>Clicks</th>
<th>Voiced Implosives</th>
<th>Ejectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilabial</td>
<td>V</td>
<td>E</td>
</tr>
<tr>
<td>Dental</td>
<td>V</td>
<td>E</td>
</tr>
<tr>
<td>Alveolar</td>
<td>V</td>
<td>E</td>
</tr>
<tr>
<td>Palatal</td>
<td>V</td>
<td>E</td>
</tr>
<tr>
<td>Posteralveolar</td>
<td>V</td>
<td>E</td>
</tr>
</tbody>
</table>

---

### OTHER SYMBOLS

- **W**: Voiced labial-velar implosive
- **H**: Voiced labial-velar ejective
- **T**: Voiced labial-velar approximant
- **K**: Simultaneous /k/ and /x/
- **Z**: Voiced alveolar approximant
- **J**: Voiced alveolar ejective
- **Q**: Voiced alveolar trill
- **P**: Voiced palatal lateral
- **L**: Voiced palatal ejective
- **M**: Voiceless labial-velar ejective
- **S**: Voiceless labial-velar implosive
- **D**: Voiceless palatal trill
Do you know where I can find pins!

Here is the bin!
But I meant
How do we perceive sounds!

- Sound waves are first collected in our outer ear. Then they travel down our ear canal and strike the eardrum, causing it to vibrate. These vibrations are in turn transmitted to our inner ear by the bones of our middle ear (hammer, anvil, and stirrup). Our inner ear plays changes the vibrations created by the sound waves into electrical impulses, or signals, which can be recognized and decoded by our brain. Our brain receives these impulses in its hearing centers and interprets them as a type of sound.

https://www2.southeastern.edu/Academics/Faculty/mrossano/perception/powerpoints/Chapter%2010.pptx
Outer ear
Collects sound waves. The configuration of the outer ear serves to amplify sound, particularly at 2000-5000 Hz, a frequency range that is important for speech.

Middle ear
Transforms the energy of a sound wave into the internal vibrations of the bone structure of the middle ear and transforms these vibrations into a compressional wave in the inner ear.

Inner ear
Transform the energy of a compressional wave within the inner ear fluid into nerve impulses which can be transmitted to the brain.
what the ear hears!

- The healthy ear can perceive frequencies from 20 Hz to 20,000 Hz. The sound frequencies between 500 Hz and 4000 Hz are those that are most important for speech.

- **Amplitude**: height of wave; measured in dB

---

https://www2.southeastern.edu/Academics/Faculty/mrossano/perception/powerpoints/Chapter%2010.pptx
Primary auditory cortex: Brodmann areas 41 and 42.
Secondary auditory cortex: Brodmann areas 22 and 37.

http://ist-socrates.berkeley.edu/~jmp/LO2/6.html
Task of the listener: Mapping sound to meaning

- Requires converting the raw signal into the kinds of sound representations that connect to meaning.

- What are those sound representations like?
  - Sequences of phonemes!
Task of the listener: Mapping sound to meaning

- Stored sound representations must abstract away from a huge amount of physical variance in the signal
  - Loudness
  - Pitch
  - Gender
  - Speaker identity
  - Accent and other variance in pronunciation
  - Etc...
Categorization inside and outside language

- What makes a chair a chair?

- What makes a ‘t’ a ‘t’?

- Narrowing down the question:
  - What makes a ‘t’ a ‘t’, as opposed to a ‘d’?
### Distinctive features

<table>
<thead>
<tr>
<th></th>
<th>( t )</th>
<th>( d )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Major class features</strong></td>
<td>+ consonantal</td>
<td>+ consonantal</td>
</tr>
<tr>
<td></td>
<td>- sonorant</td>
<td>- sonorant</td>
</tr>
<tr>
<td></td>
<td>- syllabic</td>
<td>- syllabic</td>
</tr>
<tr>
<td><strong>Manner features</strong></td>
<td>- nasal</td>
<td>- nasal</td>
</tr>
<tr>
<td></td>
<td>- continuant</td>
<td>- continuant</td>
</tr>
<tr>
<td></td>
<td>- lateral</td>
<td>- lateral</td>
</tr>
<tr>
<td><strong>Laryngeal features</strong></td>
<td>- voiced</td>
<td>+ voiced</td>
</tr>
<tr>
<td><strong>Place of articulation features</strong></td>
<td>- round</td>
<td>- round</td>
</tr>
<tr>
<td></td>
<td>+ anterior</td>
<td>+ anterior</td>
</tr>
</tbody>
</table>
Voiced

- [+ voiced] sounds are produced with the vocal folds close together, in such a way that air passing through them causes them to vibrate.

- Touch your larynx while pronouncing [z] vs. [s] to feel that [z] is voiced and [s] voiceless.
Voicing and stop consonants

- **Stop consonants**: p, t, k, b, d, g
  - Produced by causing a complete closure of the vocal tract and then releasing it.

