



# The Role of Listener Expectations on Judgments of Children's /s/ Productions, Revisited

Sarah Schellinger<sup>1,2</sup>, Marie K. Meyer<sup>2</sup>, Benjamin Munson<sup>2</sup>, Jan Edwards<sup>1</sup>, and Mary E. Beckman<sup>3</sup>

<sup>1</sup>Department of Communicative Disorders, University of Wisconsin, Madison, <sup>2</sup>Department of Speech-Language-Hearing Sciences, University of Minnesota, Minneapolis <sup>3</sup>Department of Linguistics, Ohio State University, Columbus



## Phonetic Transcription of Children's Speech: Pluses and Minuses

- **Phonetic transcription is the 'standard-of-care' tool for the assessment of speech-sound disorders.** It is fast and relatively easy to implement; Normative databases for the assessment of speech-sound disorders reference the number of sounds correctly produced; and The field has survived for 70+ years with this as the primary tool that we use.
- What's potentially problematic with it?
  - **It is ill-suited to assess the full range of productions that children produce.** Children gradually master adult-like productions of phonemes, and often produce sounds that are intermediate between adult endpoints (Baum & McNutt, 1990; Li, Edwards, & Beckman, 2008). Phonetic transcription forces users to characterize continuous development with a finite set of symbols and diacritics.
  - **It is potentially subject to biases based on listener expectations.** Listeners interpret sounds relative to expectations about the speakers who produce them, including expectations about age (Drager, 2006), regional dialect (Niedzielski, 1999), and gender (Munson, 2009).

## Schellinger, Edwards, Munson, and Beckman (2008a,b,c)

- Examined (a) whether listeners are able to rate children's productions intermediate between /s/ and /θ/ as such given an alternative assessment method, visual analog scaling, (b) whether this was biased by enforcing expectations about the child's overall level of ability, as cued by a carrier phrase.
- Stimuli were 200 /s/ and /θ/-initial CV sequences that had been produced by children participating in the παιδολογος project (<http://www.ling.ohio-state.edu/~edwards/>) (Figure 1). They had been transcribed by two native-speaker phoneticians in six categories: correct /s/, [s] for /θ/ substitutions, intermediate productions closer to /s/ (sθ), intermediate productions closer to /θ/ (θs), [θ] for /s/ substitutions, and correct /θ/.
- Twenty-one listeners were presented with the stimuli and rated how close to /s/ or /θ/ they were by clicking on the line in Figure 2.
- Click location in pixels discriminated among the six transcription categories (Figure 3). All pairwise differences among transcription categories were significantly different in a single-factor within-subjects ANOVA.

