Assessment of Phonetic Skills in Children 3: Fidelity of Responses under Different Levels of Task Delay

Benjamin Munson[1], Eden Kaiser[2], and Kari Urberg Carlson[1]

[1]Department of Speech-Language-Hearing Sciences, [2]Program in Linguistics, University of Minnesota, Minneapolis

Children acquire speech sounds gradually

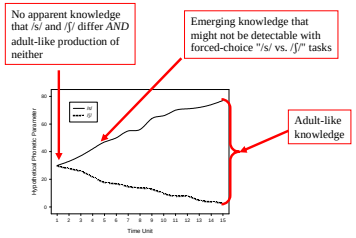
- Cross-sectional and longitudinal studies of speech-sound development often show an apparently discontinuous acquisition of speech sounds (e.g., Smith, 1973)
- Acoustic and articulatory studies suggest that children's speech sound acquisition involves the *gradual* acquisition of speech-sound contrasts. Some representative studies of this phenomenon are:
  - Macken and Barton (1980): Children gradually learn to differentiate English voiced and voiceless initial stops along the voice-onset time dimension.
  - Edwards, Gibbon, and Fourakis (1997): children with velar fronting patterns may differentiate between target /k/ and [t] for /k/ substitutions acoustically
  - Li, Beckman, and Edwards (2008): Children gradually differentiate between /s/ and /ʃ/ along two acoustic parameters, spectral center of gravity and frequency of the second formant at the onset of the following vowel.

Most assessment tools elicit categorical judgments

- In most clinical assessment regimens and experimental research protocols, adults listen to a child and transcribe their production using a phonetic symbol and an optional series of diacritic markings.
- These transcriptions do not allow us to examine directly children's gradual attainment of speech sounds.
- The widespread use of phonetic transcription is arguably the by-product of two factors
  - 1. The apparent ease and simplicity of the task of phonetic transcription
  - 2. The belief, based on studies of categorical perception, that people cannot perceive fine phonetic detail within a phoneme category.

People *can* perceive fine phonetic detail in children's speech

- Urberg Carlson, Kaiser, and Munson (2008, this conference) and Schelling, Edwards, Munson, and Beckman (2008, this conference) showed that adults can perceive fine phonetic detail within a speech-sound category when an appropriate task is used.
- Using Visual Analog Scaling tasks, we showed that adults can perceive fine phonetic detail in children's attempts to produce the /s/-/θ/ contrast (Schellinger et al.) and the /s/-/ʃ/ contrast (Urberg-Carlson et al.).
- Urberg-Carlson et al. argue that VAS tasks were superior to two other continuous-rating methods, Direct Magnitude Estimation of category goodness, and reaction times in phoneme categorization tasks.



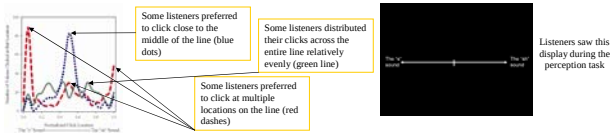
Why should we care?

•These findings are evidence that children's phonological category learning begins well before it is perceptible to adults.

•This has implications for prognosis in speech-sound therapy (Tyler, Figourski, & Langsdale, 1993), with potential implications for goal selection.

Adults vary in how categorically they rate children's speech

- The listeners in Urberg Carlson et al. varied in the extent to which they provided continuous responses in the VAS /s/-/ʃ/ rating task, as shown below

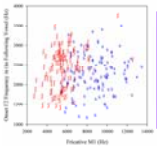


Why?

- There are numerous reasons why these differences might have occurred. This poster examines one possibility, namely, that the more-categorical listeners were those who waited longer between hearing the sound and making their ratings. This delay promoted people's parsing fine phonetic detail into categories. Our research question is **Do listeners perceive children's speech more categorically when a long delay is enforced between hearing a speech token and rating it?**

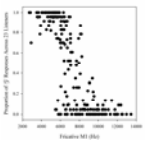
We used /s/ and /ʃ/ productions of 2- to 5-year old typically developing children, and adults

- The stimuli were elicited using a real-word repetition task, where to-be-repeated prompts were presented concurrent with pictures of the objects that they were naming. The speech tokens were collected as part of the παιδολογος project (http://www.ling.ohio-state.edu/~edwards, Jan Edwards, PI).
- Stimuli were 250 initial CV sequences excised from whole words. They included target /s/ and /ʃ/, and both [s]-for-/ʃ/ and [ʃ]-for-/s/ substitutions



This is a plot of the 250 fricatives (/ʃ/ and /s/) in the M1-by-onsetF2 auditory-acoustic space. Details of this acoustic analysis are reported in Li (2008).

