Levels of Representation

The perception of speech involves multiple levels of abstraction away from the measurable physical world.

What you “see” (in Praat, for example) is not necessarily what you “get” (perceive).

Different levels of representation: MUSIC

- **Acoustic**
  - Concerns the energy produced, shape of the sound wave

- **Phonetic**
  - What you actually pronounce, described in terms of the mechanics of articulation
  - Segments are ‘phones’
  - Notation: written inside slashes /p/

- **Phonological**
  - The mental representation of the sound segments that make up words
  - Smallest segments that can change meaning are ‘phonemes’
  - Notation: written inside square brackets [p]
  - Larger units: syllables

Different levels of representation: SPEECH SOUNDS

- **Acoustic**
- **Phonetic**
- **Phonological**
We perceive categories where there is continuous variation physically

Sometimes we perceive a /d/ sound as a [d], other times as a [t]

Sometimes we pronounce a /t/ sound as a [t], other times as a [d]

1 phoneme /t/ (sound stored in the brain)

1 phoneme 2 allophones /t/ [d]
Thai vs. English

Aspirated
Eng [spʰaːt] (‘spot’)
Thai [pʰaa] (‘forest’)

Unaspirated
Eng [spʰat] (‘spot’)
Thai [paa] (‘split’)

There are at least 2 ways to pronounce a /p/ in English

/p/

[p]  [pʰ]

But these 2 pronunciations don’t affect meaning

sound(s) actually produced
There are at least 2 ways to pronounce a /p/ in English.

Abstract unit: the PHONEME

This is a single mental category for English speakers.

Allophones of the phoneme:

/p/ [p] [pʰ]

Aspiration is NOT a distinctive feature in English.

Thai is different.

2 different PHONEMES in Thai.

These are 2 different mental categories for Thai speakers.

/p/ [p] [pʰ]

Aspiration IS a distinctive feature in Thai.

Thai is different.

These are 2 different mental categories for Thai speakers.

/p/ [p] [pʰ]

Stored	Produced

English /l/ [l] /r/ [r]

Korean /l/ [l] /r/ [r]
**Voiceless (oral) stops**

<table>
<thead>
<tr>
<th>Oral stop</th>
<th>Bilabial</th>
<th>Maxillary Dental</th>
<th>Inter-dental</th>
<th>Alveolar</th>
<th>Palatal</th>
<th>Velar</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td></td>
<td>T</td>
<td>D</td>
<td>K</td>
<td>G</td>
<td>(?)</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td>N</td>
<td>J</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td></td>
<td>F</td>
<td>θ</td>
<td>S</td>
<td>Ž</td>
<td>(h)</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td>Δ</td>
<td>Z</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td></td>
<td></td>
<td>€</td>
<td>J</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td></td>
<td></td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Developmental Differentiation**

How we perceive speech depends on whether we think we are listening to speech.

- **Universal Phonetics**
- **Native Lg. Phonetics**
- **Native Lg. Phonology**

0 months | 6 months | 12 months | 18 months

**Different levels of representation are useful for different purposes**

- At what level of representation are words stored in memory?
- At what level of representation does our auditory cortex FIRST encode language sounds?
- At what level of representation do newborns perceive language, and how does it change throughout their first year and a half?
Adults perform better at non-native contrasts if they think the sounds are not language sounds!

strated. In one of the more intriguing demonstrations, it was shown that if the critical acoustic information in the speech contrast is presented alone so that the syllables no longer sound like speech, adults can discriminate nonnative contrasts. To illustrate, Werker and Tees (1984b) presented adult English speakers with either the ejective portion alone from the Nihlakamp /k’i/-/q’i/ contrast (telling listeners that it was water dropping into a bucket and that they should signal when the bucket was switched) or a truncated portion of the Hindi retroflex/dental (/Ta/-/ta/) contrast. In each case adult English speakers discriminated the shortened pairs with ease, but they still failed to distinguish the full syllables even when tested on them immediately after being presented with the shortened portions. Furthermore, adults can be “taught” to discriminate the full

Japanese phonotactic constraints

Japanese doesn’t allow any consonant clusters!

Japanese syllable structure: (C) V (C)

epenthesis

[Du] [u]

m e k d o n a l d
Japanese syllable structure: (C) V (Č)

epenthesis

[u]

(CV CV CV CV)
məku donald

Phonemic Level: /məkdonald/

Phonetic Level: [məku donaludo]

Japanese syllable structure: (C) V (Č)

epenthesis

[o]

(CV CV CV CV)
məku donalud

Do Japanese people pronounce McDonald’s this way because that’s how they perceive the word?

Or can they actually hear the difference?
Behavioral Results

- Japanese have difficulty hearing the difference

![Graph showing perceptual judgments as a function of vowel duration](image1.png)

Dupoux et al. 1999

What level of representation does the brain use at the earliest moments of auditory processing of speech?