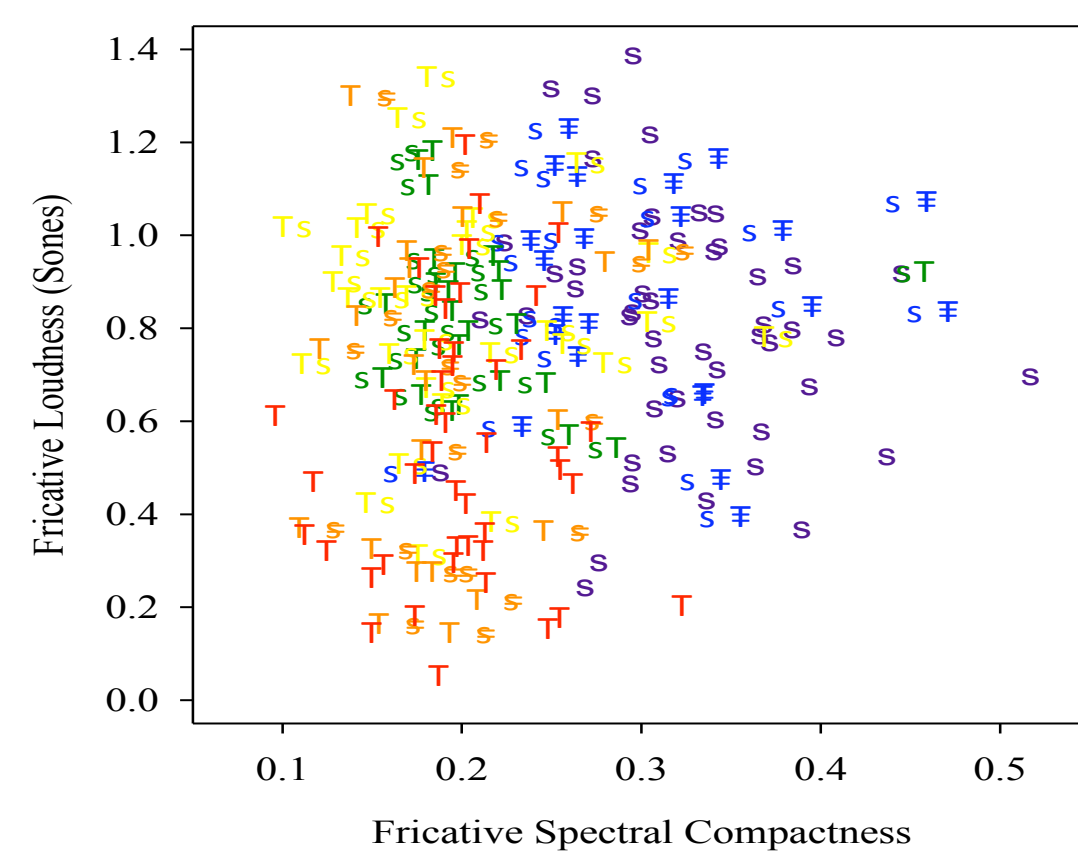


Figure 1 (left). The 200 stimuli in the two dimensional loudness-by-spectral compactness space, separated by transcription category

