Global and detailed speech representations in early language acquisition

Pierre Hallé (Labo de Phonétique et Phonologie, Paris)

Cognitive and Physical Models of Speech Production, Speech Perception and Production-Perception Interaction. Part III: Planning and Dynamics
"There has always been a tension between two ways of understanding linguistics […]. On one view (which was dominant in the first part of this century), language has a structure that can be explored independently of any efforts to figure out what particular speakers may do or think […]. On the other view of linguistics (that has come dominant in the past several decades), the goal of linguistics is to model what it is that goes inside a speaker’s head." (Goldsmith, 1999)
A few questions about representations

• RR's articulatory or acoustic targets ≈ mental representations

  - representations of basic speech units*
    rules combine basic units into higher order units

  - representations of combined units such as words

• Three questions about representations:
  - What could be the "basic units" of representation?
  - Same representations in production and perception?
  - What degree of detail is coded?

I focus mainly on the last question
Plan

• Data on prelexical children
  The syllable as the basic unit of production?
  Global rather than detailed specifications for early syllables
  Newborns "count" syllables rather than phonemes (or moras)
  Utterances with a syllable structure favor speech-mode processing
  Do infants parse syllables into segments or treat them as whole-units?

• Data on 1st words produced and on trained vs. untrained word recognition
  Simplification and regularities in produced first words
  Phonetic detail in trained word recognition at 8 mos (Jusczyk & Aslin, 1995)
  ...not in untrained familiar word recognition ca. 11 months?
  Phonetic detail from ~14 months onward (Swingley, Plunkett, Werker...)

• Cs/Vs asymmetries in adults: Cs <-> lexicon; Vs <-> syntax?
  Word reconstruction experiments (van Ooijen, Cutler...)
  Aphasic patient data (Caramazza et al. 2000)
  Theoretical account (Nespor, Peña, & Mehler, 2003)
  Segmentation and generalization experiments with adults (Meher's group)
  Cs for words and Vs for rules with 12-month-olds (Meher's group)
  Tentative conclusions, Sven Öhman's insight
Children need to discover words and how to combine them: Words and Rules

But before children can analyze speech into words, what can they do?

– discriminate languages (rhythmic classes), speech sounds like /ba/-/pa/
– start to babble ...

In all these capacities, the SYLLABLE seems to play an important role

• the syllable as a basic unit** of production?
  – implicit in the “canonical” babbling definition (adult-like syllables)
  – explicit in MacNeilage and Davis’ Frame-then-Content model
  – in articulatory phonology: syllable = time frame wherein oscillatory systems are synched to produce consonant and vowel gestures
prelexical (production)

• global rather than detailed specifications for early syllables
  
  - MacNeilage F-then-C: 'content' comes later...  
    *pure frames, front frames, back frames, nasal frames*  
    fine-tuning appears later via biomechanical maturation and input
  
  - articulatory phonology's view of early syllables:  
    gross constriction gestures of the lips, tongue tip, tongue body,  
    imprecise gestural overlap of undifferentiated units, followed by:  
    C-V phasing, differentiation/individuation of syllable's components
  
• CV co-occurrence data support 'global'
  
  - account for *single place of articulation on the entire syllable*:  
    • simplified specification for F-then-C  
    • immature gestural overlap for articulatory phonology

=> early representations: syllabic, global, common fate of Cs and Vs?
prelexical (perception)

- newborns "count" syllables rather than phonemes (or moras)
  Bijeljac-Babic et al. (1993), *Developmental Psychology*: pre-/post-shift sets HAS

(a) they discriminate CV CV (rifu, kepa...) from CV CV CV (mazopu, rekiva...)

(b) they DO NOT discriminate 4-phoneme (rifu, iblo...) from 6-phoneme disyllabic items (treklu, suldri...)

(a) 2→3 or 3→2 syllables
(b) 4→6 or 6→4 phonemes
prelexical (perception)

• *salient* syllable structure favors speech-mode processing


(a) infants do not discriminate [tʃp]-[pʃt] (*these were not Tashlihyt infants...*)

(b) they discriminate [utʃpu]-[upʃtu]  
    as [utʃpu]-[upʃtu]?

