Clinical Transcription: Old Concerns, New Solutions

Jan Edwards
University of
Wisconsin, Madison

Benjamin Munson
University of
Minnesota, Minneapolis

Why a talk about transcription?

- Technology has changed, but have we?
  - 1960s-70s: cheap portable technology for making audio recordings
  - 1990s: inexpensive digital technology for recording and analysis.
  - 2000s: free signal-processing software, cheap fast computers
- We continue to rely on transcription of elicited single words.
  - In research studies
  - In clinical practice (assessment and treatment)

Advantages of transcription

- Time-efficient
- The basis for standardized assessments
- Results easy to understand and to explain to clinicians and researchers.

Problems with transcription: 1

- Listener expectations: Listener judgments are influenced by factors other than the acoustic signal.
  - Native Language (Li, Munson, Beckman, Edwards, Yoneyama, & Hall, 2008)
  - Regional dialect (Hay, Warren, & Drager, 2006; Niedzielski, 1999)
  - Age (Drager, forthcoming)
  - Gender (Johnson, Strand & D'Imperio, 1999; Munson, 2009)
  - Race (Staum Casasanto, 2008)
  - Presence of disorder (Podol & Salvia,1976; Munson, Edwards, Schellinger, Beckman, & Meyer, forthcoming)

Problems with transcription: 2

- Children's productions do not always progress directly and categorically from incorrect to correct productions.
  - Existence of covert contrast (Li, Beckman, & Edwards; Macken & Barton, 1980; Munson et al., forthcoming): subphonemic differences between two sounds that are not perceptible to adults.
  - Other types of intermediate productions (Stoel-Gammon, 2001)

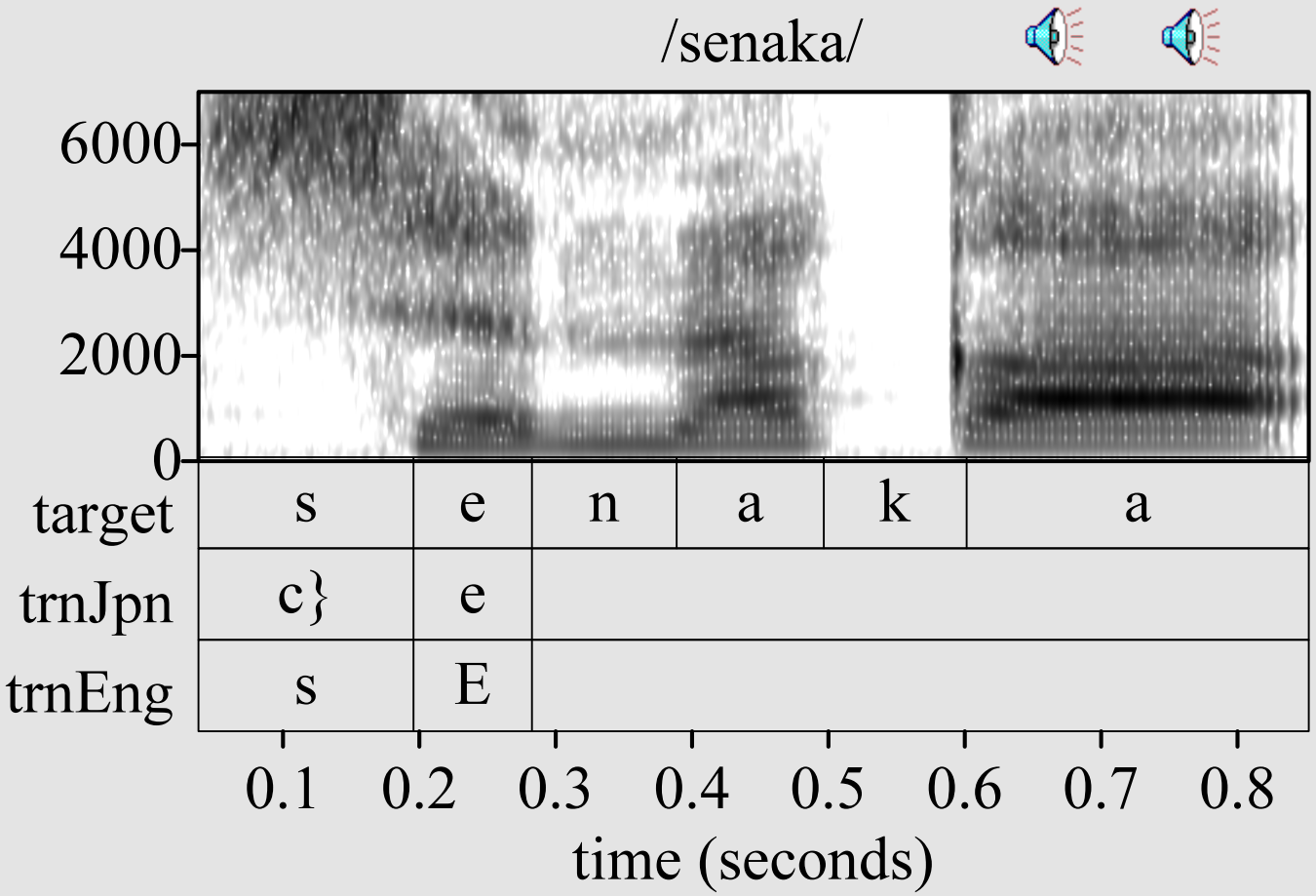
Illustrations of these problems from the παιδολογος database