Figure 3 (below). Average VAS ratings for the six fricative types

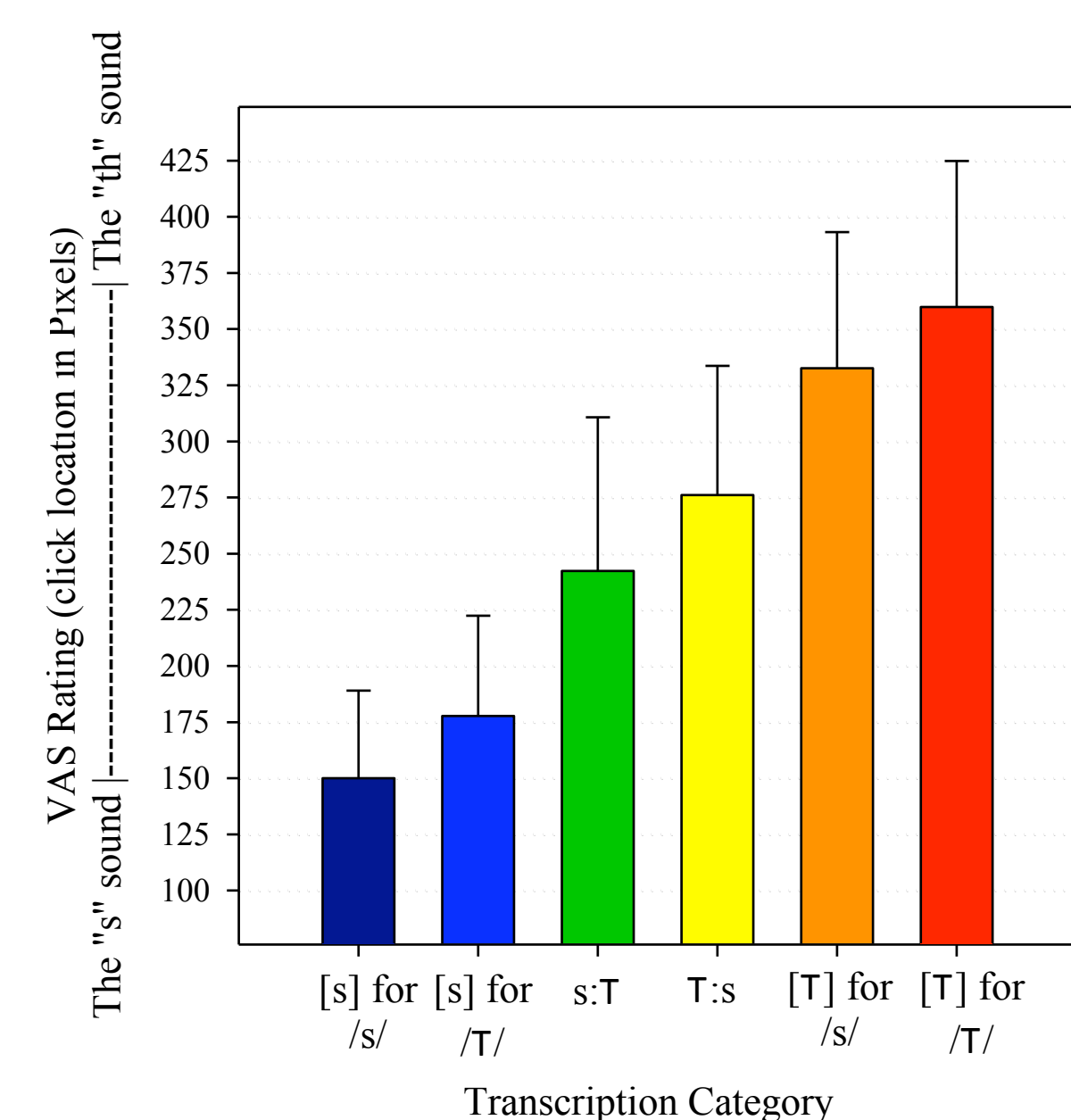


Figure 2 (above). The VAS rating display.

## Schellinger et al. (continued)

- A separate group of listeners were presented with the stimuli in a new perception task, preceded by one of two carrier phrases.
  - A child's production of "I really like" with the formant frequencies and fundamental frequency scaled lower, consistent with those of an older child.
  - The same child's production of "I weawwy yike" with the formant frequencies and fundamental frequency scaled higher, consistent with those of a younger child.
  - The carrier phrases were intended to bias the listeners to believe that the child was younger and had disordered speech, or older and had normal speech
- Each stimulus was paired with both carrier phrases. A single block of stimuli with the two carrier phrases was presented. Stimuli were presented in random order. Listeners were asked if the child produced the "s" sound correctly.
- There was a tendency for listeners to rate the intermediate productions as more accurate when biased to believe that the child was younger (Figures 4 and 5)