(c) (control) they discriminate [tɒp]-[pɒt] alright

interpretation: same acoustic [tʃp] and [pʃt] processed differently:
- surrounding vowels let emerge a *salient* syllabic structure
- easy discrimination on those utterances parsed into syllables
- infants may tend to parse anything into spoken syllables...
Do infants parse syllables or treat them as whole-units?

“. . . My inclination is to suppose that the preliminary auditory segmentation (if any) is syllabic rather than phonemic and that within-syllable segmentation may often be synonymous with classification.” (Studdert-Kennedy 1979, ICPhS, Copenhagen)

Do infants discriminate /ba/-/pa/ because they discriminate /b/-/p/?

in other words:

(a) Do infants extract consonants from syllables?

Logic: if infants notice the common /b/, they won't react to an added /bu/

<table>
<thead>
<tr>
<th>Habituation test</th>
<th>2-m-olds react to whatever change</th>
<th>newborns only react to Vowel change (not to the da change)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ba bi be bo ber</td>
<td>bu new V</td>
<td></td>
</tr>
<tr>
<td></td>
<td>da new C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>du new C &amp; V</td>
<td></td>
</tr>
</tbody>
</table>

Bertoncini et al. (1988), *JEP:General*: HAS with preshift and postshift sets
prelexical (perception)

(b) Do infants extract vowels?

*Logic: if infants notice the common /i/, they won't react to an added /di/

\[
\begin{array}{c|c|c}
\text{Habituation} & \text{test} & 2\text{-m-olds react to whatever change} \\
\text{bi si li mi} & \{\text{di} & \text{new C} \\
& \text{ma} & \text{new V} \\
& \text{da} & \text{new C & V} \\
\end{array}
\]

newborns only react to Vowel change (not to the di change)

interpretation: infants do not parse syllables into phonemes; they rather treat syllables as whole-units; vowels seem more salient than consonants, just *for newborns*

caveat: recent data suggesting feature-generalizations (Cristiá)

We now turn to words: How do children represent words?
First words in production

Children "select" those words among adult words that correspond to the articulatory patterns, or routines, of their own babbling. (Vihman: "vocal motor schemes", "articulatory filter")

*Individual variability is the rule:* holistic vs. analytic children

<table>
<thead>
<tr>
<th></th>
<th>Emilie (14 mos, 15 words)</th>
<th>Marie (14 mos, 15-20 words)</th>
</tr>
</thead>
<tbody>
<tr>
<td>holist</td>
<td><strong>ba</strong> balle</td>
<td><strong>aettæe</strong> attend</td>
</tr>
<tr>
<td></td>
<td><strong>bo</strong> bouton</td>
<td><strong>hato</strong> bateau</td>
</tr>
<tr>
<td></td>
<td><strong>bebe</strong> bébé</td>
<td><strong>bebe</strong> bébé</td>
</tr>
<tr>
<td></td>
<td><strong>poe</strong> pomme</td>
<td><strong>dodo</strong> dodo</td>
</tr>
<tr>
<td></td>
<td><strong>po</strong> chapeau</td>
<td><strong>tebo</strong> c’est beau</td>
</tr>
<tr>
<td></td>
<td><strong>popo</strong> petit pot</td>
<td><strong>ebotsa</strong> c’est beau ça</td>
</tr>
<tr>
<td></td>
<td><strong>ka</strong> canard</td>
<td><strong>ta:tin</strong> tartine</td>
</tr>
<tr>
<td></td>
<td><strong>ke</strong> clef</td>
<td><strong>papitza</strong> papillon</td>
</tr>
<tr>
<td></td>
<td><strong>kki</strong> cuillère</td>
<td><strong>voajy</strong> voiture</td>
</tr>
<tr>
<td></td>
<td><strong>kRe</strong> Mickey</td>
<td><strong>hemjetsa</strong> mimichat</td>
</tr>
<tr>
<td></td>
<td><strong>qa</strong> sac</td>
<td><strong>popi</strong> poupée</td>
</tr>
</tbody>
</table>
From whole-word units to segments

Well illustrated by Macken's (1979) case study: Mexican Spanish "Si" from 1;6 to 2;5 (also see Macken 1992; Vihman 1997)

