# Using corpora to study size effects

Mary E. Beckman, Kathryn Campbell-Kipler, Kiwako Ito,
Elizabeth McCullough, Eric Ruppe, Tsz-Him Tsui

Department of Linguistics
The Ohio State University
http://ling.osu.edu/research/groups/soundsize

# Why use corpora ?

*"[C]orpora help us bridge the gap between the analyst's conception of the data and the actual data themselves. Phonologists have formulated generalizations, some of which, as we know now, thanks to corpus-based research, do not do full justice to the data. Language appears to be much more complex than is generally assumed and this complexity is important for theories of phonology as well as for theories of speech production and comprehension."*

Ernestus & Baayen (2011)

# Example 1 – Discontinuity between babbling and speech ?

*"During the babbling period the child easily produces the widest variety of sounds (for example, clicks, palatalized, rounded or pharyngealized consonants, affricates, sibilants, etc.) almost all of which he eliminates upon passing to the 'few words' stage..."*

Jakobson (1939)

# Example 1 – Discontinuity between babbling and speech ?

*"During the babbling period the child easily produces the widest variety of sounds (for example, clicks, palatalized, rounded or pharyngealized consonants, affricates, sibilants, etc.) almost all of which he eliminates upon passing to the 'few words' stage..."*

Jakobson (1939)

- Vihman, Macken, Miller, Simmons & Miller (1985) use the Stanford Child Phonology Study corpus to show "striking parallelism between babbling and words within each child, across time and within time period."

# Example 2 – Universality of "Expressive palatalization" ?

*[a] type of palatalization that is not phonologically conditioned, but ... [is] associated with "smallness", "childishness", or "affection" (Nichols 1971, Ferguson 1977, Ohala 1994)... [and] is used cross-linguistically in sound symbolism, diminutive morphology, hypocoristics, and in "babytalk".... In Japanese, for example, palatalization is associated with "childishness" ... in sound-symbolic (mimetic) vocabulary (Hamano 1998). Palatalization of certain consonants also serves as a marker of the Japanese babytalk register (Chew 1969).*

Kochetov & Aldetere (2011)

# Example 2 – Universality of "Expressive palatalization" ?

*[a] type of palatalization that is not phonologically conditioned, but ... [is] associated with "smallness", "childishness", or "affection" (Nichols 1971, Ferguson 1977, Ohala 1994)... [and] is used cross-linguistically in sound symbolism, diminutive morphology, hypocoristics, and in "babytalk".... In Japanese, for example, palatalization is associated with "childishness" ... in sound-symbolic (mimetic) vocabulary (Hamano 1998). Palatalization of certain consonants also serves as a marker of the Japanese babytalk register (Chew 1969).*

Kochetov & Aldetere (2011)

Despite showing many other hallmarks of infant-directed speech (Igarashi & Mazuka, 2008), few of the recordings of mothers in the RIKEN Japanese Mother-Infant Conversation Corpus show any instances of Expressive Palatalization (Mazuka, p.c.)

# Using corpora to ask questions about sound size effects

Overview

- Sample research questions we can address related to sound size effects
- Methodological questions related to initial corpus building
- Methodological questions related to subsequent corpus reuse

# Specific questions we could ask about ...

The relationship between vocabulary size and phonological acquisition

- Japanese-learning children stereotypically substitute [ʃ] for /s/
- English-learning children stereotypically substitute [s] for /ʃ/
- Can these patterns be explained by the relative frequencies of the two sounds in the earliest words learned by the children ?

# Specific questions we could ask about ...

The relationship between vocabulary size and phonological acquisition

- Japanese-learning children stereotypically substitute [ʃ] for /s/
- English-learning children stereotypically substitute [s] for /ʃ/
- Can these patterns be explained by the relative frequencies of the two sounds in the earliest words learned by the children ?

The relationship between patterns observed in infant-directed speech and acquisition of phonemes

- Expressive palatalization of /s/ and /ts/ is said to be a marker of Japanese babytalk register.
- How often is such palatalization used, and how is that related to the stereotypical consonant substitution pattern ?

# Specific questions ... (cont.)

The interpretation of "fronting" errors such as [s] for /ʃ/

- English-learning children also stereotypically substitute [θ] for /s/
- How do English-speaking adults interpret these patterns relative to the perceived size/age of the talker ?

# Specific questions ... (cont.)

The interpretation of "fronting" errors such as [s] for /ʃ/

- English-learning children also stereotypically substitute [θ] for /s/
- How do English-speaking adults interpret these patterns relative to the perceived size/age of the talker ?

The relationship to possible size effects in older talkers

- In at least some English-speaking cultures, a less extreme "fronting" is interpreted as a "gendering" of /s/ in adult talkers
- Is this a talker size effect ? Also, how is it interpreted relative to the interpretation of "fronting errors" in younger talkers ?

# Designing the recordings

Questions about sampling methods

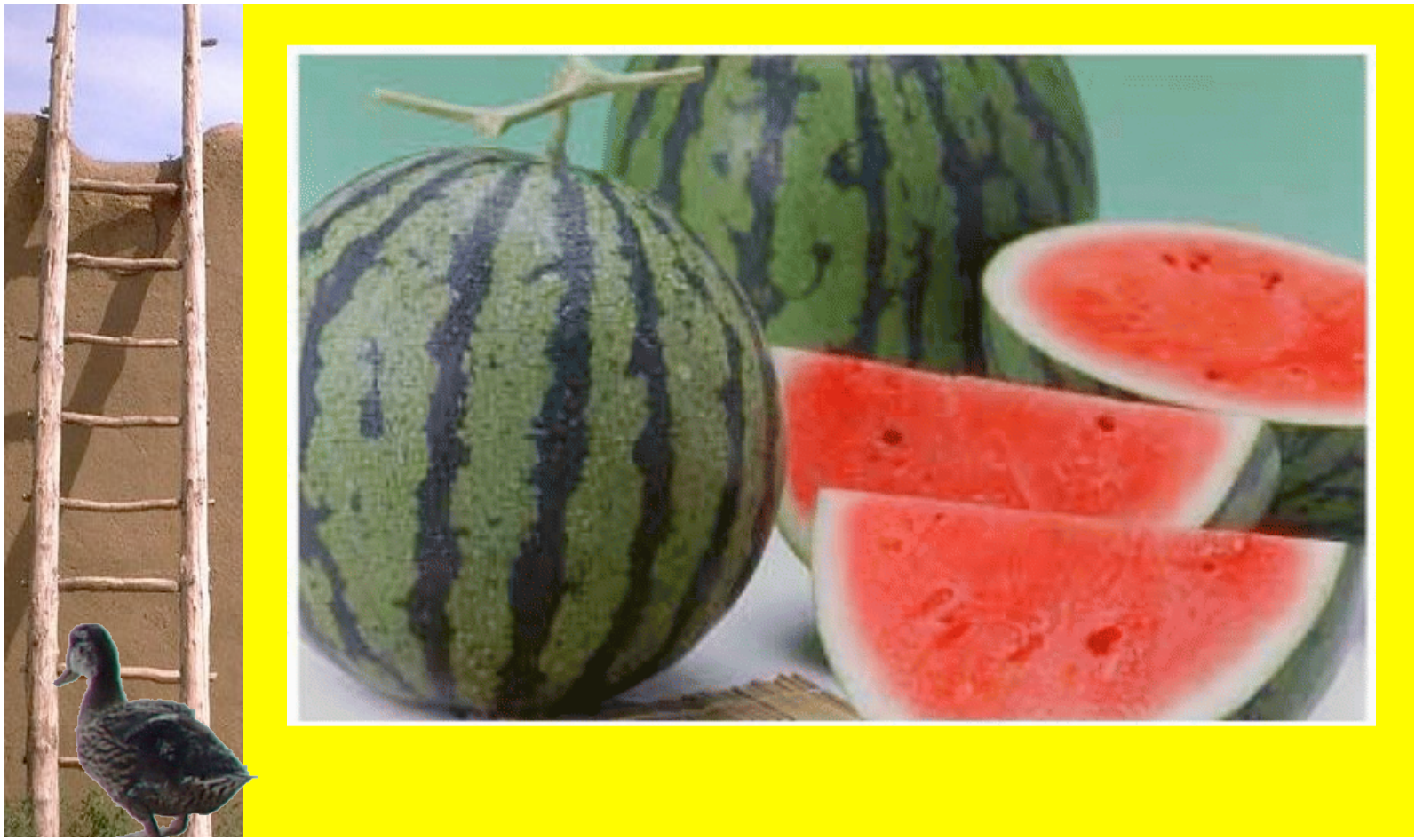
- One language / dialect versus several
- Cross-sectional versus longitudinal
- Many talkers versus dense sampling (compare Iowa / Nebraska Articulation Norms Project versus The Human Speechome Project)

Questions about elicitation protocols

- Lena recordings
- Sociolinguistic interviews versus serendipitous reuse (e.g. call-in radio programs)
- Structured and unstructured play sessions
- Targeted productions in picture-naming or repetition tasks, etc.

# Example : picture-prompted word repetition task

Child sees picture of *suika* 'watermelon', hears audio prompt (word in child-directed voice), and repeats word.



# Methodological issues re this sampling protocol

Edwards & Beckman (2008) on the problems ...

- Does not control for word familiarity
- Samples repetitions rather than spontaneous productions

And on the advantages ...

- Controls for spontaneity, since word is modeled for older children as well as younger ones
- Can elicit a large number of target word productions very quickly
- Can design word list to control for vowel context and overcome consonant type frequency
- Can elicit words that may not be familiar to youngest children
- Can even elicit nonwords – e.g., to sample sequences that are very frequent in one language but phonotactically illegal in another

# More general questions about sampling methods

Issues that we would like the panelists to help us address include ...

