Infant Speech Perception (cont.)
Outline

• Introduction to Phonology
• Problem of Speech Perception
• Testing two theories of speech perception
• Becoming a native listener
Phonemes are bundles of features

p: bilabial, voiceless, stop

s: alveolar, voiceless, fricative

<table>
<thead>
<tr>
<th>Manner of Articulation</th>
<th>Bilabial</th>
<th>Labiodental</th>
<th>Interdental</th>
<th>Alveolar</th>
<th>Palatal</th>
<th>Velar</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stop</td>
<td>p</td>
<td>b</td>
<td></td>
<td>t</td>
<td>d</td>
<td>k</td>
<td>g</td>
</tr>
<tr>
<td>Fricative</td>
<td>f</td>
<td>v</td>
<td>θ</td>
<td>s</td>
<td>z</td>
<td>š</td>
<td>ž</td>
</tr>
<tr>
<td>Affricate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ċ</td>
<td>ţ</td>
<td></td>
</tr>
<tr>
<td>Nasal</td>
<td>m</td>
<td></td>
<td></td>
<td></td>
<td>n</td>
<td>ň</td>
<td></td>
</tr>
<tr>
<td>Lateral Liquid</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>l</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retroflex Liquid</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>r</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glide</td>
<td>w</td>
<td>w</td>
<td></td>
<td></td>
<td>y</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Speech sounds vary along a continuum
Yet we perceive them as discrete categories
### Two contrasting theories

**The Motor Theory**

1. **SP is based on production**
   - Process: Determine what articulatory gestures a speaker made

2. **SP is species specific**
   - Speech production and perception co-evolved

3. **SP is innate**
   - Tacit knowledge of articulation given by evolution

**The Auditory Theory**

1. **SP by generic auditory mechanisms**
   - Process: Auditory system transforms sound wave so phonemes available

2. **Not species specific**
   - Production system evolved to make use of existing auditory capacities

3. **SP may be innate**
   - Innate properties of the auditory system may be enough OR perceptual learning may be required
1. Is Speech Perception Innate?
Stimuli from different phonological categories are distinguished.

SPM sucks per minute

Eimas, et al., 1971

“PA₁”

“BA”

Mean number of sucking responses per minute

Time

Before

Minutes before and after change in speech sound

After

REVIEW
But stimuli from the same phonological category are not

SPM sucks per minute

Eimas, et al., 1971
Speech Perception is Innate

Predicted by Motor Theory
- Speech Perception driven by innate knowledge of articulation

Consistent with Auditory Theory
- Speech perception due to innate structure of auditory system
2. Is Speech Perception Species Specific?
Do other animals perceive phonemes categorically?

- Kuhl & Miller, 1978: test chinchillas and humans with identical stimuli
- Human Task: identification (b or p)
- Chinchillas: avoidance conditioning
Avoidance Conditioning Procedure

Shock paired with speech sound at one end of continuum (A)

0  VOT in milliseconds  PA
Avoidance Conditioning Procedure

Sound at other end (B) paired with safety

BA

VOT in milliseconds

80
Avoidance Conditioning Procedure

Animal learns to run to other side of cage when it hears sound A but stay after B.
What will they do for sounds in between?

stay  ?  ?  ? run
Predictions

Voice Onset Time (ms)

Percent labelled [b]

Categorical Perception
Graded Perception
Kuhl & Miller, 1978

Voice Onset Time (ms)

Percent labelled [b]

English Speakers

Chinchillas
Some aspects of speech perception not species specific

Contrary to Motor Theory

• Claim: only humans have knowledge (innate or learned) of articulation

Consistent with Auditory Theory

• General auditory abilities adequate for (some aspects of) speech perception
3. Is Speech Perception Affected by Knowledge of Articulation?
McGurk Effect
McGurk Effect

• Vary visual information about place of articulation
  – palatal: GA
  – dental: DA
  – bilabial: BA

• Vary auditory stimulus (BA→DA→GA)

• Task: phoneme identification
McGurk Effect

• Finding: phoneme category boundary affected by visual information
  – sound BA + visual GA = percept DA

• Adult speech perception affected by visual cues
  – Did they learn it?

