Cross-language perspectives on the interaction between production and perception in phonological acquisition

26th General Meeting of the Phonetic Society of Japan

by Mary E. Beckman, Ohio State University

Acknowledgments:

- National Institute of Deafness and Communication Disorders grant RO1 02932 (to Jan Edwards)
- Ohio State University Cognitive Science Fellowships (to Kiyoko Yoneyama, Fangfang Li, & Patrick Reidy)
- Ohio State University Targeted Investment in Excellence award (Eun Jong Kong & Fangfang Li)
- National Science Foundation grants BCS 0729306, BCS 0729140, & BCS 0729277 (to Mary Beckman, Jan Edwards, & Benjamin Munson)

Acknowledgments, cont.

杉藤美代子先生を偲びて。。。

Phonological acquisition (音韻獲得) involves ...

developing skills for both:
- production (音声生成)
- perception (音声知覚)

It takes place in the context of social exchange, where children listen to:

1. their own productions
2. others' productions
3. others' productions addressed to them in response to their own productions

Plan for this talk

Goal: to explore what cross-language comparisons can tell us about the relationship between speech production (音声生成) and perception (音声知覚) in phonological acquisition (音韻獲得) via 3 case studies

1. Different perceptual valuation of “unmarked” (無標) feature in the acquisition of stop voicing contrasts
2. New methods for evaluating differences in production and perceptual weighting of cues in the acquisition of sibilant fricative place contrasts
3. Applying methods to understand stop place contrasts
4. and to see effect of perception on production response

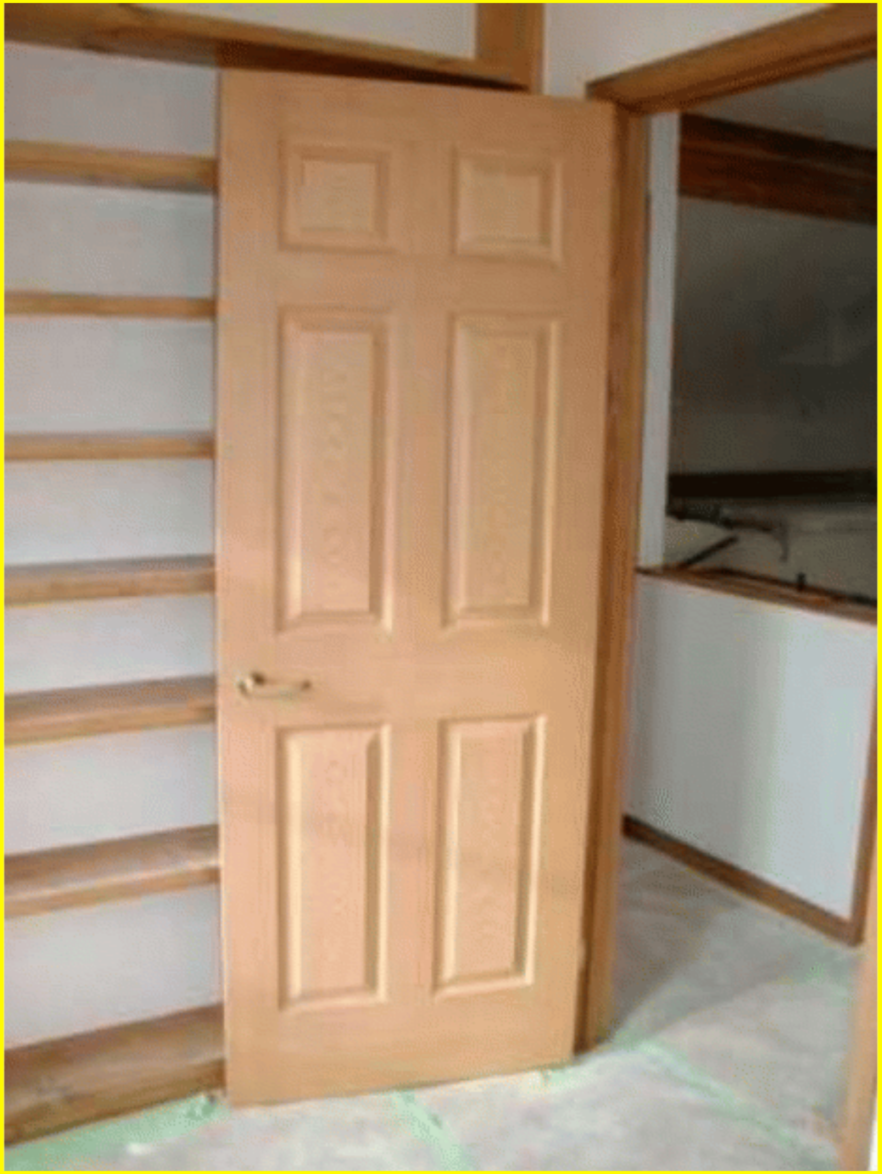
To begin: introduce the παιδολογος project

What is the παιδολογος project?

- Ongoing comparison of phonological acquisition across languages, with data collection for Cantonese, English, Greek, and Japanese starting in 2003. Later: Korean, Mandarin & Min Nan Chinese, French, Drehu
- Data: Productions of analogous sounds in analogous word positions across languages, elicited using the same task and the same recording equipment.
- Subjects: 100-130 children for each target language, covering the same age range (2 through 5 years), and 20-25 adults (same age as the children's parents).
- Transcribed using the same transcription protocol.
- Recordings of initial four languages available at http://childes.psy.cmu.edu/data/PhonBank/

The παιδολογος elicitation task

- Materials are familiar words containing the target consonants in word-initial position.
- Productions of the words are elicited in a picture-prompted repetition task, in which the child …
  - sees a picture depicting the target word,
  - hears a female voice producing the word in a child-directed speech style,
  - and repeats the prompted word.
- Between trials, the child is rewarded by seeing an animal walk up a ladder to the left of the picture.



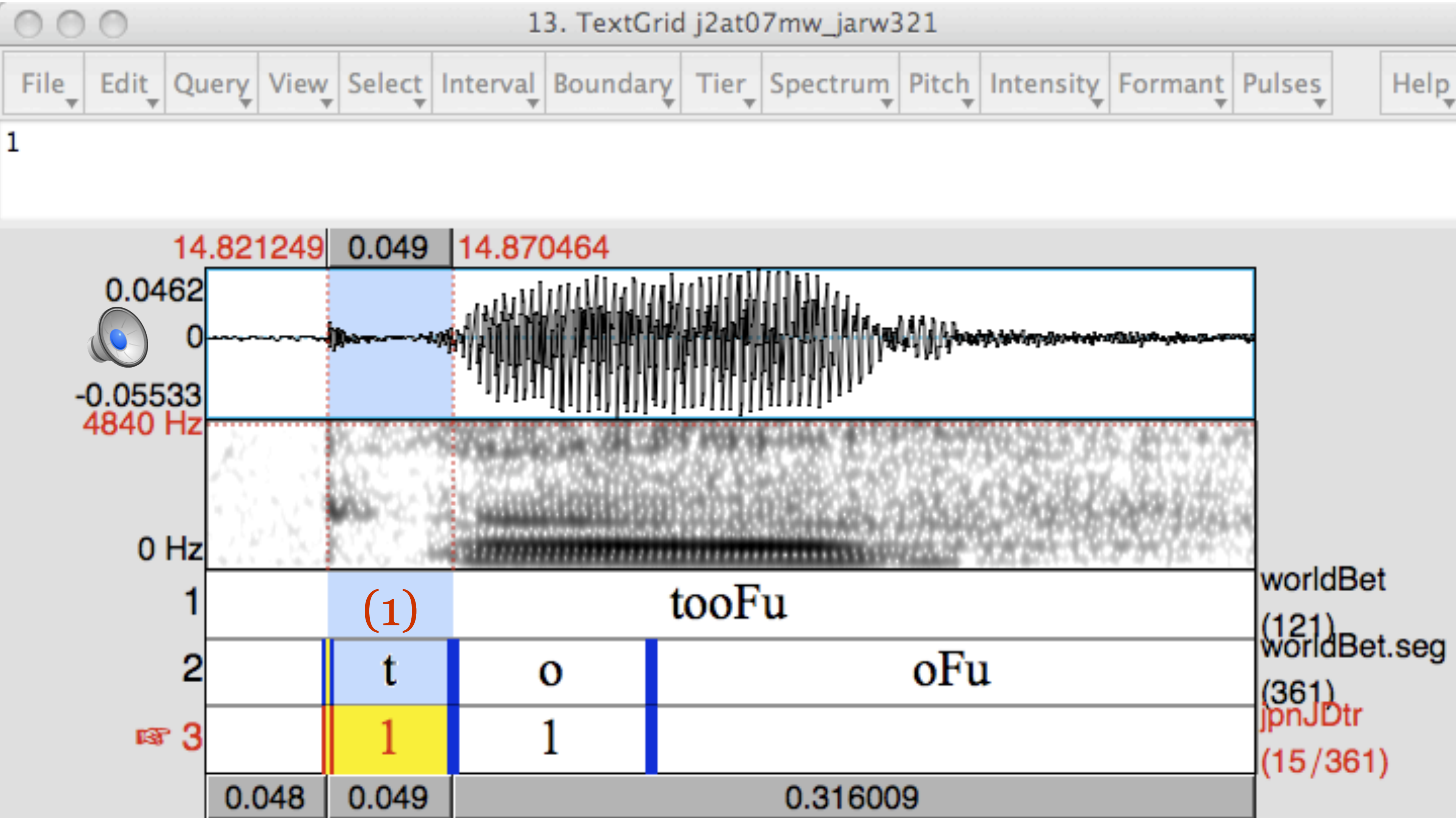
target: [doa] ドア

target: [doŋguri] どんぐり

target: [to:ϕu] 豆腐

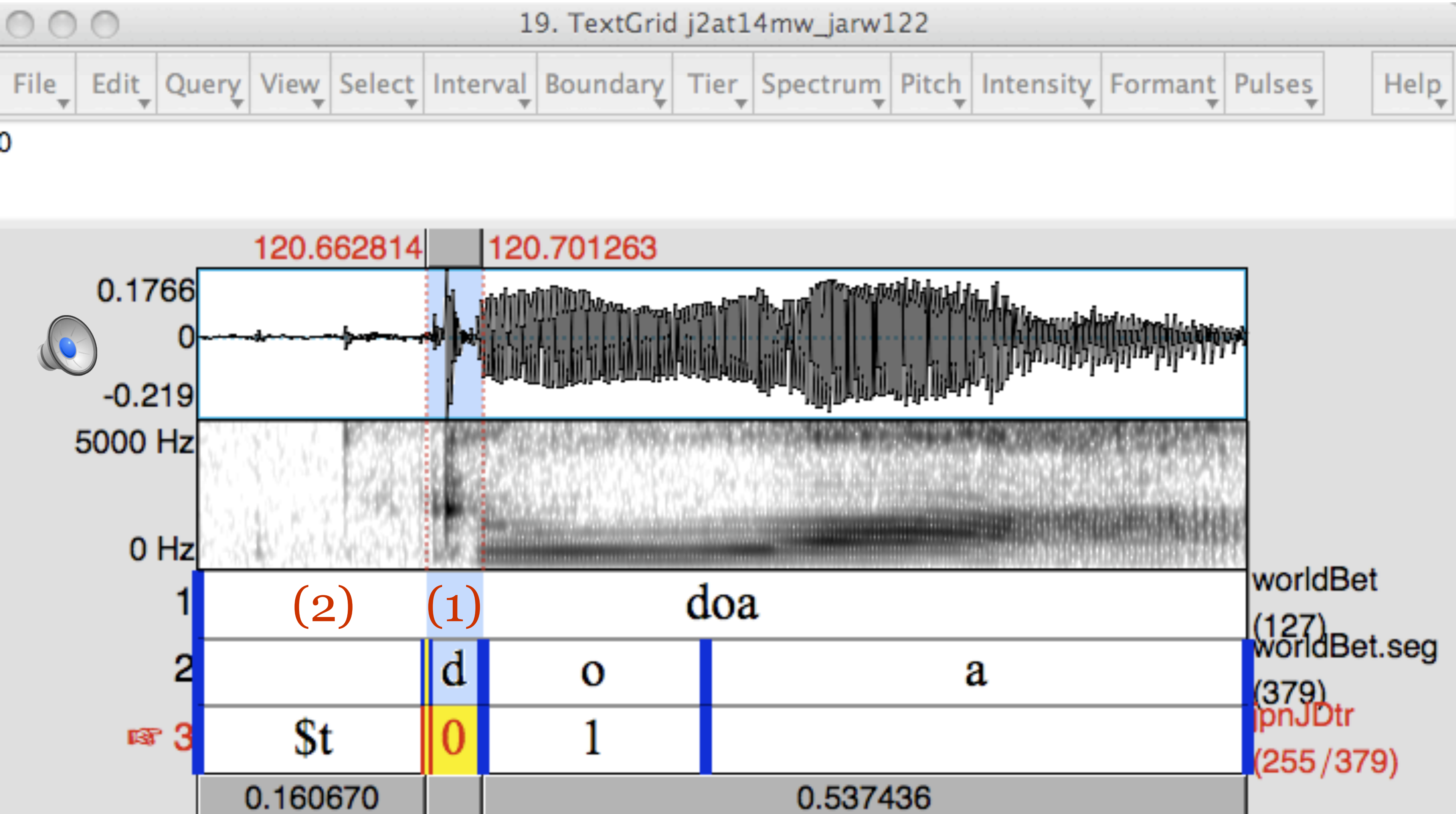
The paidologos transcription protocol

Two-stage transcription protocol for: (1) correct or not and (2) ..



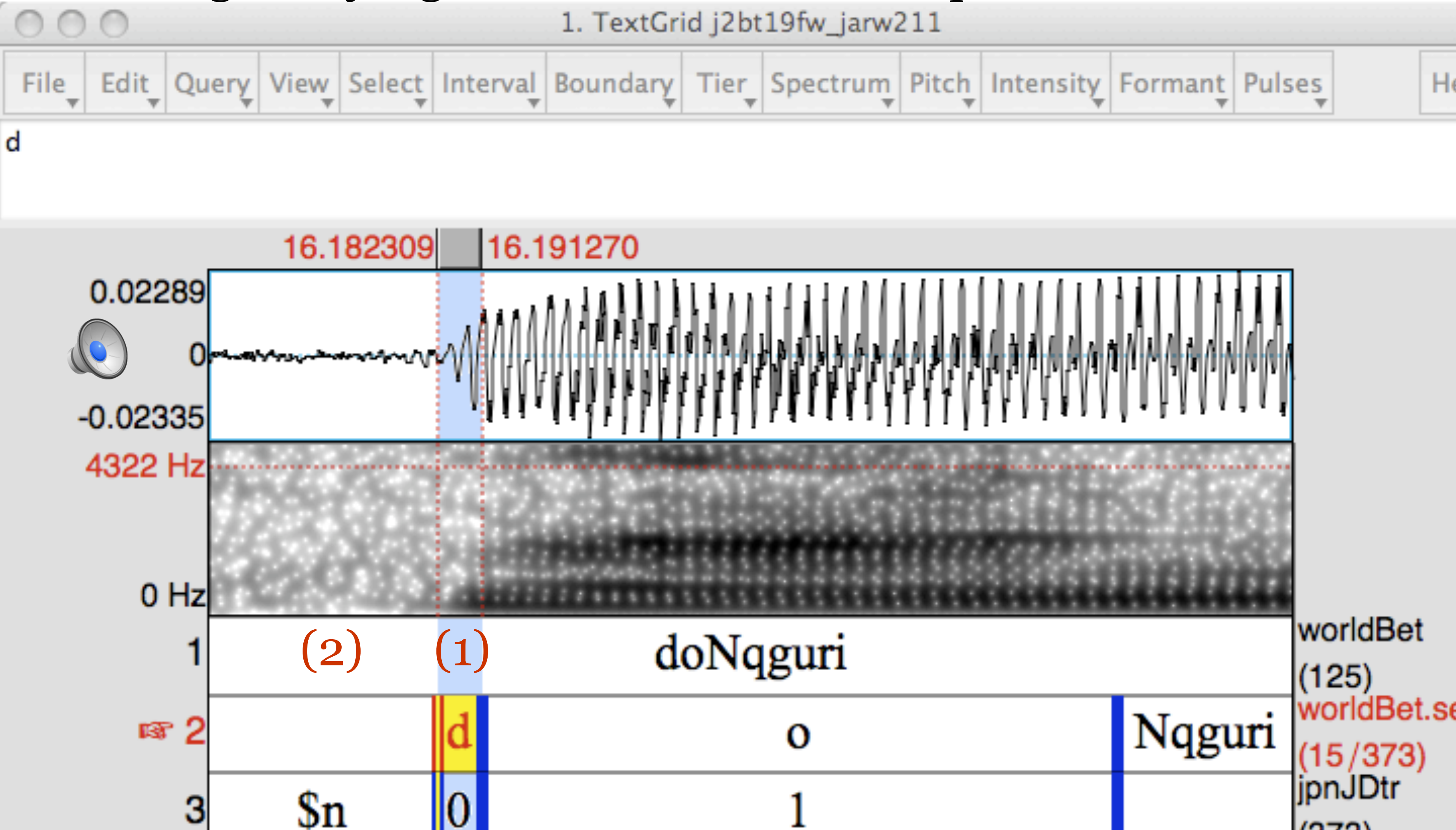
The paidologos transcription protocol

If at stage (1) judged not correct, then (2) perceived substitution



The paidologos transcription protocol

If at stage (1) judged not correct, then (2) perceived substitution



Production: age-typical misarticulations

devoicing or nasalization of voiced stops

target [doa]	ドア	transcribed as:	/d/ → [t]	🔊
target [gasu]	ガス		/g/ → [k]	🔊
target [donguri]	どんぐり		/d/ → [n]	🔊

backing of coronal stops

target [tamago]	卵	transcribed as :	/t/ → [k]	🔊

palatalization (and assibilation) of stops and fricatives

target [k^jimono]	着物	transcribed:	/k^j/ → [tʃ]	🔊
target [semi]	セミ		/s/ → [ʃ]	🔊

Production: age-typical misarticulations

Children who are learning English also have age-typical misarticulations, although details are specific to English.