- Are there other advantages and disadvantages of the picture-prompted word repetition task ?
- How can we overcome the disadvantages ? For example, can we refine other sampling sampling protocols to control for spontaneity ?
- What are the advantages and disadvantages of the different protocols for sampling child-directed speech ?
- How can we overcome the disadvantages ?
- How can we contribute to the community of corpus builders and users who participate in corpus sharing through PhonBank so as to develop a forum for continued discussion of these issues ?
- Other questions related to "reproducible science" here ?

# Designing the tags

Annotation conventions

- Tags for talker and context properties
- Orthographic transcription of parent and child turns (cf. CHAT)
- Symbolic transcription of phonological targets
- Symbolic transcription of "phonetic" productions

Importance of keeping a trail from the transcriptions / tags back to the primary recordings, etc., other issues discussed in report of Annotation Standards Working Group at :
http ://cyberling.elanguage.net/page/Working+Groups

# Example : child-directed speech corpus

Transcriber segments the mother's utterances in Praat, while making a orthographic transcription with associated time stamp.

(1208.3010177710485) しおちゃん　お鼻出ちゃったよ　おいで

(1211.6015863935875) うん　きれいきれいしよ

(1218.0114621302794) おいしょ　ほらさっぱり　よいしょ

transliteration and gloss:

(1208.3010177710485) shiochan ohana dechattayo oide

Shio-chan, your nose is running. Come here.

(1211.6015863935875) un kireekireeshiyo

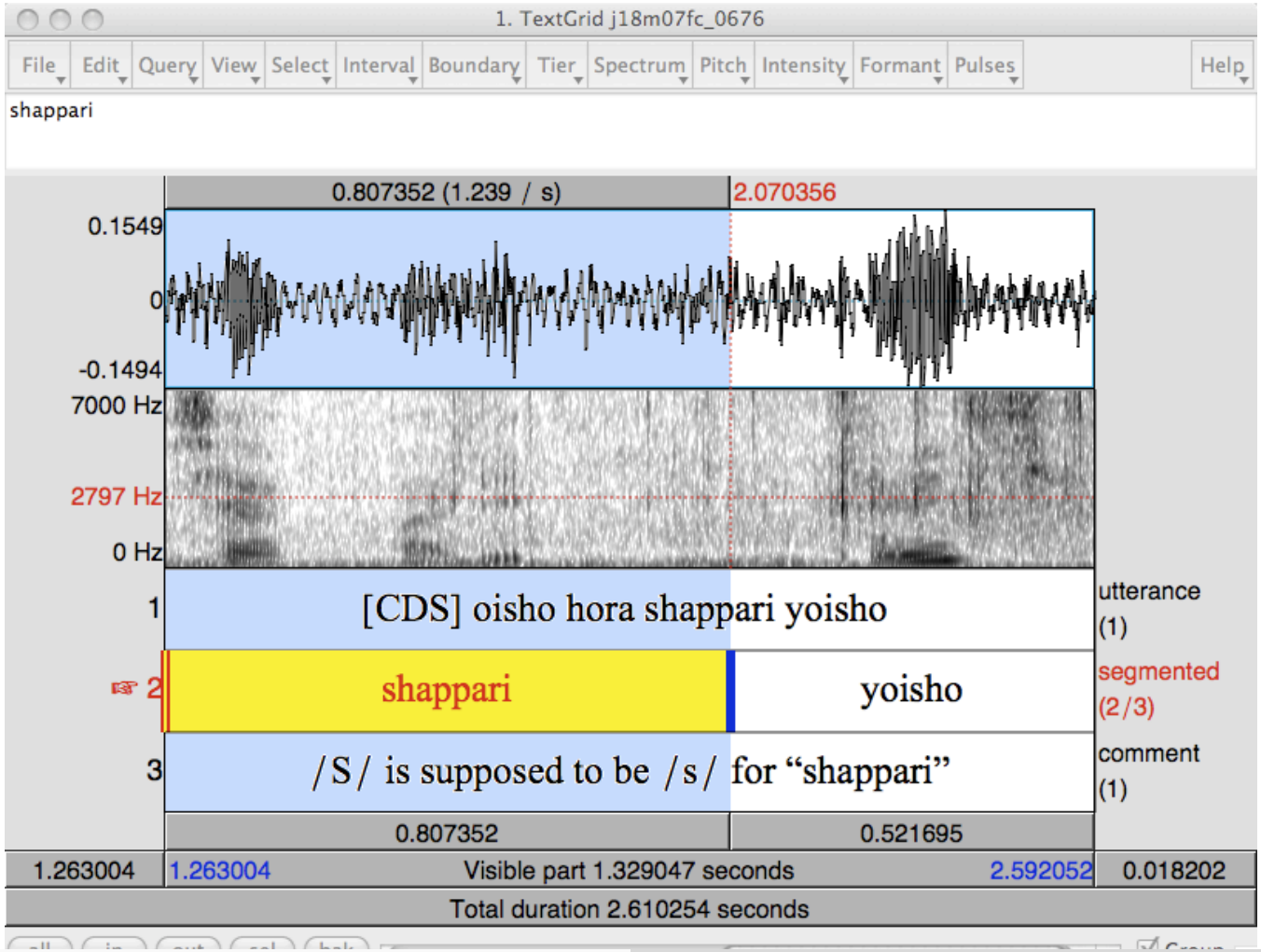
Yes, let's make it clean.

(1218.0114621302794) oisho hora shappari yoisho

Umph.  See, isn't that better? Umph.

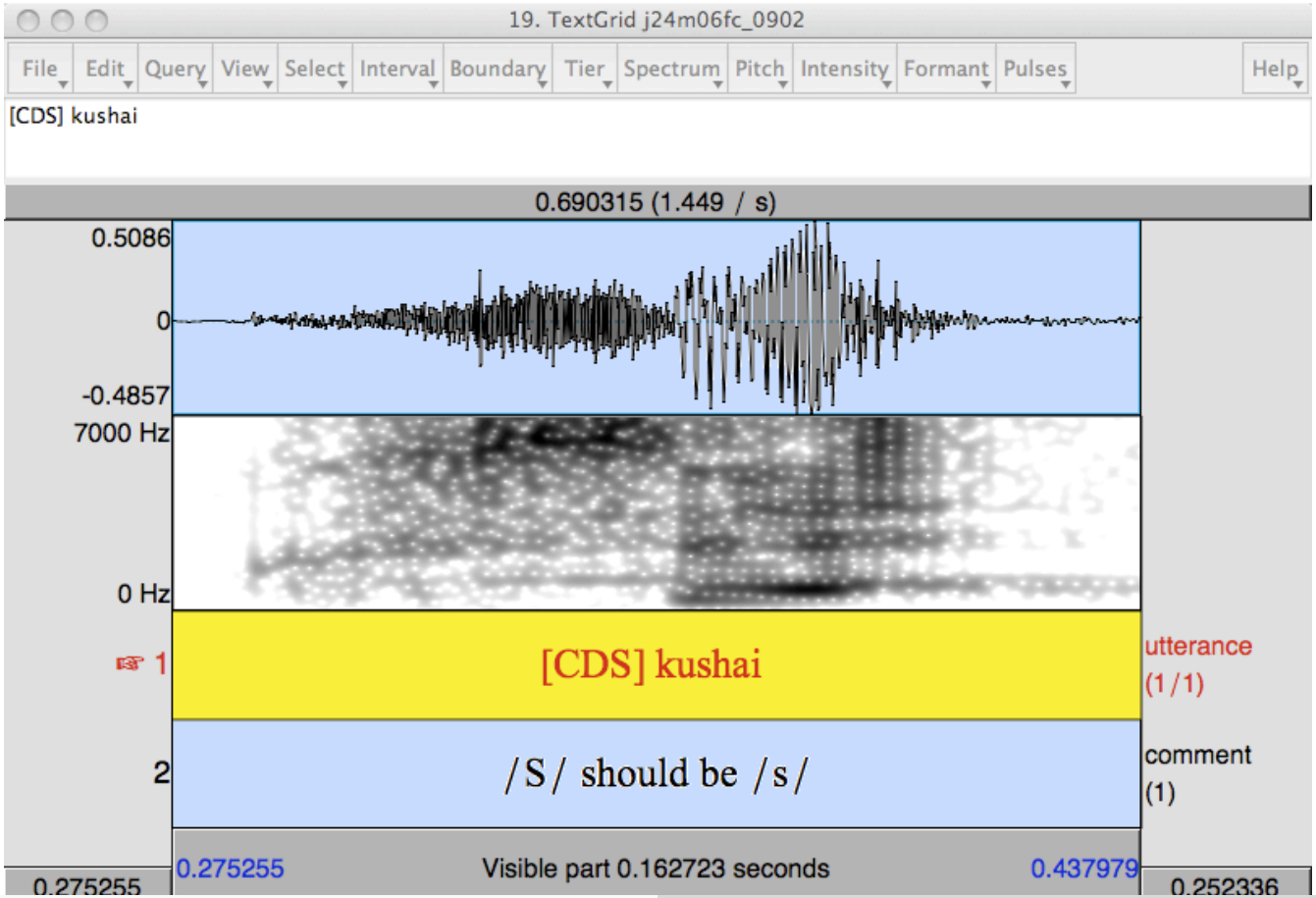
# Transcribing child-directed speech corpus, cont.