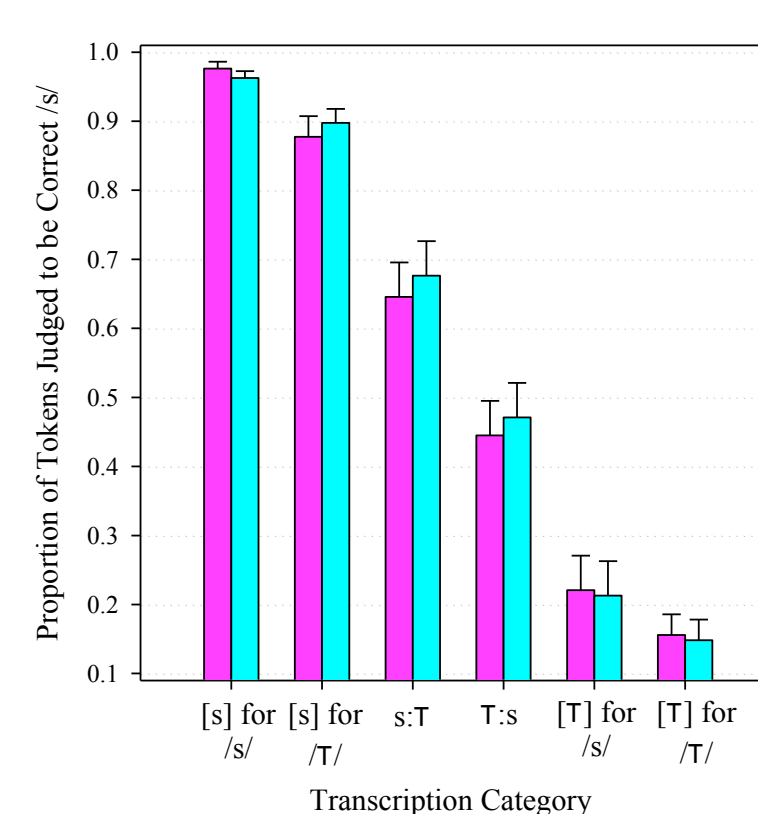


Figure 4 (left). Proportion of tokens judged to be a correct /s/ in Schellinger et al. (2008), separated by carrier-phrase type (pink="really like", blue="weawwy yike"), and transcription category.

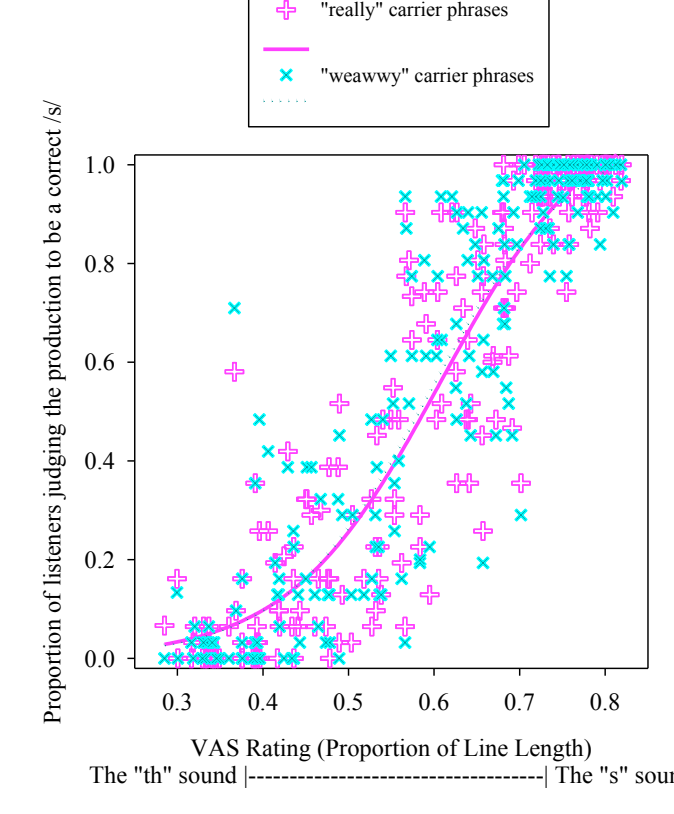


Figure 5 (right). Proportion of listeners who judged each token to be a correct /s/, plotted against the average VAS rating of the fricative. Lines represent logistic regression functions.

## This study further examines the role of bias on the perception of children's /s/ and /θ/ productions

- The relatively small effect of carrier-phrase type on ratings in Schellinger et al. may have been a consequence of different choices made in the design of the study. Three experiments examined whether a stronger effect of carrier phrase could be elicited with different experiment designs and instructions

## Stimuli Common to the Experiments

- The same two-hundred stimuli used in Schellinger et al., shown in Figure 1.

## Listener Population

- Listeners were recruited from the University of Minnesota community. All were native, monolingual speakers of North American English.

## Procedures Common to the Experiments

- Experiments were run with E-Prime. Responses and response times (not reported here) were measured. Participants were instructed to respond as quickly as possible whether the initial sound was a correct "s" sound".