“voicing” (deaspiration) of voiceless stops

target: [t^hɔl] ‘tall’ transcribed as: /t^h/→[t] (=/d/)

fronting of dorsal stops

target: [k^hot] ‘coat’ transcribed as: /k^h/→[t^h]

fronting of the voiceless post-alveolar fricative

target: [ʃ] ‘shoe’ transcribed as: /ʃ/→[s]

☞ How should we account for these misarticulations?

Generative Phonology: child-specific rules

In Generative Phonology, these misarticulations are described in terms of “rules” that change feature values.

[+voice] → [-voice]

target: [t^{h}ɔl] ‘tall’ transcribed as: /t^{h}/→[t] (=/d/)

[-coronal] → [+coronal]

target: [k^{h}ot] ‘coat’ transcribed as: /k^{h}/→[t^{h}]

[-anterior] → [+anterior]

target: [ʃ] ‘shoe’ transcribed as: /ʃ/→[s]

☞ Why would a child’ s phonological grammar have such seemingly dysfunctional “rules”?

Explaining age-typical misarticulations

Generative Phonology model assumes:

- The target consonant and substituted consonant differ by the value specified for some distinctive feature
- The value that is changed in production is the more difficult “marked” (有標) value for the feature
- The child accurately perceives the distinction between the target consonant and the substituted consonant
- Accurate perception allows adult-like representations of the target consonant in a word’ s “underlying form”
- The child’ s “rule” transforms the underlying form with the difficult-to-produce “marked” feature value into a “surface form” with the “unmarked” (無標) value.

Case study 1: Stop voicing contrasts

Jakobson (1939): “So long as stops in child language are not split according to the behavior of the glottis, they are generally produced as voiceless and unaspirated.”

- As predicted: devoicing of voiced stops in Japanese

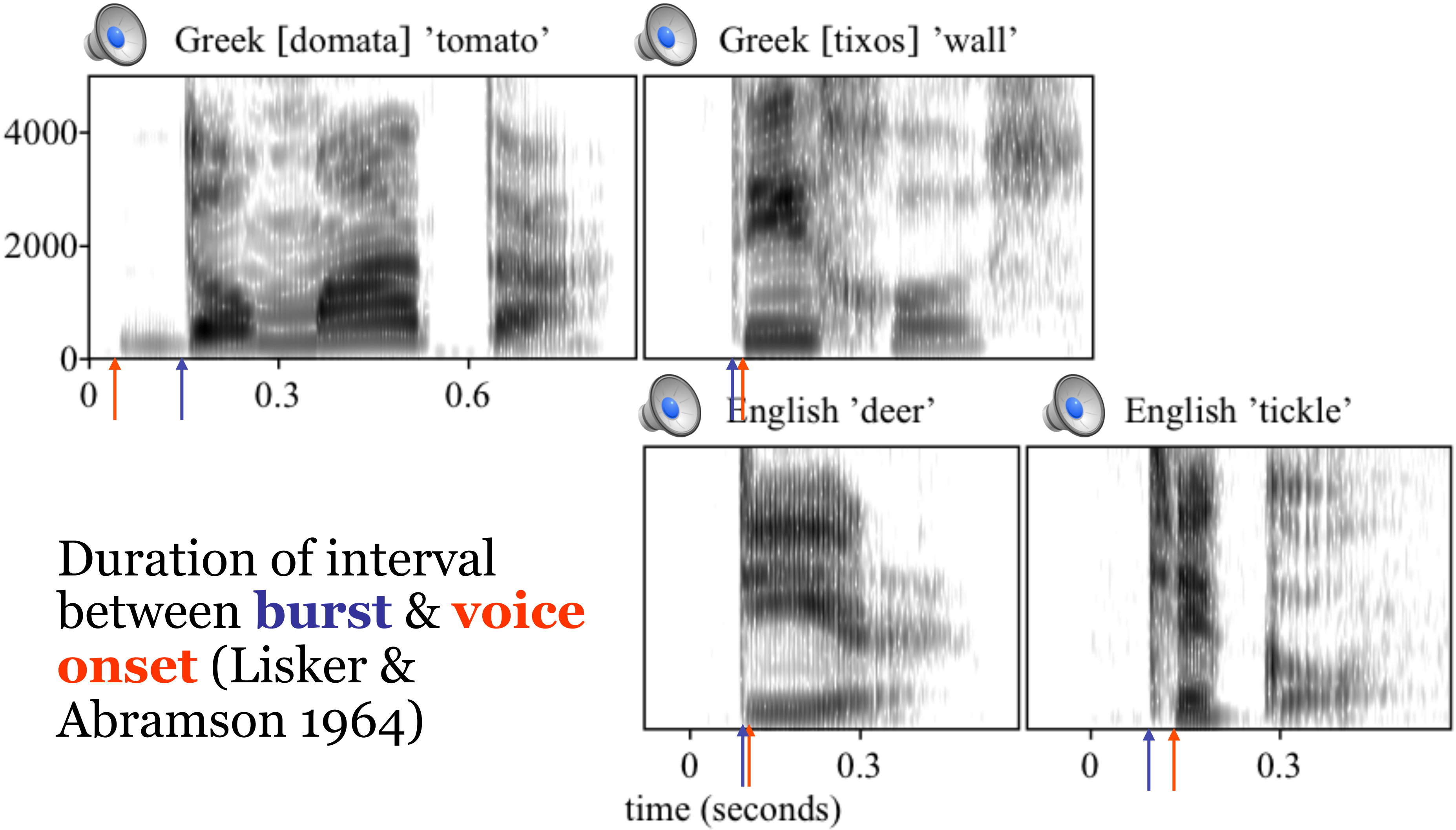
target [doa]	ドア	/d/ → [t]
target [gasu]	ガス	/g/ → [k]

- Exception? “voicing” of voiceless stops in English

  target: [t^{h}ɔl] ‘tall’ /t/ → [

In earlier literature on English as well, stops produced in children’s babbling and first words were perceived (and written) as “b”, “d”, “g” rather than as “p”, “t”, “k” (cf. Darwin 1877, Velten 1943, Ingram 1974)

Voice onset time as a measure of voicing

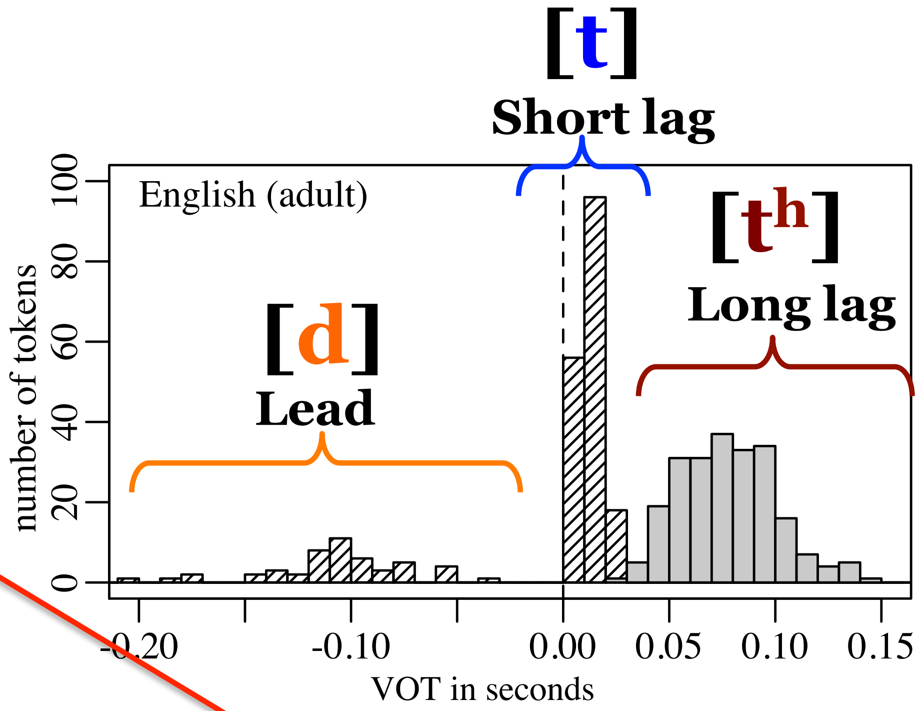


Duration of interval between **burst** & **voice onset** (Lisker & Abramson 1964)

VOT and acquisition of voicing contrasts

Category with short lag VOT first, because it requires the least precise articulation (Kewley-Port & Preston 1974).

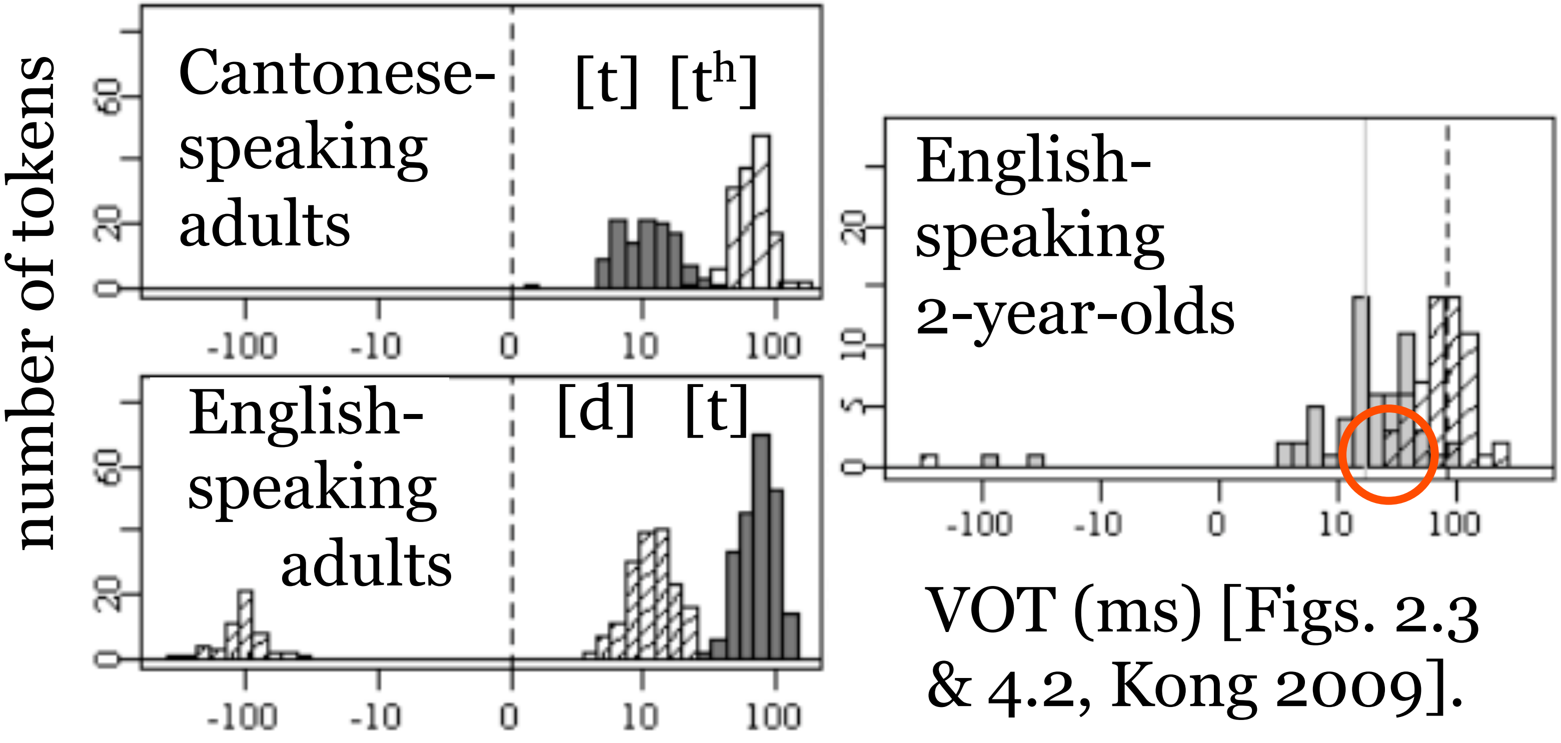
short lag achieved by glottis opening at any time during the oral occlusion: easy to produce!



Language	Lead	Short lag	Long lag
English (Macken & Barton 1980a)	voiced	voiced	voiceless
French (Allen 1985)	voiced	voiceless	
Cantonese (Clumeck et al 1981)		unaspirated	aspirated
Thai (Gandour et al 1986)	voiced	unaspirated	aspirated

Acquisition of "marked" feature of voicing

- English seemed to be an exception, until use of VOT explained transcribed [d, g] for /t, k/ substitutions in (Kewley-Port & Preston 1974, Macken & Barton 1980).



VOT (ms) [Figs. 2.3 & 4.2, Kong 2009].

Other substitution patterns

Japanese-speaking children substitute voiceless stops for voiced stops, ...

target [doa]　ドア　transcribed as:　/d/ → [t] 🔊

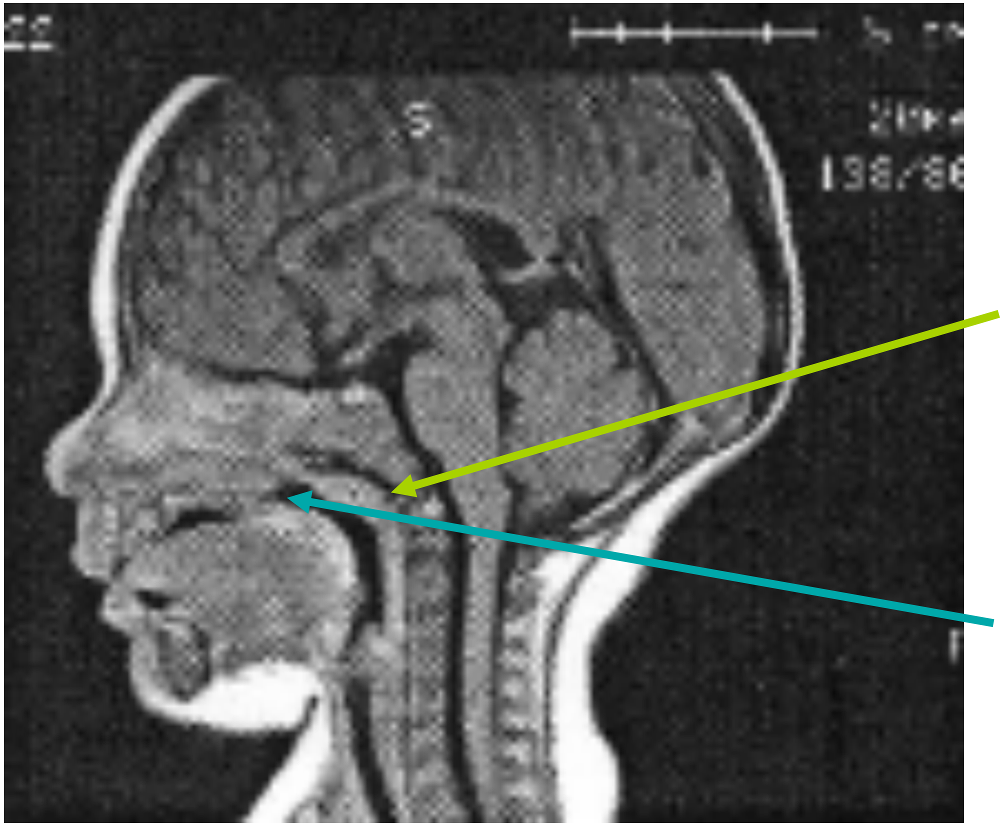
target [gasu]　ガス　/g/ → [k] 🔊

but they also substitute nasals sometimes:

target [donguri] どんぐり　/d/ → [n] 🔊

Why are [d] and [g] difficult for children?