Transcriber also inserts a transliterated version of the orthographic transcription, with notes about expressive palatalization, etc.



# Checking instances of Expressive Palatalization

TextGrid files can be searched using grep or Patrick Reidy's R functions. (This instance of tagged Expressive Palatalization was more ambiguous.)



# Methodological issues re this annotation protocol

Advantages

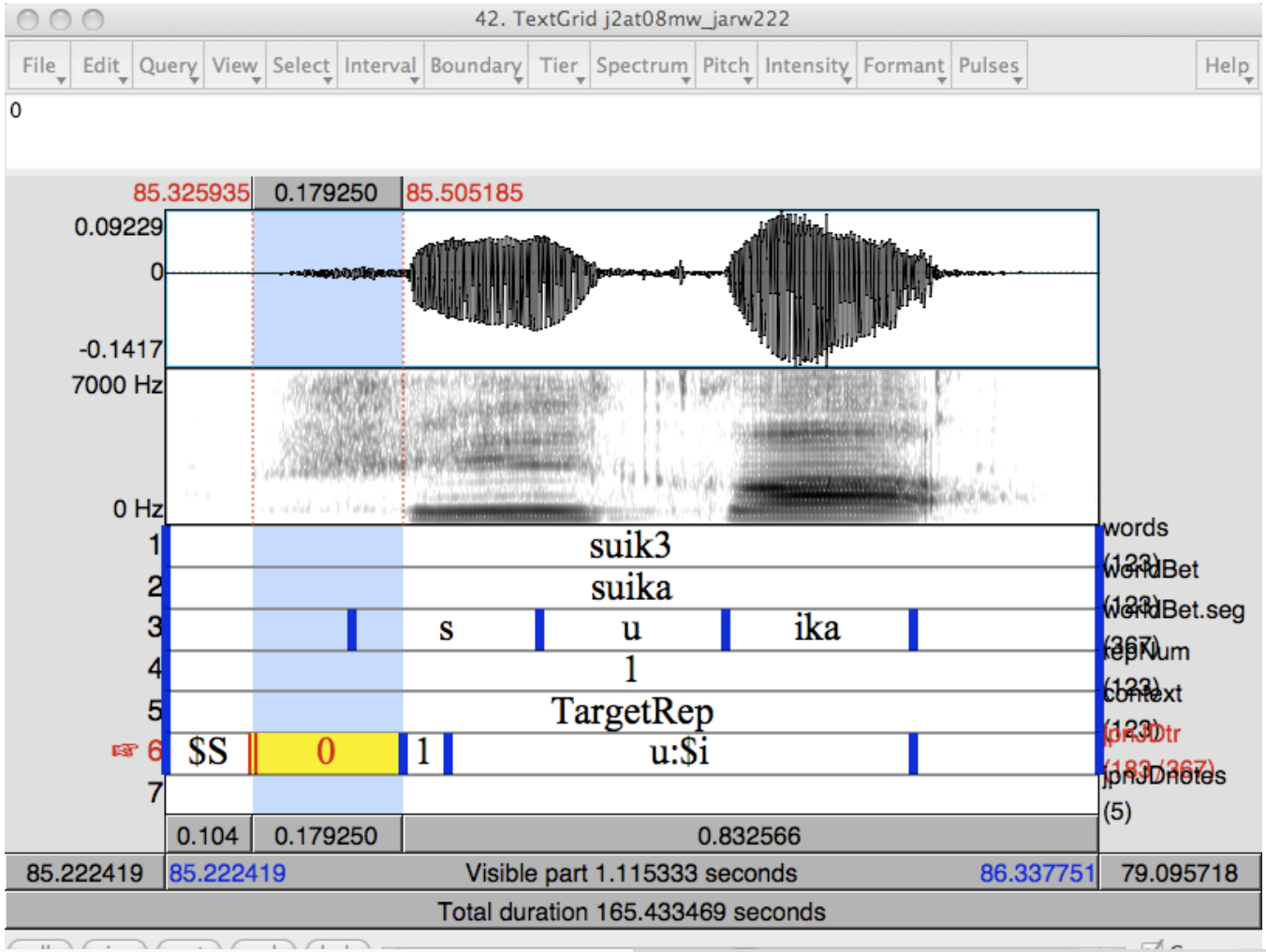
- Transcriber can see wave form and spectrogram
- Could begin transcribing before working out conventions for segmenting into smaller constituents such as Prosodic Words
- Can use orthographic transcription files and Praat TextGrid files in scripts for automatic dictionary look up of target phoneme string (English) and Word segmentation (Cantonese)

Problems

- Didn't first think through the annotation conventions, so ...
- Not taking advantage of work already done on issues of segmentation, etc., as encoded in the CHAT conventions
- Tagging of features of interest (such as Expressive Palatalization in Japanese) is very ad hoc and error-prone
- There is a lot of work left to do to clean up the transcription before we can contribute the recordings to CHILDES

# Example : corpus of elicited word repetitions

Segmentation first. Transcriber then tags each target as correct (1) or incorrect (0) before transcribing substitution if 0.



# Methodological issues re this annotation protocol

The advantages

- Separates (less skilled) segmentation task from (much more skilled) transcription task
- Clearly separates use of symbolic transcription to judge "phonemic" value from use to note possible "phonetic" value
- Use of ASCII version of IPA makes tag files platform independent, searchable in R, etc.

The disadvantages

- Annotation conventions are highly specific to original research question, and hence ...
- Much work to translate into CHAT format when contributed to PhonBank
- Only target C and following V transcribed, so not useful for developing tools such as Preston et al. (2011) Weighted Measure of Speech Sound Accuracy

# More general questions about annotation methods

Issues that we would like the panelists to help us address include ...

- Yvan, Mits, Hélène, can you help us fill this in at the session ?

# Adding tags for phonetic analysis

Tags for various measurements, such as ...

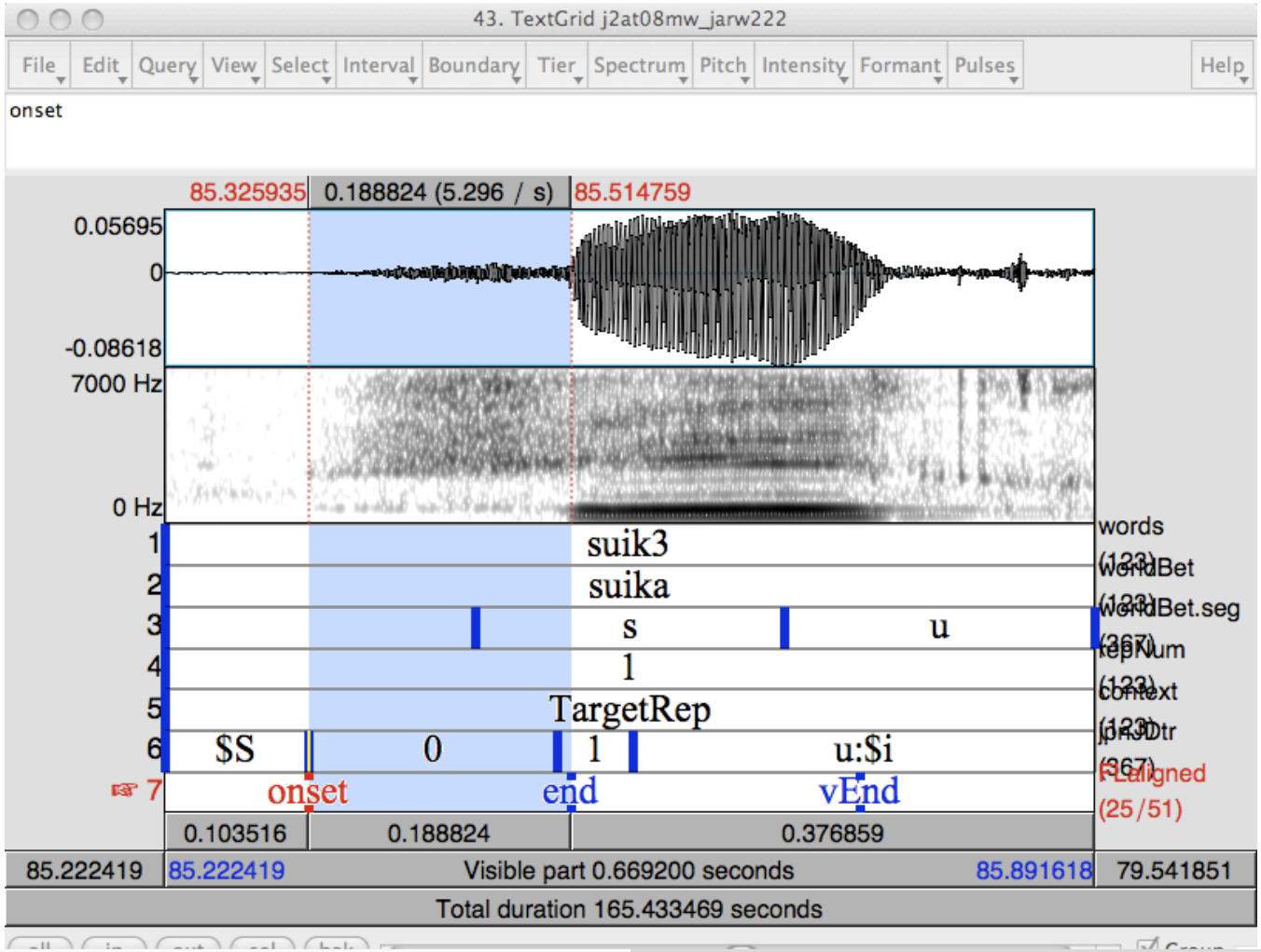
- point of stop release and first glottal pulse in vowel, to measure voice onset time
- interval of frication noise to extract windows for spectral analysis
- beginning and end of initial CV, to be able to extract that interval for use as a stimulus in some kind of perception experiment

Design considerations

- choose tags for easy search
- create tagging protocols (e.g., scripts) to minimize typing errors, etc.
- develop criteria that can be transmitted to new taggers
- test tags and calibrate criteria in inter-tagger reliability tests

# Example : aligning segment boundary tags

Transcriber provides rough identification of segment boundaries.
Trained "aligner" then carefully tags segment edges.