A separate experiment collected forced choice /s/ vs. /ʃ/ judgments. These are shown in this figure, plotted against fricative M1. Further details about that experiment can be found in Urberg Carlson et al. (this conference) and Li (2008).



Details of the experiment…

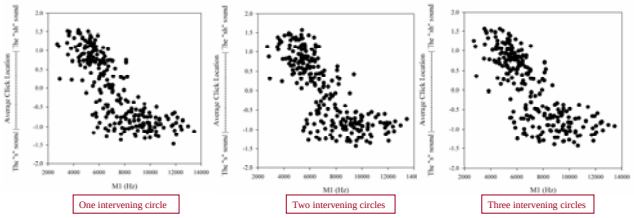
1. Listeners (24 native, monolingual speakers of American English, including both university students and members of the community) heard a speech token

2. Listeners saw either one, two, or three screens with red circles that they were instructed to click on. One third of the speech tokens were followed by one circle, one-third by two circles, and one-third by three circles. The assignment of tokens to condition was counter-balanced across listeners.

3. Listeners clicked on the above line to indicate where they thought the token fell on the above line.

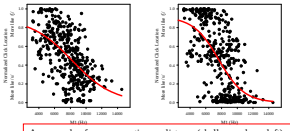
What did the data look like?

- The three plots below show the average standardized (z-score) click locations (pooled across the 24 listeners) plotted against the fricative's first spectral moment. This plots showed that click location was very strongly related to M1.



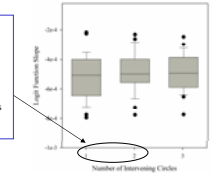
We assessed degree of categoricity of response using logit functions

- For each of the 24 listeners, separate logit functions were fit to the M1-by-click location plots for the one-, two-, and three-circle conditions. The slope of the logit function was taken as a measure of categoricity of response, with steeper slopes indicating more categorical responses. See the figures below for examples of more- and less-categorical responders.
- Slopes were submitted to a non-parametric Friedman's ANOVA by ranks. The slopes did not differ significantly across delay intervals ($\chi^2_{[df=2,\ n=24]} = 1.75$, $p = 0.417$, see figure below). Crossover points also did not differ as a function of delay interval.



An example of a more-continuous listener (shallower slope, left) and a more-continuous listener (steeper slope, right). These listeners are from Urberg-Carlson et al.

The difference between the one- and two-circle conditions did approach significance ($z = -1.174$, $p = 0.086$ in a Wilcoxon Signed Ranks test), but it was in the opposite-than-expected direction.



Our conclusion: perception of fine phonetic detail in children's speech is robust across different levels of task delay

- We can be confident that listeners' perception of phonetic detail would be robust to the kinds of delays that happen in real-world assessments of speech production.
- We continue to research why there are reliable differences across listeners in how categorically they respond to speech tokens.

Acknowledgements

This research was supported by NSF Grant BCS0729277 to Benjamin Munson, and by NIH grant R01 DC02932 to Jan Edwards. We thank Fangfang Li for conducting the acoustic analysis of the tokens. We thank Alyssa Zitzman and Celina Marsie for help running subjects. We thank Rana Heretz and Eric Foster-Lanzer for feedback on this experiment and for help contracting the logit analyses. We thank Mary Beckman and Jan Edwards for their extremely useful input on this work.

References

Edwards, J., Gibbon, F. and Fourakis, M. (1997). On discrete changes in the acquisition of the alveolar/velar stop consonant contrast. *Language and Speech*, *40*, 203-210.

Li, F. (2008). Dissertation in Progress. Department of Linguistics, Ohio State University.

Li, F., Beckman, M., and Edwards, J. (2008, in press). Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers. *In press in Journal of Phonetics*.

Macken, M., and Barton, D. (1980). A longitudinal study of the acquisition of the voicing contrast in American English word-initial stops, as measured by voice onset time. *Journal of Child Language*, *7*, 41-74.

Smith, N. (1973). *The Acquisition of Phonology: A Case Study*. Cambridge: Cambridge University Press.

Tyler, A., Figueroa, G., and Langdale, T. (1993). Relationships between acoustically determined knowledge of stop place and voicing contrasts and phonological treatment progress. *Journal of Speech, Language, and Hearing Research*, *36*, 746-759.