- Targets: word-initial lingual obstruents in 5 vowel contexts in:
  - Phase 1: Cantonese, English, Greek, Japanese
  - Phase 2: above 4, Korean, Mandarin, Taiwanese, French, ….
- Participants:
  - Phase 1: 10 two- and 10 three-year-olds for four languages
  - Phase 2: ~20 of 2-, 3-, 4-, 5-year-olds, adults / language
- Procedure:
  - Elicit single word repetitions of target CVs in familiar words and nonwords.
  - Data collected in Hong Kong, Tokyo, etc.
- Measures:
  - Native-speaker transcriptions of target word-initial consonants
  - Acoustic measures
  - Naïve listeners' perceptions

Transcription analysis

- Transcription
  - Trained native-speaker phonetician for each language
  - Phonemic transcription: Initial consonants transcribed as correct or incorrect
  - Phonetic transcription: Errors transcribed as:
    - Within-inventory substitutions
      - /kha:55thoN55phi:n35/ (*cartoon*)– [t^h] for /k^h/
    - Outside-inventory substitution
      - /sinefo/ (*cloud*) – [t͡ɕ] (alveolopalatal affricate) for /s/
    - Intermediate between two sounds
      - *tube* – /tj/ or [kj]
    - Distortion
      - *gumdrops*
    - Deletion
      - *sister*

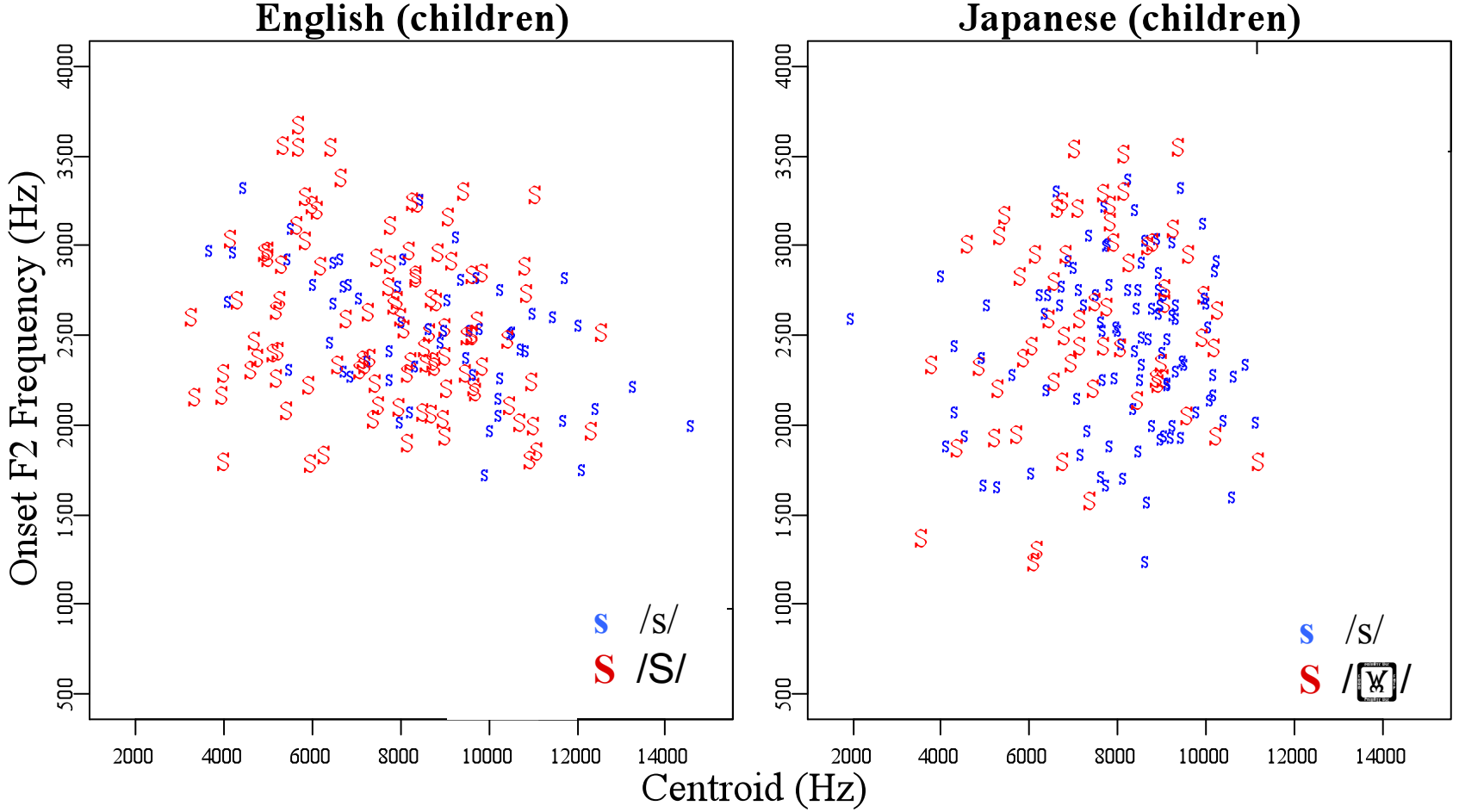
I. Cross-linguistic differences: /s/ in English and Japanese