(1) whole-word units

- 1;7–1;9: only one word template (pattern, gabarit): labial–dental

  zapato → p\textipa{w\textipa{t}\textipa{\textbar{o}}} \quad \textit{Fernando} → \textipa{w\textipa{\textbar{n}\textipa{\textbar{o}}}}
  
  manzana → m\textipa{\textbar{\v{e}n\v{n}}}na \quad \textit{Ramon} → m\textipa{\textbar{\v{e}n}
  
  sopa → pw\textipa{\textbar{\v{e}t}a} \quad \textit{perro} → b\textipa{\textbar{\d{e}}\textipa{\d{e}}}
  
  reloj → bu\textipa{\textbar{\d{d}}}do \quad \textit{gato} → *kako (harmonic pattern)

- 1;10–1;11: new templates: m\textbar{s\textbar{\_}}, f\textbar{n\textbar{\_}}, p\textbar{l\textbar{\_}}, b\textbar{\v{e}n\textbar{\_}}, k\textbar{\_t\_}, \eta\textbar{\_t\_}...

(2) adult-like word-forms as strings of segments

- from 2;1; all Cs appear in words, with principled simplifications
<table>
<thead>
<tr>
<th>Phonological processes at 3-years (Vihman &amp; Greenlee, 1987)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syllable deletion</td>
</tr>
<tr>
<td>animals &gt; ˈæmz</td>
</tr>
<tr>
<td>...</td>
</tr>
<tr>
<td>Final consonant deletion</td>
</tr>
<tr>
<td>because &gt; piˈkʌ</td>
</tr>
<tr>
<td>...</td>
</tr>
<tr>
<td>Consonant harmony</td>
</tr>
<tr>
<td>yellow &gt; ˈlēlou</td>
</tr>
<tr>
<td>...</td>
</tr>
<tr>
<td>Cluster reduction</td>
</tr>
<tr>
<td>flower &gt; ˈfawr</td>
</tr>
<tr>
<td>...</td>
</tr>
<tr>
<td>Segment substitution</td>
</tr>
<tr>
<td>– velar, palatal fronting</td>
</tr>
<tr>
<td>cow &gt; tau, show &gt; sou</td>
</tr>
<tr>
<td>...</td>
</tr>
<tr>
<td>think &gt; fink /sink</td>
</tr>
<tr>
<td>...</td>
</tr>
<tr>
<td>– stopping</td>
</tr>
<tr>
<td>some &gt; tʌm</td>
</tr>
<tr>
<td>...</td>
</tr>
<tr>
<td>– gliding</td>
</tr>
<tr>
<td>love &gt; jʌv</td>
</tr>
<tr>
<td>red &gt; wed</td>
</tr>
<tr>
<td>...</td>
</tr>
</tbody>
</table>
Phonetic detail in *trained* word recognition at ~8 months

(Jusczyk & Aslin [1995], *Cognitive Psychology*)

6- and 7½-month-olds, HPP paradigm with familiarization

**Familiarization:** lists of tokens of either *{feet, bike}* or *{cup, dog}*

**Test:** passages with *{feet, bike}* vs. passages with *{cup, dog}*

7½- but not 6-m-olds recognize (=prefer) familiarized words

7½-m-olds fail to recognize "mispronounced" familiarized words:

```
dog → bawg
   cup → tup
feet → zeet
  bike → gike
```

**Conclusion:** rather detailed coding of newly learned word-forms
untrained familiar word recognition

- infant's own firstname: 4 ½ months (Jusczyk & Mandel, 1996)

- sound-image matching at 6 months:
  - father's face matched with dad or daddy,
    mother's face with mum (Tincoff & Jusczyk, 1999)

  NB. not very robust: 8/14 infants succeeded; not replicated

- words assumed to be familiar: 10-11 months (no-training HPP)

  found for: French, Japanese, British English, Welsh, Dutch

  (e.g., Hallé & de Boysson-Bardies, 1994, Infant Behavior & Development)
Looking times at familiar vs. rare words: British 9- & 11-m-olds

examples: familiar: button, balloon; rare: maiden, taboo

Vihman, dePaolis, Nakai, & Hallé (2004), J. Memory and Language
Phonetic detail of word representation in early receptive lexicon

- at the 11-month onset point, representation seem rather flexible
  method: no-training HPP with *mispronounced* words (MPs)

French 11-m-olds recognize: \textit{canard, chaussure}...

tolerate C1 mispronunciations: \textit{ganard, kaussure} ...

but not C1 deletion: \textit{anard, aussure}

but not C2 mispronunciations: \textit{calard, chauture} ...