• Do infants know visual cues to articulation?
  – No evidence of visual cues influencing categorization
  – But can infants match sound with the correct mouth shape?
Infants are sensitive to visual-auditory correspondences
(Kuhl & Meltzoff, 1982)

[a]-face

[i]-face

Preferential looking paradigm
a...a....a...a..a

[a]-face

[i]-face
McGurk Effect in infants (5m)

(Rosenblum, Schmuckler, & Johnson, 1997)

Figure 1. Mean duration of looking at the audiovisual stimuli of Experiment 1 and the audio-alone stimuli of Experiment 3 (see text for details).
Infants integrate visual information in speech perception

• By 1m of age they match speech sounds to faces
• By 5m visual information affects speech perception
• Very hard to test younger infants in visual tasks
Speech Perception is influenced by information about articulation

Predicted by Motor Theory
• Innate connection btw production & perception

Inconsistent with Auditory Theory
• A visual-auditory theory could explain
• Would require perceptual learning
Provisional Conclusions

• Speech Perception makes use of auditory mechanisms which evolved prior to language
  – These abilities are innate

• Speech Perception also makes use of our knowledge of articulation
  – This ability MAY be innate
Becoming a Native Listener

What develops during speech perception?
Newborns are universal listeners

- Infants perceive speech categorically
- Newborns are sensitive to every phonological distinction yet tested
- What happens to the categories that are not used in the language that they learn?
Developmental changes

• English versus Japanese

Figure 12.2. Test of the categorical perception of /ra/ and /la/ by American and Japanese adults. American listeners show the characteristic peak in discrimination at the phonetic boundary; Japanese listeners do not. (From Miyawaki et al., 1975.)
Developmental changes

- English versus Spanish VOT

Figure 4.11. AX discrimination functions obtained by Williams (1977) for English and Spanish monolingual listeners judging the same labial VOT stimuli. A 20-ms step size was used. (Selectively redrawn from Figures 1 and 2 of Williams [1977].)
Testing Across the Lifespan

• Habituation paradigms work only with young infants
• Adults & children can be asked to detect a change
• The Conditioned Head Turn Paradigm: parallel task for older infants
Visual Reinforcer (VR)
Toy that lights up and moves at the experimenter’s command

Controls for sound stimuli and the VR
Conditioned Head Turn

- Infant Trained with 2 clearly different auditory stimuli (bell and whistle)
  - each time the sound changes the toy is activated
  - next toy is activated only if the sound changes AND the baby turn to look at it
Conditioned Head Turn

- Child hears Stimulus 1 (/ba/) repeatedly
- Then Stimulus 2 is presented (/da/)
- If child detects difference, they should turn to look at the toy when the stimulus changes
- If they do not detect it, they shouldn’t turn until after the toy is activated
Conditioned Head Turn Paradigm

Kuhl Lab, U Washington, 1992
Conditioned Head Turn Paradigm

Kuhl Lab, U Washington, 1992
Results for 2 non-native phonological contrasts

Werker, 1995
Ability to perceive non-native contrasts declines in 1st year

Werker, 1995

Longitudinal Data
Contrasting Views

- Structure-adding
  - Native-language phonology
    - 'Universal Phonetics' (initial state)
- Structure-changing
  - Native-language Phonetics
    - non-native boundary (erased)

Acoustics
Maintenance or Loss Model

• If you don’t use a perceptual ability, you lose it
• Parallel to aspects of early visual development
Functional Reorganization

• The newborn has perceptual categories
• Those which are meaningful in the native language become speech categories
• The remainder are perceived but not recruited in speech perception
Evaluating the models

Problems with maintenance-loss

• Children acquire new languages without accents
• Many of the relevant sounds appear in child’s input but not meaningful
• Adults can be trained to make distinctions
• Perceptual distinction is readily available for non-linguistic tasks