/s/ in /senaka/

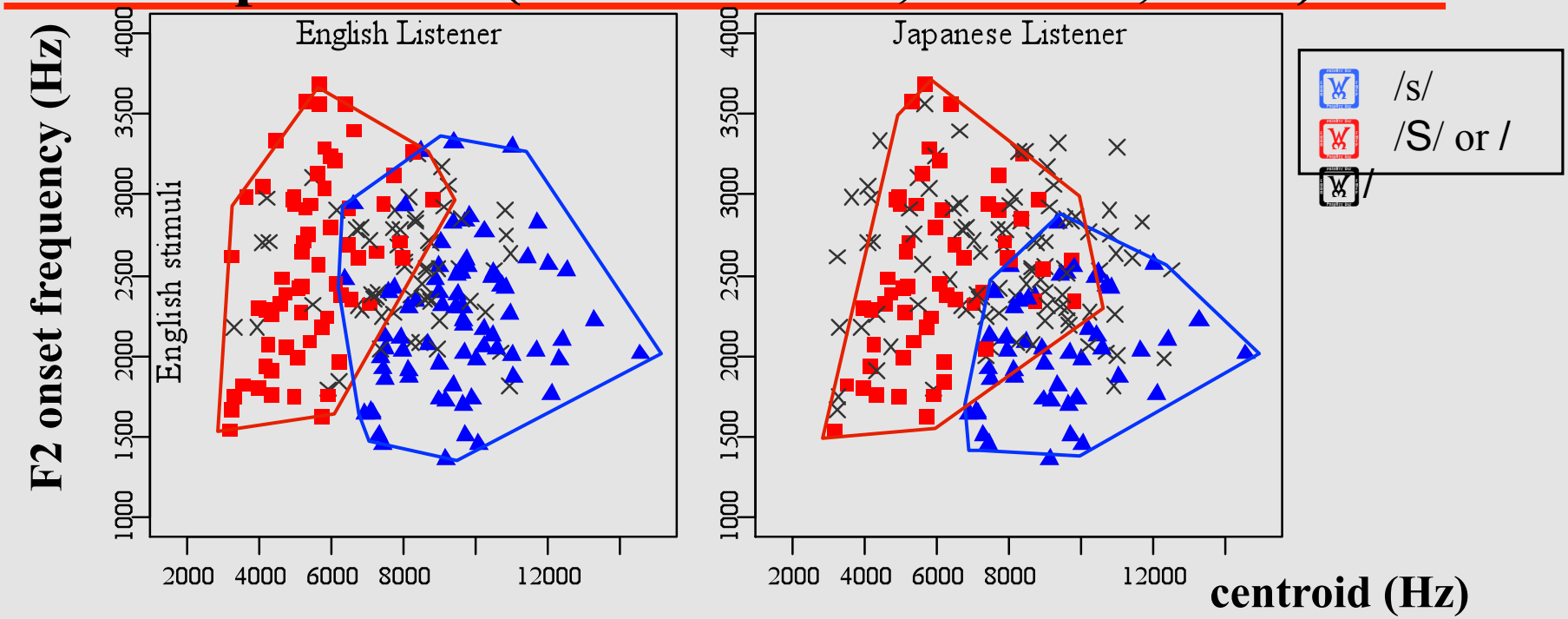
- transcribed as [ɯ] for /s/ substitution by Japanese speakers.
- judged as correct /s/ by English speakers.

I. Cross-linguistic differences: Acoustic analysis of children's productions (from Li, 2008)



•Children's productions show many intermediate tokens for both languages and for both phoneme categories.

I. Cross-linguistic differences: Perception experiment (from Munson, Li et al., 2008)



English listeners:

- acceptable range for /s/ is larger than acceptable range for /S/

Japanese listeners:

- acceptable range for / / is larger than acceptable range for /s/

II. Non-categorical nature of development: Intermediate productions

- Children do not always progress directly from incorrect to correct productions.
- Some productions are intermediate between two sounds.
  - Covert contrast
  - Other intermediate productions:
    - English: [k] or [g] 
      [f] or [T] 
    - Greek: [k] or [t] 
      [s] or [T] 

II. Intermediate productions: Perception experiments (Schellinger et al., 2008)

- Schellinger, Edwards, Munson, & Beckman (2008, and in preparation) asked two questions:
  - Are intermediate productions a valid transcription category?
  - Are intermediate productions more susceptible to listener bias?

