Speech perception and categorization

What’s going on in the first year

innate auditory abilities govern categorization of speech

then, infants start sorting out which sounds of a particular language are meaningfully different (meaningfully for the language, not for the infant, not yet)

at the same time, infants begin building a vocabulary of words (mostly forms only, no meanings)

by 10-12 months, infants say their first words
What’s in the toolbox from the start?

1. Infants like speech.
2. Infants can tell their language from (some) other languages.
3. Infants match audiovisually.
4. Infants have biases in categorization of speech sounds.
5. Infants can learn categories -- and they do.
But first a methodological interlude

“High-amplitude sucking”

baby given pacifier with pressure-sensitive device
when baby sucks (hard), computer plays sound
compute when the baby’s sucking tails off, and then change sound
does sucking rebound?

50
25
baseline

play test / control stim.

30
15
30
15
baseline

habituated!

test

habituated!
“High-amplitude sucking”

Figure 4.1 Apparatus for presenting artificially manipulated speech sounds to young infants. The infant sucks on a pacifier connected to recording instruments as speech-like sounds are presented from a loudspeaker just above the Raggedy Ann display.
Infants like speech

Newborns prefer speech to:
- white noise
- Sinewave speech
Infants like speech

Newborns don’t prefer speech to:
- rhesus monkey calls!

37-hour-olds (range 18 - 120 hours)

monkey calls all from “positive social contexts such as affiliative interactions (girneys, coos) and the discovery of food (coos, warbles).”
human calls nonwords, 3 women, “ploo”, “keev”, “yut”, etc., IDS register
Method: HAS (pure preference, no habituation)
Infants like speech (but:)

Vouloumanos, Hauser, Werker, & Martin, 2010

**newborns:**
speech vs monkey, n.s.
monkey > noise
human > noise

**3 month olds:**
speech > monkey.
(1-screen vis. pref. procedure)
Newborns distinguish different languages, to some degree

Basic method:
- record sentences in languages 1 and 2
- play a subset of these, from one language, to habituation
- then some infants hear new sentences from L1;
  other infants hear (new) sentences from L2
- evaluate sucking rate (during habituation and test)
Newborns distinguish different languages, to some degree

Basic method:
- record sentences in languages 1 and 2
- play a subset of these, from one language, to habituation
- then some infants hear new sentences from L1;
  other infants hear (new) sentences from L2
- evaluate sucking rate (during habituation and test)

Mehler, Jusczyk, Lambertz, Halsted, Bertoncini, & Amiel-Tison, 1988

1. French 2-day-olds prefer French to Russian (more sucking during habituation, and rebound Rus->Fr)
2. Not found for Paris-born but foreign-parent 2-day-olds
3. No differences if you play the speech backwards
4. Low-pass-filter 400Hz: still discriminate Rus / Fr (suggests basis in intonation or rhythm of languages)

So far: hearing French makes infants like it, distinguish it from Rus.
Newborns distinguish different languages, to some degree

Mehler, Jusczyk, Lambertz, Halsted, Bertoncini, & Amiel-Tison, 1988

5. US 2-month-olds discriminate English from Italian
5b. Still true if low-pass filtered
6. But don’t discriminate French from Russian

Nazzi, Bertoncini, & Mehler, 1998. 4 talkers; habit to 2, test on other 2
1. French 2-day-olds discriminate English and Japanese
2. But not English and Dutch
3a. Mix Eng. & Dut., Ital. & Sp.; ED≠IS, IS≠ED
3b. Mix ES & ID or EI & DS; no dishabituation.

The claim, then: newborns can distinguish languages if they belong to different rhythmic classes.

Follow-ups: saltanaj resynthesis, testing on tamarins & rats...
e.g. Toro, Trobalon, Sebastián-Gallés, 2005
Intermodal integration in early speech perception

Kuhl & Meltzoff, 1982

One face says /i/ ... /i/ ....
Other face says /a/ ... /a/ ...
Infant hears either /i/ or /a/
19-week-olds (~4.4 months)
Looking to match: 73.6%, 24/32 Ss

Extension to 2 month olds: Patterson & Werker 2003


/a/ vs /u/, /i/ vs /u/, /i/ vs /wi/
Intermodal integration in early speech perception

Chen, Striano, & Rakoczy, 2004, 1- to 7-day-olds

Model says: aaahhh ... aaahhh ... (4x / trial, 8 trials); then, model says: mmmm ... mmmm ... (4x, 8 trials) {or reverse order}

Q: does infant make like mouth movements?
Q: does infant make like mouth movements?