# Example : acoustic analyses of fricatives

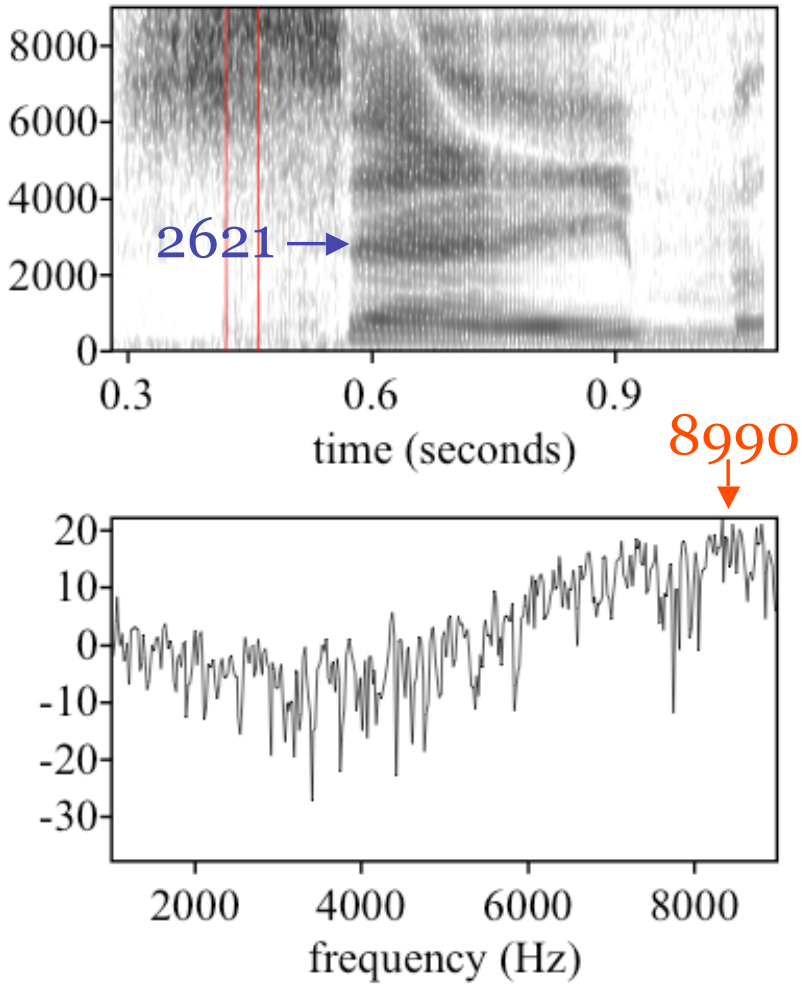
With segment edges marked, can estimate ....

spectral moments

- Choose a representative window during the fricative's turbulent part.
- Calculate a spectrum and treat it as a pdf, by ...
- Calculating moments, such as the mean (or centroid) frequency

formant transitions

- Measure F2 at voice onset



# Comparing acoustic measures across children

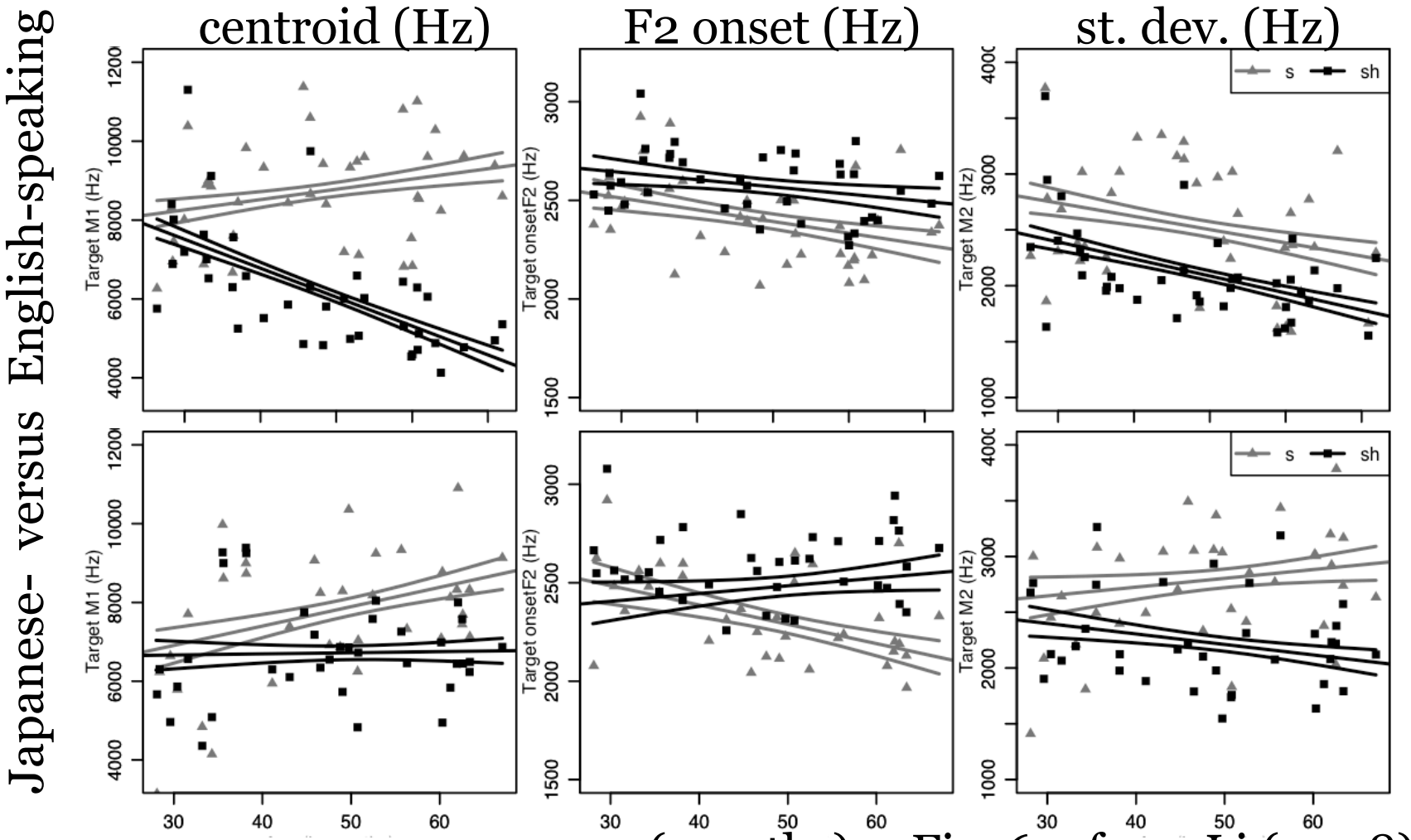
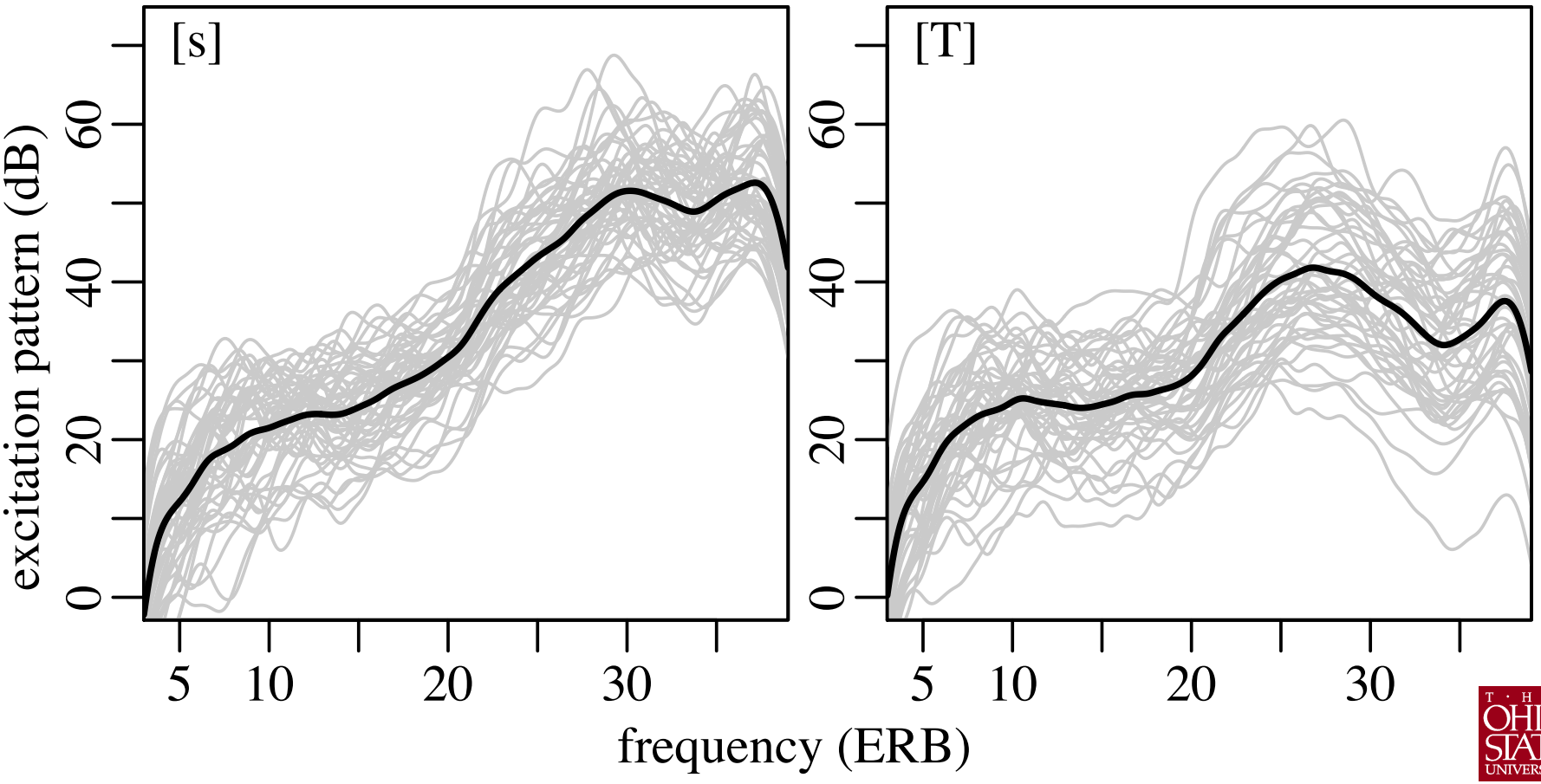


