

WHY DO ADULTS VARY IN HOW CATEGORICALLY THEY RATE THE ACCURACY OF CHILDREN'S SPEECH?

Eden Kaiser¹, Benjamin Munson², Fangfang Li^{3,4}, Jeffrey Holliday⁴, Mary E. Beckman⁴, Jan Edwards⁵, and Sarah Schellinger^{2,5}

¹Program in Linguistics, University of Minnesota; ²Department of Speech-Language-Hearing Sciences, University of Minnesota; ³Department of Psychology, University of Lethbridge; ⁴Department of Linguistics, Ohio State University; ⁵Department of Communicative Disorders, University of Wisconsin-Madison

Background

Children acquire speech sounds gradually

- Phonetically transcribed speech samples suggest that speech sounds are acquired discontinuously (e.g., Smith 1973)
- However, instrumental studies show that acquisition is in fact gradual (e.g., Macken & Barton, 1980; Edwards, Gibbon, & Fourakis, 1997; Li, Beckman, & Edwards, 2008)
- Phonetic transcription, the most common analysis tool used to study phonological acquisition, limits the degree to which speech sounds can be represented continuously

Most assessment tools elicit categorical judgments

- Clinical assessments and experimental research on phonological development often use broad transcriptions made by native speakers
 - Easier, simpler, and faster than the alternatives
 - Based on the assumption that people can't perceive within-category distinctions
- These transcriptions don't allow tracking of children's gradual acquisition of speech sounds
- This poster is part of a larger set of experiments examining novel methods for measuring children's speech production

People can perceive fine phonetic detail in children's speech

- Many people can perceive fine differences in sounds' acoustic characteristics within categories (e.g., Kong, 2009; Munson, Li, Yoneyama, Hall, Beckman, Edwards, & Sunawatari, 2008; Urberg Carlson, Munson, & Kaiser, 2009; Urberg Carlson, Kaiser, & Munson, 2008)
- Most people use fine phonetic detail within categories to identify attributes about a speaker, such as age, social class, gender, sexuality, social group identification, and mood (Crocker & Munson, 2006; Drager, 2008; Munson, McDonald, DeBoe, & White, 2006; Smith, Munson & Hall, 2008)

Motivation for the current study

In Munson, Kaiser, and Urberg Carlson (2008), adults judged the accuracy of Japanese- and English-acquiring children's speech

- Participants listened to two- to five-year-old children's and adults' CV productions of /s/ and a postalveolar sibilant fricative (/ʃ/ or /ç/)
- They rated how "s-like" or "sh-like" each sound was by clicking on a double-headed arrow anchored with "the 's' sound" and "the 'sh' sound" at either end
- They indicated their judgments on a visual analog scale (VAS), similar to the one in Figure 3

Two different types of response patterns emerged

- Categoricity of response was assessed by logit functions predicting click location from fricatives' M1, as illustrated in Figures 1 and 2 (shallower slopes = less categorical)
- Some participants utilized all points on the VAS line for their responses (Figure 1)
- Others grouped responses around discrete locations on the line (Figure 2)
- This led to the question: why are some more categorical than others?

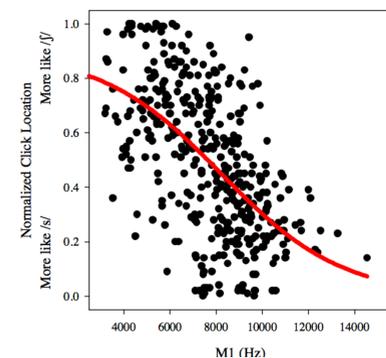


Figure 1: The second-most continuous responder on /s/-/ʃ/ discrimination judgments (slope = -0.000144)

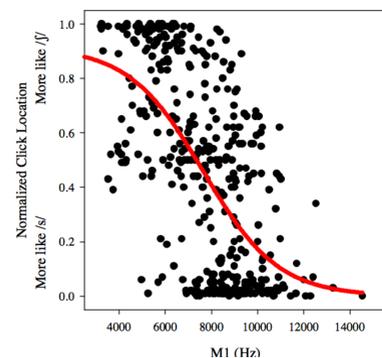


Figure 2: The most (realistically) categorical responder on /s/-/ʃ/ discrimination judgments (slope = -0.000301)

Why are some adults more categorical and some more continuous when they rate children's speech?