- Stops cannot be heard until the closure is released.

- All vowels are voiced.

- When the stop is released quickly, and the voicing of the vowel starts quickly, the stop is considered [+ voiced].

- The difference between [d] and [t] is that the onset of voicing of the following vowel takes longer for [t] than for [d].
Acoustically, voiced and voiceless stops differ in their Voice Onset Time (VOT).

- VOT: The time it takes for the voicing of the vowel to start (voicing = vibration of the vocal folds).
What do we know about the brain bases of phonological categorization/phoneme representations?
**Evidence:**

- Lesion data from patients with PWD (pure word deafness)
- FMRI studies
- MEG studies
- Auditory mismatch (ERP) studies
Lesion Data

- Early studies of PWD provided the first evidence that speech sounds are processed in the temporal lobe.

- Lesions causing specific phoneme perception impairment was distinguished from lesions causing other forms of “word deafness”, linking the former to damage in the superior temporal gyrus (STG) (Henschen, 1918-1919).

- Lesion data from patients with PWD strongly suggest some degree of bilateral phoneme processing, because the vast majority of cases have bilateral STG lesions (Buchman et al., 1986; Poeppel, 2001)

*This term was used to designate any speech comprehension impairment.
Lesion Data

- *(Wada) testing shows that patients maintain much of their ability to discriminate phonemes even after anesthetization of the language dominant hemisphere (Boatman et al., 1998; Hickok et al., 2008).

- https://youtu.be/N5_nX_LZ834
- https://youtu.be/SBKc_ncPzOo

- Data suggest that **phoneme perceptual processes are much more bilaterally represented than are later stages of speech comprehension.** Nevertheless, PWD after unilateral damage is occasionally reported, and the damage in these instances is nearly always on the left side, suggesting a degree of asymmetry in the bilateral representation (Buchman et al., 1986; Poeppel, 2001).

*Wada testing* is a test used to identify the dominant hemisphere for Language
FUNCTIONAL IMAGING STUDIES

- **Methodological limitation:**
  - The loud noises produced by rapid gradient switching present a challenge for fMRI studies of auditory processing, because these noises not only mask experimental stimuli but also activate many of the brain regions of interest.

- **Solution:**
  - separating successive image volume acquisitions with a **silent period** during which the experimental stimuli are presented. If the acquisitions are separated by a sufficient amount of time (i.e., .7 seconds), then the measured BOLD responses primarily reflect activation by the experimental stimuli rather than the noise produced by the previous image acquisition.
**FMRI Studies**

- Compared with a silent baseline, speech sounds (whether syllables, words, pseudowords, or reversed speech) activate most of the STG bilaterally (Binder et al., 2000; Wise et al., 1991).

- The activated regions include Heschl’s gyrus and surrounding dorsal STG areas that contain primary and belt auditory cortex (Rauschecker & Scott, 2009), suggesting that much of the activation is due to **low-level, general auditory processes**.
FMRI Studies

- A variety of nonspeech control stimuli (frequency-modulated tones, amplitude-modulated noise) were introduced to identify areas that might be more speech-specific (e.g., Binder et al., 2000; Zatorre, Evans, Meyer, & Gjedde, 1992).

- **Findings:** primary and belt regions on the dorsal STG respond similarly to speech and simpler nonspeech sounds, whereas areas in the ventral STG and superior temporal sulcus (STS) respond more to speech than to simpler sounds.

- The pattern suggests a hierarchically organized processing stream running from the primary auditory cortex on the dorsal surface to higher level areas in the lateral and ventral STG.
FMRI Studies

- In contrast to the data from Wada testing and patients with PWD, the fMRI studies suggest that, at least in the intact brain, the **left auditory system** performs most of the computation related specifically to phoneme perception.
MEG Studies

An auditory evoked neuromagnetic field

M100 (N1 in ERP terms)

- The auditory N100 is generated in the primary and association auditory cortices in the superior temporal gyrus and in Heschl's gyrus bilaterally (Zouridakis et al., 1998).

- It has been found to be sensitive to things such as the predictability of an auditory stimulus, and special features of speech sounds such as voice onset time.

For example:

- Frequency
  - higher frequency → shorter latency
ERP Studies

- Event-related potentials data were obtained simultaneously with fMRI, showing that the difference between phonemic and nonphonemic processing begins at approximately 180 ms after stimulus onset and peaks at approximately 200 ms, manifesting as a larger P2 deflection in the phonemic condition (Liebenthal et al., 2010).

- P2 component may reflect the sensation-seeking behavior of an individual, and recently it’s associated with attention and working memory.