The successful production of a stop involves release of air pressure that builds up in oral cavity after ….

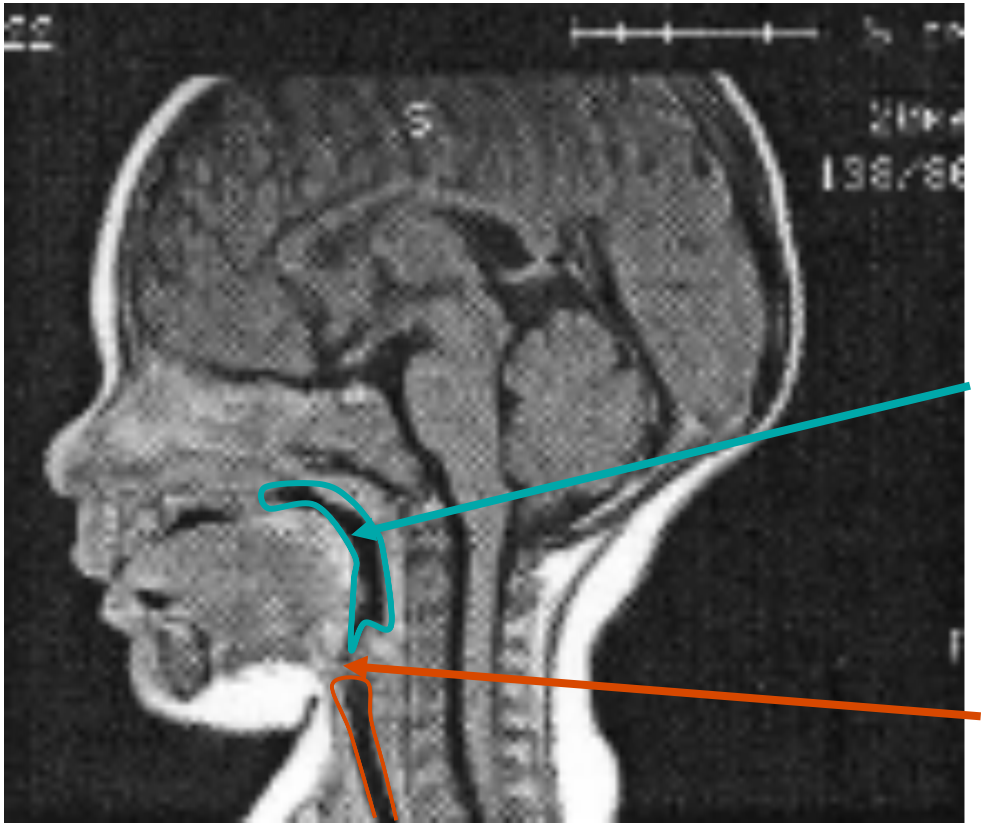


momentary seal at nasopharynx to block air passage through nose coupled with momentary seal in oral cavity to block air passage through mouth

Figure 1 in Vorperian, Kent, Lindstrom, Kalina, Genry, & Landell (2005): mid-sagittal MRI of 7-month old female.

Why do children substitute [t] and [k]?

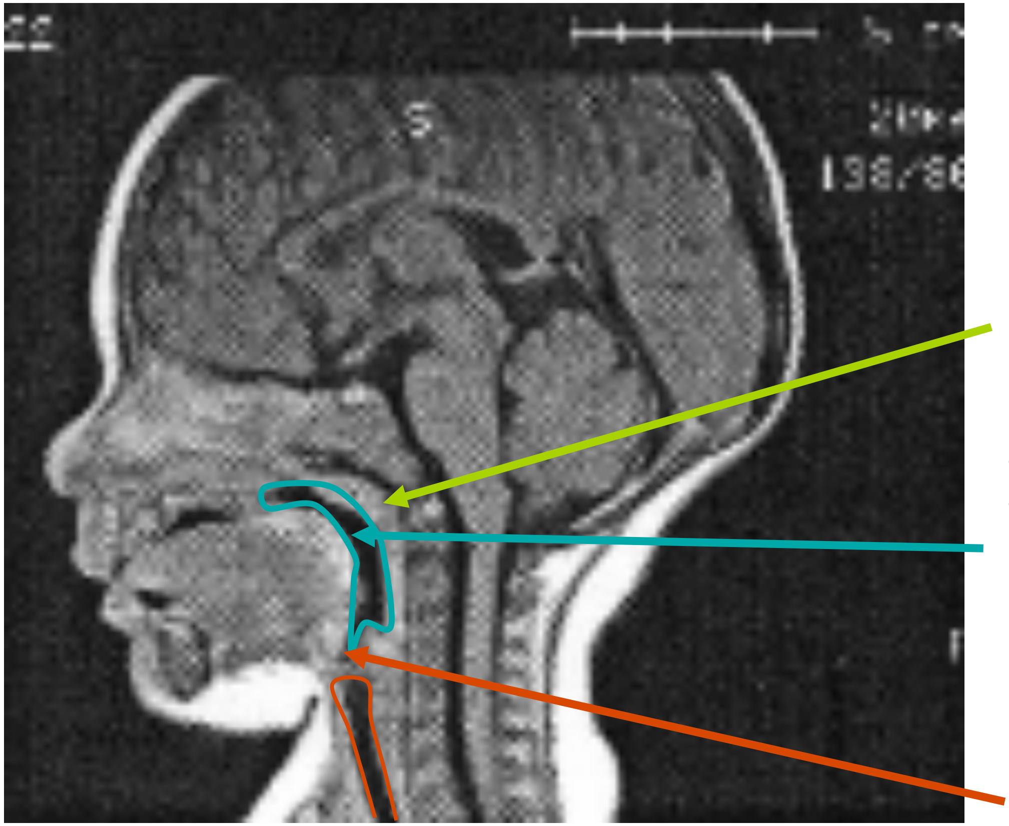
Voicing happens when air flow pushes vocal folds apart.
Requires air pressure in oral cavity < pressure below glottis



voiceless stop variants [t, k] happen if air pressure in oral cavity builds up to impede air flow across the glottis

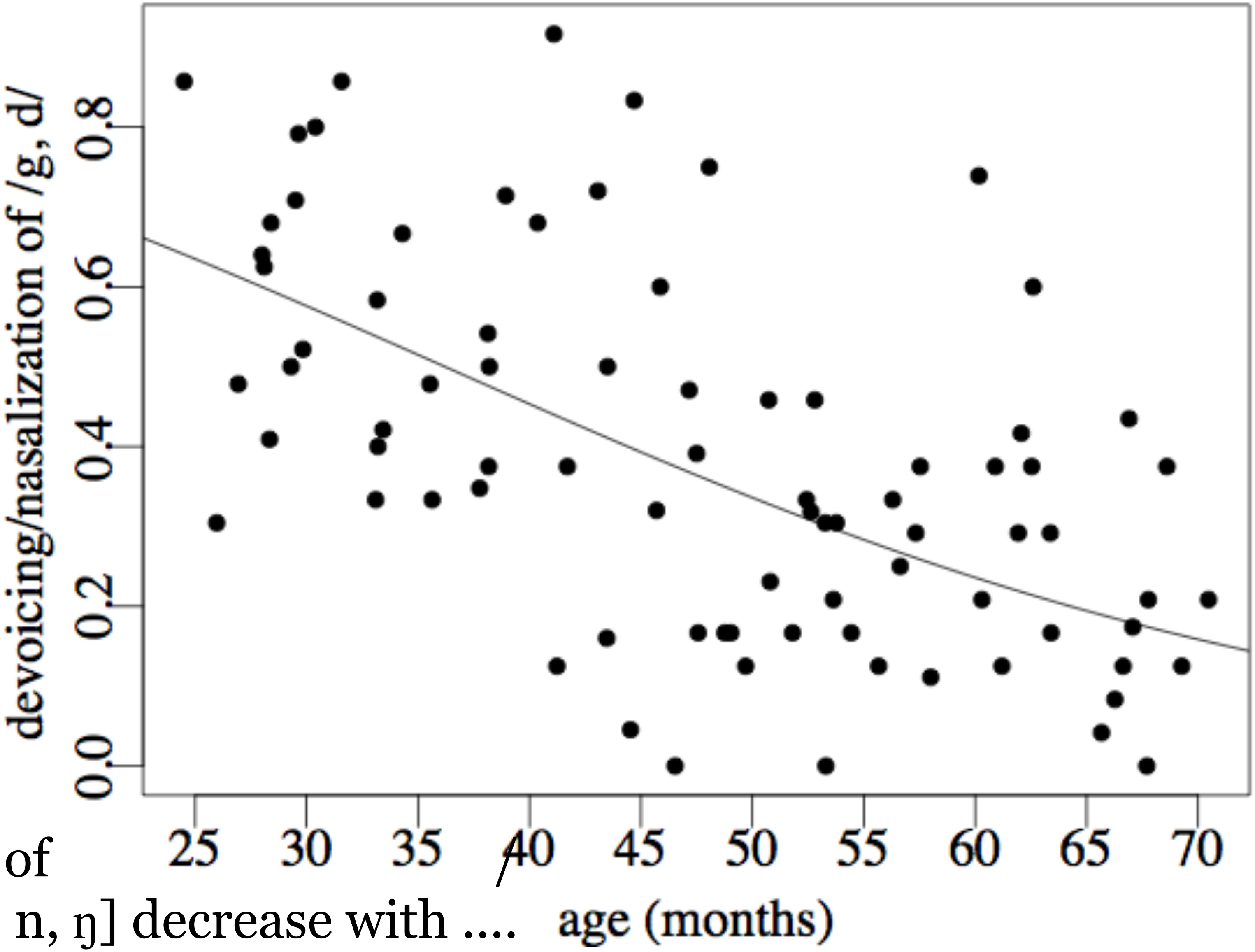
Why do children substitute [n] and [ŋ]?

Voicing requires air pressure below glottis > oral cavity



nasal variants [n, ŋ] if nasopharynx partially opened to prevent air pressure build-up in oral cavity so that air can continue to force air through the glottis

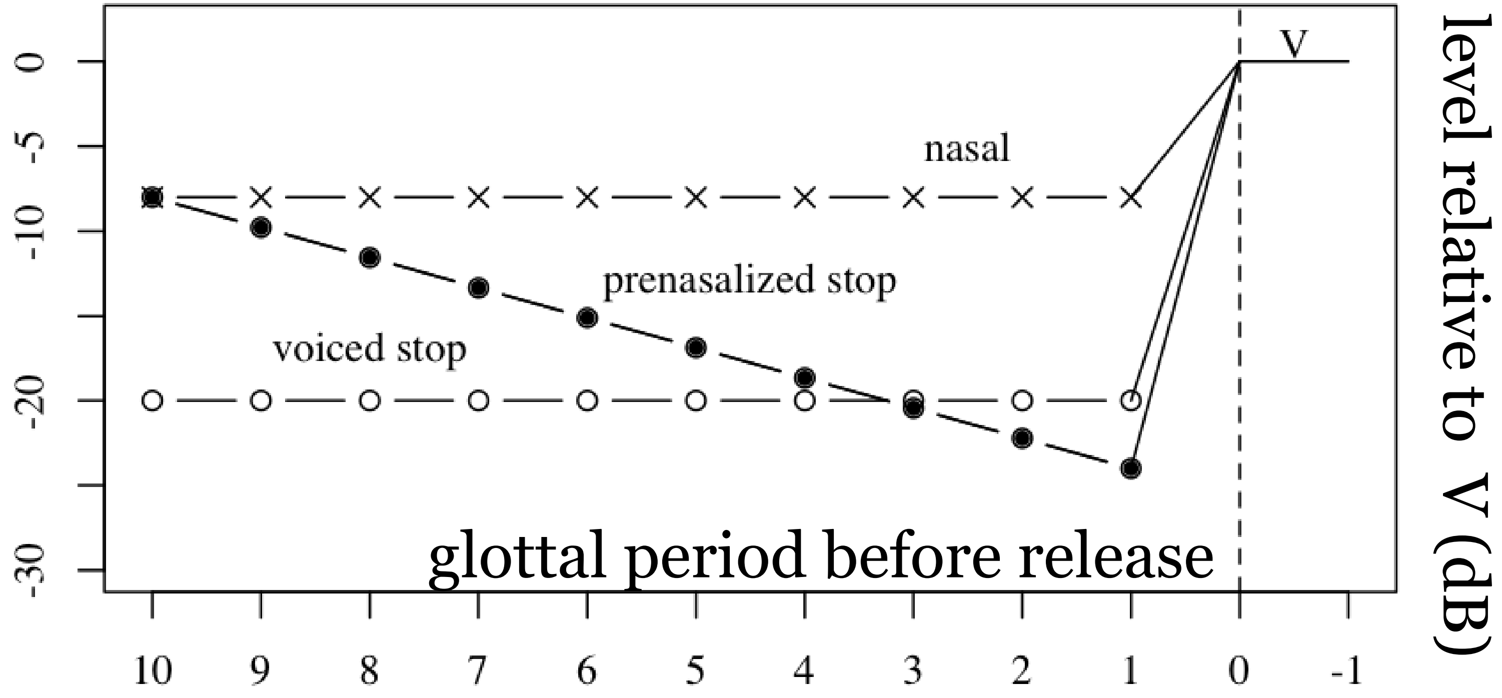
Progressive mastery of Japanese [d] and [g]



Productions of
/d, g/ as [t, k, n, ŋ] decrease with ....

Explaining another apparent exception

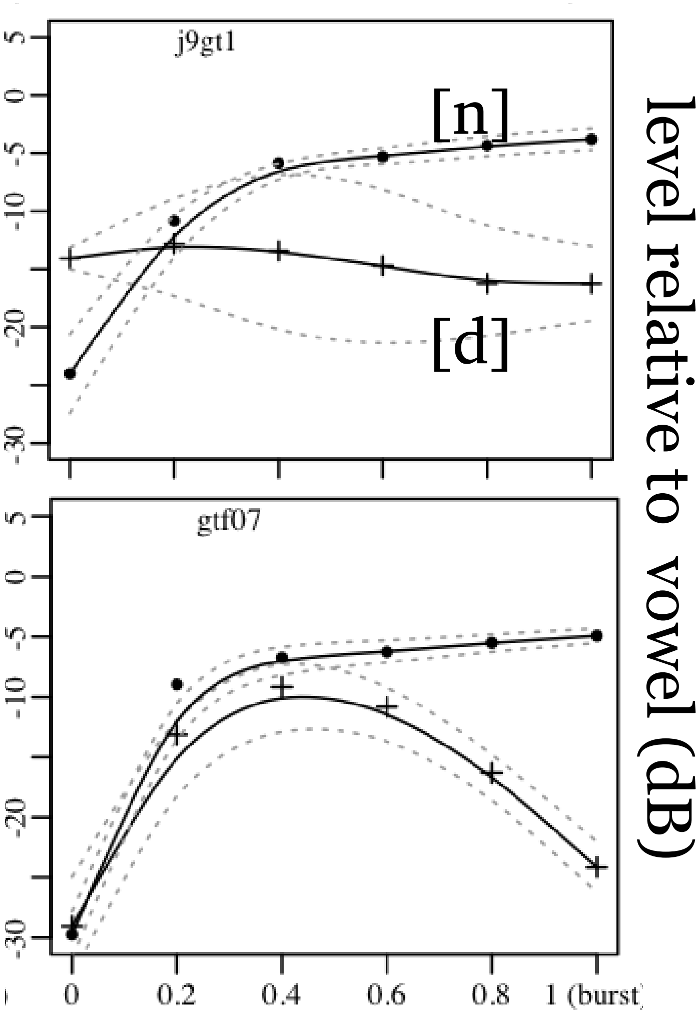
- Japanese children produce lead VOT values at 4 years.
- Greek children have **lead** VOT values as early as 2 years.
- Kong et al (2012) adapted the acoustic model from Burton, Blumstein, & Stevens’s (1972) study of the Moru language contrasts among [n], prenasalized [^{n}d], & [d].