Some hypotheses:

- Differences in overall response strategies: individuals tend to respond similarly across tasks
- Differences in processing speed (see Munson, Kaiser, & Urberg Carlson, 2008)
- Differences in attention to indexical information

Munson, Kaiser, and Urberg Carlson (2008) concluded that the second hypothesis cannot explain these response patterns

- We hypothesized that listeners would be more categorical when there was a long delay between the presentation of the stimulus and the response than when the response was immediate
- We enforced one-, two-, and three-second delays between presentation of the stimulus and elicitation of the response
- Fine phonetic detail in children's speech was perceived equally well across different task-delay conditions; the categoricity of responses didn't change
- This reflects the kinds of delays that occur in real-world assessment of children's speech

So we created an experiment to test the third hypothesis

- Indexical variables, like gender, are conveyed through fine phonetic detail within categories.
- Hence, we predict that people will be more continuous when they attend to indexical information than when attending to categorical linguistic information
- To encourage *categorical* responses we interleaved forced-choice vowel identification questions between the /s/-/θ/ judgments; vowels were embedded in CV syllables
- To encourage *continuous* responses we interleaved sex-identification judgments between the /s/-/θ/ judgments

Procedure

The experiment consisted of three blocks; the first and last were randomized between subjects

- Lexical (categorical) task: /s/-/θ/ judgments alternating with vowel identification (vowels were embedded in CV syllables)
- Short distractor task: judging emotion of speaker based on distorted signals
- Indexical (continuous) task: /s/-/θ/ judgments alternating with judgments of speaker sex (see Figures 3 and 4)

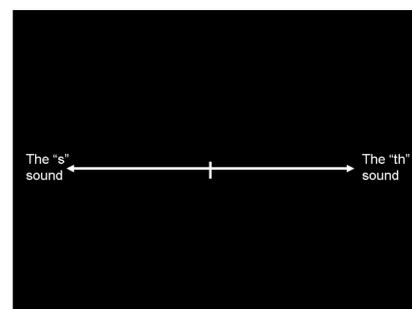


Figure 3: VAS rating screen for /s/-/θ/

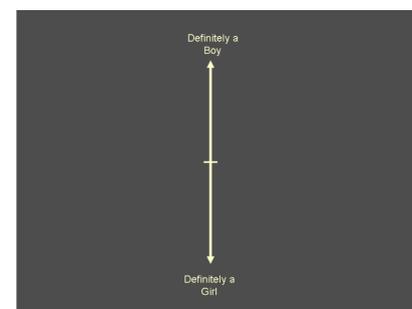


Figure 4: VAS rating screen for gender

We used the same stimuli that were used in Schellinger, Edwards, Munson, and Beckman (2008) to examine adults' ratings of children's /s/-/θ/ productions

- Stimuli were two- to five-year olds' productions of target /s/ and /θ/ in CV syllables
- Stimuli were randomly assigned to the first and last blocks
- We counterbalanced the assignment of stimuli to condition

Results

No significant difference was found between the two conditions

- Relative intensity was used as a predictor for VAS click locations on the /s/-/θ/ continuum (see Figure 5)
- Figure 6 is scatterplot of the subjects' /s/-/θ/ slopes in the two blocks
- A Wilcoxon Signed-Ranks test showed no significant difference between the two conditions, Lexical and Indexical ($z = -0.114$, $p = 0.91$)
- The median slope for the Lexical condition ("Vowel Blocks") was -6.11
- The median slope for the Indexical condition ("Gender Blocks") was -6.96