Chen, Striano, & Rakoczy, 2004

Infants who kept eyes **closed**:

<table>
<thead>
<tr>
<th>Model</th>
<th>Opening</th>
<th>Clutching</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/</td>
<td>9.5</td>
<td>5.1</td>
</tr>
<tr>
<td>/m/</td>
<td>1.5</td>
<td>10.4</td>
</tr>
</tbody>
</table>

Infants who kept eyes **open**:

<table>
<thead>
<tr>
<th>Model</th>
<th>Opening</th>
<th>Clutching</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/</td>
<td>12.3</td>
<td>1.5</td>
</tr>
<tr>
<td>/m/</td>
<td>7.5</td>
<td>7.2</td>
</tr>
</tbody>
</table>

All model /a/ vs /m/ comparisons $p < .05$
Language differentiation from visual information

Weikum, Vouloumanos, Navarra, Soto-Faraco, Sebastián-Gallés, & Werker, 2007

4, 6, 8 months old, English background

Videos of 3 bilingual French/English speakers reading sentences; No audio presented to infants

1. Show clips from each speaker, 1 language, till visual habituation.
2. Switch language (expt) or not (control).
Language differentiation from visual information

Weikum, Vouloumanos, Navarra, Soto-Faraco, Sebastián-Gallés, & Werker, 2007

4, 6, 8 months old, English background

Videos of 3 bilingual French/English speakers reading sentences; No audio presented to infants

1. Show clips from each speaker, 1 language, till visual habituation.
2. Switch language (expt) or not (control).

4, 6 mos: increase to switched language, decrease to same language
8 mos: no significant difference

(But: Fr/Eng bilingual 6, 8mos did dishabituate to change.)
What else is in the toolbox?

Biases in speech sound discrimination

Example: onset consonant voicing

the sound [p] in, say, “a pack”:

1. stop vocal fold vibration
2. put lips together
3. release lips and let air through
4. start vocal fold vibration

time between #3 and #4: Voice Onset Time (VOT)

VOT is a primary cue listeners use to distinguish [p] from [b] at syllable onset.
VOT variation in English

Thai uses all three: ba: ‘crazy’; pa: ‘aunt’; pâa: ‘cloth’
English speakers can distinguish synthesized b/p sounds differing only in VOT, if tested using 20 and 40 msec. But they can’t (or perform poorly) if tested using 0 and 20 or 40 and 60.

*Categorical perception* (strict definition): listeners can only discriminate sound pairs they can give different labels.

*Categorical perception* (loose definition): listeners are much better at discriminating between categories than within categories.
Testing VOT perception in infants: 
*Eimas, Siqueland, Jusczyk, & Vigorito 1971*

participants
18 1-month-olds, 18 4-month-olds

procedure
HAS. Habituation criterion 20% decline over 2 min relative to preceding min.

stimuli
6 syllables: VOT -20, 0, +20, +40, +60, +80
(Adults perceive -20, 0, and 20 as [b], others [p])

VOT manipulated by
(a) delaying F1 onset;
(b) adding noise to F2, F3 during F1 delay.
Both of these lead to percept of longer VOT.
How do infants perceive speech sounds?

Eimas et al.: 1- and 4-month-olds; habituation procedure.

VOT: +20/+40 (ba/pa) 0/+20, +60/+80 control

Data: 4 month-olds
Uniquely human genetic innate language capacity!...or not

chinchilla
macaque
J. quail
Further study: replications testing discrimination of many speech sounds.

Under ideal conditions, young infants can tell apart (almost) any two speech sounds that are used in any language for conveying different meanings.

<table>
<thead>
<tr>
<th>some contrasts infants discriminated in studies</th>
<th>some contrasts that seem more difficult</th>
</tr>
</thead>
<tbody>
<tr>
<td>b/p</td>
<td>S/Z (Eilers &amp; Minifie 1975)</td>
</tr>
<tr>
<td>b/d</td>
<td>f/sh (Eilers, Wilson, &amp; Moore 1977)</td>
</tr>
<tr>
<td>r/l</td>
<td>#n/#ng (Narayan, 2006)</td>
</tr>
<tr>
<td>b/w</td>
<td>d/D (Polka, Colantonio, &amp; Sundara 2001)</td>
</tr>
<tr>
<td>b/m</td>
<td></td>
</tr>
<tr>
<td>w/j</td>
<td></td>
</tr>
</tbody>
</table>
Other speech capacities: rate-determined categorization

Eimas & Miller 1980

[b] and [w] differ, in part, in how speedy the transition is

Fast transition: [b]
Slow transition: [w].

But “fast” vs “slow” are not absolute; they depend on speaking rate (or syllable duration; short syllable = fast rate)

80 ms long syllable: transition 16 ms = b, 40 ms = w
296 ms long syllable: transition 40 ms = b, 64 ms = w

Do infants show the same dependency?
Eimas & Miller 1980

80 ms long syllable: transition 16 ms = b, 40 ms = w
296 ms long syllable: transition 40 ms = b, 64 ms = w

HAS method. Familiarization rather than habituation.