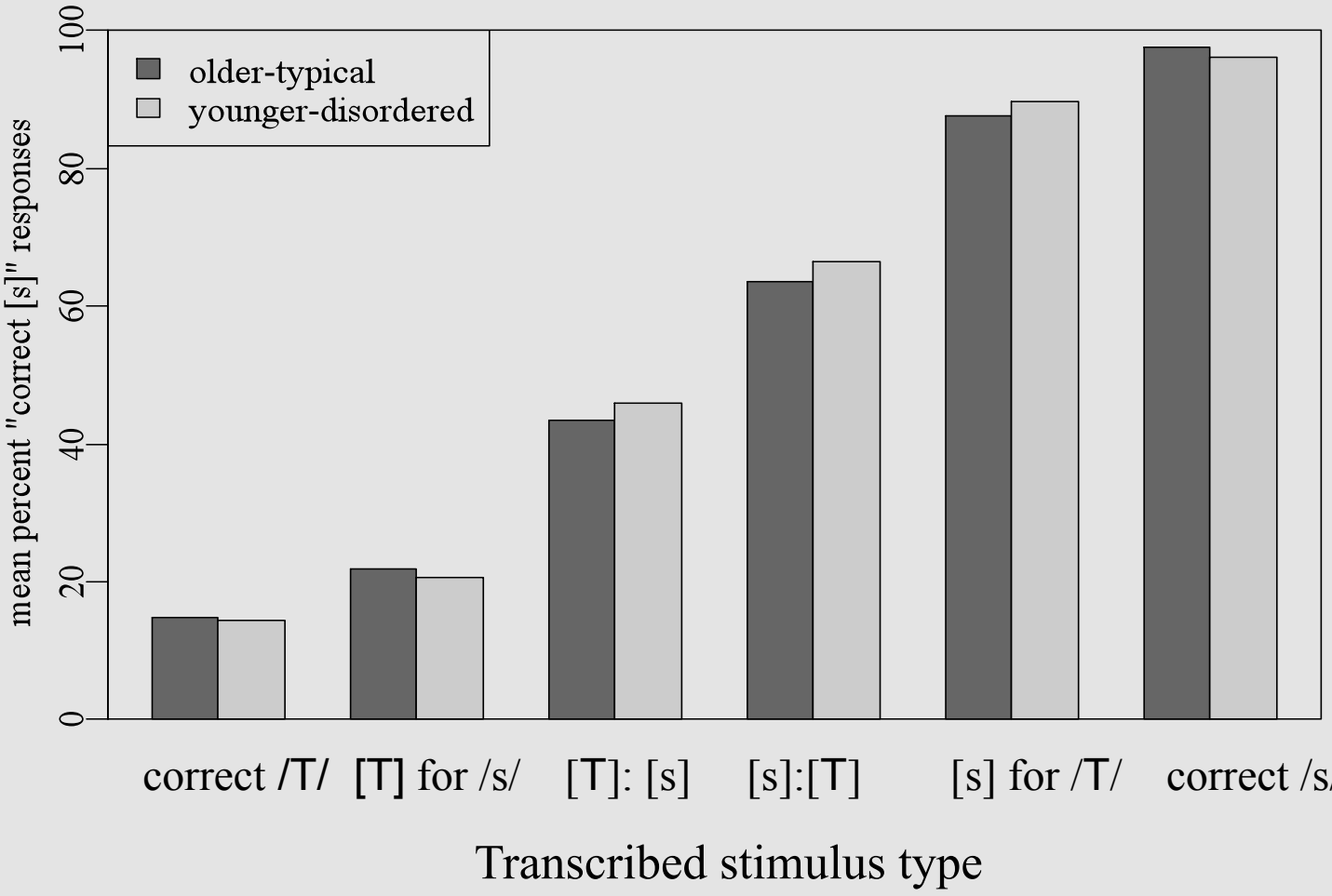
II. Intermediate productions: Perception experiment (Schellinger et al., 2008)

- Stimuli:
  - 200 CV sequences from single-word productions of English-speaking children, aged 2 through 5 years.
    - correct /s/
    - [s] for /T/
    - intermediate: closer to /s/ than /T/
    - Intermediate: closer to /T/ than /s/
    - [T] for /s/
    - correct /T/

II. Intermediate productions: Perception experiment (Schellinger et al., 2008)

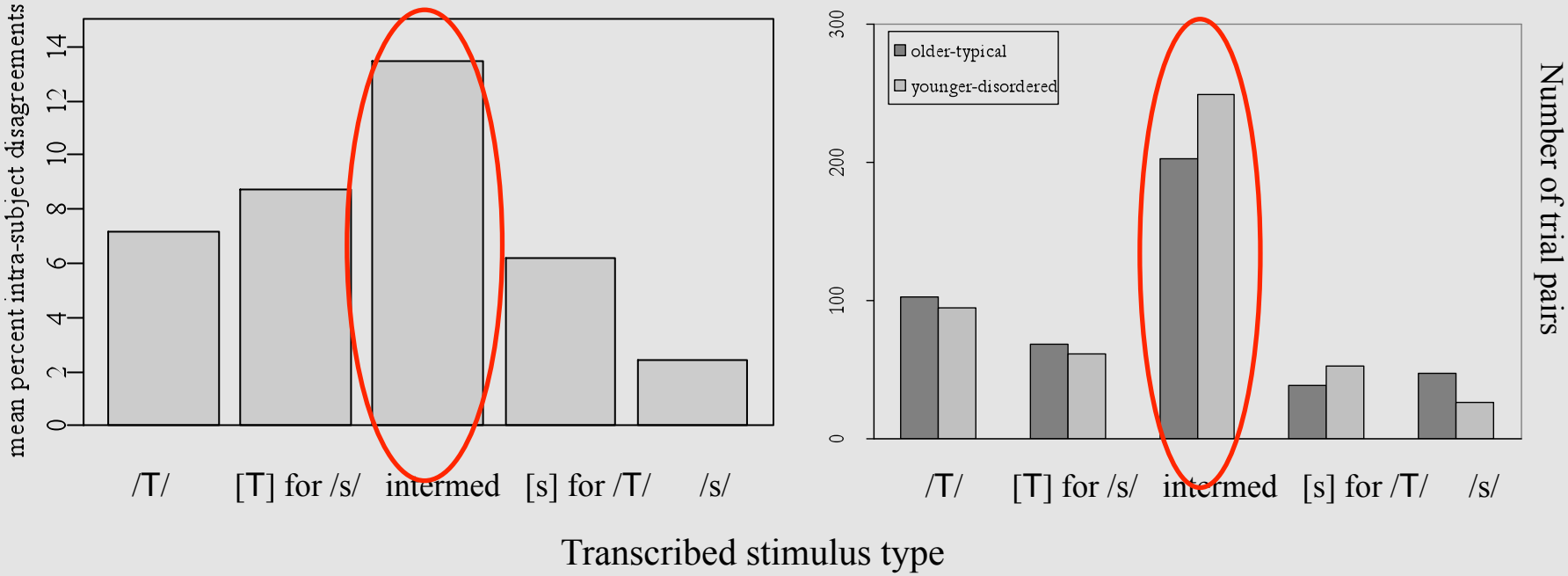
- Participants:
  - 30 naïve listeners
- Task:
  - Is it a correct /s/?
  - Respond as quickly as possible
- All CV sequences presented two times, once with each of the following two carrier phrases
  - "I really like" (older-typical)
    - No articulation errors
    - Formants/F0 altered to resemble an older child
  - "I weawwy yike" (younger-disordered)
    - Articulation errors
    - Formants/F0 altered to resemble a younger child