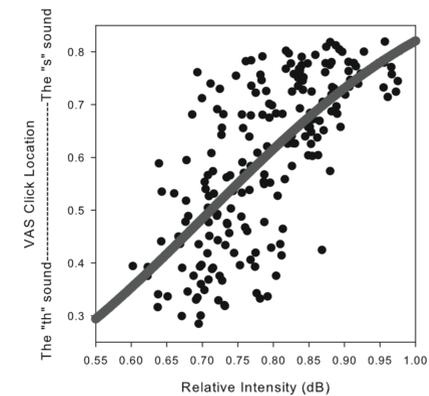


Figure 5: /s/-/θ/ judgments pooled across the two conditions

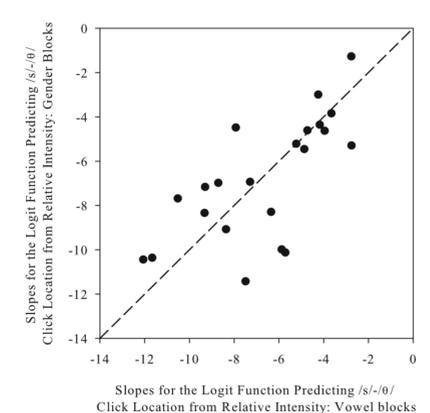


Figure 6: Comparison of slopes between the two conditions

Conclusions and future work

People are consistent across different types of tasks in rating children's speech

- We tested whether people would respond categorically to fricatives if they are made to pay more attention to categorical aspects of the stimuli (like the identity of phonemes), and conversely, whether people would respond continuously to fricatives if they are made to pay more attention to indexical information like speaker gender
- This didn't happen, so we can conclude that differences in attention to indexical information (the third hypothesis) cannot explain why some adults are more categorical and others more continuous when they rate the accuracy of children's speech
- People's ability to perceive fine phonetic detail in productions is robust across different tasks

Future work

- Test the first hypothesis
- Continue to refine our measure of categoricity of response
- Explore other factors that will predict VAS ratings on the /s/-/θ/ continuum; maybe relative intensity isn't the best predictor for all listeners

References

- Crocker, L. and Munson, B. (2006). Speech characteristics of gender-nonconforming boys. Presented at the Conference on New Ways of Analyzing Variation in Language, Columbus, OH.
- Drager, K. (2008). Sensitivity to grammatical and sociophonetic variability in perception. Presented at the Conference on Laboratory Phonology, Wellington, New Zealand.
- 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.
- Kong, E. (2009). The Development of Phonation-type Contrasts in Plosives: Cross-linguistic Perspectives. Dissertation in preparation at the Ohio State University.
- Li, F., Edwards, J., and Beckman, M. E. (2009). Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers. *Journal of Phonetics*, 37(1): 111-124.
- Munson, B., Li, F., Yoneyama, K., Hall, K. C., Beckman, M. E., Edwards, J., and Sunawatari, Y. (2008). Language-specific production and perception of sibilant fricatives in Japanese and English. Presented at the 82nd Annual Meeting of the Linguistics Society of America, Chicago, IL.
- Munson, B., McDonald, E., DeBoe, N., and White, A. (2006). The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech. *Journal of Phonetics* 34: 202-40.
- Macken, M. and Barton, D. (1980). The acquisition of the voicing contrast in English: a study of voice onset time in word-initial stop consonants. *Journal of Child Language*, 7, 41-74.
- Munson, B., Kaiser, E., and Urberg Carlson, K. (2008). Assessment of Phonetic Skills in Children 3: Fidelity of Responses under Different Levels of Task Delay. Presented at the ASHA Convention, Chicago, IL.
- Schellinger, S., Edwards, J., Munson, B., and Beckman, M. E. (2008). The role of listener expectations on judgments of children's /s/ productions. Presented at the Symposium on Research in Child Language Disorders, University of Wisconsin - Madison.
- Smith, E. A., Munson, B., and Hall, K. C. (2008). Rethinking the meaning of Minnesotan [æ]: sexual orientation or personal well being? Presentation given at the Conference on New Ways of Analyzing Variation in Language, Houston, TX.
- Smith, N. (1973). *The acquisition of phonology: A case study*. Cambridge: Cambridge University Press.
- Urberg Carlson, K., Kaiser, E., and Munson, B. (2008). Assessment of Phonetic Skills in Children 2: Testing Gradient Measures of Children's Productions. Presented at the ASHA Convention, Chicago, IL.
- Urberg Carlson, K., Munson, B., and Kaiser, E. (2009). Gradient measures of children's speech production: Visual analog scale and equal appearing interval scale measures of fricative goodness. Presented at the Meeting of the Acoustical Society of America, Portland, OR.

Acknowledgments

This work was supported by NSF Grant BCS0729277 to Benjamin Munson, and by NIH grant R01 DC02932 to Jan Edwards. We gratefully acknowledge the generous assistance of Kari Urberg Carlson, Eric Foster-Lussier, Ilana Heintz, and Eunjong Kong.