Language Comprehension

Speech Perception
Meaning Representation
Speech Comprehension and Production

- Formulate communicative intentions
  - Select lexical-semantics
  - Retrieve phonological code
    - Articulatory planning (syllabification)
    - Execute motor articulatory routines
  - Sentence construction
- Integrate with knowledge about speaker and world
  - Retrieve word meanings
  - Spoken word recognition
  - Recognise speech sounds
- Sentence parse and comprehension
- Early auditory processing

Sound
Speech Perception

- The first step in comprehending spoken language is to identify the words being spoken, performed in multiple stages:

1. Phonemes are detected (/b/, /e/, /t/, /e/, /r/, )
2. Phonemes are combined into syllables (/be/ /ter/)
3. Syllables are combined into words (“better”)
4. Word meaning retrieved from memory
Spectrogram: I owe you a yo-yo
Speech perception: two problems

• Words are not neatly segmented (e.g., by pauses)

• Lack of phoneme invariance
  • Coarticulation = consecutive speech sounds blend into each other due to mechanical constraints on articulators

• Speaker differences; pitch affected by age and sex; different dialects, talking speeds etc.
How Do Listeners Deal with Variability in Acoustic Input?

• Use of visual cues:
  – McGurk effect

• Use of semantic cues:
  – Phonemic restoration

• **Categorical perception**: continuous changes in input are mapped on to discrete percepts
Phonemic restoration

<table>
<thead>
<tr>
<th>Auditory presentation</th>
<th>Perception</th>
</tr>
</thead>
<tbody>
<tr>
<td>Legislature</td>
<td>legislature</td>
</tr>
<tr>
<td>Legi_lature</td>
<td>legi lature</td>
</tr>
<tr>
<td>Legi*lature</td>
<td>legislature</td>
</tr>
</tbody>
</table>

- It was found that the *eel was on the axle. wheel
- It was found that the *eel was on the shoe. heel
- It was found that the *eel was on the orange. peel
- It was found that the *eel was on the table. meal

McGurk Effect
Perception of auditory event affected by visual processing

Demo 1
AVI: http://psiexp.ss.uci.edu/research/teachingP140C/demos/McGurk_large.avi
MOV: http://psiexp.ss.uci.edu/research/teachingP140C/demos/McGurk_large.mov

Demo 2
MOV: http://psiexp.ss.uci.edu/research/teachingP140C/demos/McGurk3DFace.mov

McGurk Effect

• McGurk effect in video:
  – lip movements = “ga”
  – speech sound = “ba”
  – speech perception = “da” (for 98% of adults)

• Demonstrates parallel & interactive processing: speech perception is based on multiple sources of information, e.g. lip movements, auditory information.

• Brain makes reasonable assumption that both sources are informative and “fuses” the information.
Categorical Perception

- **Categorical perception**: high level cognitive processes (i.e., categorization) can influence perceptual processes.

![Diagram showing the relationship between Categorization, categorical perception, and Perception of Sounds/Images.](Image)
Differences among items that fall into different categories are exaggerated, and differences among items that fall into the same category are minimized.

(from Rob Goldstone, Indiana University)
Examples

• from “LAKE” to “RAKE”
  – http://www.psych.ufl.edu/~white/Cate_per.htm

• from /da/ to /ga/

Good /da/

Good /ga/

1  2  3  4  5  6  7  8
Identification: Discontinuity at Boundary

% of /ga/ response vs Token

0% 100%

Token: 1 2 3 4 5 6 7 8
Pairwise discrimination

Discriminate these pairs
(straddle the category boundary)
Pairwise Discrimination
(same/different)

% Correct Discrimination

Pair of stimuli

1_2 2_3 3_4 4_5 5_6 6_7 7_8
What Happened?