Japanese versus Greek “voiced” stops

typical Japanese adult male has the Moru voiced stop pattern

typical Greek adult male has the Moru prenasalized stop pattern



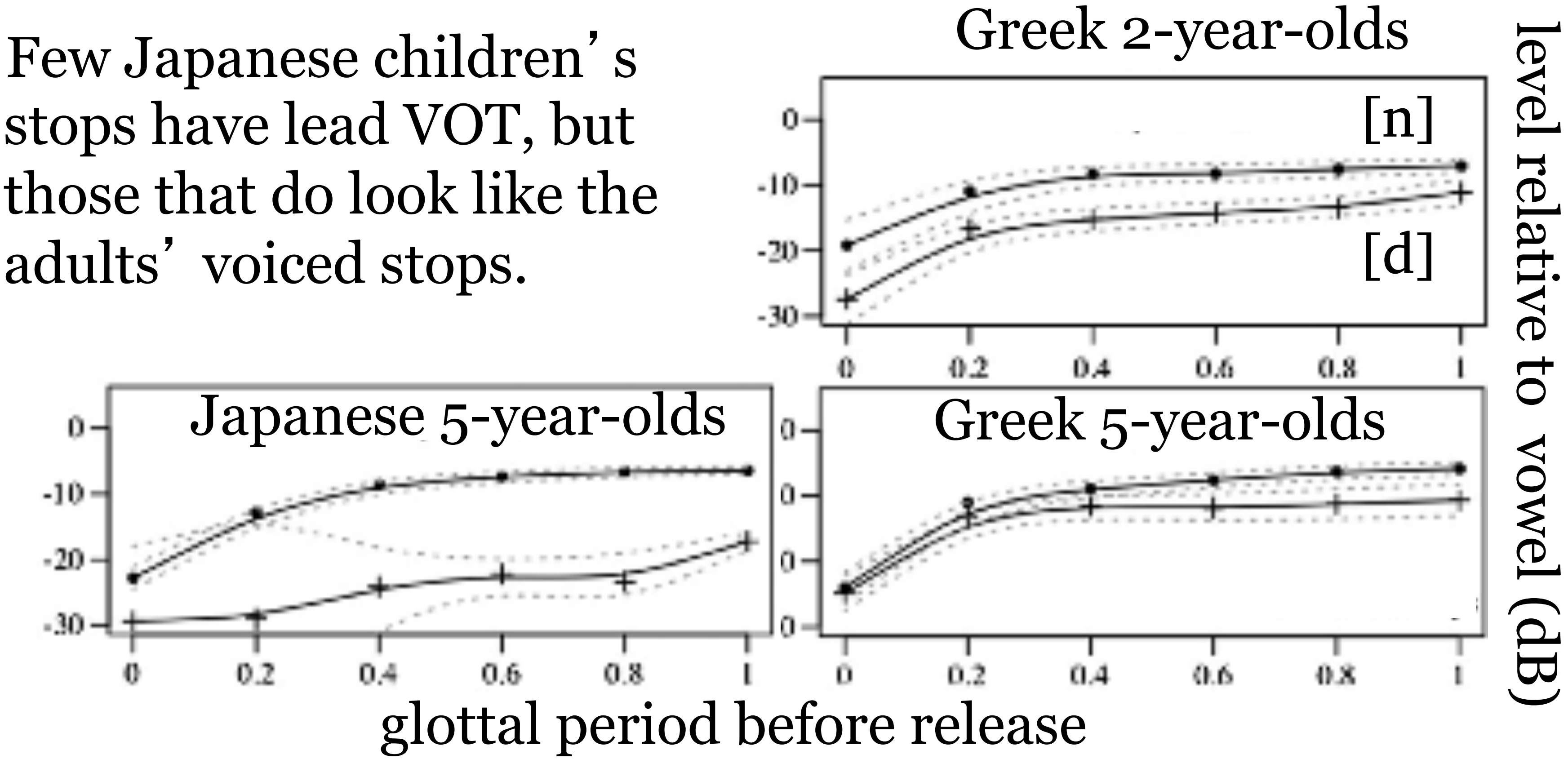
level relative to vowel (dB)

glottal period before release

- Greek voiced stops developed only recently, from clusters such as [mp] and [nt], and in some dialects they are consistently prenasalized (Arvaniti & Joseph 2000; 2004).

Japanese versus Greek "voiced" stops

- Most Greek children's stops have lead VOT, and they look even more nasalized than the adults'.
- These are perceived as correct productions by the Greek-speaking transcriber (but as nasals by Japanese transcriber).
- Few Japanese children's stops have lead VOT, but those that do look like the adults' voiced stops.



Summary 1: Measuring voicing contrasts

- Children must develop skills for producing even very difficult “marked” sounds such as voiced stops.
- Phonological acquisition evident in the way in which perceived early mispronunciations give way to perception of consistently adult-like pronunciations.
- English and Greek are apparent exceptions to the claim that voiced stops have the “marked” feature.

☞Apparent exceptionality resolved when continuous measures of voicing/aspiration and pre-nasalization are developed and applied to understand that:

  - English voiceless stops are aspirated (marked)
  - Greek voiced stops are prenasalized (less marked)

Case study 2: Fricative place contrasts

- English and Japanese contrast [±anterior] sibilants
- In English, typical error is [-anterior] → [+anterior]

  target: [ʃ] ‘shoe’ transcribed as: /ʃ/→[s]
- In Japanese, typical error is [+anterior] → [-anterior]

  target [semi] セミ transcribed as: /s/ → [ʃ]

  target [suika] 水瓜

Generative Phonology model assumes:

- The changed feature in production is the more difficult “marked” (有標) value for the feature.

☞ How can [+anterior] (in English) and [-anterior] (in Japanese) both be the unmarked value?

Explaining age-typical misarticulations

Generative Phonology model also assumes:

- The child accurately perceives the phonetic property that distinguishes the target and substituted sounds.
- Accurate perception allows adult-like representations of the target consonant in a word’ s “underlying form”
- The child’ s “rule” transforms the underlying form with the difficult-to-produce “marked” property value into a “surface form” with the “unmarked” (無標) value.
- Distinctive features are universal. If two languages use a distinctive feature, it has the same phonetic property.

☞ What is the basis for these assumptions about accurate perception and uniform phonetic properties?

Accurate perception: the [fɪs] phenomenon

Children who are transcribed as substituting one sound for another in their own productions might reject that substitution in others’ imitations of their productions.

> One of us, for example, spoke to a child who called his inflated plastic toy fish a *fis*. In imitation of the child’ s pronunciation, the observer said: “This is your *fis*?” “No,” said the child, “my *fis*.” He continued to reject the adult’ s pronunciation until he was told, “This is your fish.” “Yes,” he said, “my *fis*.”
>
> (Berko & Brown, 1960, p. 531)

Accurate (and uniform) perception?

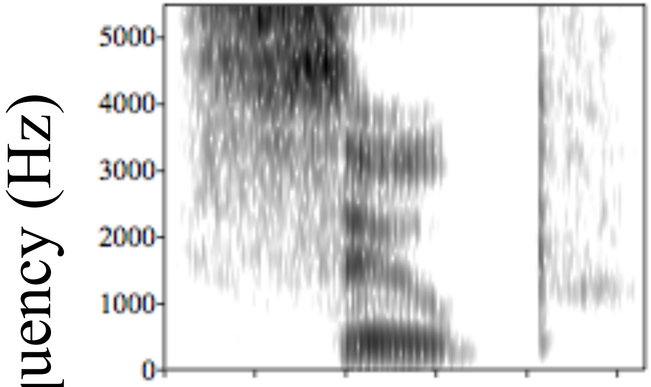
Generative Phonology model also assumes:

- The child accurately <u>perceives</u> the phonetic property that distinguishes the target and substituted sounds.
- Distinctive features are <u>universal</u>. If two languages use a distinctive feature to distinguish two sounds, the contrast taps the same phonetic property in both.

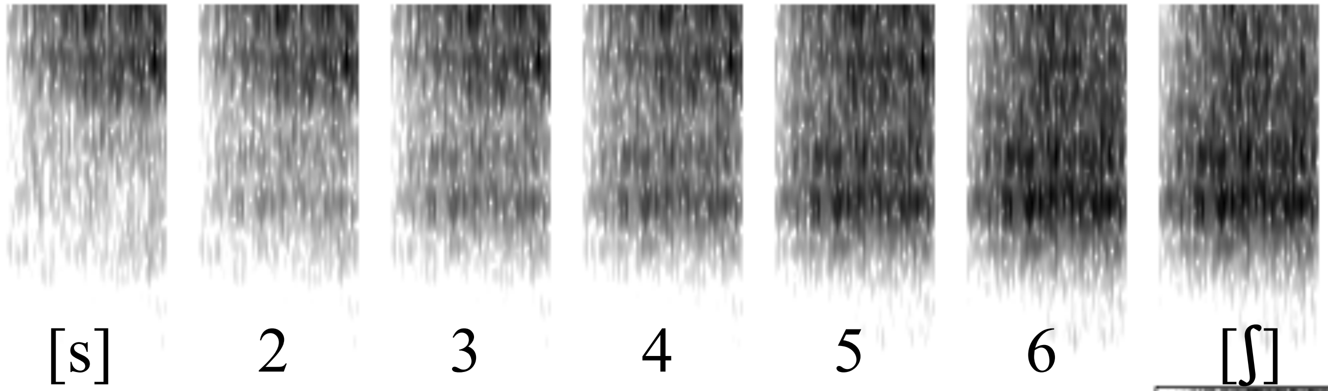
☞What is the basis for these assumptions about accurate perception and uniform phonetic properties?

1. What is the phonetic property that differentiates between the [±anterior] sibilants [s] and [ʃ]?
2. Is it the same for speakers of English and Japanese?
3. Is it the same for the boy who said [fɪs] and the adult?

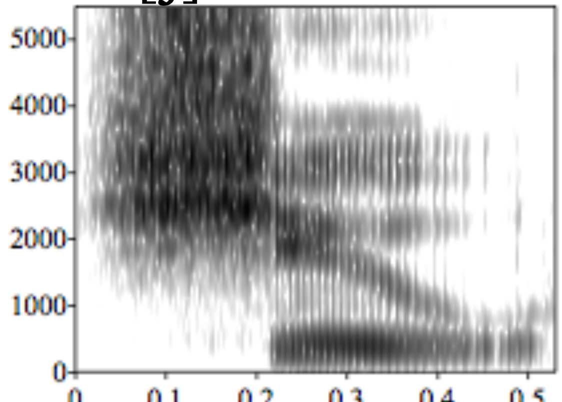
Major cue to the English [s]:[ʃ] contrast



- Major cues to the [s]:[ʃ] contrast are present in spectrum of the fricative itself.
- Evidence for this in perception tests that manipulate the distribution of energy there.

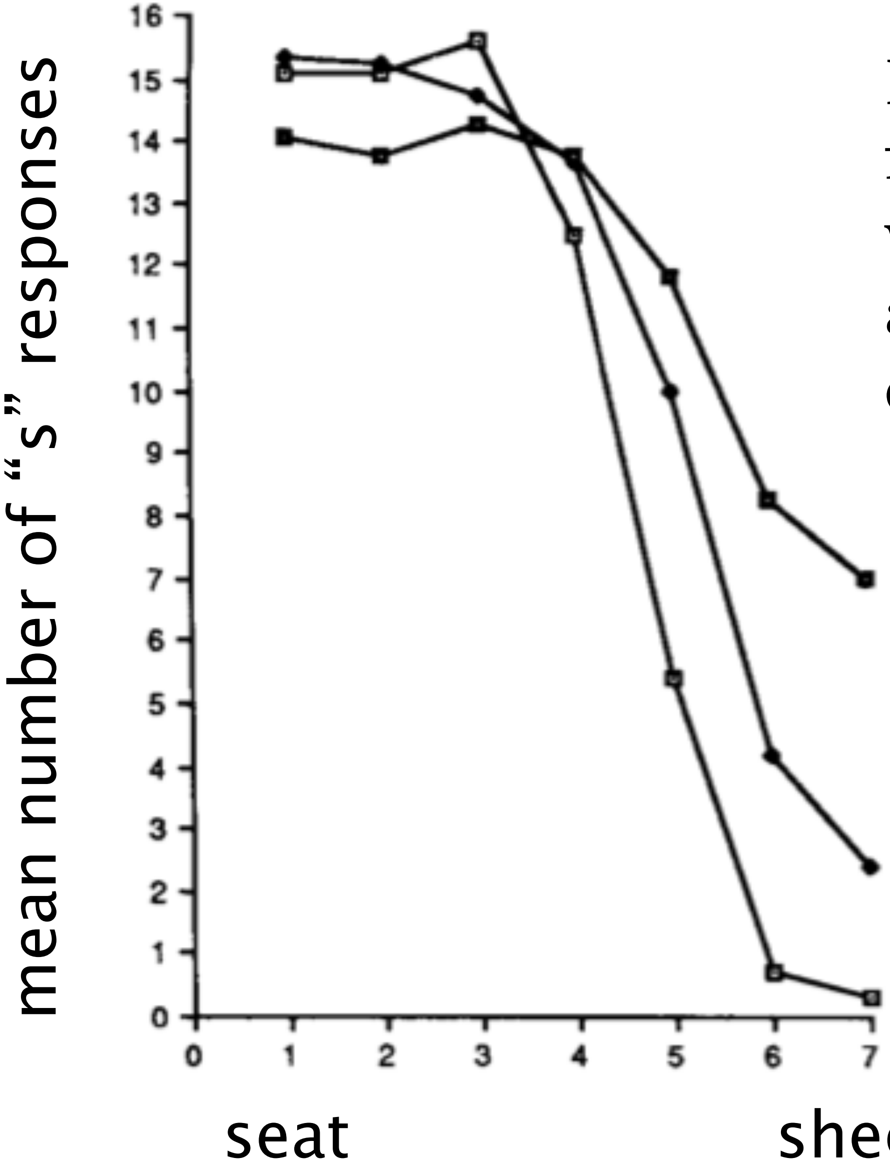


- Rvachew and Jamieson (1989) used a 7-step continuum
- Adults show a sharp step-like identification curve.