British 11-m-olds recognize: \textit{button, dirty, balloon}...

tolerate C2 mispronunciations: \textit{busson, dirny}...

tolerate mis-stressed forms: \textit{buTTON, BAlloon}...

but not C1 mispronunciations: \textit{vutton, nirty}...
Phonetic detail according to syllable stress in English vs. French (Hallé & de Boysson-Bardies 1996, Vihman et al. 2004)

**interpretation**: less flexible (more detailed) coding of stressed syllables
Dutch infants on monosyllabic familiar words
Swingley (2005), *Developmental Science*

11-month-old Dutch infants, HPP, familiar monosyllabic words

Example:

\[
\begin{align*}
\text{mont} ('\text{mouth}') & \quad \text{(CP)} \\
(\text{highly familiar}) & \quad \left\{ \\
\rightarrow \text{nont} & \quad \text{(onset MP)} \\
\rightarrow \text{monk} & \quad \text{(offset MP)}
\end{align*}
\]

CPs and MPs compared with (Dutch) nonwords

- preference for CP over nonwords
- for CP over onset MP (but not over offset MP)

\[\Rightarrow \text{fine detail on consonants coded for familiar words?}\]

CP: correct pronunciation; MP: mispronunciation
Global representations at 11 months?

• limited degree of flexibility:
  (1) initial consonants of stressed syllables strictly specified
    (Vihman et al. 2004; Swingley 2005)
  (2) consonant skeleton must be there (Hallé & de Boysson-Bardies 2006)

• underspecification:
  (1) initial consonants of non-stressed syllables loosely specified
  (2) stress pattern: baLOON ≈ BAlloon (Vihman et al. 2004)

• different coding for newly learned words?
  C1 contrast ignored in some word-learning studies:
  Example: word pair dih-bih not learned at 14 months
    (Stager & Werker 1997, Nature)

=> review of this line of research
"switch" procedure with visual fixation (VF) (learning word–object associations)

habituation

lif lif lif lif lif lif

dishabituation

neem ...

expected if association learned: VF recovery after switch
Phonetic detail in lexical representation: from 14 month onward  
(Werker, Swingley, Plunkett, Nazzi, etc.)

(a) *Werker's group: switch procedure, word-learning*

- "easy" switch: *lif* → *neem* detected at 14 months  
  (not before 14 mos: Werker et al. 1998, *Dev. Psy.*)

- one-feature change *dih* → *bih* detected at 17 mos, not at 14 mos  

- two-feature change *pin* → *din* not detected at 14 months  
  Pater, Stager, & Werker 2004, *Language*)

- */d/* → */b/* switch detected *for familiar words* (*doll* → *ball*) at  
  14 months (Fennel & Werker 2003, *Language & Speech*)

=> suggests coding in word-learning harder than in known words
Phonetic detail in lexical representation: from 14 month onward

(b) *Swingley*: preferential looking procedure, word-recognition
   correct pronunciation (CP) and mispronunciation (MP) of a word,
   visual choice between matched picture and distractor picture

• *dog* and *tog*: longer looking time (LT) to target than distractor
  *but longer LT for dog than tog*
  at 18-23 months (Swingley & Aslin 2000, *Cognition*)

(c) *Plunkett* (same procedure)
   MP unrecognized: children don't look at the dog's picture for *tog*

=> phonetically detailed representations
back to phonetic detail from 14 months onward

• Werker and Swingley: only looked at consonant MPs *

• Nazi, then Plunkett, looked at phonetic detail in vowels

  - "name-based categorization" data (Nazzi 2005, Cognition):

    performance <-> success at learning contrasted word-object pairings

    **C-contrast > V-contrast**  
    (65% > 54% ≈ chance)

    Example:  
    pize-tize, pize-pyze  
    (French 20-month-olds)

  

=> Vs coded more loosely than Cs in children's lexicon?
name-based categorization

(1) Learning 3 word-object pairs:
   - two objects with identical label
     (here, two 'toto's and one 'dada')

(2) Test phase:
   - one 'toto' shown to child:
     'donne celui qui "va avec"'
Impact of vowel versus consonant mispronunciation

– Plunkett's group: PL, familiar words' CP vs. MP (Mani & Plunkett 2007)

Example: *bib* (CP), *bab* (MP-V), *dib* (MP-C)

15-, 18-, and 24-month-old English children fail to look at target when hearing MPs; 15-m-olds tend (ns) to look at target for MP-Vs

\[\Rightarrow\] Contradicts Nazzi 2005?

dev. trend: Vs first ignored, later as important as Cs

measures = increase of %looking at target (PTL); of longest looking time (LLK) at target, after target is named.

shown: LLK data
Discrepancies between Nazzi and Plunkett's data: Why?

