Sensation & Perception

Ch. 13: Speech Perception

© Takashi Yamauchi (Dept. of Psychology, Texas A&M University)

Main topics

Phoneme
Speech spectrogram
The segmentation and variability problems
Categorical perception
Multimodal speech perception, Top-down process
Speech perception and the brain
• Can you understand this?
Visual stimuli

- Units
  - Primary colors / 3 types of cones (short, medium, long wavelengths)
  - Visual features (lines, dots)
  - Objects (geons)
What are the perceptual units of speech?
Stimuli

• **Phonemes:**
  • The smallest unit of the sound system in a language
    – consonants and vowels
    – If you change a phoneme in a word, the meaning of the word is also changed. (e.g., hip → tip; cat → hat; dog → dig)

• English has about 40 different phonemes
  – Every English word is produced by a combination of phonemes.

• Different languages have different phonemes
  – E.g., Japanese does not distinguish r and l.
  – E.g., rice / lice; pray / play; election / erection;
• Babies can distinguish speech sounds initially, but lose their abilities as they learn a particular language.

• Critical period and phoneme learning: 3:22
Table 9.1  
Phonetic Symbols.
The major consonants, vowels, and diphthongs of English and their phonetic symbols.

<table>
<thead>
<tr>
<th>Consonants</th>
<th>Vowels</th>
<th>Diphthongs</th>
</tr>
</thead>
<tbody>
<tr>
<td>p  pill  θ thigh</td>
<td>i  beet</td>
<td>ay  bite</td>
</tr>
<tr>
<td>b  bill  ő thy</td>
<td>i  bit</td>
<td>æw  about</td>
</tr>
<tr>
<td>m  mill  š shallow</td>
<td>e  bait</td>
<td>o y  boy</td>
</tr>
<tr>
<td>t  till  ž measure</td>
<td>e  bet</td>
<td></td>
</tr>
<tr>
<td>d  dill  č chip</td>
<td>æ  bat</td>
<td></td>
</tr>
<tr>
<td>n  nil  ʃ gyp</td>
<td>u  boot</td>
<td></td>
</tr>
<tr>
<td>k  kill  l lip</td>
<td>u  put</td>
<td></td>
</tr>
<tr>
<td>g  gill  r rip</td>
<td>æ  but</td>
<td></td>
</tr>
<tr>
<td>η  sing  y yet</td>
<td>o  boat</td>
<td></td>
</tr>
<tr>
<td>f  fill  w wet</td>
<td>o  bought</td>
<td></td>
</tr>
<tr>
<td>v  vat  m whet</td>
<td>a  pot</td>
<td></td>
</tr>
<tr>
<td>s  sip  h hat</td>
<td>æ  sofa</td>
<td></td>
</tr>
<tr>
<td>z  zip</td>
<td>i  marry</td>
<td></td>
</tr>
</tbody>
</table>

Distinguishing phonemes

• Phonemes are distinguished by the way they are articulated.
Place of articulation

Table 9.2 The Place of Articulation.

1. the two lips together (bilabial)
2. the bottom lip against the upper front teeth (labiodental)
3. the tongue against the teeth (dental)
4. the tongue against the alveolar ridge of the gums just behind the upper front teeth (alveolar)
5. the tongue against the hard palate in the roof of the mouth just behind the alveolar ridge (palatal)
6. the tongue against the soft palate, or velum, in the rear roof of the mouth (velar)
7. the glottis in the throat (glottal)

This division breaks up the consonants into seven major groups:

1. bilabial: p, b, m, w
2. labiodental: f, v
3. dental: θ, ð
4. alveolar: t, d, s, z, n, l, r
5. palatal: š, ž, č, j, y
6. velar: k, g
7. glottal: h

Manner of articulation

**Table 9.3 Manner of Articulation.**

1. **stops**: p, b, t, d, k, g
2. **fricatives**: f, v, θ, s, z, ñ, ň, h
3. **affricates**: č, ʡ
4. **nasals**: m, n, ň
5. **laterals**: l
6. **semivowels**: w, r, y

1. **stops**: complete closure at a point of articulation
2. **fricatives**: construction at a point of articulation which can produce sustained turbulence or vibration
3. **affricates**: sequence of complete closure followed by a fricative-like rush of air through a constriction
4. **nasals**: complete closure of the mouth at some point of articulation along with an opening of the nasal cavity to let air rush through the nose
5. **laterals**: shaping the tongue so that the main opening is at the sides of the tongue
6. **semivowels**: shaping the tongue so that the main opening is at the middle of the tongue
Consonants and vowels (two major phonemes)

- Consonants and vowels are distinguished by how they are articulated.
  - Consonants obstructs air, but vowels do not.
- E.g.,
  - p, b, m, n, t, d, k, g
  - e, i, a, o.
- Consonants differ in three primarily attributes:
  - place of articulation, manner of articulation, and voicing ([b] [p]).
  - process phonemes by attending to these articulatory features.
Articulation