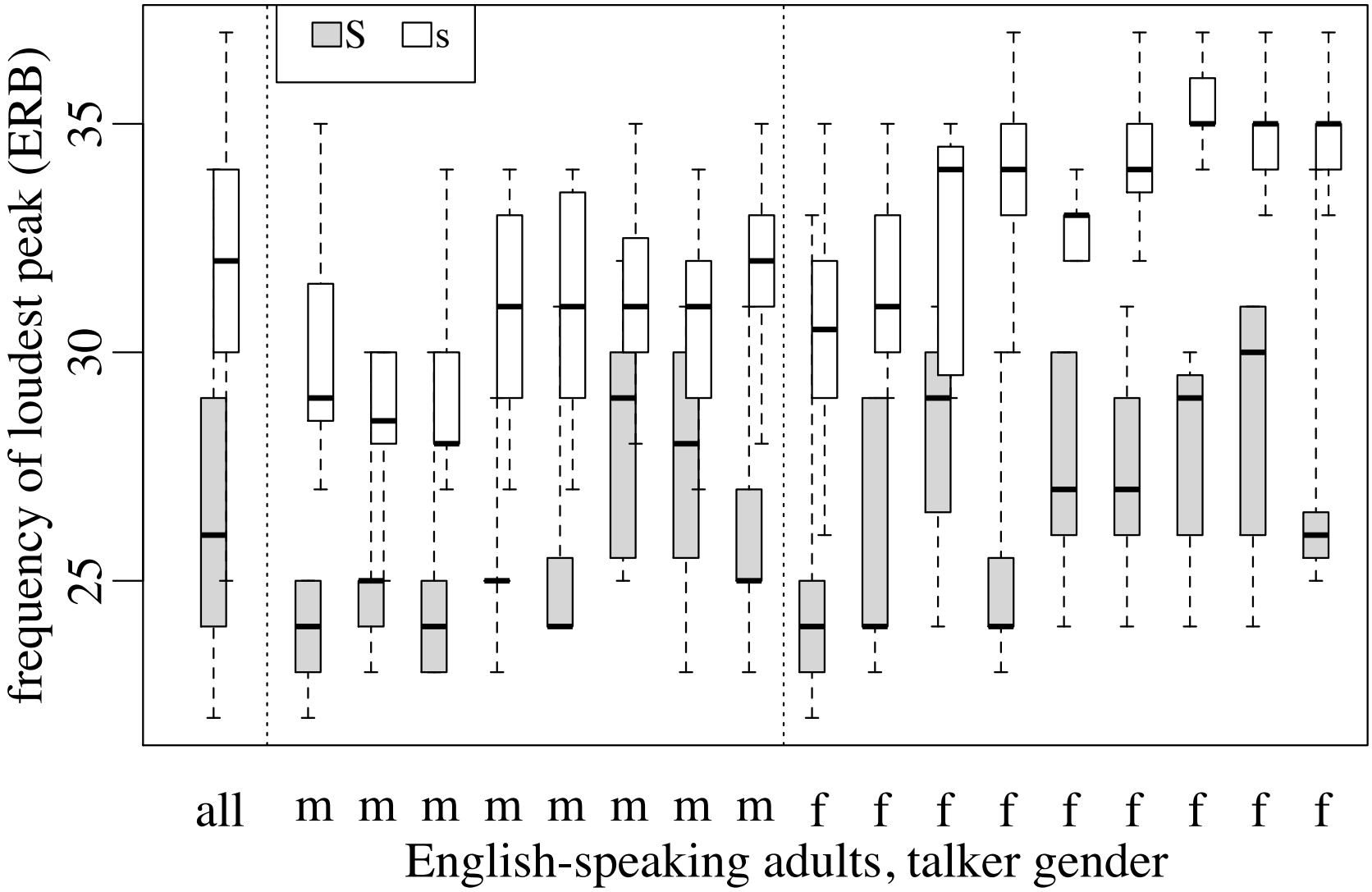

Fig. 6.3 from Li (2008).

# Example : auditory analyses of fricatives

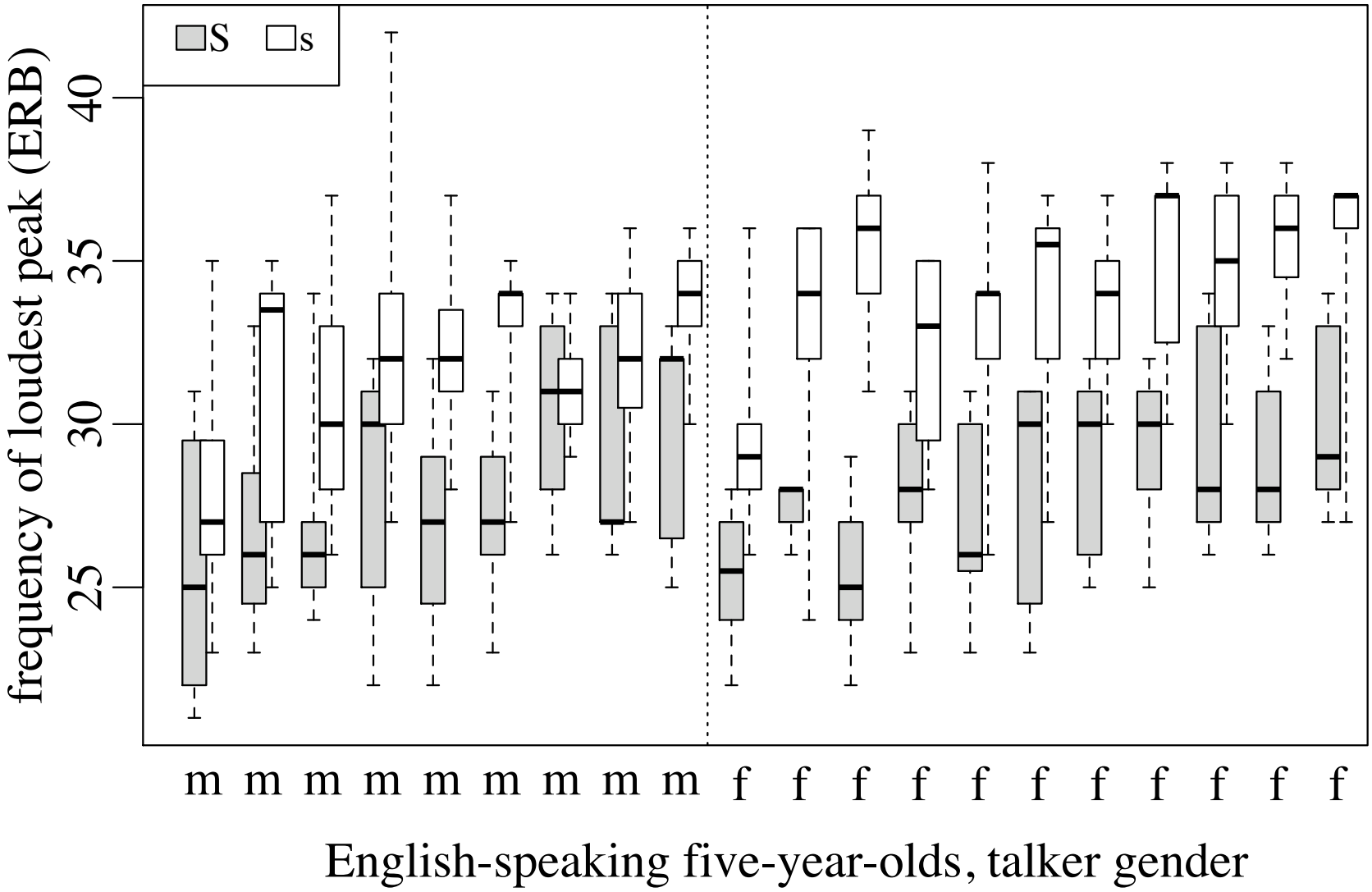
Can apply gamma filters to transform estimated spectrum to excitation pattern ....