This poster is available online at <http://www.tc.umn.edu/~munso005>

**Acknowledgements** This research was supported by NIH grant R01 DC02912 to Jan Edwards, by NSF Grant BCS0729277 to Benjamin Munson, and by a University of Minnesota Undergraduate Research Opportunity Program (URUP) grant to Marie Meyer and Benjamin Munson. We thank Kai Urbig-Carlson and Edin Kaiser for help testing subjects.

**References** Baum, S., & McNutt, J. (1990). An acoustic analysis of frontal misarticulation of /s/ in children. *Journal of Phonetics*, 18, 51-63.

Drager, Katie (2006) From Bad to Good: the relationship between perceived age and vowel perception in New Zealand English. *The Res* 45, 35-68.

Li, F., Edwards, J., & Beckman, M. (2009). Contrast and covert contrast: The phonetic development of voiceless obstruent fricatives in English and Japanese toddlers. *Journal of Phonetics*, 37, 111-124.

Munson, B. (2009). Gender biases in fricative identification revisited. Oral presentation at the annual meeting of the Linguistic Society of America, San Francisco, CA.

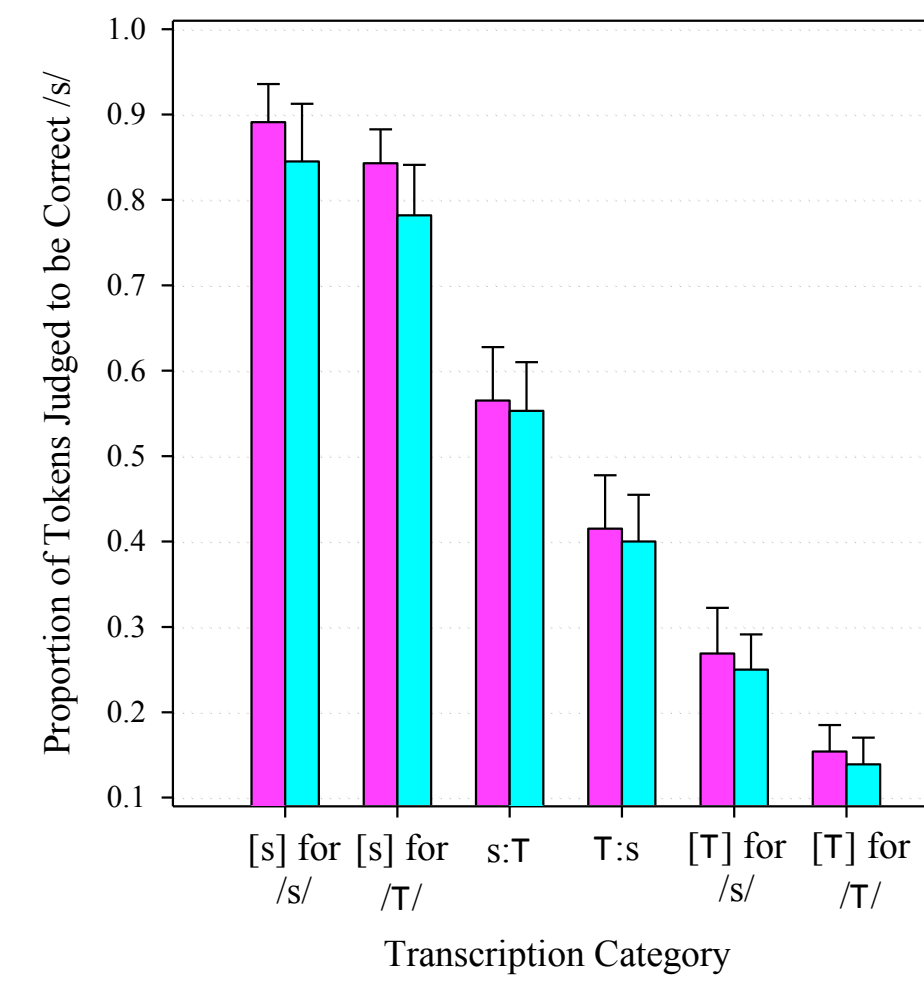
Nadzhak, S. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Social Psychology*, 70, 62-85.