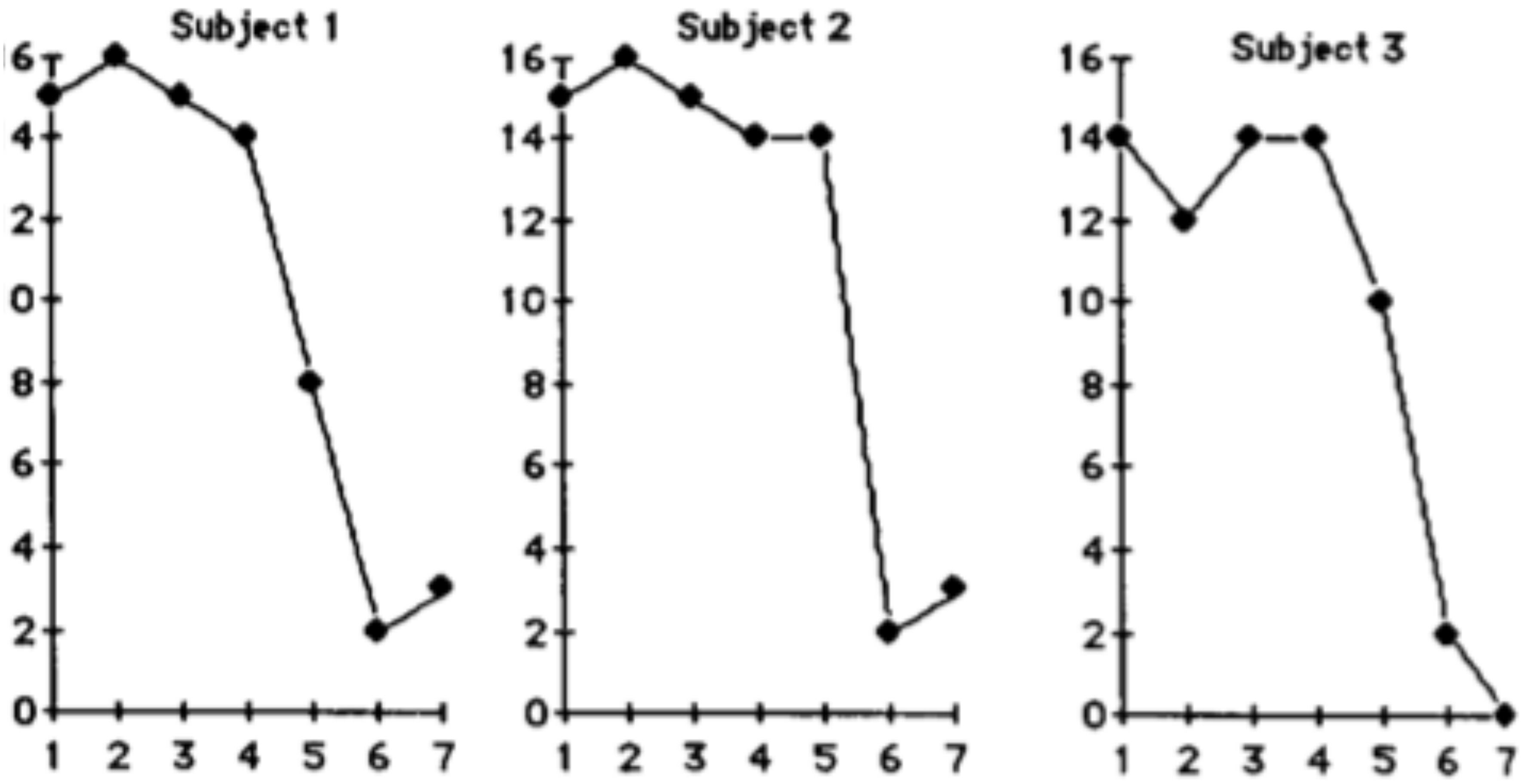
Exceptions to the [fɪs] phenomenon

Rvachew & Jamieson (1989) find that at least some children who have persistent (age-atypical) substitution patterns do perceive contrast differently.



Rvachew & Jamieson (1989), Fig. 3

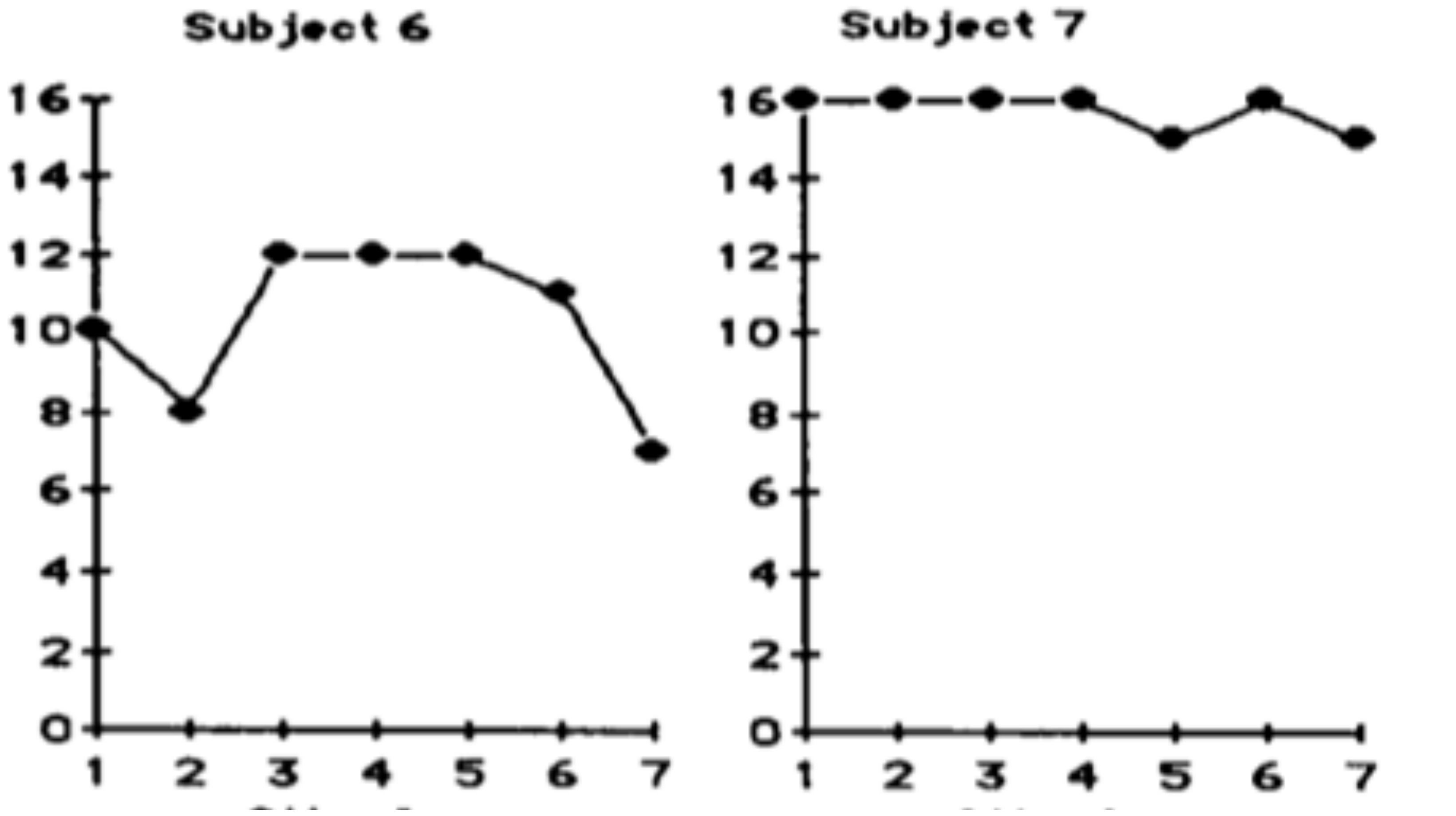
mean number of “s” responses



seat sheet

three of the children with age-typical articulation

Rvachew & Jamieson (1989), Fig. 4



mean number of "s" responses

seat

sheet

two of the children with speech-sound disorder

The questions again ...

- Both English and Japanese have a contrast in coronals between a more anterior [s] and more posterior [ʃ] .
- English [s] mastered earlier than [ʃ] and [s] substitutes for [ʃ] (Smit et al. 1991) -- i.e., a “fronting” pattern.

  [ʃu] ‘shoe [sup] ‘soup’

- Japanese [ʃ] mastered earlier than [s] and [ʃ] substitutes for [s] (Nakanishi et al., 1972) -- i.e., a “backing” pattern.

  [ʃu:kuri:mu] シュークリーム [suika] 水瓜

☞ If substitution patterns reflect markedness, how can [+anterior] and [-anterior] both be the unmarked value?

☞ Could it be that the perceptual cues to the [±anterior] contrast are not actually the same in the two languages?

Articulation of English [s] and [ʃ]

- English [s] (left) is alveolar and often apical.
- English [ʃ] (right) has a short constriction in the post-alveolar region, and it is rounded.

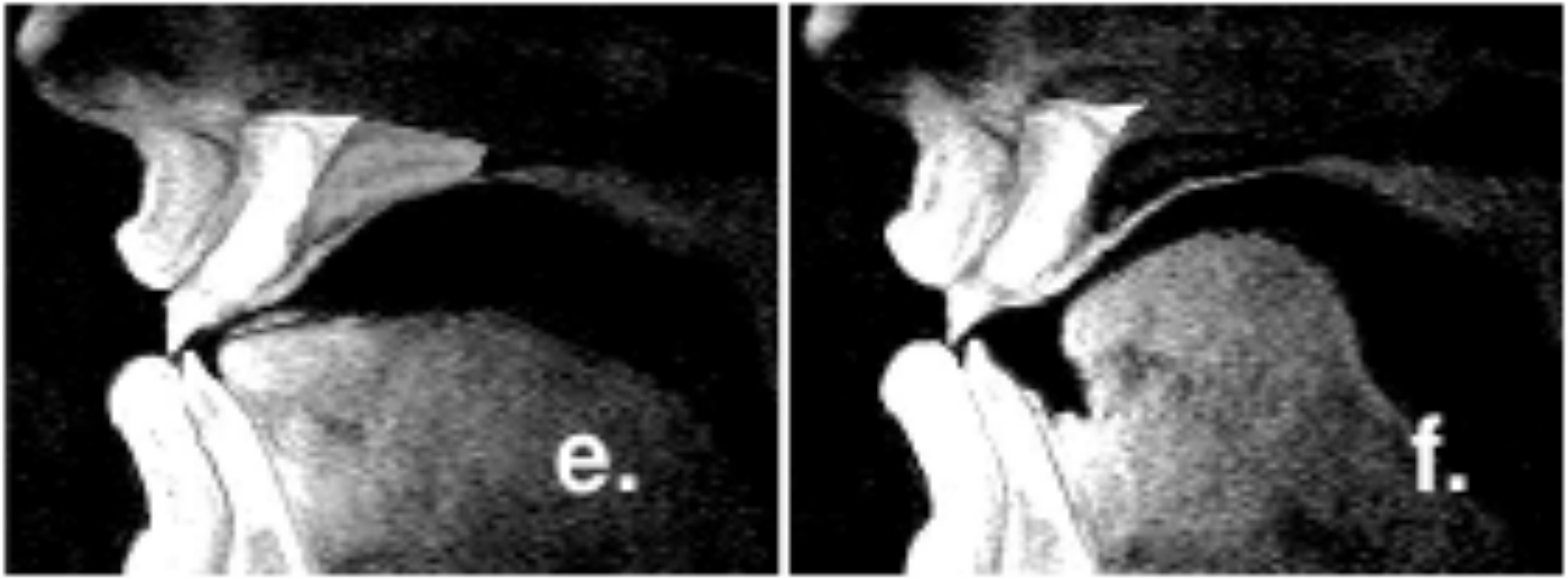


Fig. 2 from Toda and Honda (2003).

Articulation of Japanese [s] and [ʃ]

- Japanese [s] (left) is dental and typically laminal.
- Japanese [ʃ] (right) is alveolo-palatal, with a long constriction channel, and the lips are spread.

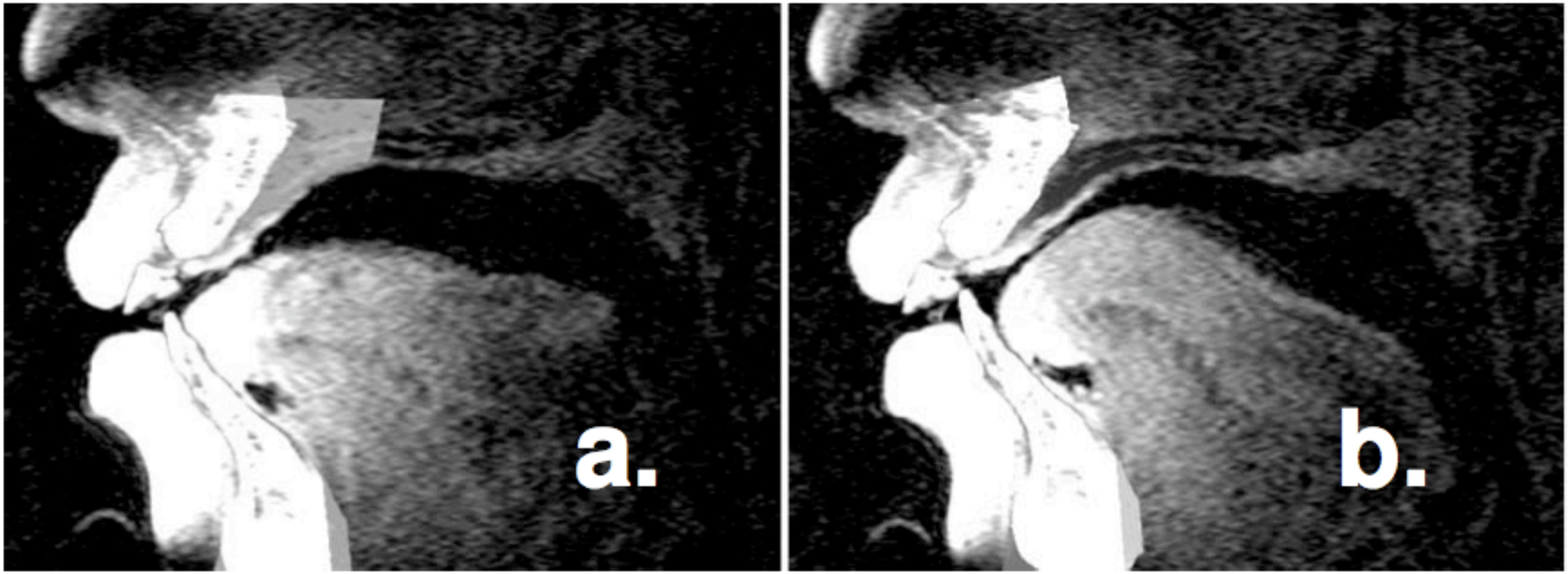


Fig. 2 from Toda and Honda (2003).

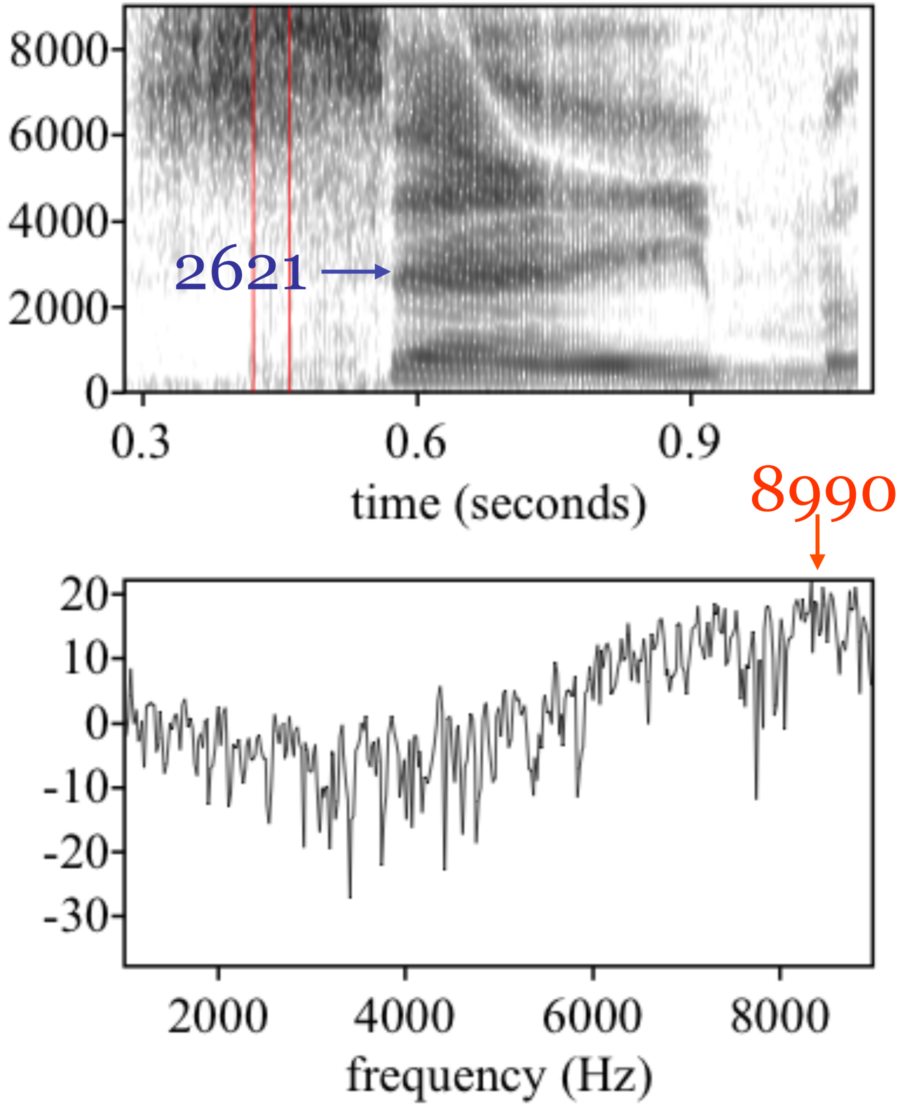
Acoustic measures for sibilant contrasts

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



Adult productions of [s] versus [ʃ]

Much better differentiation by centroid value for English.