- very different experimental procedures (PL vs. NBC)
- word-recognition vs. word-learning
- one vs. two syllable words, English vs. French

- vowel MP effects found in word-learning, English 14-m-olds
  *padge* --> *poudge, mot* --> *mit* (Mani & Plunkett 2008)

- vowel *switch* in *deet–dit* (but not *deet–doot*) detected by English 14-m-olds (Curtin et al. 2009; also see Dietrich et al. 2007)

- *V contrast* ignored and *C contrast* learned in French 16-m-olds (simplified NBC: Havy & Nazzi 2009); 30-m-old French and English children tolerate better *V* than *C* variation (Nazzi et al. 2009)

  => unclear picture, methodological issues...
More support for a lesser weight of Vs in the lexicon

• AE and JE (Jamaican English) differ mainly on vowels:
  hid, head, hood: /ɪ, ɛ, ʊ/ (AE) vs. /i, e, u/ (JE)
  had, hawed, hod: /æ, ɔ, a/ (AE) vs. /a, a, a/ (JE)

• Preference for familiar words across AE and JE:
  found with 19- but not 15-month-old AE children

=> At 19 months, children ignore dialectal variation in vowels

developmental trend:
young children sensitive to V variations, older ones tolerate V variation in recognizing words
(Best et al. 2009 Psych. Science: phonological constancy)
Consonant/Vowel asymmetry in adults

- **word reconstruction (WR) data:**
  Cs more important to maintain than Vs (or Vs' greater mutability)

  *kebra* --> *cobra* more often than *zebra* (Cs not Vs maintained)
  (Cutler at al. 2000; van Ooijen 1996: *Memory and Cognition*)
  found for English, Spanish, Japanese, Dutch: ~70% > 30%

- PET study of WR
  *unsane* --> *insane/*unsafe
  more activation for C than V reconstruction in anterior left IFG (BA 45 & 47) and in PMC (BA 6)
  (Sharp et al. 2005, *B&L*)
Consonant/Vowel asymmetry in adults

- *aphasic patient data* (double dissociation)

AS: more errors on Vs than Cs; IFA: more errors on Cs than Vs (Caramazza et al. 2000, *Nature*)
Consonant/Vowel asymmetry in adults

• **quantitative facts**
  - acoustically, Vs much more variable than Cs
  - categorical perception of C quality, continuous of V quality

• **theoretical considerations**
  (Nespor, Peña, & Mehler 2003, *Lingue e Linguaggio*)
  - more Cs than Vs across languages => Cs bear more info
  - Cs tend to disharmonize, Vs to harmonize *within words*
  - Cs tend to alternate in quality, not often in quantity as Vs
  - Vs not required to alternate in quality (e.g., *banana*)
  - Cs, not Vs may constitute lexical roots (Semitic languages)
  - **C tier has a lexical motivation**, **V tier motivation** is prosodic in nature
    (McCarthy 1985; Goldsmith 1976)
Consonant/Vowel asymmetry in adults 
(word segmentation data from detection of recurrent patterns)

Saffran et al. 1996: learning "words" from a stream of syllables with manipulation of syllable transition probabilities (TPs)

S-words: $S_2S_3S_4$ $S_5S_6S_7$ ... defined by TP "dips"

... $S_1$ $S_2$ $S_3$ $S_4$ $S_5$ $S_6$ $S_7$ ...

...bidakupadotigolabubidakugolabubidakupadoti...