Schellinger, S., Edwards, J., Munson, B., & Beckman, M. E. (2008a). Does "yike" count as transcription as well as in /s/ productions? Poster presented at the Symposium on Research in Child Language Disorders, 5-7 June, University of Wisconsin - Madison.

Schellinger, S., Edwards, J., Munson, B., & Beckman, M. E. (2008b). Assessment of children's speech production 1: Transcription categories and listener expectations. Poster presented at the 2008 ASHA Convention, November 20-22.

## Experiment 1: Blocked Carrier Phrases

- Schellinger et al. mixed together "really like" and "weawwy yike" carrier phrases in a single block. This potentially decreased the extent to which listeners interpreted the carrier phrase and CV stimulus as having been produced by the same child.
- Experiment 1 examined whether a stronger influence of carrier phrase could be found if the "really like" and "weawwy yike" carrier phrases were presented in separate blocks.
- Fifteen listeners participated in this experiment. The design was identical to Schellinger et al. except the two carrier phrase types were presented in separate blocks. The order of the carrier phrases was randomized across listeners. Results are in Figure 6.

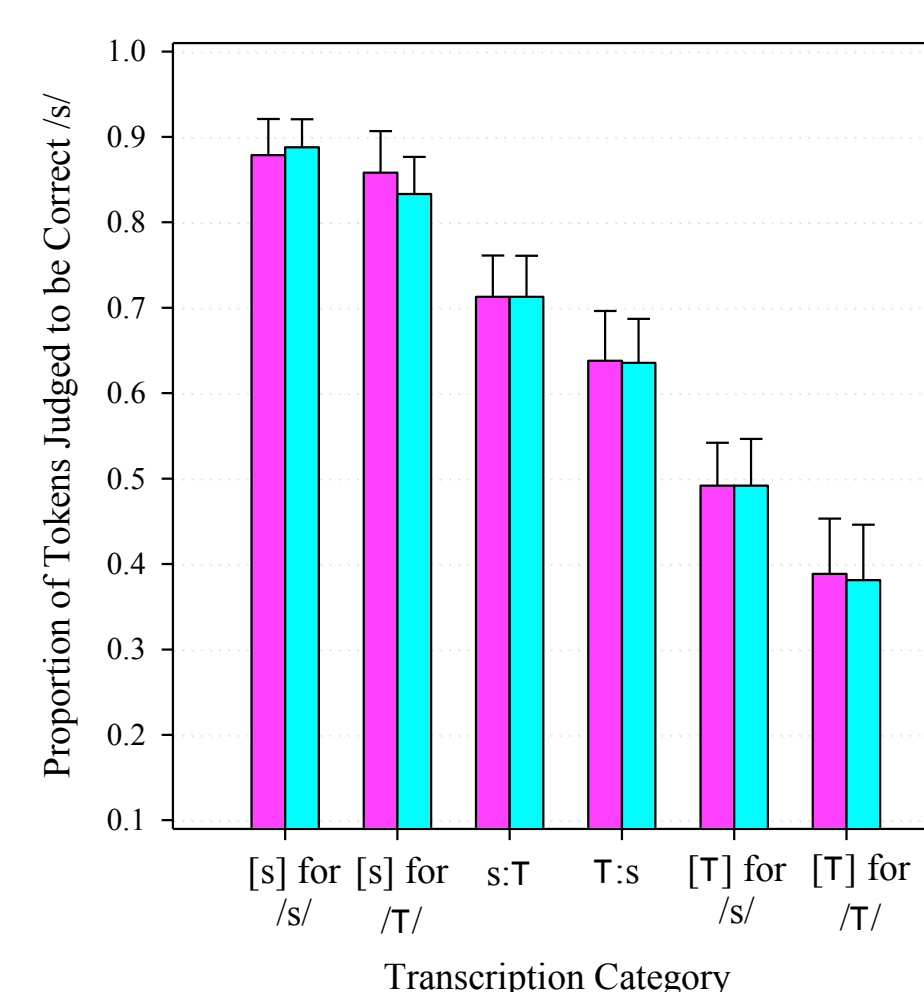