Physical World

Perceptual Representation
Categorical Perception depends on language

• In one language a difference in sound may make a difference between words; in another, it might not

• Example
  – The Japanese language does not distinguish between /l/ and /r/
  – These sounds belong to the same category for Japanese listeners
  – They find it very hard to discriminate between them (Massaro, 1994)
Non-English Contrasts

Hindi
- Dental Stop
- Retroflex Stop

Salish
(Native North American—Canadian—language)
- Uvular
- Velar
Models of Spoken Word Identification

• The **Cohort** Model
  – Marslen-Wilson & Welsh, 1978
  – Revised, Marslen-Wilson, 1989

• The **TRACE** Model
  – Similar to the Interactive Activation model
  – McClelland & Elman, 1986
Online word recognition: the cohort model

<table>
<thead>
<tr>
<th>Time</th>
<th>“s”</th>
<th>“sp”</th>
<th>“spee”</th>
<th>“speed”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salmon</td>
<td>Speed</td>
<td>Speed</td>
<td>Speed</td>
<td>Speed</td>
</tr>
<tr>
<td>So</td>
<td>Sesame</td>
<td>City</td>
<td>Sun</td>
<td>Special</td>
</tr>
<tr>
<td>Special</td>
<td>Species</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>etc ...</td>
<td>etc ...</td>
<td>etc ...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Recognizing Spoken Words: The Cohort Model

- All candidates considered in parallel

- Candidates eliminated as more evidence becomes available in the speech input

- **Uniqueness point** occurs when only one candidate remains
Evidence for activation of spurious words

- If we recognize words by recognizing a cohort of possibilities, then the cohort words should exert some influence.

- Shillcock (1990). Test for semantic priming of cohort words:

  "He picked up the trombone"

  "trom" "bone"

  Semantic priming for the word "rib"
TRACE model

• Similar to interactive activation model but applied to speech recognition

• Connections between levels are bi-directional and excitatory
  → top-down effects

• Connections within levels are inhibitory producing competition between alternatives

(McClelland & Elman, 1986)
TRACE Model

Features over time

(McClelland & Elman, 1986)
Human Eye Tracking Data

‘Pick up the beaker’

Eye tracking device to measure where subjects are looking

Alloppena, Magnuson & Tanenhaus (1998)
Human Eye Tracking Data

Human eye tracking data highly similar to TRACE predictions

Alloppena, Magnuson & Tanenhaus (1998)
Representing Meaning
Representing Meaning

• Mental representation of meaning as a network of interconnected features

• Evidence comes from patients with *category-specific impairments*
  – more difficulty activating semantic representation for some categories than for others
Category Specific Semantic Deficits

• Warrington and Shallice (1984) reported a patient called JBR who following an acute lesion to the left temporal lobe (as a result of herpes encephalitis) had a selective deficit when asked to name pictures from just one semantic category – living things.

• By contrast JBR was able to name non-living objects very well including those with low frequency names such as ‘accordion’ that were matched for the number of letters in the name and the visual complexity of the object.

• Other patients have shown opposite pattern
## Summary of patient data

<table>
<thead>
<tr>
<th>Living</th>
<th>Nonliving</th>
<th>Artefacts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animal</td>
<td>Fruit</td>
<td></td>
</tr>
<tr>
<td>x</td>
<td>x</td>
<td>√</td>
</tr>
<tr>
<td>√</td>
<td>√</td>
<td>x</td>
</tr>
<tr>
<td>x</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>√</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>√</td>
<td>x</td>
<td>√</td>
</tr>
</tbody>
</table>
Regions of the brain involved in representation of word meaning

Damage to these areas affects understanding the meanings of certain words, sometimes in the form of category-specific impairments. The patterns of impairment suggest that words are represented in semantic networks that include various types of information, including perceptual aspects. Some of this information is also represented in similar areas of the right hemisphere.
Implications

• Different types of objects depend on different types of encoding
  → perceptual information
  → functional information
Sensory-Functional Approach

• Category specific effects on recognition result from a correlated factor such as the ratio of visual versus functional features of an object
  – living more visual and nonliving more functional.

• Farah & McClelland (1991) report a dictionary study showing the ratio of visual to functional features for living things and nonliving things:
  – living things was 7.7:1 and nonliving was 1.4:1.
A neural network model of category-specific impairments

- A single system with functional and visual features. Model was trained to discriminate 20 living and nonliving things.

- Two main layers: semantic and input. Semantic has 3:1 ratio of visual and functional properties.

- Objects have visual and functional property codes:
  - 7.7:1 for living things
  - 1.4:1 for nonliving things

Farah and McClelland (1991)
Simulating the Effects of Brain Damage by “lesioning” the model

Visual Lesions: selective impairment of living things

Functional Lesions: selective impairment of non-living things

Farah and McClelland (1991)