Developing acoustic measures to evaluate the emergence of phonological contrast

Mary E. Beckman (Ohio State University)
Jan Edwards (University of Wisconsin -- Madison)
Benjamin Munson (University of Minnesota)

Acknowledgments:
• NIDCD grant RO1 02932 (Jan Edwards)
• Ohio State University Targeted Investment in Excellence award (Eun Jong Kong & Fangfang Li)
• McKnight Presidential Fellowship (Benjamin Munson)
• NSF grants BCS 0729306 (Mary Beckman), BCS 0729140 (Jan Edwards), & BCS 0729277 (Benjamin Munson)
• OSU Center for Cognitive Science seed grant (Mary Beckman, Mikhail Belkin, & Eric Fosler-Lussier)

Even before there were tape recorders ...

Symbolic transcription of young children’s productions:
1) uncovered several common cross-linguistic trends.
   • for example, for voicing or aspiration contrasts, ... “p” / “t” mastered before “b”/“d” or “pʰ” / “tʰ” so that, e.g., French-learning child is transcribed as saying touche ‘tag’ for douche ‘shower’
2) confirmed robust language-specific “exceptions”.
   • for example, in English, the stereotypical stop in canonical babbling and early stop-initial words is transcribed as “b” or “d” rather than “p” or “t” (cf. Darwin 1877)

Early acoustic analyses explains both

Category with short lag VOT first, because it requires the least precise articulation (Kewley-Port & Preston 1974).

What VOT has taught us ...

• Development is much more gradual than would seem from transcription data alone.
• Children may be perceived as incorrect even when they are beginning to make a distinction:
  • Macken and Barton (1980) use VOT to show “covert contrast” between English short-lag [d] and “not quite so short” [t] transcribed as [d] at 18-22 months.
  • We need to look across languages in order to understand how community perceptual norms (as well as the intrinsic “articulatory difficulty”) influence the emergence of contrast.

<table>
<thead>
<tr>
<th>Language</th>
<th>Short lag Achieved</th>
<th>Long lag Achieved</th>
</tr>
</thead>
<tbody>
<tr>
<td>English (see below)</td>
<td>Vowled</td>
<td>Vowled</td>
</tr>
<tr>
<td>French (after 7 mos)</td>
<td>Vowled</td>
<td>Vowled</td>
</tr>
<tr>
<td>Cantonese (see below)</td>
<td>Vowled</td>
<td>Vowled</td>
</tr>
<tr>
<td>Thai (Statement et al)</td>
<td>Vowled</td>
<td>Vowled</td>
</tr>
</tbody>
</table>

The παθολογος project data
The παιδολογός project data design

- Productions elicited of analogous sounds in analogous word positions across languages, using same task and same recording equipment.
- Large number of children (100+) for each target language, covering same age range (2 through 5 years).
- Transcribed using comparable two-stage transcription protocol: (1) correct vs. incorrect & (2) perceived substitution (θ for s), with intermediate types (θ:s).
- Recordings available for continuing acoustic analysis and as a source of stimuli for perception experiments, shared at http://childes.psy.cmu.edu/data/PhonBank/

Explaining other apparent exceptions

- Japanese children produce lead VOT values at 4 years.
- Greek children have lead VOT values as early as 2 years.
- Kong (2009) adapted the acoustic model from Burton, Blumstein, & Stevens’s (1972) study of the Moru language contrasts among [n], prenasalized [d], & [d].

Fricative development (from Li et al., 2009)

- Both English and Japanese have a contrast between alveolar / dental [s] and postalveolar / alveolopalatal [ʃ].
- English [s] mastered earlier than [ʃ] and [s] substitutes for [ʃ] (Smit et al. 1991) — i.e., a “fronting” stereotype.
- Japanese [ʃ] mastered earlier than [s] and [ʃ] substitutes for [s] (Nakanishi et al., 1972) — i.e., a “backing” stereotype.

Articulation of Japanese [s] and [ʃ]

- Whereas English [s] is alveolar and often apical, Japanese [s] is lamino-dental (left panel).
- Whereas English [ʃ] is a rounded apical postalveolar, Japanese [ʃ] is a lip-spread alveolopalatal (right).

Acoustic measures for sibilant contrasts

- While centroid not ideal (cf. Shadle et al., this session), it has helped to explain language-specific stereotypes.
Differences in children’s productions

<table>
<thead>
<tr>
<th>Age (months)</th>
<th>Centroid (Hz)</th>
<th>F2 Onset (Hz)</th>
<th>St. Dev. (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japanese</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 6.3 from Li (2008).

Adult “parsing” of children’s productions

- The community norms for the articulations and acoustic cues to the [s]-[ʃ] contrast differ somewhat between English and Japanese.
- Some English-speaking children who are transcribed as substituting [s] for target [ʃ] produce F2 onset frequencies that are appropriate for Japanese [ʃ].
- Could differences in community norms for adult perceptual parsing of the children’s productions also contribute to the different stereotypical substitutions?

Effect of experience on language-specific consensus responses (70%+ “yes”)

- Pattern of relationship to cues differs between English- and Japanese-speaking adult listeners for same tokens.

Related language-specific covert contrast

- At least some English-acquiring children show higher F2 onset in their stereotypical [s] substitutions for [ʃ].

Visual Analog Scale

- The Li et al. (2011) paired questions method requires two trials per stimulus.
- Also, the interpretation of “no” responses is difficult.
- Urberg Carlson, Kaiser, and Munson (2008) developed an alternative method that uses a Visual Analog Scale (VAS) to probe adult perception continuously.

The participant responds by clicking appropriately on the VAS responses related to acoustic cues
Moving beyond moments (Reidy, in progress)

- Excitation pattern as filtered by outer and middle ear at middle 40 ms of “correct [s]” versus “correct [θ]”.

Moving beyond moments (Reidy, in progress)

- Measure the compact quality of [s] versus diffuse [θ] by computing dB above 20 ERB relative to dB below.

Relating to transcription and to VAS

- Relating the transcriptions, the acoustic analyses, and the results of perception studies with the ραιδολογις recordings shows value of cross-language comparison across children recorded at a wide range of ages.
- Work is in progress on developing psychoacoustic measures that might be a closer match to the adult community norm responses to children’s productions.
- Work is also in progress to explore acoustic measures in relationship to age-appropriate articulatory models.
- Work is beginning to create a longitudinal database.

Summary and what’s next

Watch for results at http://www.learningtotalk.org