• Five of the six transcription categories were rated significantly differently from one another. Only the two types of /s/ ([s] for /s/ and [s] for /θ/) did not differ from each other. Carrier phrase had a marginally significant influence on ratings ( $F[1,14] = 3.68, p = 0.077$ ). Listeners tended to rate fewer /s/ tokens as accurate when preceded by the 'weawwy' carrier phrases. This did not interact with transcription category

Figure 6 (left). Proportion of tokens judged to be a correct /s/ in Experiment 1, separated by carrier-phrase type (pink="really like", blue="weawwy yike"), and transcription category.

## Experiment 2: No Mention of Disorder

- Experiment two was differed from Experiment 1 only in that it deleted mention of speech-sound disorders in the instructions. Listeners were told only that they would be rating the accuracy of children's speech. Fifteen listeners participated. Results are in Figure 7.
- Both Schellinger et al. (2008) and Experiment 1 explicitly instructed participants that the purpose of the study was to examine the perception of the speech of children who made developmental speech-sound disorders. The explicit mention of disorder, and the use of the term 'lisp' in the instructions, may have activated social stereotypes about children's productions that suppressed listeners' integration of the carrier phrase and the target.



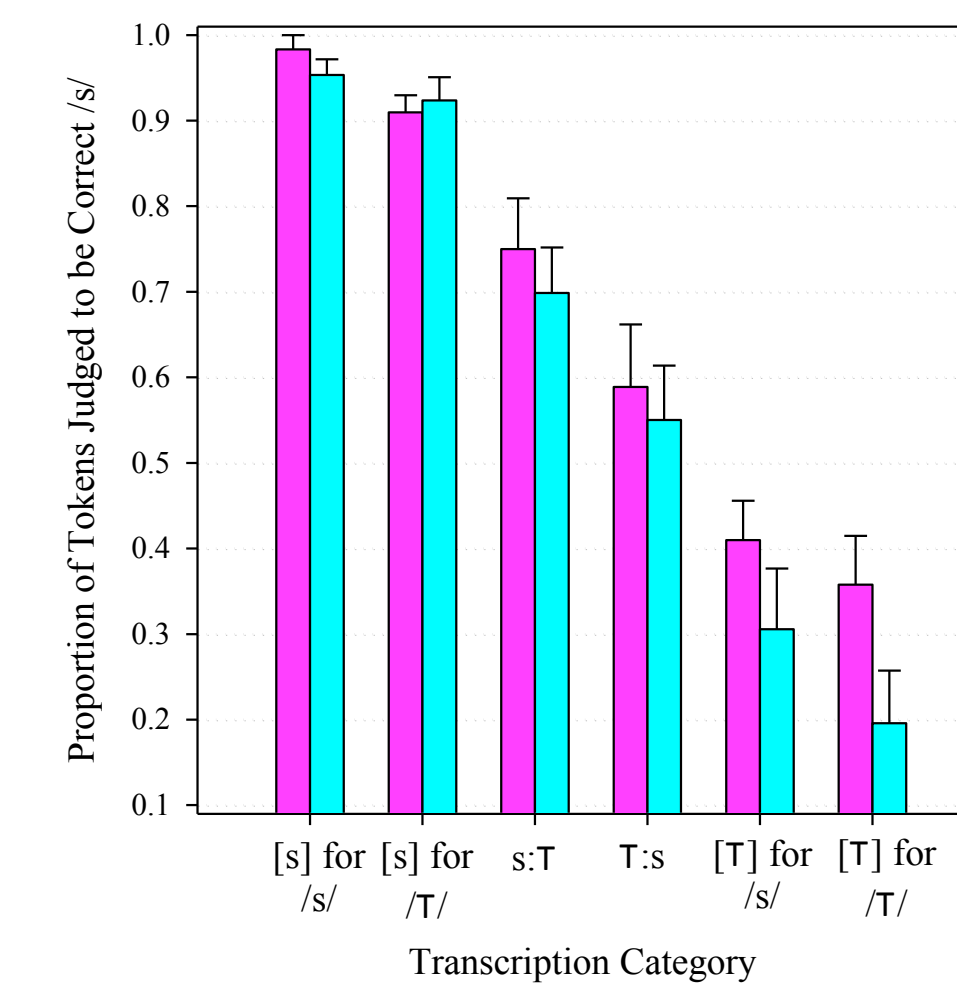
• Transcription category affected ratings significantly. Only the two types of /s/ ([s] for /s/ and [s] for /θ/) did not differ from each other. Carrier phrase did not affect rating of intermediate tokens.