II. Intermediate productions: Perception experiment (Schellinger et al., 2008)



- Significant main effect of transcription category.
- No main effect of carrier phrase.

II. Intermediate productions: Perception experiment (Schellinger et al., 2008)



•Intermediate productions were more likely to be rated differently across the two carrier phrase conditions than other transcription categories.

•On these intermediate productions, listeners were more likely to hear a correct /s/ when the CV was preceded by a "younger-disordered" carrier phrase.

Summary: Problems with transcription

I. Transcription influenced by listeners' expectations and experience.
   - English and Japanese speakers processed the same acoustic-auditory space differently.
   - Intermediate productions of /s/ more likely to be judged as correct for a "younger-disordered" child.

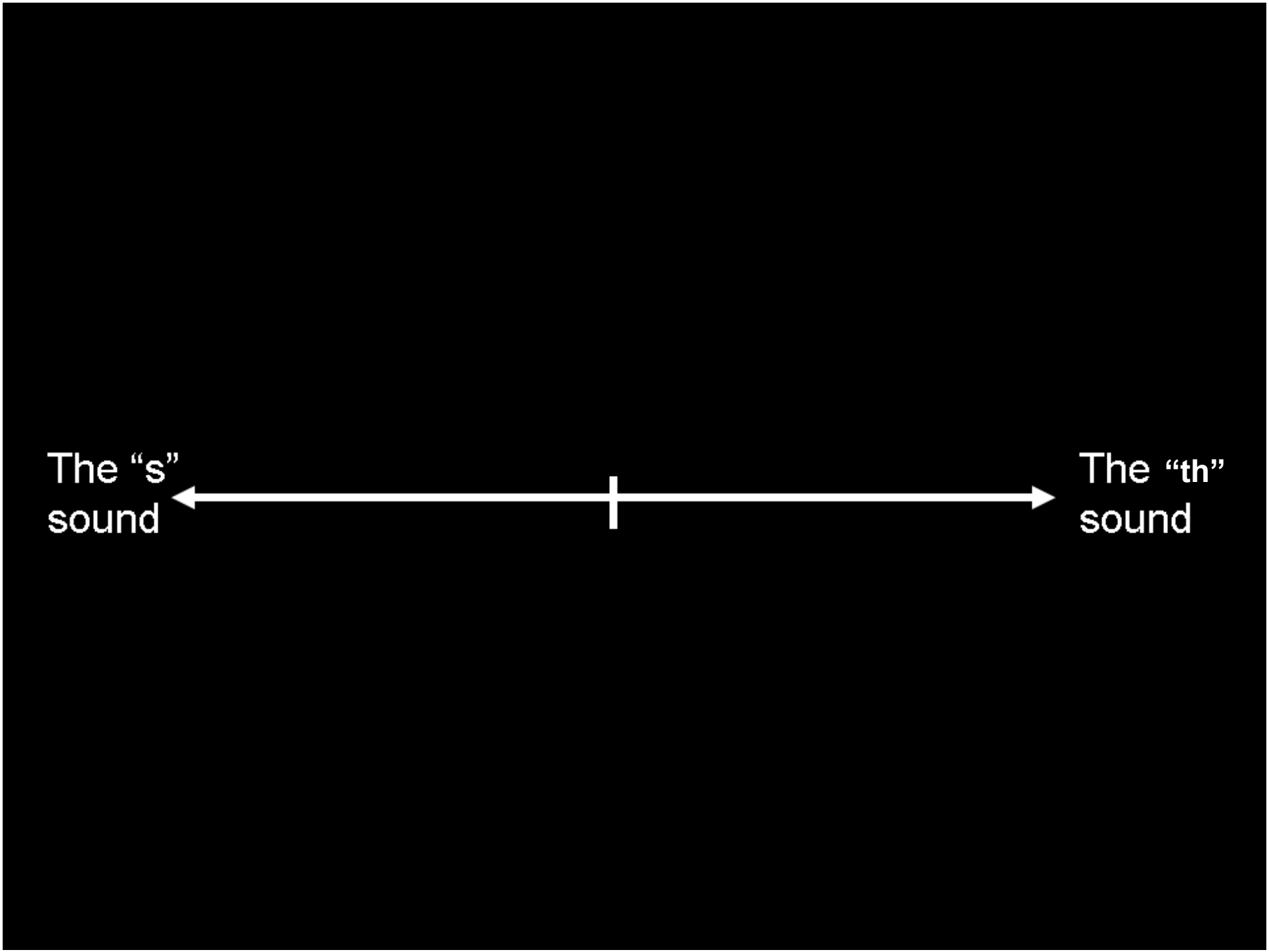
II. Children don't proceed directly and categorically from incorrect to correct productions.
   - Intermediate productions perceived as such by listeners.
   - Intermediate productions more susceptible to listener bias.