7 min opportunity to suck to hear syllable. Then switch to other syllable for 4 min.

5 groups:
• short syll, 16 ms -> 40 ms \{b-w\}
• short syll, 40 ms -> 64 ms \{w-w\}
• long syll, 16 ms -> 40 ms \{b-b\}
• long syll, 40 ms -> 64 ms \{b-w\}
• control: no change
Eimas & Miller 1980

80 ms long syllable: transition 16 ms = b, 40 ms = w
296 ms long syllable: transition 40 ms = b, 64 ms = w

HAS method. Familiarization rather than habituation.

7 min opportunity to suck to hear syllable. Then switch to other syllable for 4 min.

5 groups:
• short syll, 16ms -> 40ms [b-w] +4.0 *
• short syll, 40ms -> 64 ms [w-w] -2.5
• long syll, 16 ms -> 40 ms [b-b] -3.8
• long syll, 40ms -> 64 ms [b-w] +4.3 *
• control: no change -4.6

Change in sucking rate after switch
At least for some speech features, infants appear well adapted to language.

But not all languages are the same.
language-specific refinement

Werker & Tees 1984: test discrimination of Hindi dental and retroflex /t/, and discrimination of Nthlakampx velar and uvular consonants [k’] and [q’], using CHT.

镫 velar 镖 uvular
Headturn / Conditioned Headturn

standard headturn setup

“behavioral observation audiometry”
Conditioned Headturn  ("visual reinforcement audiometry")

FIG. 12.1. A schematic diagram of the conditioned head-turn procedure.
Conditioned Headturn

FIG. 12.2. Madeleine performing in the conditioned head-turn procedure.
Conditioned Headturn

1. training: background + target; automatic reinforcer
2. conditioning: background + target; gradually increase reinforcer delay
3. testing: background + test stimuli; reinforcer only if baby turns

Werker & Tees procedure:

background `k_2....`k_i_3....`k_i_2...
change `q_i_1....`q_i_2.....`q_i_3....

Move from conditioning to testing after 3 correct anticipatory HTs in a row
Criterion to “pass”: 8/10 correct responses
If not passing within 25 change trials: test on /ba-da/;
    Have to succeed b/d to “count”; else, excluded, n=2
% infants reaching discrimination criterion

stimuli

Werker & Tees 1984

native Hindi, Salish

n$_{tot}$=3

longitudinal
(n=6, tested on both lgs)
Vowels: *Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992*

The phenomenon: it’s harder to tell prototypes from outliers, than outliers from one another.
(This phenomenon is only possible if you’ve learned the category.)

6mos; CHT; background is Engl. or Swed. prototype

Train HT to variants from the 4th ring; 7/8 criterion.

Test HT to all 32 variants.

Prediction: infants should fail to discriminate variants of their own -vowel prototypes more often than for other-lang prototypes
For 6 month olds, the similarity space around the /i/ and /y/ vowels differs for US and Swedish infants.
language-specific refinement: more vowels

Polka & Werker 1994, using visual habituation procedure

German /u/ vs /y/, 4 months discriminate
Canadian Eng. infants 6 months don’t
10-12 months don’t
language-specific refinement: vowels

Do Spanish infants discriminate Catalan /e/ and /E/?

Spanish and Catalan

Bosch & Sebastián-Gallés 2003
Methodological interlude

Headturn Preference procedure

1. trial starts with green light
2. once baby’s ready, flash one side light
3. when baby turns to light, start playing sounds from speaker
4. when baby turns away, stop sounds and start new trial

dependent measure: listening time to a given kind of auditory material
Headturn Preference procedure with familiarization (Bosch & Sebastián-Gallés 2003)

familiarization (2 min exposure):
1/2 infants:  dedi, dedi, dedi, dedi, dedi, dedi,...
1/2 infants:  dEdi, dEdi, dEdi, dEdi, dEdi, dEdi,...

test:
dedi, dedi, dedi, dedi, dedi, dedi, dedi,...
dEdi, dEdi, dEdi, dEdi, dEdi, dEdi,...

Results:
all 4.5 month olds discriminate /e/ and /E/;
Catalan 8.0 month olds discriminate them too;
Spanish 8.0 month olds don’t.
So: infants listen to the sounds of their language, and are able to form language-specific phonetic categories.

How?

Is it phonology?
How can you learn phonetic categories?

1. The semantic route:

   - very long VOTs
   - short VOTs
   - prevoicing (negative VOT)
How can you learn phonetic categories?