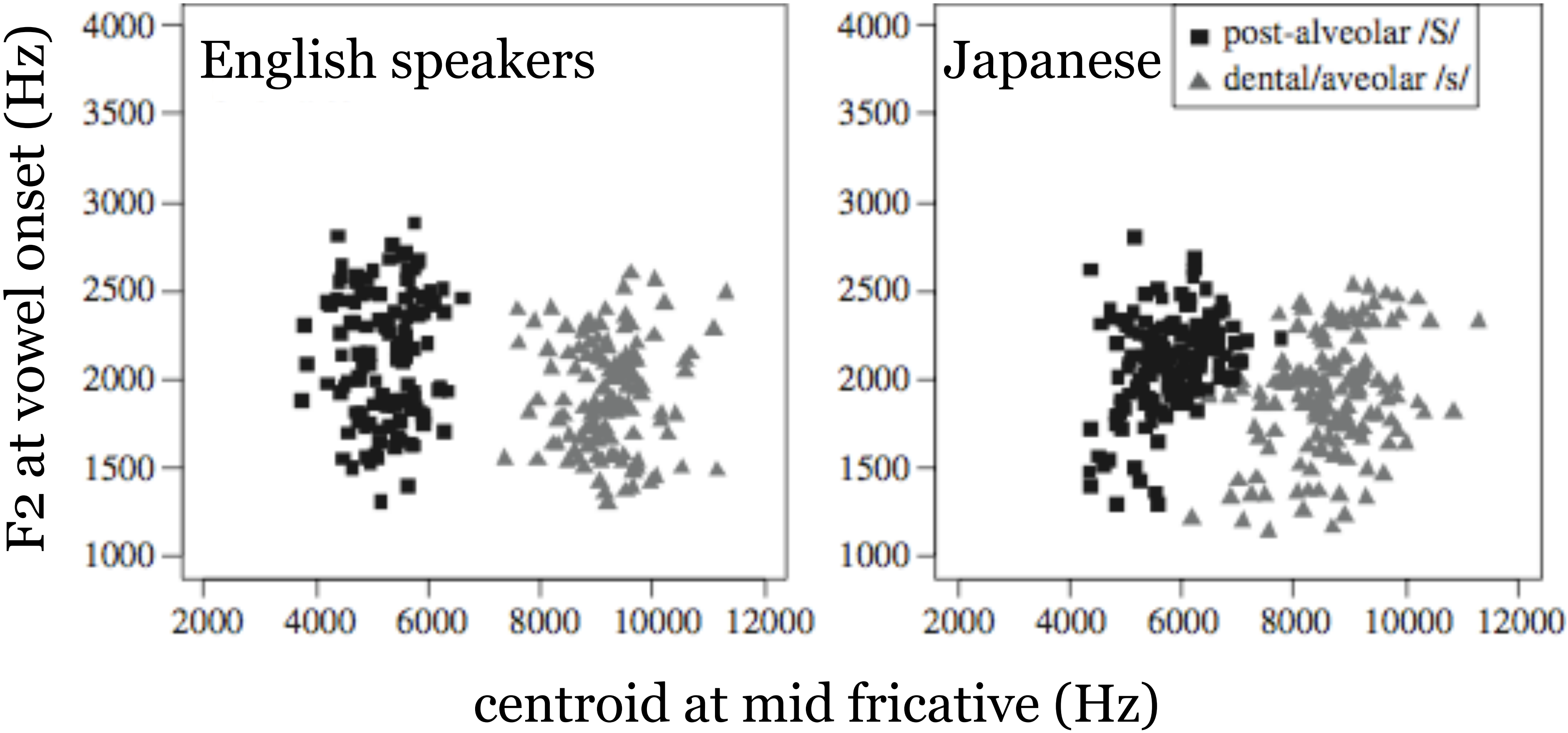
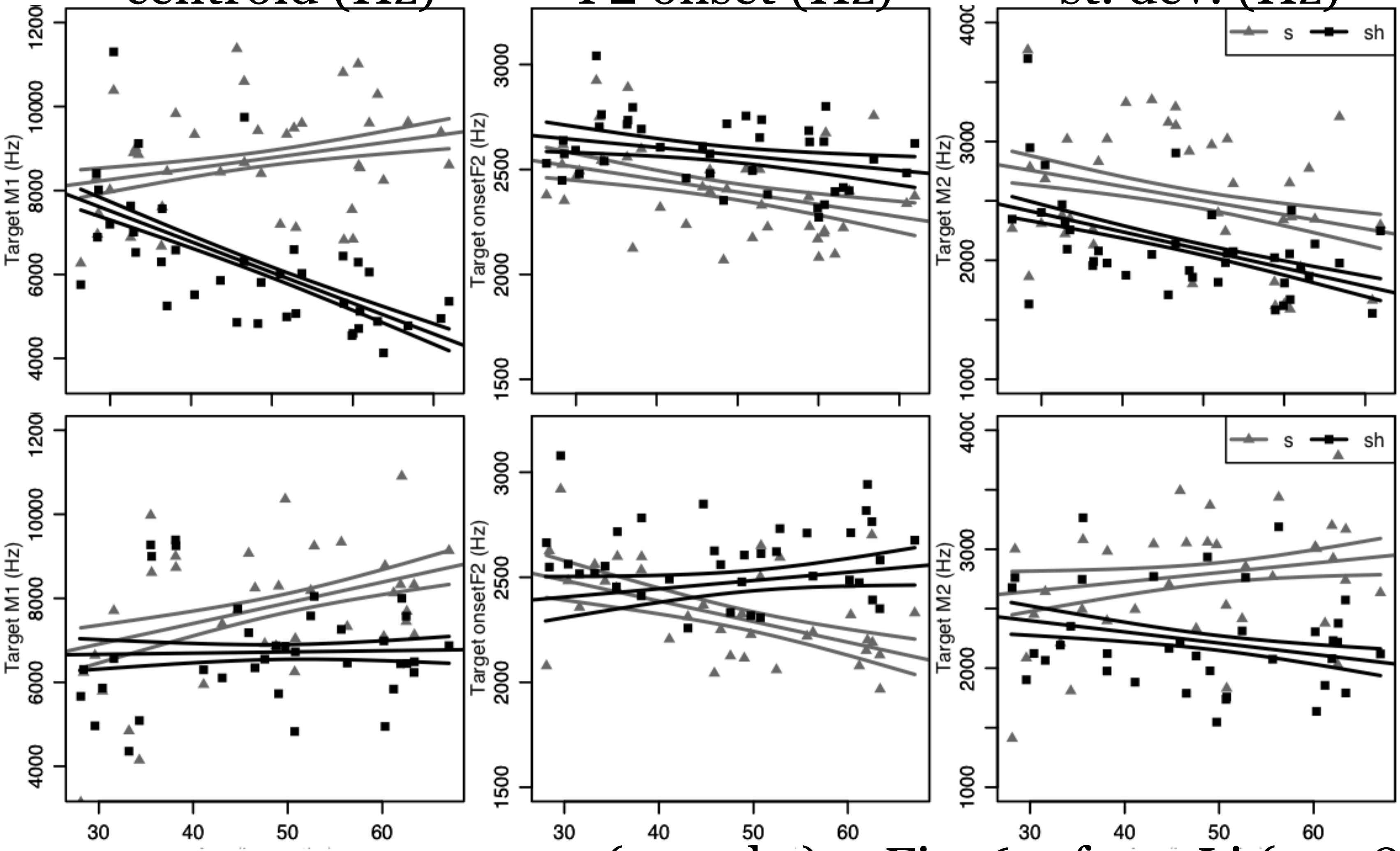


Fig. 3 from Li, Edwards, & Beckman (2009).

Cross-language differences in child productions

Japanese- versus English-speaking

centroid (Hz) F2 onset (Hz) st. dev. (Hz)



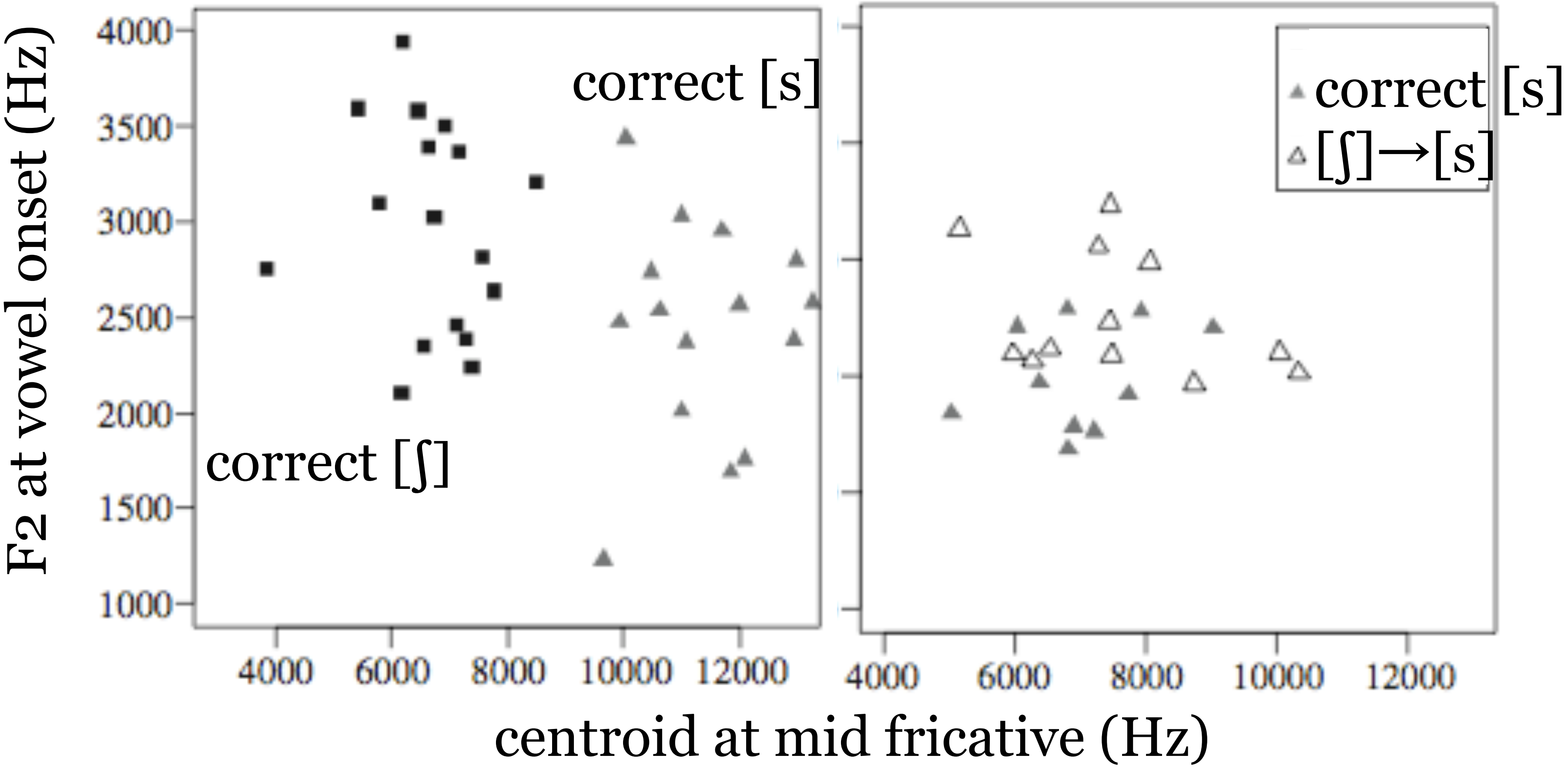
age (months)

Fig. 6.3 from Li (2008).

Different child productions in English

child with robust contrast
(no [s] for [ʃ] substitutions)

child with covert contrast
(many substitutions)



Figs. 4 & 6a from Li, Edwards, & Beckman (2009).

How do adults perceive child productions?

- Cross-language differences in the acoustic cues to the [s]:[ʃ] contrast …
- Some English-speaking children who are transcribed as substituting [s] for target [ʃ] produce F2 onset frequencies that are appropriate for Japanese [ʃ].
- How does adults' perception of the children's productions contribute to the different stereotypical substitutions?
- Li, Munson, Edwards, Yoneyama, and Hall (2011):
  - Delexicalize by splicing out initial CV as stimuli
  - Ask 19 English- and 20 Japanese-speaking adults: (1) "Is it the 's' sound?" and (2) "Is it the 'sh' sound?"

Cross-language differences in adult perception

- Consensus responses (70%+ “yes”) differ between English- and Japanese-speaking listeners <u>for the same English CV tokens</u>.

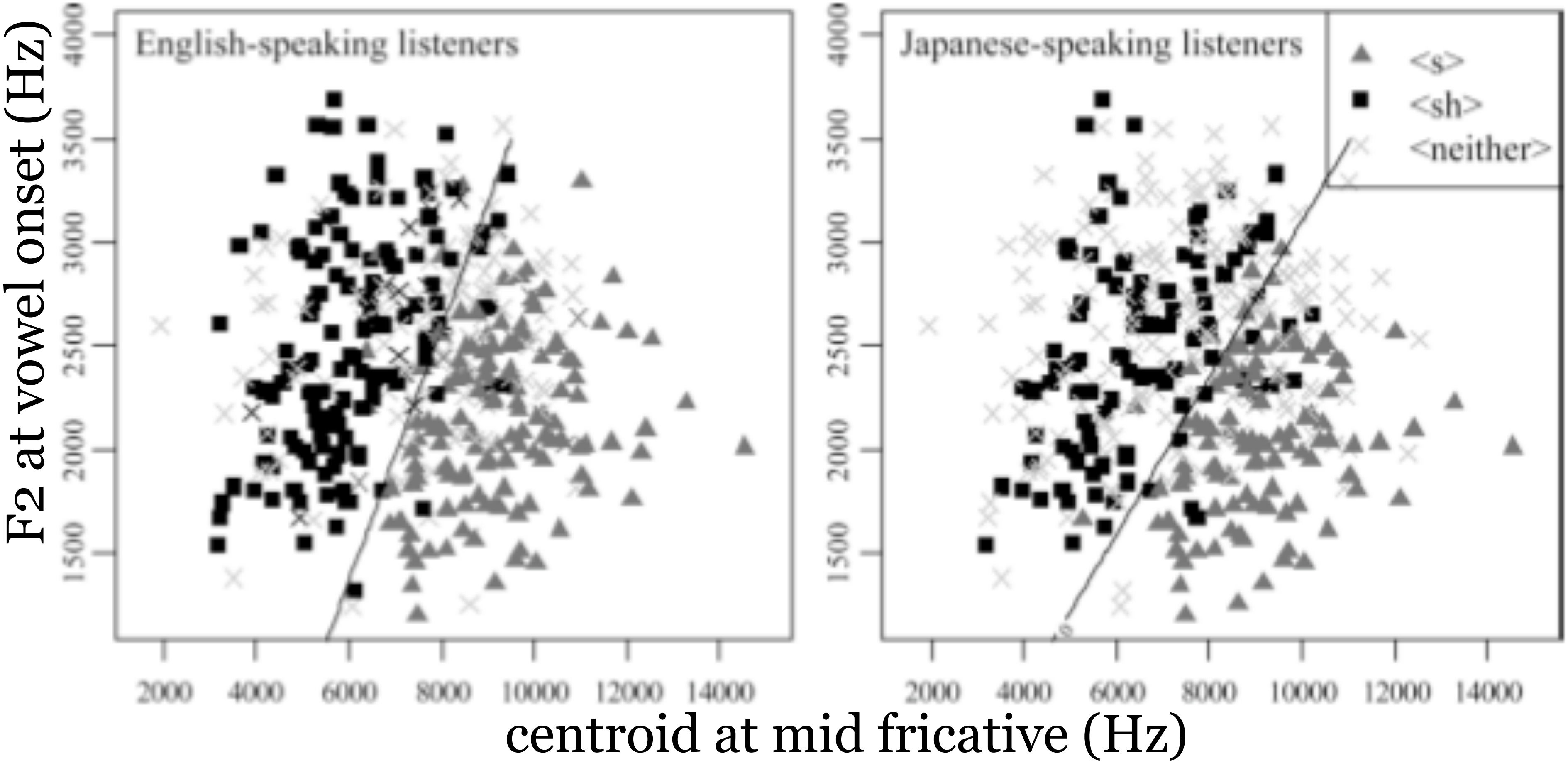
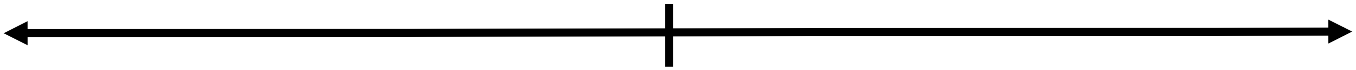


Fig. 2 from Li, Munson, Edwards, Yoneyama, & Hall (2011).