Both 8-month-olds (HPP) and adults (forced-choice) succeed in "segmenting" S-words: (e.g., golabu >> datigo)

they have learned $S_2S_3S_4$ rather than e.g. $S_4S_5S_6$
Word segmentation experiments

- Elaboration: TPs between Cs or between Vs (Bonatti et al. 2005)

C-words: \texttt{p_r_g_} \texttt{b_d_k_} \texttt{m_l_t_}

success

(87.7 > chance)

V-words: \texttt{_ô_i_а} \texttt{_o_ẽ_y} \texttt{_u_e_ã_}

failure

(54.2 \approx \text{chance})
segmentation versus generalization

- *segmentation into “words” seems much easier for C- than V-words*

Fits well with the idea that Cs are essential for coding lexical items

- pattern generalization (e.g., ABA: 1st element = last element)

  example

  Familiarization: b_d_k_ words with ABA V pattern (e.g., badeka, bedoke)

  Segmentation test: C-word vs. part-word (e.g., badeka vs. naboda)

  Generalization test: rule-word vs. nonrule-word (e.g., budiku vs. biduku)

  generalization found for ABA V patterns

  NOT found for ABA C patterns or even AAA C patterns

Hochman, Benavides, Nespor, & Mehler (submitted)

Experiment A: word-learning (12-months)

(1) **familiarization**: kuku <-> \(A\) and dede <-> \(B\)

(2) **test with novel "words"**: keke and dudu

**notations**

- \(\text{C}_C\_\) learned => keke with \(A\) and dudu with \(B\) (C-looks)
- \(\text{V}_V\_\) learned => keke with \(B\) and dudu with \(A\) (V-looks)

Score: \((\text{C-looks} - \text{V-looks})/(\text{C-looks} + \text{V-looks})\)

**RESULTS**: Score = 0.23 >> 0 => \(\text{C}_C\_\) learned more than \(\text{V}_V\_\)
Experiment B: extraction of a regularity

(1) familiarization: \{lula, lalo, fufa, fofu, dado, dodu\} <-> \{dala, dolo, fudu, fodo, lafa, lufu\} <->

(2) test with novel "words": \{meke, kimi\} and \{kike, memi\}
(new Cs and Vs)

- for \{kike, memi\}, infants do not look more to (A) than (B):

  \[
  \frac{A-B}{A+B} = -0.16 \(^\circ\)
  \]

- for \{meke, kimi\}, infants look more to (B) than (A):

  \[
  \frac{B-A}{A+B} = +0.60 (***)
  \]

=> V1=V2 is learned, not C1=C2
Two mechanisms seem to help discovering *words and rules*

- *statistical learning* helps to find *words*
  - seems to rely on the distribution of *C* rather than *V* co-occurrences in the input; *large input required* (e.g., each "word" repeated 50 times)*

- a *generalization mechanism* helps to find *rules*
  - seems to rely on the regularities across *Vs* rather than *Cs*
  - precise implementation unknown, needs further research
  - *fast extraction* of regularities (Peña et al. 2002)**
Some thoughts on Cs and Vs and representations

• the C-V asymmetry line of research is consistent with the autosegmental account of two independent C and V tiers

**Vowel tier:** domain/locus of prosodic processes such as tone spreading

**Consonant tier:** main specification of lexical items*

• also consonant with the idea that the speech stream is a stream of vowels on which consonants are coproduced

The essential features of the coarticulation properties of Swedish dental stops in vowel-consonant-vowel contexts can be described by the formula $s(x; t) = v(x; l) + k(t)[c(x) - v(x; t)]wc(x)$ ... Vocal tract shapes measured from x-ray motion pictures of a set of Swedish vowel-consonant-vowel utterances compare well with shapes generated by the formula. This result is consistent with our earlier conclusions about coarticulation, viz., that the vowel and consonant gestures are largely independent at the level of neural instructions. (Sven Öhman (1967), JASA)
Merci pour votre attention
Sketch

-- Whether articulatory or acoustic “goals,” the goals of speech production are what psycholinguists would call “representations.” A debated issue is of whether the representations for perception are similar, an issue, I believe, that will be much addressed during this summer school. In this talk, I will address the issue of speech representations in infants but will mainly focus on perception and lexical access.

-- Most researchers agree that the syllable is a basic unit of production in prelexical children; this is implicit in many descriptions of the babbling stage, and quite explicit in MacNeilage and Davis’ Frame-then-Content model. The articulatory phonology approach would also hold that the syllable is the time frame wherein oscillatory systems are synched to produce consonant and vowel gestures.