• The intermediate tokens and the /θ/ productions were more likely to be rated as accurate /s/ productions than in previous studies. This is consistent with studies showing considerable overlap between adults' /s/ and /θ/ productions.

Figure 7 (left). Proportion of tokens judged to be a correct /s/ in Experiment 2, separated by carrier-phrase type (pink="really like", blue="weawwy yike"), and fricative type.

## Experiment 3: Carrier Phrase Matched for f0

- The failure of carrier phrase to influence perception in Schellinger et al. and in Experiments 1 and 2 was perhaps due to listeners not interpreting the carrier phrase and the target CV to be produced by the same child.
- Experiment 3 matched the f0 of the carrier phrase to the f0 of the target in an attempt to make them sound more like they were produced by the same child. Listeners were also explicitly instructed to imagine that the carrier phrase and the CV were produced by the same child.
- The analysis of Experiment 3 is preliminary and is based on six participants' responses, shown in Figure 8.



• Transcription category significantly affected ratings. All post-hoc paired comparisons among fricative types were significant. Carrier phrase had a marginally significant influence on ratings ( $F[1,5] = 4.69, p = 0.083$ ) and interacted significantly with transcription category,  $F[5,25] = 4.62, p = 0.004$ . Less /s/-like fricatives were rated as disproportionately less accurate when preceded by f0-matched 'weawwy' carrier phrases.

Figure 8 (left). Proportion of tokens judged to be a correct /s/ in Experiment 3, separated by carrier-phrase type (pink="really like", blue="weawwy yike"), and transcription category.

## Conclusions

1. Naïve listeners identify intermediate productions as such across a variety of tasks.
  2. Listeners' ratings in Experiments 1 and 2 were not strongly biased by carrier phrases. The carrier phrases were intended to suggest to the listeners that the children producing the fricatives were either older and more advanced in their phonological development, or younger and less advanced.
  3. Listeners were more likely to rate /θ/ productions as correct /s/ tokens if the instructions didn't make reference to speech disorders and didn't use the word "lisp."
  4. Listeners in Experiment 3 were more likely to rate tokens of /θ/ as instances of correct /s/ when the carrier phrase implied that the child producing the /θ/ was older and had more advanced phonological development.
- **Overall, our findings suggest that listeners are willing to accept /θ/ productions as instances of correctly produced /s/ when they perceive the children to be older and not disordered. Perhaps the listeners interpret the [θ]-like /s/ tokens as representing intentional socioindexical variation, which they tacitly associate with older children. More generally, these findings underscore the complex influences that category prototypicality, listener expectations, and task-related variables have on the assessment of the accuracy of children's speech**
- Ongoing research examines the extent to which these biases change when more-natural biasing conditions are introduced, and the extent to which they hold for a variety of different speech-sound contrasts.
    - Are listeners biased when we use strings of words instead of carrier phrases?
    - Does biasing exist for sound contrasts that are not the locus of sociophonetic variation in the adult language (/t/-/k/, /s/-/ʃ/)?