# English gendering of /s/, using an auditory measure



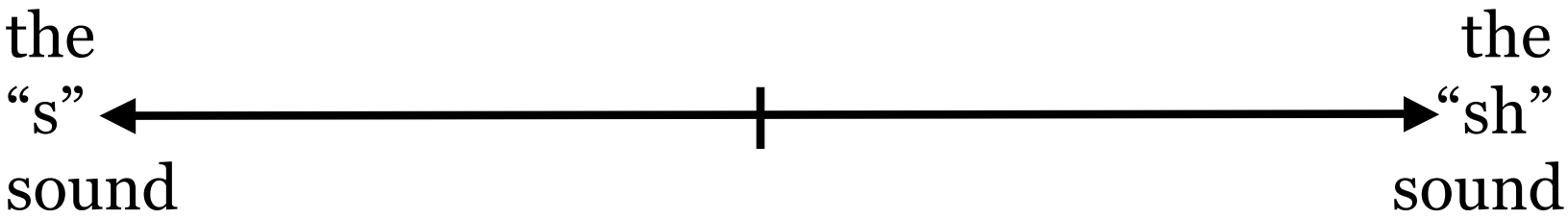
# English gendering of /s/ in five-year-olds



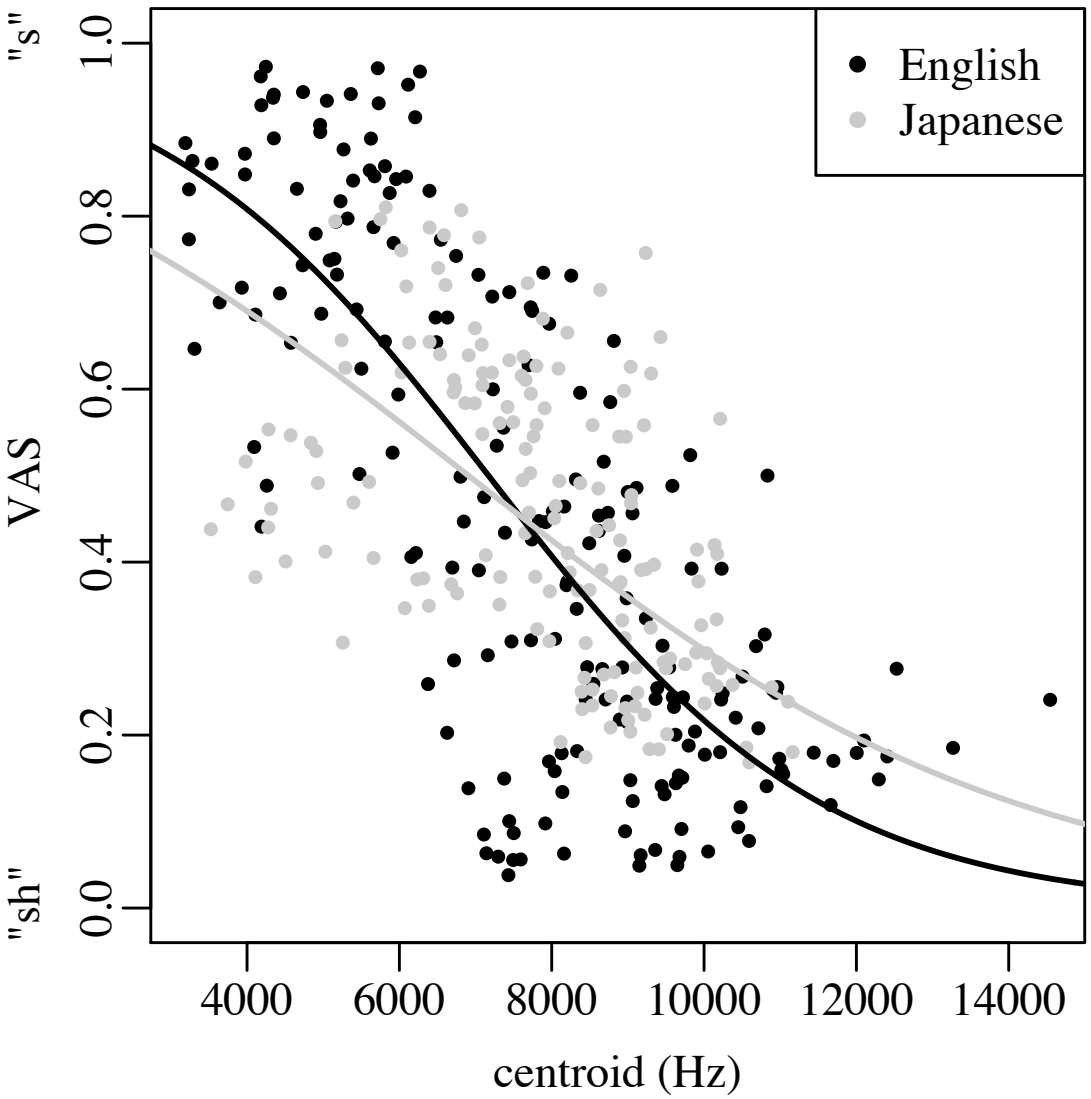
# Example : Use of tags to extract stimuli

Li et al. (2011) use tags for beginning of target C and end of following V to extract stimuli for experiment where listener makes speeded response to "Is this 's' ?" versus "Is this 'sh' ?"
They use "phonemic" transcription for words beginning with /ʃ/ or /s/ to choose an equal number of correct and incorrect productions, balancing for age and following vowel environment.

Urberg Carlson et al. (2008) present the same set of English and Japanese stimuli to English-speaking adults for rating along a visual analog scale (VAS).



# Plotting VAS results against centroid frequency




English-speaking adults rate stimuli extracted from Japanese-children's productions as less good exemplars of "the s sound" and "the sh sound".

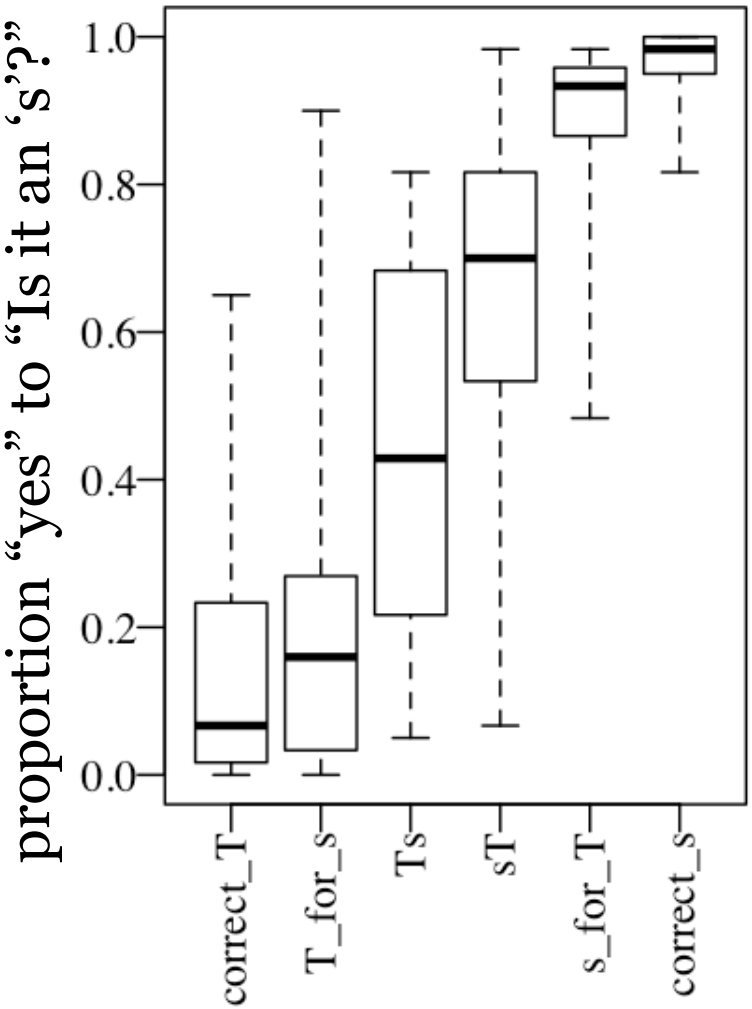
# Example : Use of tags to extract stimuli