Where do we go from here?

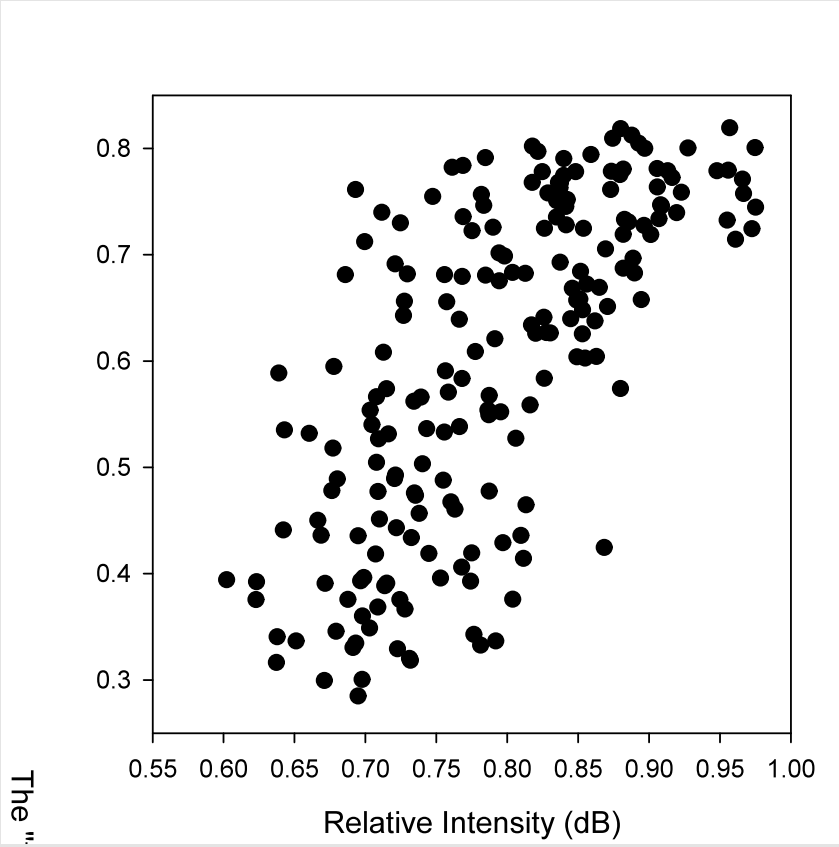
- There are problems with transcription, but…
- Transcription as a data analysis tool is here to stay ☺
- How can we improve transcription as a data analysis tool?
  - Don't transcribe live voice. Record with a digital recording device.
  - Use waveform editor and play word or part of word multiple times.
  - Use *intermediate* as a transcription category in assessment and treatment.
  - Supplement transcription with continuous rating scale.

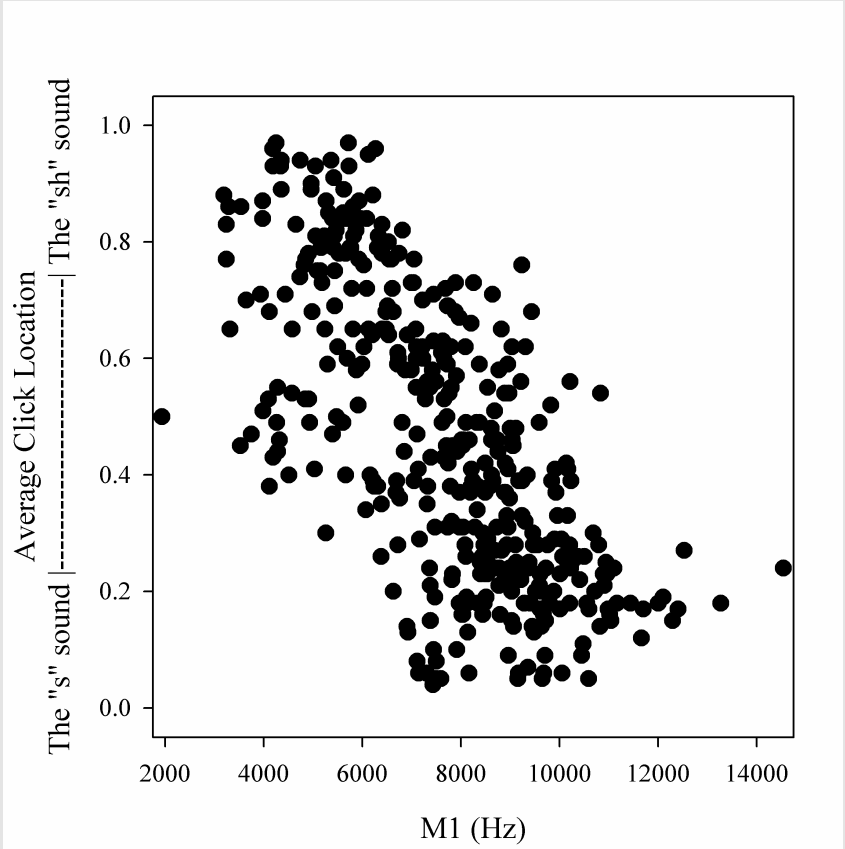
Supplement Transcription with Continuous Rating Scales