• The place of articulation
  – The two lips together (p, b, m, w)
  – The bottom lip against the upper front teeth (f, v)

• The manner of articulation
  – Stops: p, b, t, d, k, g
  – Fricatives: f, v, s, z
• Monkeys

• *Can chimps learn human-like language? (4:16)*
  – [http://www.youtube.com/watch?v=utVXZAflESo](http://www.youtube.com/watch?v=utVXZAflESo)

• *Can chimps learn human-like language? (2:38)*
  – [http://www.youtube.com/watch?v=wRM7vTrIIis&NR=1](http://www.youtube.com/watch?v=wRM7vTrIIis&NR=1)
Visualizing speech sound

• /h/

Speech Spectrogram

- Y-axis: frequency (Hz)
- X-axis: time
- Darkness: amplitude
Some major problems in speech perception

• The segmentation problem
  – How do we segment the individual words?

• The variability problem
  – Variability from different speakers
  – Variability from a phoneme’s context
    • E.g., /d/ differs in “dig” “dug”
Question:

• How does the perceptual system handle these problems, and give us a coherent speech perception?
Speech perception

• How does our perceptual system handle these problems?

Use other senses (e.g., visual infor.)
Use background knowledge (top-down process)
Using visual cues in speech perception

• The McGurk effect (demonstration)
  – http://www.youtube.com/watch?v=I1XWDOwH47Y
  – http://www.youtube.com/watch?v=jtsfidRq2tw&NR=1

Using visual cues in speech perception

• The McGurk effect (demonstration)
  – http://www.media.uio.no/personer/arntm/McGurk_english.html

  – **What am I saying?** Play the clip several times, alternating between looking at the talking head while listening, and listening with your eyes shut. Most adults (98%) think they are hearing "**DA**" - a so called "fused respons" - where the "**D**" is a result of an *audio-visual illusion*. In reality you are *hearing* the sound "**BA**", while you are *seeing* the lip movements "**GA**". The "McGurk effect" was first described by Harry McGurk and John MacDonald in "Hearing lips and seeing voices", *Nature* **264**, 746-748 (1976).
Background knowledge

• Meaning and segmentation
  – Word segmentation is much easier if it is made in a coherent sentence.

• Segmentation is affected by context and meaning
  – I scream you scream we all scream for ice cream
One phenomenon

• Categorical perception
  – /da/ vs. /ta/ or /bu/ vs. /pu/ are distinguished by when you vibrate your vocal codes --> VOT (voice onset time)

• VOT (voice onset time)
  – The time delay between a consonant is released and when the vibration of the vocal codes begin.

• /da/ vs. /ta/
  – /da/ → VOT = 17ms
  – /ta/ → VOT = 91ms
Experiment

Created artificial speech sounds (/da/ /ta/) that continuously varied in VOT.

Tested the ability to distinguish the artificial speech sounds.

Two sounds were presented in sequence, and Ss judged if they were the same or different.

Subjects could not distinguish the sounds within the same category.
What does this tell?

• The sensitivity to distinguish speech sounds is lost within the phonetic boundary.

• Speech perception
  – discard physical differences, and capture them in one category???
Meaning and Word Perception

• Experiment by Miller and Isard
  – Stimuli were three types of sentences:
    • Normal grammatical sentences
    • Anomalous sentences that were grammatical
    • Ungrammatical strings of words
  – Listeners were to *shadow* (repeat aloud) the sentences as they heard them through headphones
Meaning and Word Perception - continued

• Results showed that listeners were
  – 89% accurate with normal sentences
  – 79% accurate for anomalous sentences
  – 56% accurate for ungrammatical word strings
Speech perception and the brain

- Broca’s area
- Wernicke’s area
Speech Perception and the Brain

• Broca’s aphasia - individuals have damage in Broca’s area (in frontal lobe)
  – Labored and stilted speech and short sentences but they understand others
    • http://video.google.com/videoplay?docid=9178936581276081395&q=Broca%27s+aphasia&total=6&start=0&num=10&so=0&type=search&plindex=0

• Wernicke’s aphasia - individuals have damage in Wernicke’s area (in temporal lobe)
  – Speak fluently but the content is disorganized and not meaningful
  – They also have difficulty understanding others
    • http://video.google.com/videoplay?docid=7590914168187986085&q=Wernicke%27s+aphasia&total=3&start=0&num=10&so=0&type=search&plindex=0
Experience Dependent Plasticity

• Before age 1, human infants can tell difference between sounds that create all languages
  – Japanese infants can distinguish /r/ and /l/.
  – Japanese adults cannot.
• The brain becomes “tuned” to respond best to speech sounds that are in the environment
• Other sound differentiation disappears when there is no reinforcement from the environment