Visual Analog Scale responses

- Li et al. (2011) paired-questions method requires two trials per stimulus.
- Also, the interpretation of double-“no” responses is difficult.
- Urberg Carlson, Kaiser, and Munson (2008) adapted an alternative method that uses a Visual Analog Scale (VAS) to probe adult perception continuously.

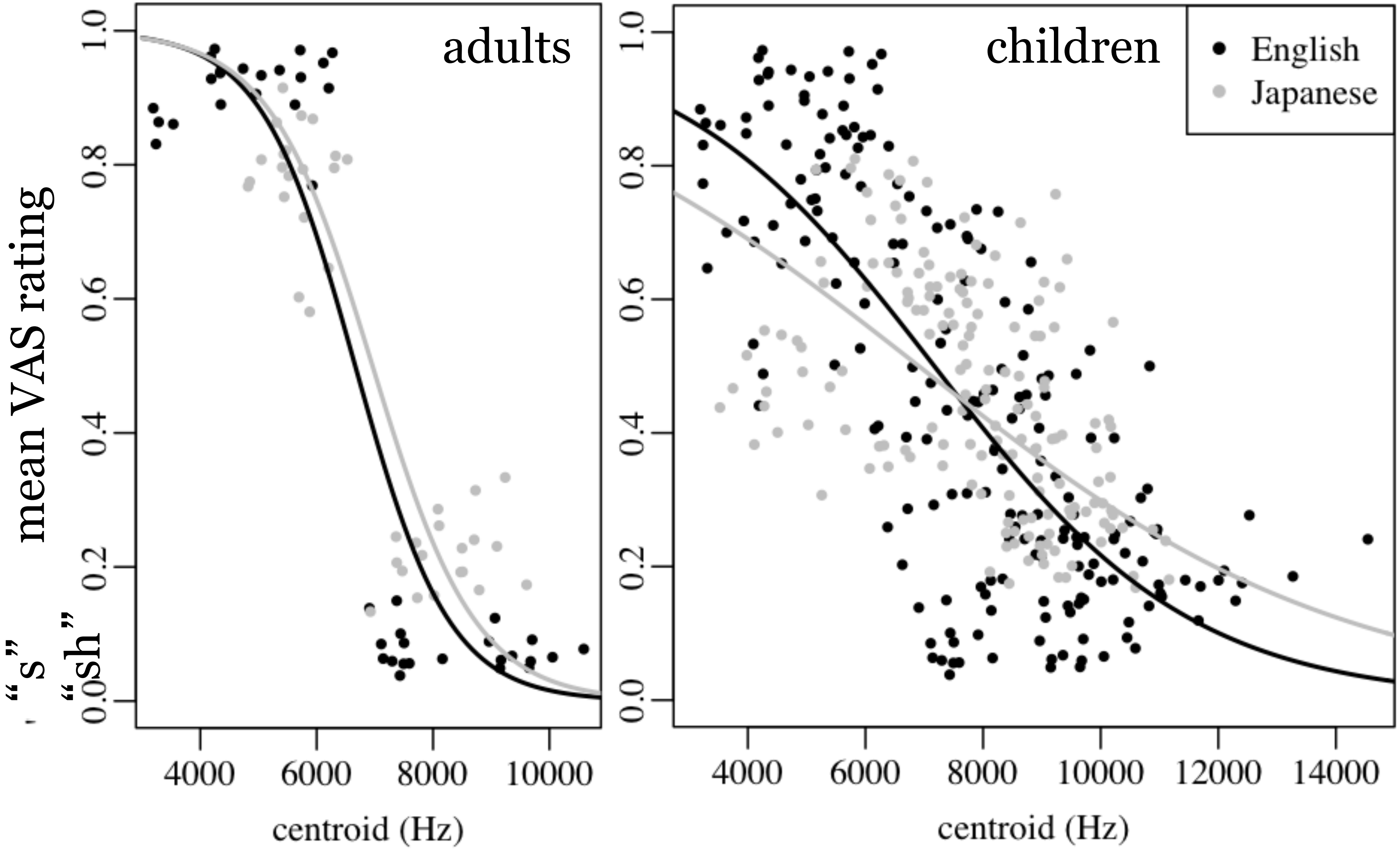


the “s” sound

the “sh” sound

participant responds by clicking appropriately on arrow

English-speaking listeners’ VAS responses



Summary 2: Measuring fricative place

- Sounds that are identified as participating in “the same contrast” in different languages can differ in:
  - the acoustic cues that differentiate their productions
  - the community norms for perceiving those cues
- The English [s]:[ʃ] contrast differs from the Japanese [s]:[ʃ] both in production details and in the perceptual weights assigned to the acoustic cues.

☞Apparent contradiction between the unmarked value being [+anterior] in English but [-anterior] in Japanese begins to be resolved when continuous measures of perception used to hypothesize that:

- English production and perception are biased to [s]
- Japanese production and perception are biased to [ʃ]

Case study 3: Stop place contrasts

English and Japanese both contrast [±coronal] stops

- In English, typical error is [-coronal] → [+coronal]

  target: [k^{h}ot] ‘coat’ transcribed as: /k^{h}/→[t^{h}]

- In Japanese, this substitution is seen in front vowel contexts, where substituted coronal is also [-anterior]

  target [k^{j}imono] 着物 transcribed: /k^{j}/ → [tʃ]

  Before back vowels, error is [+coronal] → [-coronal]

  target [tamago] 卵 transcribed as : /t/ → [k]

☞Why is English “fronting” all the way to [+anterior] and why is it not conditioned by a front vowel?

☞In back contexts, how can [+coronal] (in English) and [-coronal] (in Japanese) both be the unmarked value?

Yoneyama, Beckman, & Edwards (2003)

Participants:

- 47 typically developing children, screened for normal hearing, ranging in age from 2(years);3(months) to 5;3

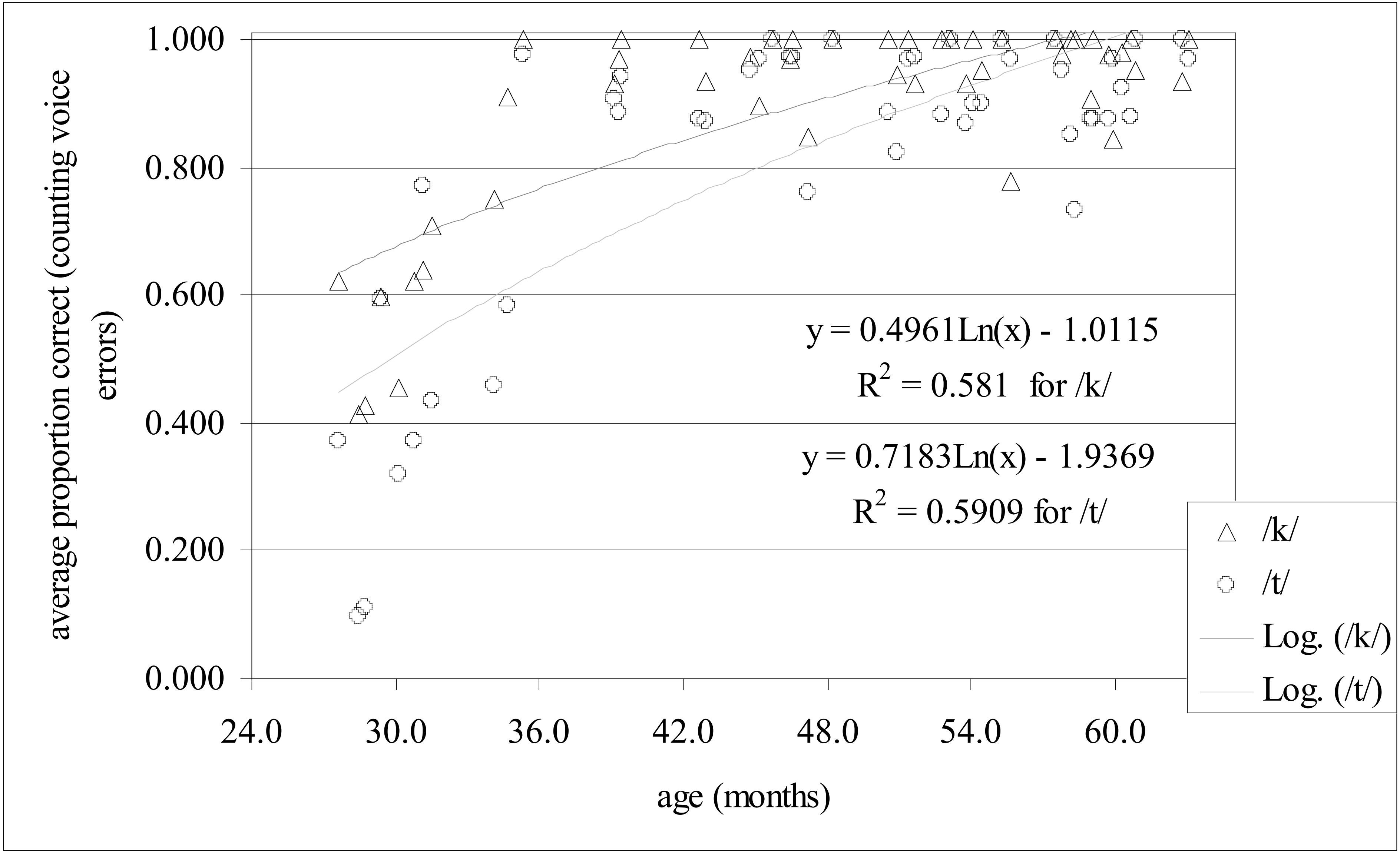
Stimuli:

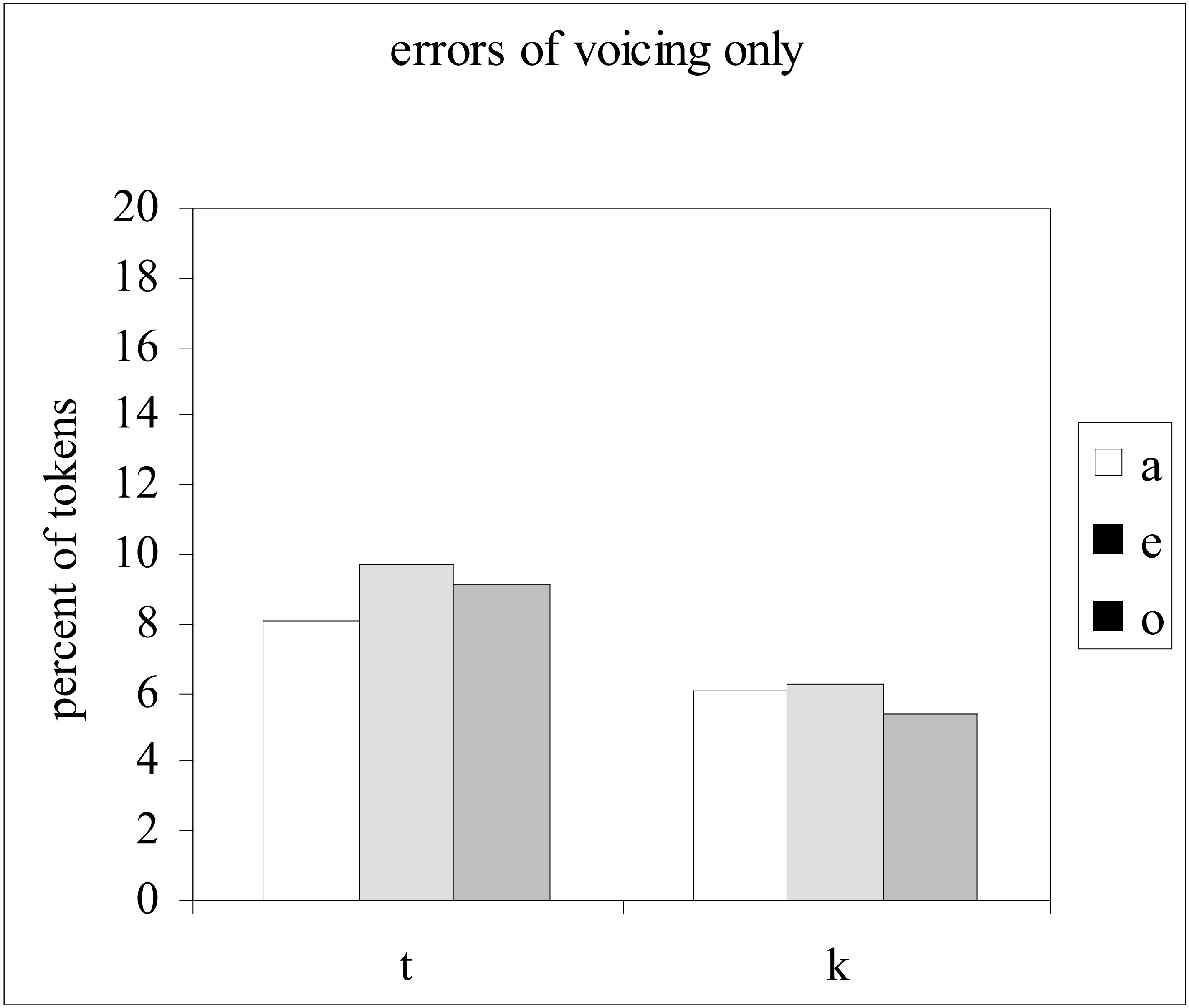
- 18 target words with word-initial /t/ or /k/, followed by vowels /a/, /e/, /o/.

