How our brains perceive categories of sounds
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?
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…

Task of the listener: Mapping sound to meaning
Categorization inside and outside language

- What makes a chair a chair?

- What makes a ‘t’ a ‘t’?
  - For the speaker?
  - For the listener?

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

<table>
<thead>
<tr>
<th>Major class features</th>
<th>$t$</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td></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>Manner features</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>Laryngeal features</td>
<td>- voiced</td>
<td>+ voiced</td>
</tr>
<tr>
<td>Place of articulation features</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.
VOT continuum for ta-da

60 msec

English boundary
Categorical perception of t-d

Eimas & Corbitt: (1973): Perception of synthesized sounds with VOTs increasing in equal steps from 0 to 80ms.

Listeners to judge whether they hear a [da] or [ta].
Within category discrimination is poor.
How do we learn to group sounds in this way? What is it good for?
A change within these VOTs … or within these VOTs is never associated with a change in meaning in English. So being able to discriminate between 10 and 20ms VOTs or 60 and 70ms VOTs won’t help you understand English.
Different words

boundary

“doll”

Short VOT

“tall”

Long VOT

Could not be different words
Phoneme

- A category of sounds.

- A mental abstraction over a group of sounds within which interchanging one token for another can never induce a change in meaning.

- Since going from /t/ to /d/ in English can induce a change in meaning, /t/ and /d/ are distinct phonemes in English.
- Requires converting the raw signal into the kinds of sound representations that connect to meaning.

  - What are those sound representations like?

  - Sequences of phonemes!
What do we know about the brain bases of phonological categorization?
An auditory evoked neuromagnetic field

M100 (N1 in ERP terms)

- Generated in auditory cortex bilaterally.

Affected by:
- Intensity
  - higher intensity $\rightarrow$ shorter latency, larger amplitude
- Frequency
  - higher frequency $\rightarrow$ shorter latency
- Phonetic identity
  - e.g., shorter latencies for a than for u (explainable in spectral terms)

*Function of M100 activity in speech processing unclear.*
Auditory Mismatch Field

- A change detector.
  - Not part of the basic processing of sounds.
Auditory Mismatch Field

- Elicited in auditory cortex at ~180ms post stimulus onset by deviant stimuli in an oddball paradigm.

XXXYXXXXXYXXXXXYXXX

X = “standard”
Y = “deviant”
Auditory Mismatch Field

- A tool for investigating what counts as a change for auditory cortex.
- Has been heavily used in the study of the neural bases of categorical perception.
Mismatch field as a function of frequency change

Sams et al. 1985
The Mismatch Field and Phonemic Categorization

Would crossing a phonemic boundary elicit a mismatch field?

If yes, this would tell us

- that auditory cortex has access to phonemic categories
- that category information is accessed by 180ms.
Voice Onset Time (VOT)

60 msec
Phillips et al. (2000, *Journal of Cognitive Neuroscience*)

**Acoustic representation**

**Standards**

**Deviants**

**Perceptual boundary**
**Phillips et al.** (2000, *Journal of Cognitive Neuroscience*)

**Phonological representation**
If the auditory cortex is sensitive to phonological categories, deviant Ts should elicit a Mismatch Field.
What should happen if we keep everything else the same, but lift all the VOTs by 20ms?
The deviants should not elicit a Mismatch Field -- they are not perceived as different from the standards.
Phillips et al. (2000, *Journal of Cognitive Neuroscience*)

- Experiment 1: standards fall on one side of a perceptual boundary, the deviants on the other.

- Experiment 2: the physical distance between standards and deviants is as in Exp. 1, but now the standards and deviants are no longer differentiated by a perceptual boundary.

- If a Mismatch Field elicited in Exp 1 is driven by the category difference between standards and deviants, it should disappear in Exp 2.
Phillips et al. (2000, *Journal of Cognitive Neuroscience*)

- A Mismatch Field is elicited only when standards fall on one side of a perceptual boundary, the deviants on the other.
- Auditory cortex has access to phonological category information at 180ms (shortly after the M100).

- In this example, category membership depended on voicing.
- Different stop consonant have different perceptual boundaries for voicing.
- If the mismatch response is sensitive to voicing, we should be able elicit it just for that feature, even across phonemic categories.
Multiple places of articulation

[pæ tæ tæ kæ dæ pæ kæ tæ pæ kæ bæ tæ pæ kæ gæ]

(Phillips, Pellathy & Marantz 2000)
Multiple places of articulation

\[ \text{pæ tæ tæ kæ dæ pæ kæ tæ pæ kæ bæ tæ pæ kæ gæ} \]

- - - - [+voi] - - - - - [+] - - - [+voi]

- Voiceless phonemes are in many-to-one ratio with [+voice] phonemes.
- No other many-to-one ratios in this sequence

(Phillips, Pellathy & Marantz 2000)
Multiple places of articulation

\[ p \ddot{\alpha} \varepsilon \ t \ddot{\alpha} \ e \ k \ddot{\alpha} \quad d \ddot{\alpha} \ e \ p \ddot{\alpha} \ e \ k \ddot{\alpha} \ e \ t \ddot{\alpha} \ e \ p \ddot{\alpha} \ e \ k \ddot{\alpha} \ e \ g \ddot{\alpha} \]

\[ \text{[+voi]} \quad \text{[+voi]} \quad \text{[+voi]} \]

(Phillips, Pellathy & Marantz 2000)
Stimuli

Acoustic/Phonetic Representation
<table>
<thead>
<tr>
<th>Stimuli</th>
<th>- VOICE</th>
<th>+ VOICE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>42 ms</td>
<td>54 ms</td>
</tr>
<tr>
<td></td>
<td>25 ms</td>
<td>30 ms</td>
</tr>
</tbody>
</table>

**Phonological Representation**
<table>
<thead>
<tr>
<th>Stimuli</th>
<th>VOT</th>
<th>Acoustic/Phonetic Representation</th>
<th>Phonological Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>K</td>
<td></td>
<td></td>
</tr>
<tr>
<td>42 ms</td>
<td>54 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27 ms</td>
<td>27 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24 ms</td>
<td>25 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 ms</td>
<td>12 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 ms</td>
<td>18 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 ms</td>
<td>15 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>09 ms</td>
<td>09 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>00 ms</td>
<td>05 ns</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>00 ns</td>
<td>09 ns</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>K</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 ms</td>
<td>54 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 ms</td>
<td>09 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 ms</td>
<td>15 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>00 ns</td>
<td>09 ns</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ VOICE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Mismatch response elicited by voicing

[Diagram showing brain activity with graphs and labels such as 'Left Anterior', 'Right Anterior', 'Left Posterior', 'Right Posterior', 'Phonological Feature Condition', 'Field Strength', 'Time (ms)', 'MMF', 'Generator of MMF', 'Standards', 'Deviants']
Left laterality of language related Mismatch Field

- Nonlinguistic sounds elicit a bilateral Mismatch Field.
  - I.e., both auditory cortices are “surprised.”

- Crossing a phoneme boundary in an oddball task elicits a Mismatch Field in the left hemisphere only.
  - The left hemisphere auditory cortex is specialized for phoneme perception.
Mismatch Summary

- By 180ms, auditory cortex has extracted from the input the relevant features needed for categorizing sounds into phonemes.

- Since the Mismatch field is a surprise response, the necessary computational steps needed for categorization must occur prior to 180ms. How exactly this happens, we don’t currently know.