1. The semantic route:

- very long VOTs
- short VOTs
- prevoicing (negative VOT)

...But this assumes children know enough minimal-pair words to get the job done. They almost certainly don’t, by 6 to 8 months.
How can you learn phonetic categories?

2. The distributional route:

Categories as emerging clusters of experienced sounds
-sounds with correlated properties, that
-are nearby in perceptual “space”
Can infants do this with sounds?

*Maye, Werker, & Gerken 2002*

6-8 month old English learners discriminate voiced unaspirated vs voiceless unaspirated stops (10-12 mos don’t; Pegg & Werker 1997)

(like Romance /t/, /d/, or the short-lag /t/ in Eng. “stock”)

Game plan: expose infants to a bimodal distribution of sounds (indicates 2 categories), or a monomodal one (1).

Then test discrimination of the categories.
Maye, Werker, & Gerken 2002

Materials.
- Record Am.Eng. woman saying “sta” and “da” (& “ma”, “la”)
- Edit off the [s] of “sta”
- “digitally edited and re-synthesized to form an 8-point continuum from [da] to [ta]” 4 from [da] token, 4 from [(s)ta]. Differences: prevoicing for [da]; more change in formants for [da] (steeper formant trajectory) -- mostly a POA difference

Procedure. 6 mos, 8 mos (24 each group)

Familiarize to elements of continuum + ma + la, using HPP
6 blocks of 24 syllables (16 from continuum, + 8 ma/la)
Total famil. time 2.3 minutes
1/2 of infants: bimodal continuum; 1/2 monomodal
Then test: listening to alternating vs non-alternating lists
Maye, Werker, & Gerken 2002

Familiarization freqs. vs Continuum of [da]-[ta] Stimuli
Test trials:

4 Non-alternating trials: either stim (3), or stim (6)
4 Alternating trials: stim (1) and stim (8)
Maye, Werker, & Gerken 2002

Results: listening times

1. Longer listening after bimodal famil.
2. No-alt > alt for bimodal only (p < .04)
(no age diffs)
Maye, Weiss, & Aslin, 2008

Do infants generalize learning to different phones?

Materials.

• Record Hindi man, [da], [ga] (prevoiced); [ta], [ka] (unaspirated)
• from voiced (d/g), take prevoicing, splice onto t/k, -100, -75, -50, -25
• from voiceless (t/k), shrink VOT: 0, 7, 14, 21 ms

Procedure. 8 month olds

Familiarize to elements of continuum, velar or alveolar
Total famil. time ~3 minutes, while watching silent animation
Expt 1, 1/3 of infants: bimodal continuum; 1/3 monomodal;
1/3 controls hear tone sequence rather than speech
Expt 2, all infants bimodal continuum

Then test: habit. to +7ms tokens, change trials -50ms tokens
Expt 1: test trained stim; Expt 2, test other place-of-articulation
Maye, Weiss, & Aslin, 2008

Results: did infants dishabituate to category change?

Controls, monomodal group: no;
Bimodal groups: **yes**
   change-POA group dishabituates just as much as
   same-POA group

What does it mean?

Under some conditions, infants rapidly learn (some)
phonetic distinctions
This learning may be represented over features rather
than phones
Where we are so far

Infants’ innate similarity space for speech is like ours, in relevant ways.

Infants’ speech-sound categorization starts to align with the adult system during the first year.

The only remaining theory of this process is some form of distributional learning.

For voicing, at least, infants show distributional learning in the lab.
What’s still problematic

Actual infant-directed speech does not seem to present such clean-cut categories.

- English infant-directed speech, 4 moms
  - /p,t,k/
- English adult-directed speech, 4 women
  - /b,d,g/
mothers are talking about toy objects. (“speak naturally and use each word at least 3 times.”)

P. Kuhl et al., 1997, Science
Some more English vowels

women’s citation forms, Hillenbrand et al., 1995
2000 English vowels in F1F2 space

11 phonological types, speaker s04, Buckeye corpus

a, ae, E, eI, I, i, u, U, o, c, ∧/@
600 English vowels, infant-directed speech

Brent corpus, mom F1
monophthongs only

Brent corpus, mom F1
Norwegian *infant-directed* long point vowels
words in “focal position” only

Englund & Behne
2005 J.Psylx. Rsch
Looks a lot like a ball.
How can we help out the baby?
(or: what must babies be doing, if this approach is right?)

1. Use a better description of the vowels. 
   (duration, dynamic features, F3)

2. Give the baby a way to separate wheat from chaff. 
   (filter out all but the most emphatic vowels?)

3. Use contextual information.
/i/ and /I/ vowels, frequent words marked
Some unknowns

How do infants learn phonetic categories?

How do they know the role these categories play in lexical contrast?

How does perceptual knowledge contribute to the development of speech production? (how) do these processes interact?