Reorganization seems more likely
Speech Perception Before Reorganization

Behavior

Auditory → Phonetic → Articulatory

Innate & Universal
Creation of Phonological Representations

- Behavior
- Constructed & Language Specific
- Innate
- Auditory
- Phonetic
- Articulatory

Phonology
Reorganization Strikes Again

The creation of lexical representations
Children’s word perception

• Toddlers fail to learn words that vary by one feature
  – “bear” versus “pair”
  – “Piglet” versus “*Biglet”

• Why? They have phonological representations.
Do infants represent precise phonological forms of words? (Stager & Werker, 1995)

- Habituation study measuring looking time
- Referent for word

- 14 m.o.
- Habituation study measuring looking time
- Referent for word
Experiment 1: Stager & Werker, 1995

- If precise representation: dishabituation to switch
- Imprecise representation (or no “meaning”): continued habituation in switch
Experiment 1

- Do not distinguish familiar and novel pairing
Maybe the task was too hard…
(experiment 2)

- 14m & 8m
- Learn only one word

- If precise representation: dishabituation to switch
- Imprecise representation: continued habituation in "switch"
Can infants notice change in a single word?

- 8m succeed
- 14m do not
- Rules out complexity as explanation
Can 14m infants ever distinguish between words? (Experiment 3)

- Words have no phonological overlap

<table>
<thead>
<tr>
<th>Habituation phase</th>
<th>Test phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>'Lif'</td>
<td>'Lif'</td>
</tr>
<tr>
<td>'Lif'</td>
<td>'Neem'</td>
</tr>
</tbody>
</table>
Can 14m infants ever distinguish between words? (Experiment 3)

- clearly distinguish between words
- Shows task sensitive to lexical reps at 14m
Do 14m distinguish phonemes?
(Experiment 4)

- One feature different between stimuli
- No plausible referent for word
Do 14m distinguish phonemes? (Experiment 4)

• Yes 14m detect phonological changes
• But they fail to use this ability when comparing words
• WHY?

14 months (checkerboard) 'Bih'-'Dih'

- Same
- Switch
Functional Reorganization

Appears around 14m

Appears around 10m

Syntactic

Lexical

Phonological

Auditory

Phonetic

Articulatory

Innate
What information do children use to create phonological representations?
Hypothesis: represent categories heard most often

- Could explain disappearance of some categories
  - Clicks rare

- Can’t explain loss of many distinctions
  - Allophones: phonetic differences that appear in a single language but are not contrastive (e.g., aspiration)
  - Range of VOT are heard even if it isn’t contrastive
Hypothesis: learning through contrast

- Children might notice that changing some features changes meaning
  - “pat” has different meaning from “bat” (pig vs big)
- But changing others does not change meaning
  - “ţip” and “tip” used interchangeably
- What would children have to know to apply this strategy?
- Is this consistent with the known data?
Hypothesis: Distributional Learning

- Native speakers production of VOT may contain information about the boundary
- Not uniform; 2 clumps

Figure 5-3. VOT productions of a single normal adult speaker of American English for words beginning with /d/ and /t/. (Figure adapted with permission from Miller, R. E. (1980). Production
Bimodal distribution could act as a cue for phonological contrast

Figure 1: Monomodal vs. Bimodal Distributions

Frequency of Occurrence

(\text{short VOT}) \quad \text{e.g., Voice Onset Time} \quad (\text{long VOT})

\begin{itemize}
\item \text{---} = \text{Monomodal distribution}
\item \text{--} = \text{Bimodal distribution}
\end{itemize}
Experimental Test of Distributional Hypothesis

Children exposed to 1 of 2 nonsense languages

Figure 2: Stimuli Presentation Frequency during Acquisition Phase

- Number of Presentations per Block of Training
- Token Number
- /da/
- /ta/

- = Monomodal group
- = Bimodal group
Results

Does seem to affect categorization

Figure 3: Results, Experimental Contrast Pairs

Percent of "different" responses to /d~//t/ pairs
Summary

• reliable phonetic features used in phonological representation
• unreliable phonetic features are suppressed
• …but can be used for non-linguistic purposes
• higher reps are subset of information available at lower level