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-Lusser)

Even before there were tape recorders …

Symbolic transcription of young children's productions:

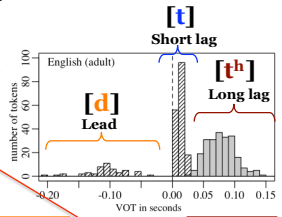
1) uncovered several common cross-linguistic trends.
   - for example, for voicing or aspiration contrasts, …
     "p" / "t" mastered before "b"/"d" or "p^h" / "t^h"
     so that, e.g., French-learning child is transcribed as saying *touche* 'tag' for *douche* 'shower'

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).

short lag achieved by glottis opening at any time during the oral occlusion: easy to produce!



Language	Lead	Short lag	Long lag
English (Macken & Barton 1980a)	voiced	voiced	voiceless
French (Allen 1985)	voiced	voiceless	
Cantonese (Clumeck et al 1981)		unaspirated	aspirated
Thai (Gandour et al 1986)	voiced	unaspirated	aspirated

What VOT has taught us …

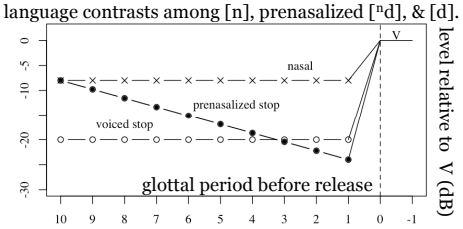
- Development is much more gradual than would seem from transcription data alone.
- Children may be perceived as incorrect even when they are beginning to make a distinction:
  - Macken and Barton (1980) use VOT to show "covert contrast" between English short-lag [d] and "not quite so short" [t] transcribed 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.

The παιδολογος 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

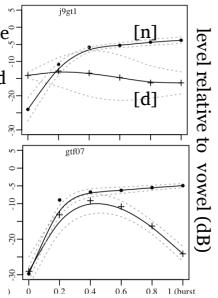
- Japanese children produce lead VOT values at 4 years.
- Greek children have **lead** VOT values as early as 2 years.
- Kong (2009) adapted the acoustic model from Burton, Blumstein, & Stevens's (1972) study of the Moru language contrasts among [n], prenasalized [ⁿd], & [d].



Japanese versus Greek "voiced" stops

typical Japanese adult male has the Moru voiced stop pattern

typical Greek adult male has the Moru prenasalized stop pattern



- Most Greek children's stops have lead VOT, and they look even more nasalized than the adults'.

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.
  - *shoe*
  - *safe*
- Japanese [ʃ] mastered earlier than [s] and [ʃ] substitutes for [s] (Nakanishi et al., 1972) -- i.e., a "backing" stereotype.
  - *shukurimu* 'cream puff'
  - *semi* 'cicada'

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).

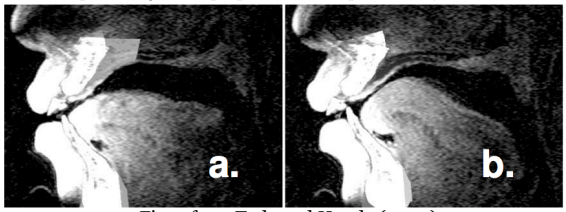


Fig. 2 from Toda and Honda (2003).

Acoustic measures for sibilant contrasts

- While centroid not ideal (cf. Shadle et al., this session), it has helped to explain language-specific stereotypes.

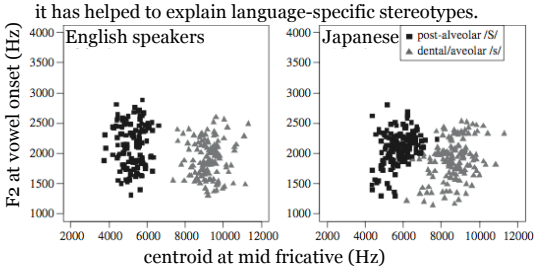


Fig. 3 from Li, Edwards, & Beckman (2009).

Differences in children's productions

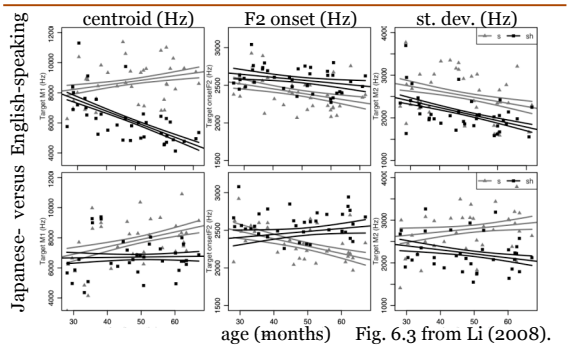


Fig. 6.3 from Li (2008).

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?
- Li, Munson, Edwards, Yoneyama, and Hall (2011) test by asking 19 English- and 20 Japanese-speaking adults (1) "Is it the 's' sound?" and (2) "Is it the 'sh' sound?"

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.