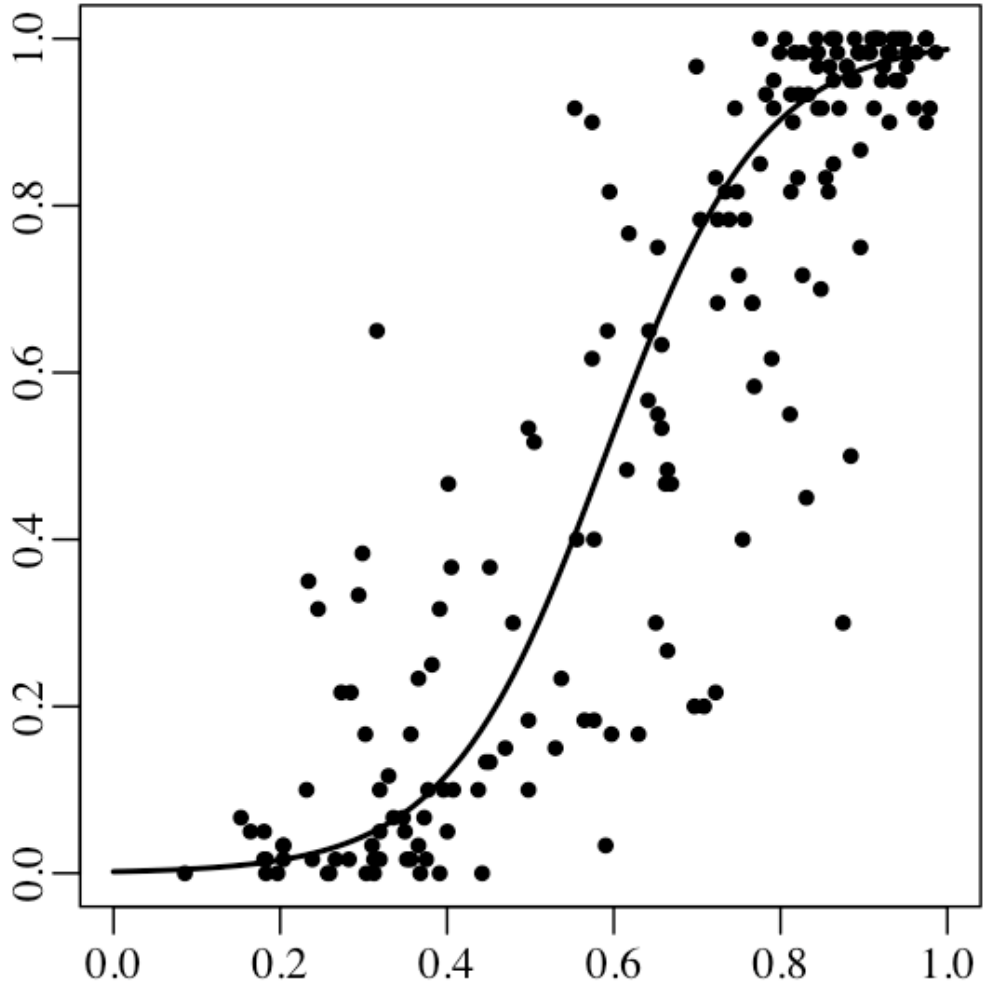
Schellinger (2008) uses tags for beginning of target C and end of following V to extract stimuli for experiment where listener makes speeded response to "Is this a correct 's' ?"
She uses "phonetic" transcription for words beginning with /θ/ or /s/ to choose tokens transcribed as substituting one of the sounds for the other.
She embeds the CV stimuli in 'I really like ____.' versus 'I weawwy yike ____.' frames.

Schellinger, Edwards, and Munson (2008) present the same set of stimuli to English-speaking adults for rating along a visual analog scale (VAS).

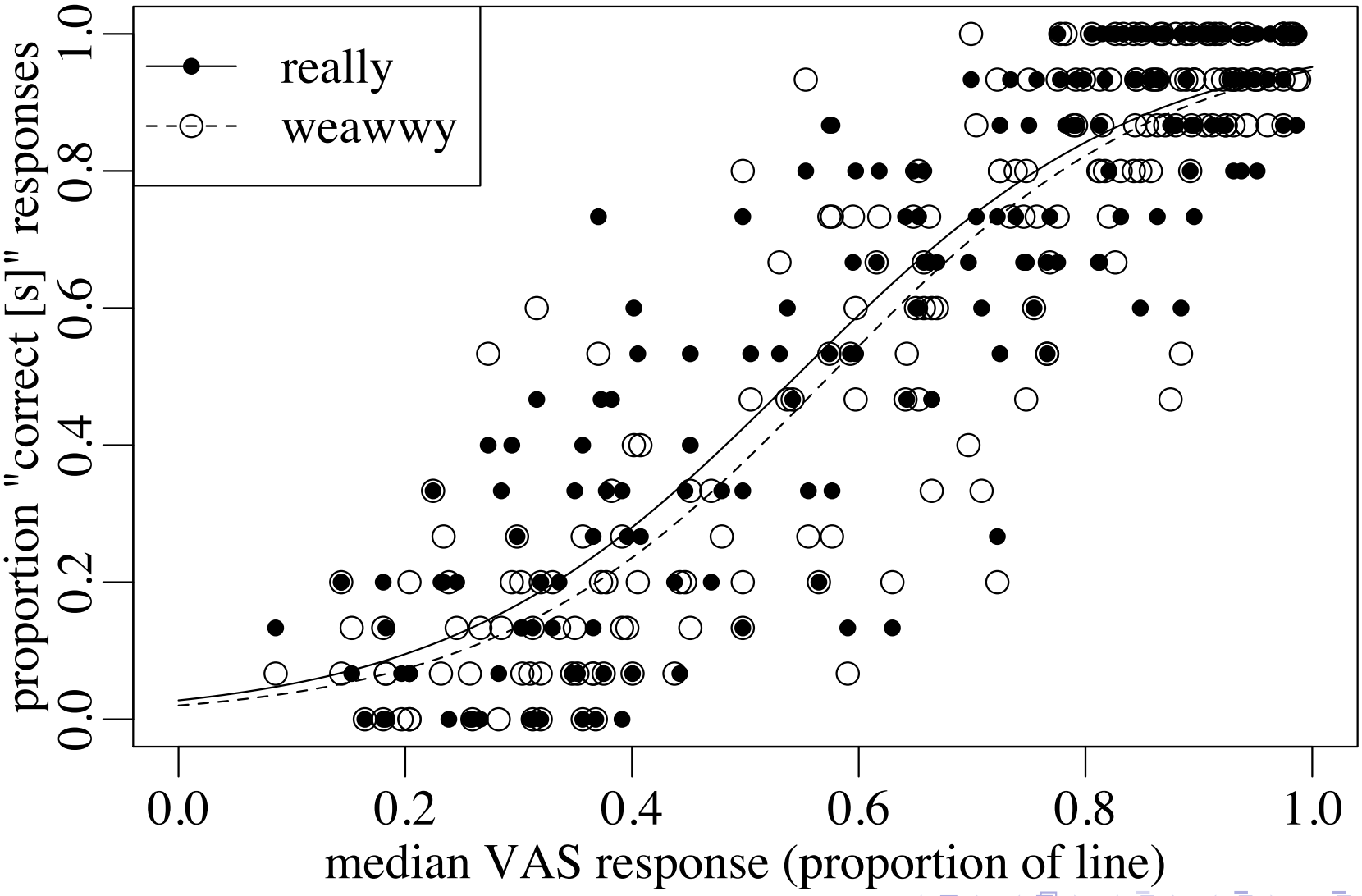
Munson, Johnson, and Edwards, and Munson (2010) present the same set of stimuli to trained speech language pathologists and to lay listeners.