Additional Finding

<table>
<thead>
<tr>
<th>Language</th>
<th>RT (ms)</th>
<th>SE</th>
<th>Err</th>
<th>RT (ms)</th>
<th>SE</th>
<th>Err</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vowel Length Contrast</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eba-ebato</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japanese</td>
<td>1082</td>
<td>45</td>
<td>7.5%</td>
<td>1187</td>
<td>75</td>
<td>3.2%</td>
</tr>
<tr>
<td>French</td>
<td>1173</td>
<td>73</td>
<td>21%</td>
<td>1002</td>
<td>54</td>
<td>5.8%</td>
</tr>
</tbody>
</table>

Table 1: Mean reaction time (ms), standard error, and error rate in ABX judgments on an epenthesis contrast and a vowel length contrast in French and Japanese participants (Experiment 3).

Electroencephalography (EEG)
Magnetoencephalography (MEG)

Brain Magnetic Fields (MEG)

Brain magnetic fields recorded fully non-invasively by arrays of SQUID detectors

Event-Related Potentials (ERPs)

Mismatch Response

Brain magnetic fields recorded fully non-invasively by arrays of SQUID detectors

John is laughing.
Mismatch Response

Latency: 150-250 msec.
Localization: Supratemporal auditory cortex
Many-to-one ratio between standards and deviants

Localization of Mismatch Response

MMN Amplitude Variation

MMN as a Function of Frequency Change

(Phillips, Pellathy, Marantz et al., 2000)

Sams et al. 1985
Different Dimensions of Sounds

- Length
- Amplitude
- Pitch
- …you name it …

Amplitude of mismatch response can be used as a measure of perceptual distance.

Impetus for Language Studies

- If MMN amplitude is a measure of perceptual distance, then perhaps it can be informative in domains where acoustic and perceptual distance diverge…

A Category

Another Category

3

III
Another Category

[t]

Another category:
Voiceless stops?

<table>
<thead>
<tr>
<th>Voiceless (oral) stops</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral stop:</td>
</tr>
<tr>
<td>p</td>
</tr>
<tr>
<td>nasal stop</td>
</tr>
<tr>
<td>Incitative</td>
</tr>
<tr>
<td>Affricate</td>
</tr>
<tr>
<td>Liquid</td>
</tr>
<tr>
<td>Glide</td>
</tr>
</tbody>
</table>

Is aspiration predictable in English? If so, how?

[pʰt]
[spit]
[lid]
[tʰæk]
[stæk]
[ræt]
[kʰæt]
[skæt]
[bæk]
### Aspiration only occurs with /p, t, k/

| /pʰt/ | /lɪd/ |
| /tʰæk/ | /stæk/ |
| /kʰæt/ | /bæk/ |

### Even with /p, t, k/, aspiration only occurs in word-initial position

| /pʰt/ | /spɪt/ |
| /tʰæk/ | /stæk/ |
| /kʰæt/ | /skæt/ |

### Aspiration is completely predictable in English:

- When /p/ is word-initial, it is pronounced [pʰ]
- When /t/ is word-initial, it is pronounced [tʰ]
- When /k/ is word-initial, it is pronounced [kʰ]

### How do we represent these “rules” of English pronunciation?

- **3 independent facts that must be memorized?**

  - When /p/ is word-initial, it is pronounced [pʰ]
  - When /t/ is word-initial, it is pronounced [tʰ]
  - When /k/ is word-initial, it is pronounced [kʰ]
How do we represent these “rules” of English pronunciation?

- /p/, /t/, and /k/ become aspirated in word-initial position
- _______ become aspirated in word-initial position

How do we represent these “rules” of English pronunciation?

- /p/, /t/, and /k/ become aspirated in word-initial position
- _______ become aspirated in word-initial position
- Voiceless (oral) stops become aspirated in word-initial position

Behavioral Finding

Your own aspiration behavior shows that you implicitly know the category “voiceless (oral) stops”

This is an unconscious mental category

Question: Where, during auditory processing of speech, do you first find evidence for this category?

Voice Onset Time (VOT)

60 msec

(Phillips et al., 2000)
Sound Groupings

[+voice]  [-voice]

(Phillips et al., 2004)

Phonological Features

Phonological Natural Classes exist because...
- Phonemes are composed of features - the smallest building blocks of language
- Phonemes that share a feature form a natural class

Effect of Feature-based organization observed in…
- Language development
- Language disorders
- Historical change
- Synchronic processes

Sound Groupings in the Brain

pæ, tæ, tæ, kæ, dæ, pæ, kæ, tæ, pæ, kæ, bæ, tæ...

(Phillips et al., 2004)
Sound Groupings in the Brain

\[ p\ae, \, t\ae, \, k\ae, \, d\ae, \, p\ae, \, k\ae, \, t\ae, \, p\ae, \, k\ae, \, b\ae, \, t\ae \ldots \]

(Phillips et al., 2004)

Feature Mismatch

(Phillips et al., 2004)

Syllables

- Sequences: *egma, egma, egma, egma, eguma*
- French have 3 mismatch responses
  - Early, middle, late
- Japanese only have late response

Dehaene-Lambertz et al. 2000

Speech Perception

Is a speaker of a language a prisoner of its ‘alphabet’ of sounds?