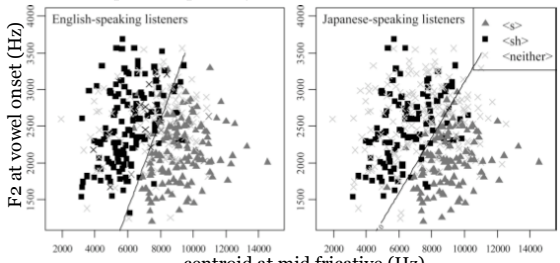
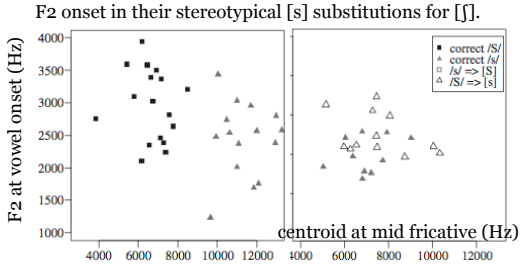


Fig. 2, Li et al. (2011).

Related language-specific covert contrast

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



Figs. 4a (left) and 6a (right) from Li et al. (2009).

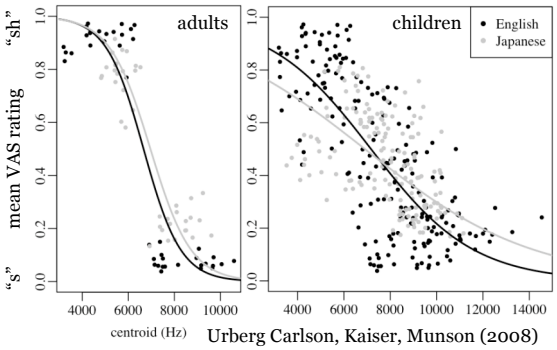
Visual Analog Scale

- 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 "s" sound ⟷ the "sh" sound

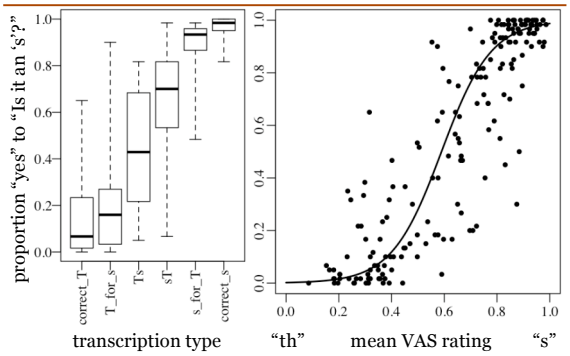
participant responds by clicking appropriately on arrow

VAS responses related to acoustic cues

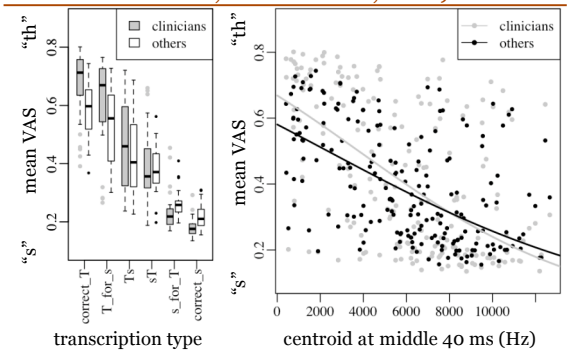


Urberg Carlson, Kaiser, Munson (2008)

[θ]~[s] contrast (Schellinger, 2008)

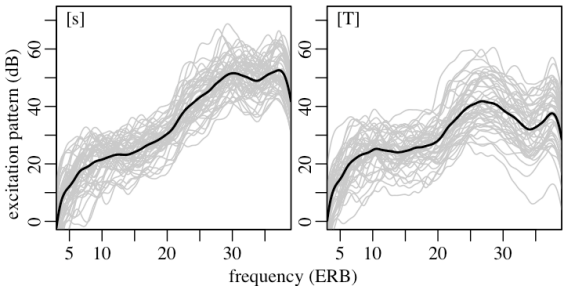


Effect of clinical experience (Munson, Johnson, and Edwards, 2010)



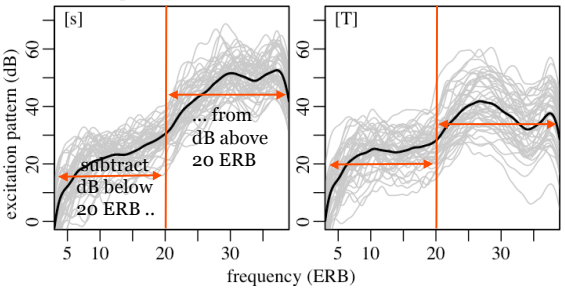
Moving beyond moments (Reidy, in progress)

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

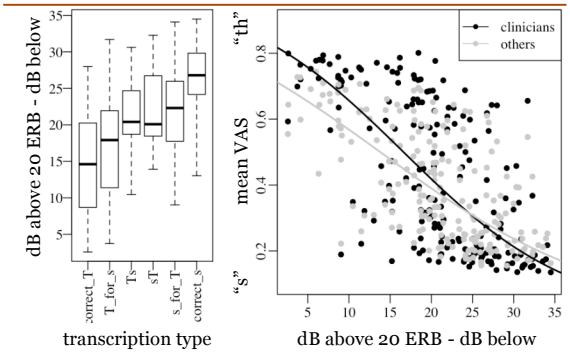


Moving beyond moments (Reidy, in progress)

- 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



Summary and what's next

- Relating the transcriptions, the acoustic analyses, and the results of perception studies with the παιδολογος 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.

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