# Relating transcription categories to VAS responses

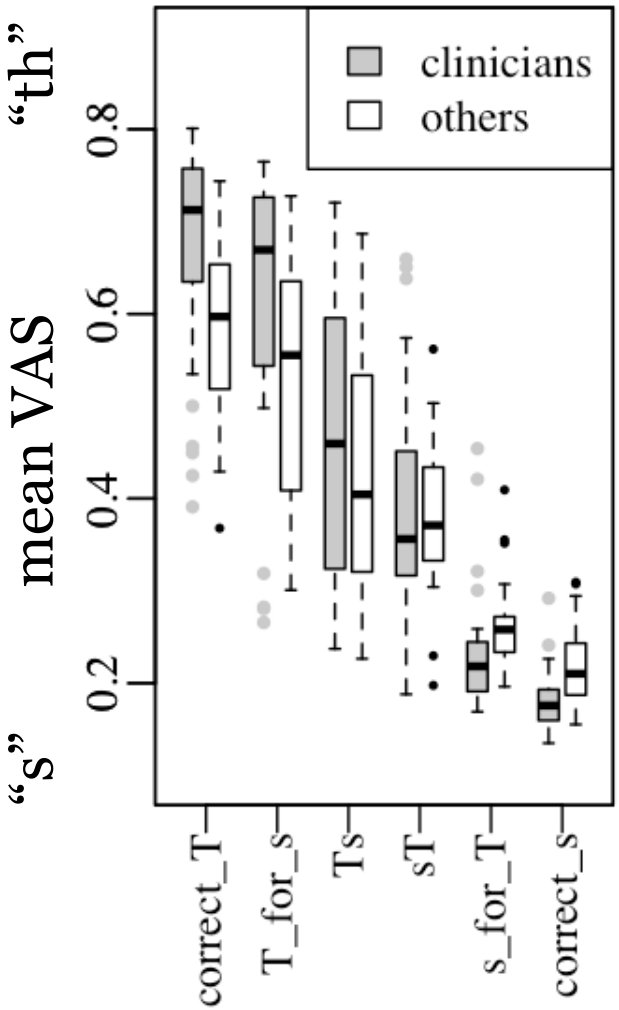


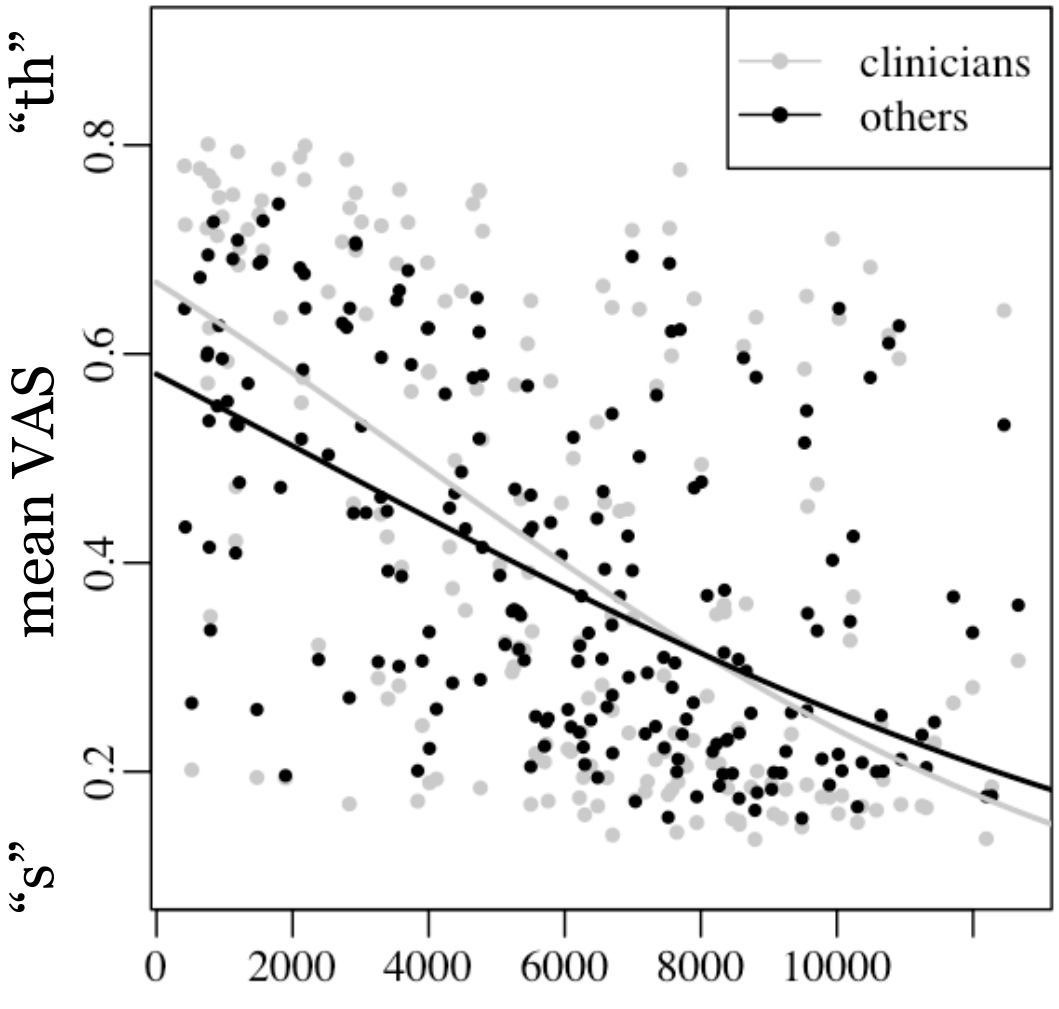



# Small effect of context cueing child size

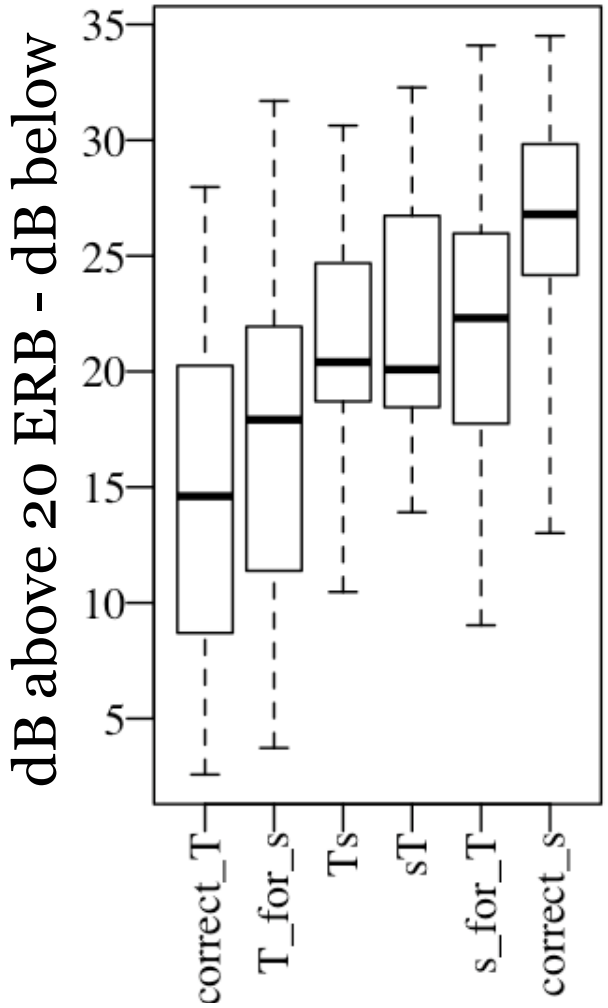


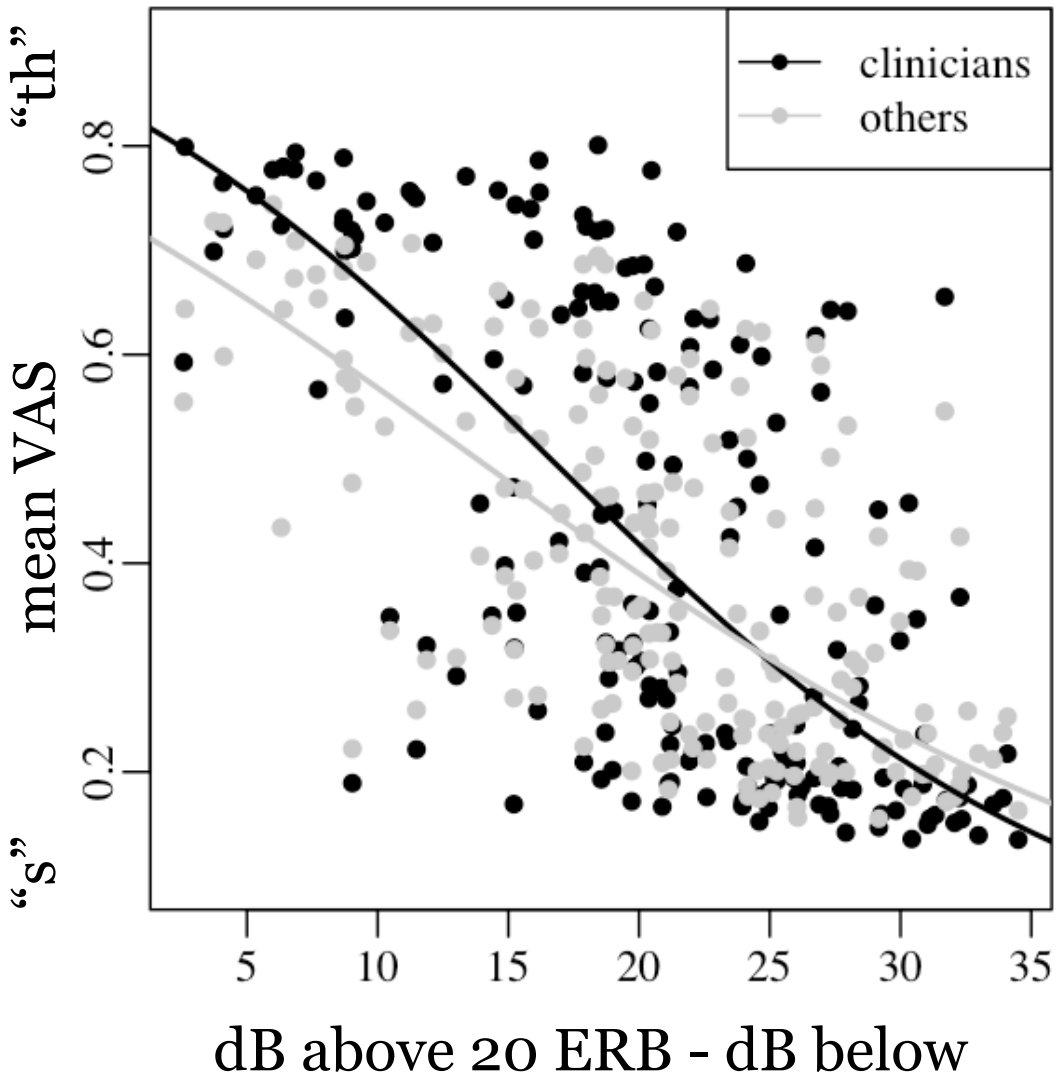
# Bigger effect of experience with children



# Relating VAS responses to an auditory measure



# Work in progress

- Develop better auditory measures that can be evaluated as cues to gendering of /s/, etc.
- Use stimuli from the paidologos corpora in perception tasks to evaluate the interpretation of those cues (see tutorial 2)
- Explore other effects of listener experience
- For example, look at effects on attention and processing of whether the listener produces a robust phonological contrast (see tutorial 3)

# Should such data be "published" with the corpus?

Suppose you've created a corpus of productions of words by English-learning children, published it on PhonBank, and then ...

- you add tags to mark the begin and end of each production of target /s/ versus /θ/ that was transcribed as some kind of more or less anterior coronal fricative
- you use these tags to analyze spectral correlates of the transcribed variation in the productions
- you use the tags to extract stimuli exemplifying the variation
- you do a VAS goodness rating task with adult English-speaking listeners using stimuli from the corpus

How useful would it be to other PhonBank users to have some or all of these added tags / data?

# More general questions about data sharing

Issues that we would like the panelists to help us address include ...

- insuring privacy / informed consent
- rewarding corpus sharing
- making mechanisms for sharing code and computational models
- sharing iterative data augmentation and developing mechanisms for conditioning on reciprocity
- other questions related to "reproducible science"