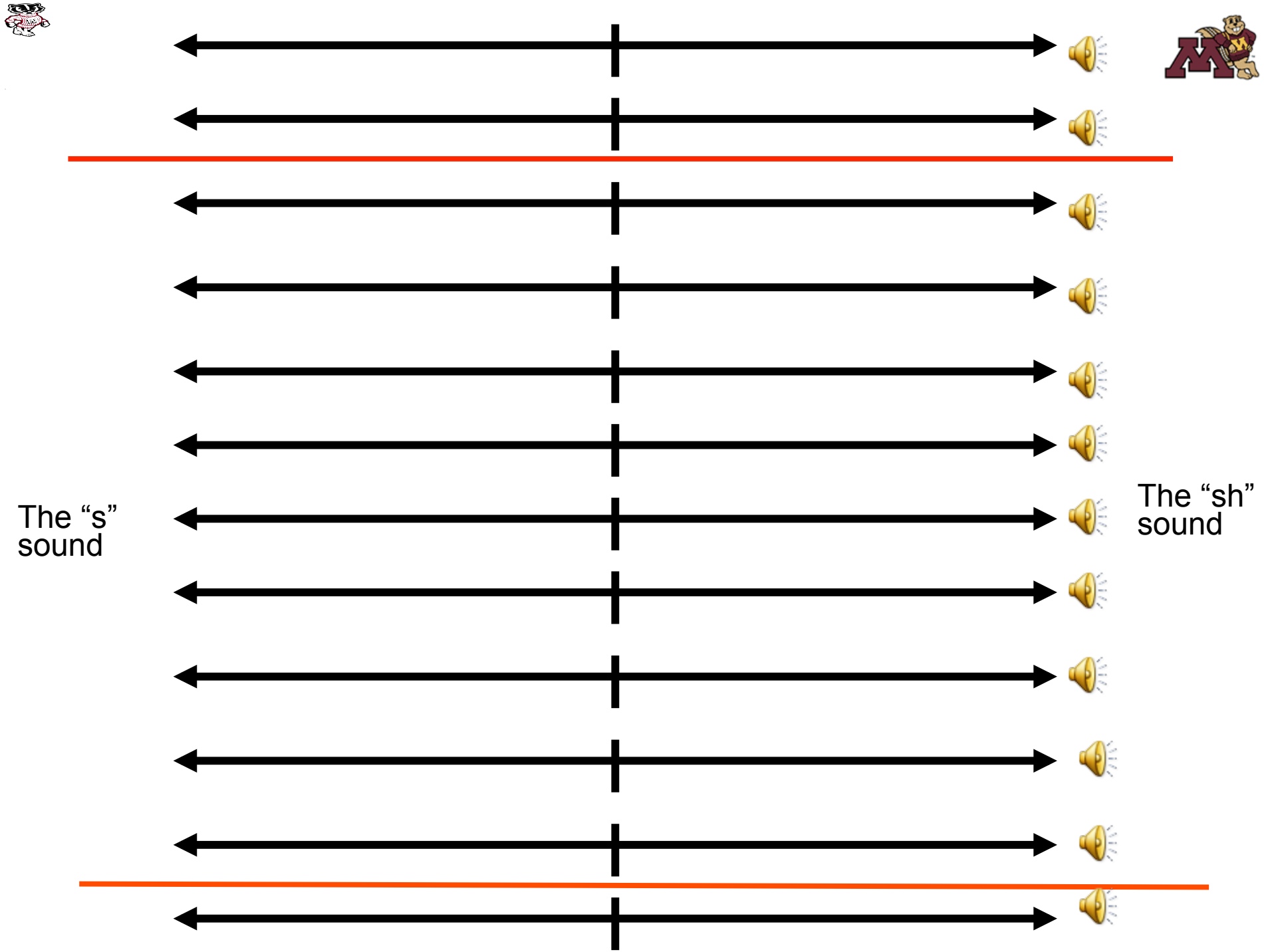
- We can supplement transcription with measures of how good an instance of the category a particular sound is.
- Urberg-Carlson et al. (2008, 2009) compared a few of these measures. The best was visual analog scaling (VAS)
  - Allows people to scale where a token falls relative to fixed endpoints.
  - The visual space is made essentially analogous to the perceptual space.
  - Listeners were explicitly instructed to click on locations that corresponded with the percept of 'proximity' to two sounds (for example /s/ and /T/).
  - VAS ratings correlated well with the acoustic parameters that differ between the endpoint sounds.