-- Likewise in perception, the syllable seems to play a primary role. The capacity of newborns to discriminate most phonemic contrasts, even nonnative ones, does not necessarily entail the phoneme is a unit of representation for them. Indeed, in the discrimination experiments using HAS, CHT, or other paradigms, infants are usually tested on syllables, for example on /ba/-/pa/: we don't know about exactly /b/-/p/!

-- Yet, a few clever experiments have tried to examine whether infants decompose syllables into consonants and vowels, or whether they “count” syllables rather than phonemes or moras. The available data point on syllables rather than anything else.

-- Now what about the lexical stage? Do young children who start a productive or a receptive lexicon represent words (as production targets or as recognized spoken items) as composed of syllables or of something else?
Child phonology studies have proposed that children first go through a whole-word stage during which the word is the basic unit (the "prosodic word": Macken, 1979; also see Vihman, 1997). Children would then develop more adult-like phonological representations,* that is, rule-like representations gradually leading to principled segmental units. For example, some children, usually during the second year, develop a few (possibly just one) consonant-vowel templates, or word patterns, followed by all their attempted words. This whole-word stage is followed by a segmental stage.

For word recognition, there is an analogous though less radical claim that infants begin with rather holistic or flexible representations and later move to phonetically more detailed representations, presumably under the pressure of a growing lexicon (Hallé & de Boysson-Bardies, 1996; Vihman et al., 2004).** Studies with older children (≥14 mos) tend to show they are sensitive to phonetic detail. A few studies suggest that Cs are more important than Vs in lexical representations. If confirmed, this trend would be fully consonant with adult studies on C/V asymmetries.

Classic word reconstruction studies, as well as recent adult studies suggest that the consonant tier mainly codes lexical units whereas the vowel tier is involved in prosody-related rule extraction (Mehler and Nespor’s group; similar findings seem to hold for 12-m-old children, Hochman et al., submitted).

At stake in this line of research is the notion that two mechanisms coexist in language acquisition: discovery of *words* via distributional statistics on Cs, and of *rules* via regularities in V patterns.
Extra materials
words (production)

pre-phonological strategies

(2) "harmonic patterns" = simplification via C (or V) harmony*

Examples:

- chapeau → papo
- gâteau → tato
- canard → nanar

* More often, regressive harmony
  (e.g., chapeau → papo rather than chapeau → chacho)
Example from de Boysson-Bardies (1996):

“Henri” (16 mos) systematically replaces /m/ with either /b/ or /p/ in /m/-voyelle-consonne-voyelle words, according to the rule:

\[
/m/ \rightarrow \begin{cases} 
/b/ & \text{if C voiced} \\
/p/ & \text{if C voiceless} 
\end{cases}
\]

voiceless /s/:  \textit{monsieur} \rightarrow \textit{peussieu}

voiceless /ʃ/:  \textit{méchant} \rightarrow \textit{pécha}

voiced /z/:  \textit{musique} \rightarrow \textit{bizik}
Increasing phonetic detail in lexicon: pressure of vocabulary size?

- **the pressure of the growing vocabulary argument**
  - irrelevant variation can be ignored in early lexicon (sparse population)
  - but rapidly many distinctions become relevant

> "...representations of lexical items may become increasingly segmented (phonemic) with development from the pressure of an increasing vocabulary size. Young children may represent *only those distinctions that are necessary* for word recognition. ... Words that have *many similarly sounding neighbors* may be forced to become *phonemically represented* chronologically earlier than words that do not have to be discriminated by many similarly sounding word neighbors." (Metsala 1997, *Memory and Cognition*, p. 161)*

- **How do representations get refined?** (parenthetical issue!)
  - item-specific coding or rule-like generalization?
  
  Example: (1) only *pin* $\Rightarrow$ *pin=kin=bin*  
  (2) *kin* learned $\Rightarrow$ *pin≠kin* but what about *bin*?
Qualification of the neighbors' pressure account

• (Dutch) bal ('ball') mispronounced dal or gal (Swingley 2003, L&S)

/d/ frequent, /g/ very rare => children likely have heard dal not gal

(a) specific-item coding view: bal ≠ dal, but still bal = gal

=> CP ≈ MP-g > MP-d

(b) generalization view: [b], [d], and [g] = different segments

(feature or gesture generalization, however infrequent is /g/)

=> CP > MP-g ≈ MP-d

found:
CP > MP-d ≈ MP-g

=> (b) rather than (a)

(18-m-olds)