- Inconsistent findings suggest different effects of intensity on the amplitudes of the auditory N1 and P2 (Paiva et., 2016).
Questions

- Is speech special?
  - Can brain damage lead to a selective loss of speech perception?
  - Are the mechanisms of speech perception different from the mechanisms of auditory perception in general?
  - Role of audio-visual integration
Disorders of auditory processing

- Rare.
- Signal from each ear is processed in both hemispheres (contra visual system).
Disorders of auditory processing:  
CORTICAL DEAFNESS

- Inability to hear sounds without apparent damage to the hearing apparatus or brain stem abnormality.
- Extensive bilateral damage to auditory cortex (BAs 41 & 42).
Disorders of auditory processing: AUDITORY AGNOSIA

- Inability to recognize auditorily presented sounds (e.g., coughing, crying) independent of any deficit in processing spoken language.
- Damage in auditory association cortex (BAs 22 & 37)
Disorders of auditory processing:
AMUSIA (a subvariety of AUDITORY AGNOSIA)

- The inability to recognize musical tones or to reproduce them (appears mainly as a defect in processing pitch)

- Relative sparing of speech and (other) non-speech perception.

- Damage generally in right temporal areas.
Disorders of auditory processing:

PURE WORD DEAFNESS

- Inability to understand spoken words while
  - auditory perception is otherwise intact &
  - other linguistic skills are intact (reading, speech production)
Disorders of auditory processing: 
PURE WORD DEAFNESS

Either
- Bilateral damage to auditory cortex

Or
- A subcortical lesion in the left hemisphere that severs both ipsilateral and contralateral projections to Wernicke’s area (Ziegler, 1952; Geschwind, 1965; Coslet et al., 1984; Jones and Dinolt, 1952).
**Disorders of auditory processing:**

**PHONAGNOSIA**

- Impairment in the ability to recognise familiar voices.
  - Speech comprehension is intact.
  - Intact ability to identify nonverbal sounds.

- Phonagnosia was first described by a study by Van Lancker and Cantor in 1982
Van Lancker et al., 1988: Double dissociation between voice recognition and voice discrimination. Some patients will perform normally on the discrimination tasks but poorly on the recognition tasks; whereas the other patients will perform normally on the recognition tasks but poorly on the discrimination tasks.

Performance on a discrimination task was impaired by a lesion to either temporal lobe.

Performance on a voice recognition task was impaired by lesions to the right parietal lobe.
Questions

- Is speech special?
  - Can brain damage lead to a selective loss of speech perception? **YES, but no one-to-one correspondence between lesion site and type of disorder. E.g., bilateral temporal damage can lead either to cortical deafness or pure word deafness.**
  - Are the mechanisms of speech perception different from the mechanisms of auditory perception in general?
    - Role of audio-visual integration
Questions

- Is speech special?
  - Can brain damage lead to a selective loss of speech perception? YES, but no one-to-one correspondence between lesion site and type of disorder. E.g., bilateral temporal damage can lead either to cortical deafness or pure word deafness.
  - Are the mechanisms of speech perception different from the mechanisms of auditory perception in general?
    - Role of audio-visual integration
Close your eyes, play this movie, and listen. Open your eyes, replay and listen again. Is he saying "ba ba" or "da da"?

https://youtu.be/aFPtc8BVdJk
What is the McGurk effect?

- The **McGurk effect** is a perceptual phenomenon that demonstrates an interaction between hearing and vision in speech perception.

- The illusion occurs when the auditory component of one sound is paired with the visual component of another sound, leading to the perception of a third sound.
Role of articulatory gestures on speech perception

- When the visual image of a person saying “ga” is combined with an audio recording of “ba”, the percept is “da” (or for some people “ga”).

- Contrasting articulatory gestures can change the auditory percept even when the signal is clear.
Is audio-visual integration specific to speech?

  - Audio–visual integration of the “plucks” and “bows” of cello playing.
    - Not only speech, but also other ecologically valid combinations of auditory and visual stimuli can integrate in a complex manner.

- Tuomainen et al. (2004, Cognition): Audio-visual integration is used in speech much more than in the perception of other sounds.
  - Investigated the effect of incongruent articulatory gestures on speech vs. nonspeech, utilizing the ambiguous status of so-called *sine-wave speech.

*sine-wave speech is a technique for synthesizing speech by replacing the formants (main bands of energy) with pure tone whistles. [http://www.lifesci.sussex.ac.uk/home/Chris_Darwin/SWS/]
Questions

- Is speech special?
  - Can brain damage lead to a selective loss of speech perception? YES, but no one-to-one correspondence between lesion site and type of disorder.
  - Are the mechanisms of speech perception different from the mechanisms of auditory perception in general?
    - This is a methodologically hard question since it’s unclear what could be a control stimulus that is physically highly similar to speech but is not speech.
    - But the top-down effects in speech perception make it possible to use ambiguous stimuli that may or may not be perceived as speech.
      - Perceiving a stimulus as speech can enhance activity around auditory cortex.
      - Role of audio-visual integration may be special for speech.
Further investigations!

- Adequate “complex” controls!
  - The tone and noise controls used in the studies are far less acoustically complex than speech sounds!!!

- The degree to which speech perception requires “special” mechanisms!