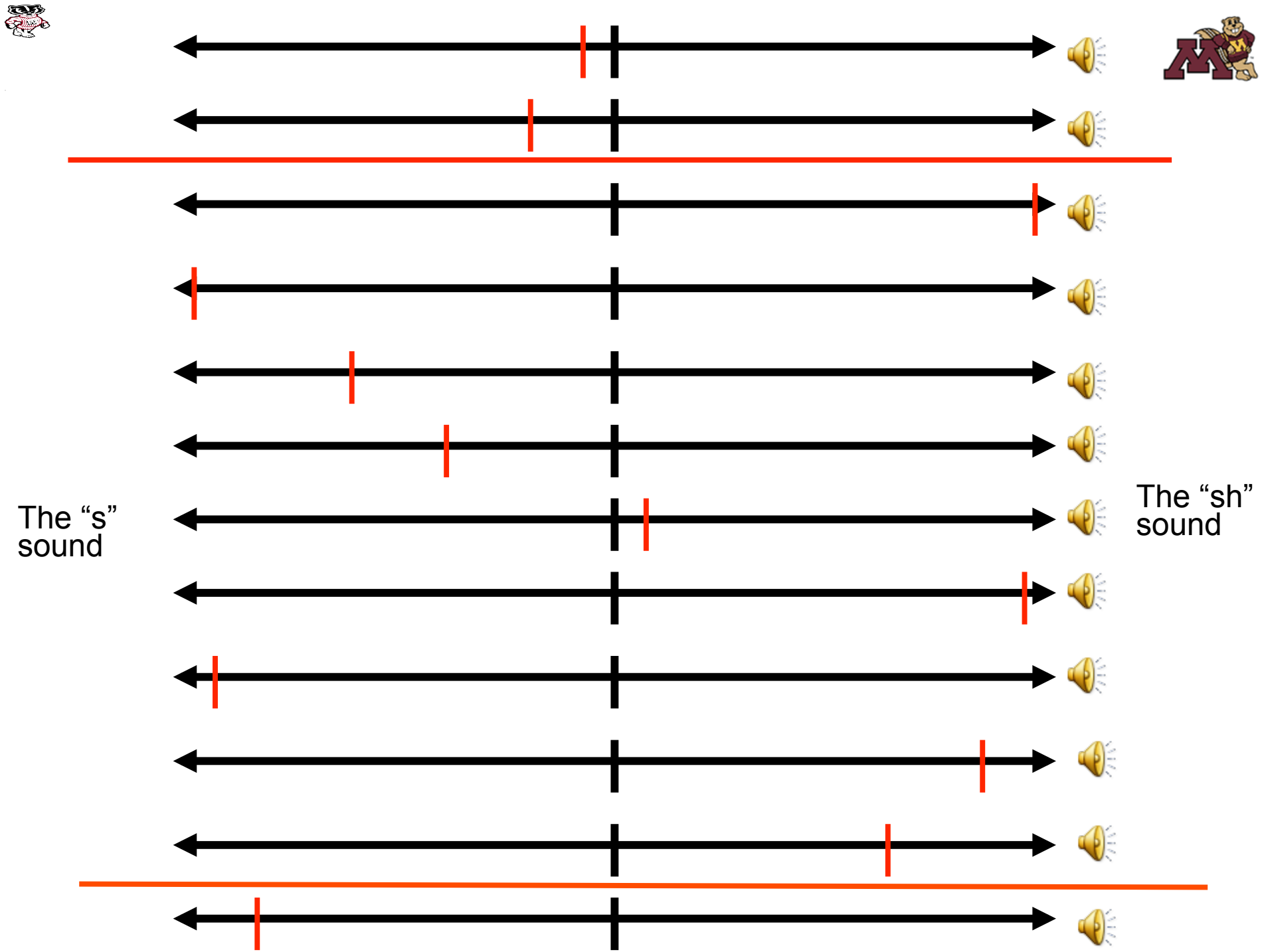

The “s” sound
The “th” sound

Supplement Transcription with Continuous Rating Scales





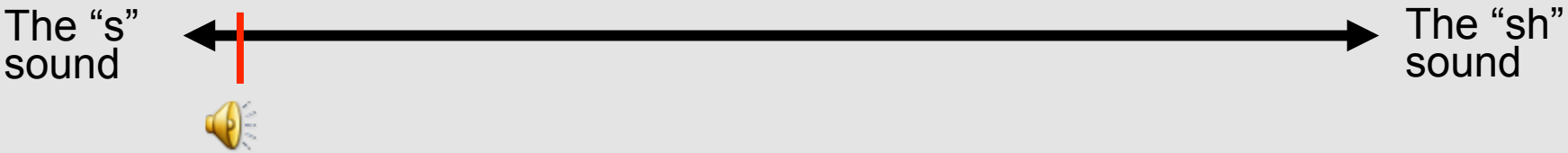

The “s” sound
The “sh” sound


The “s” sound
The “sh” sound

Supplement Transcription with Continuous Rating Scales

- These could be used in transcription to assess a child's production of sounds that appear to be neutralized
- Create VAS scales with the endpoints being the correct sound and the child's habitual substitution
- Scale productions on the line

Supplement Transcription with Continuous Rating Scales



Supplement Transcription with Continuous Rating Scales

- VAS ratings in those cases are probably better than acoustic measures, as acoustic measures necessarily focus on just a few acoustic parmameters, while clinicians' judgments use the full range of our ears (and brains!)

Let us Know how it Works!

- jedwards2@wisc.edu
- munso005@umn.edu
- If you find this useful, we would like to know. If it doesn't work out for you, we would like to know that, too!

Acknowledgments

- Co-PI: Mary E. Beckman
- Lab colleagues: Tim Arbisi-Kelm, Hyunju Chung, Jeff Holiday, Eden Kaiser, Eunjong Kong, Fangfang Li, Marie Meyer, Sarah Schellinger, Asimina Syrika, Kari Urberg-Carlson
- Help with local arrangements by Catherine McBride-Chang, Katerina Nicolaidis, Areti Okalidou, Kiyoko Yoneyama
- Support from NIDCD Grant 02932 and NSF Grant BCS-0729140 to Jan Edwards and NSF Grant BCS-0729277