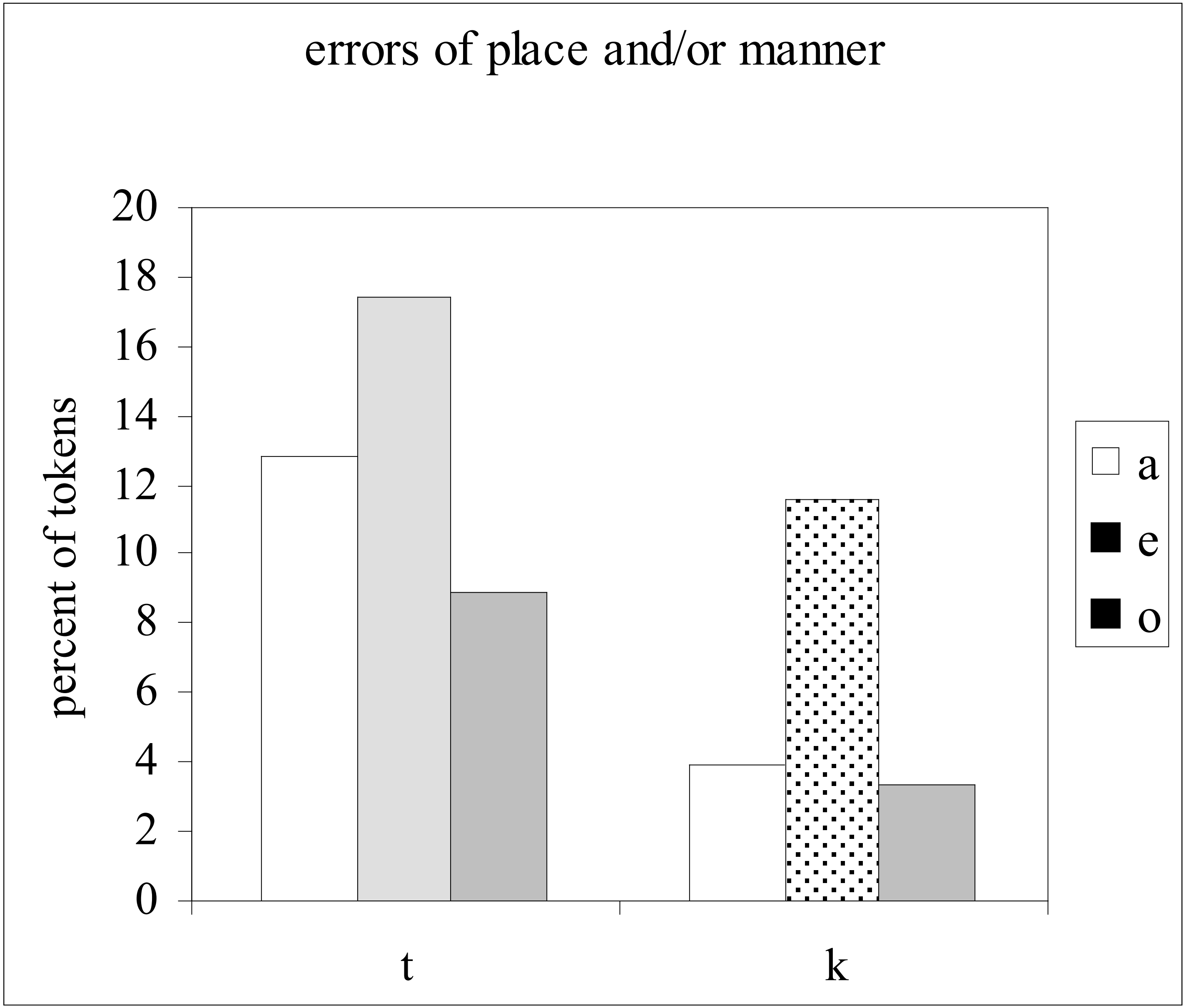
Procedure:

- elicited 5 repetitions of each target, using pictures.

Progressive mastery of stop place




errors of voicing only
percent of tokens
20
18
16
14
12
10
8
6
4
2
0
t
k
□ a
■ e
■ o


errors of place and/or manner
percent of tokens
20
18
16
14
12
10
8
6
4
2
0
t
k
a
e
o

Results from the παιδολογος corpus

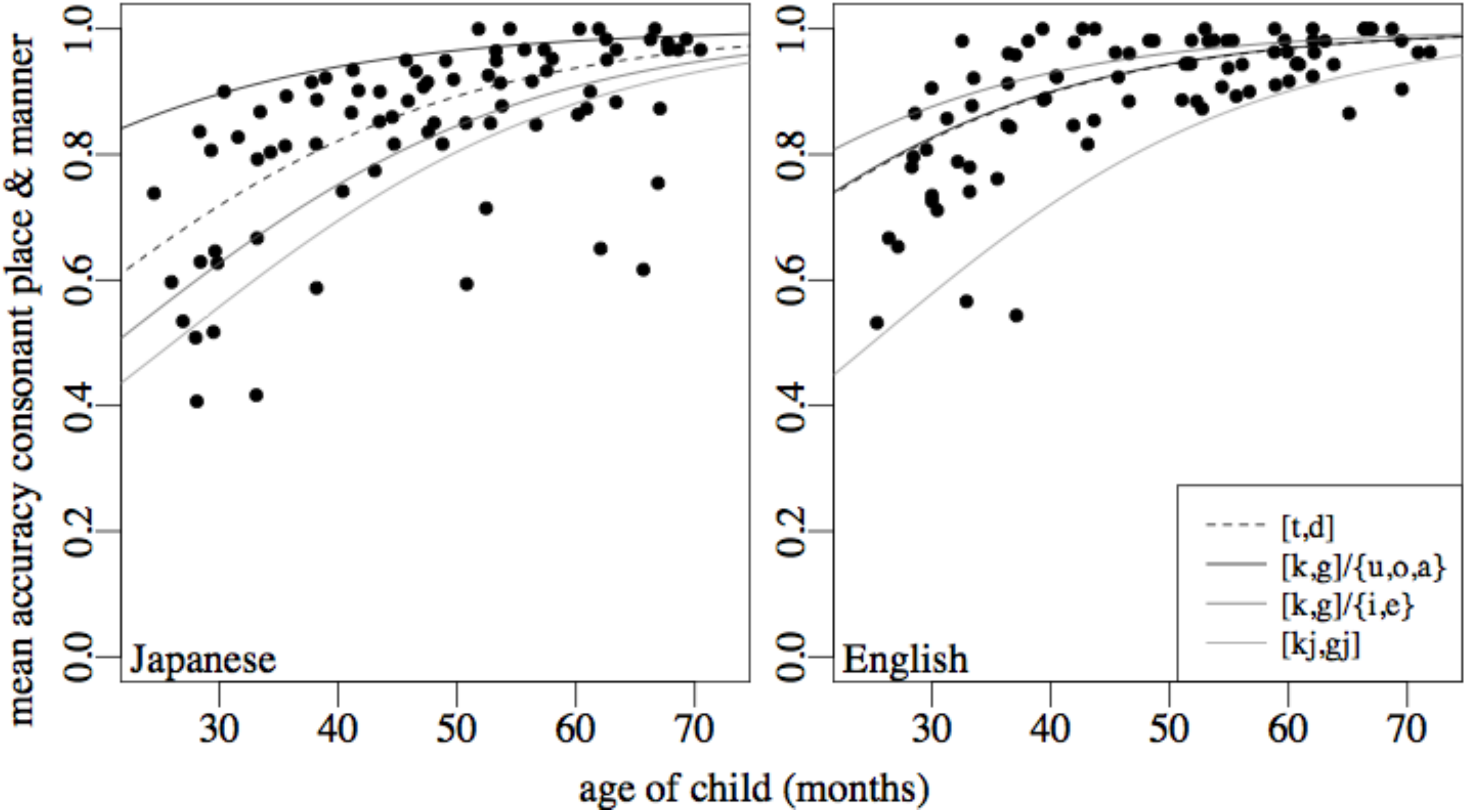
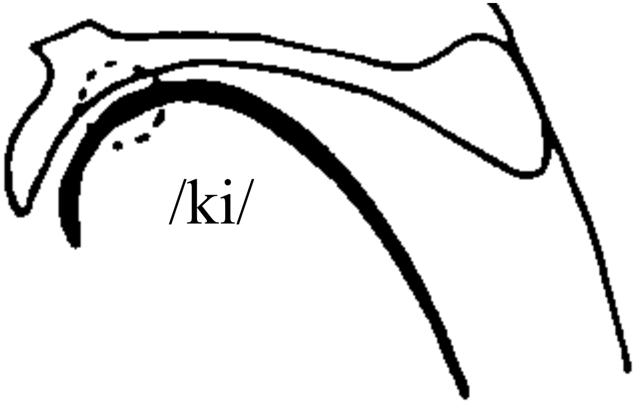
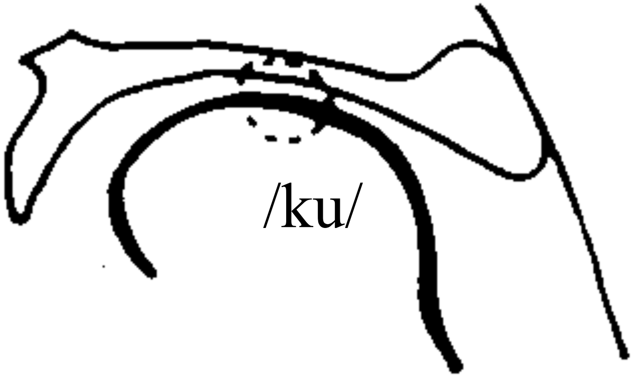
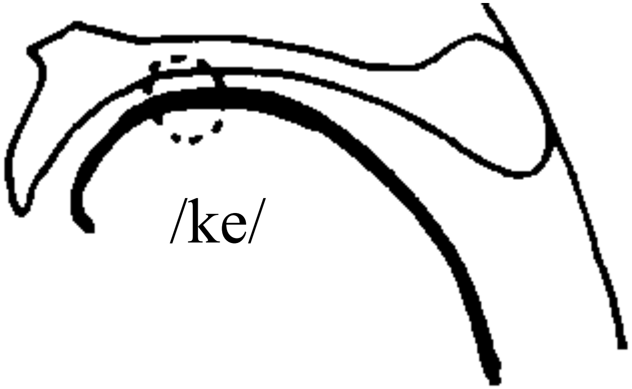


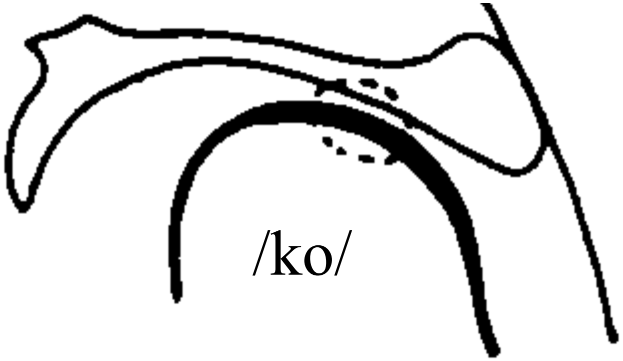
Figure from Munson, Yoneyama, & Edwards (October, 2012)

Articulation of dorsal stop in Japanese

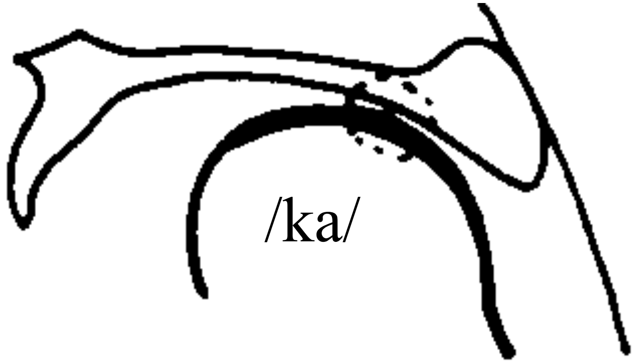




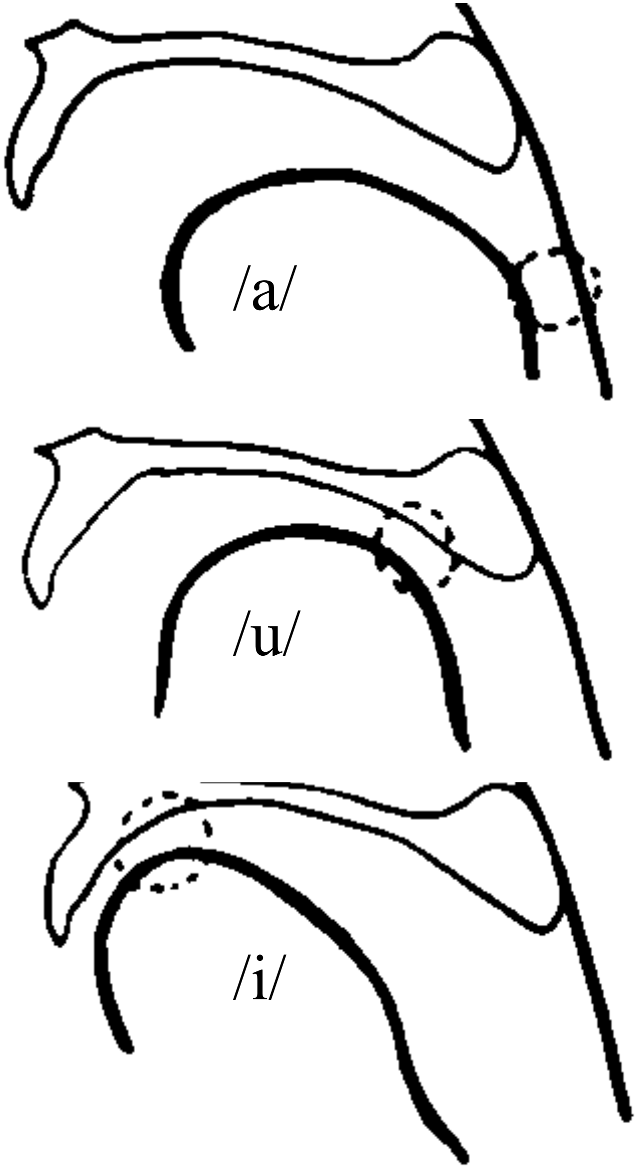




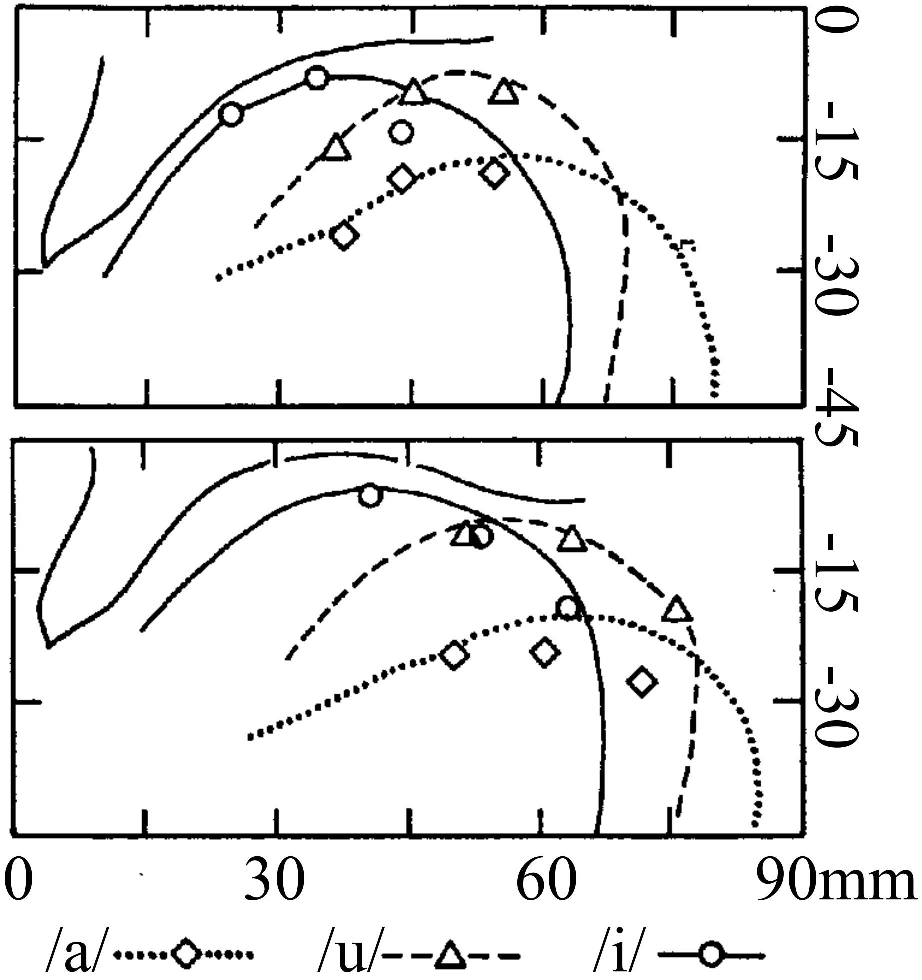
Cineflourographic midsagittal views of Japanese /k/ before each of the 5 vowels (Wada, Yasumoto, Ikeoka, Fujiki, & Yoshinaga, 1969)



English /i/ not as extreme?



Japanese point vowels, from Wada, et al. (1969)



English point vowels, from Kent & Moll (1972)

What we don’t know

What are the effects on dorsal stops of such cross-language differences in the articulation of vowels?

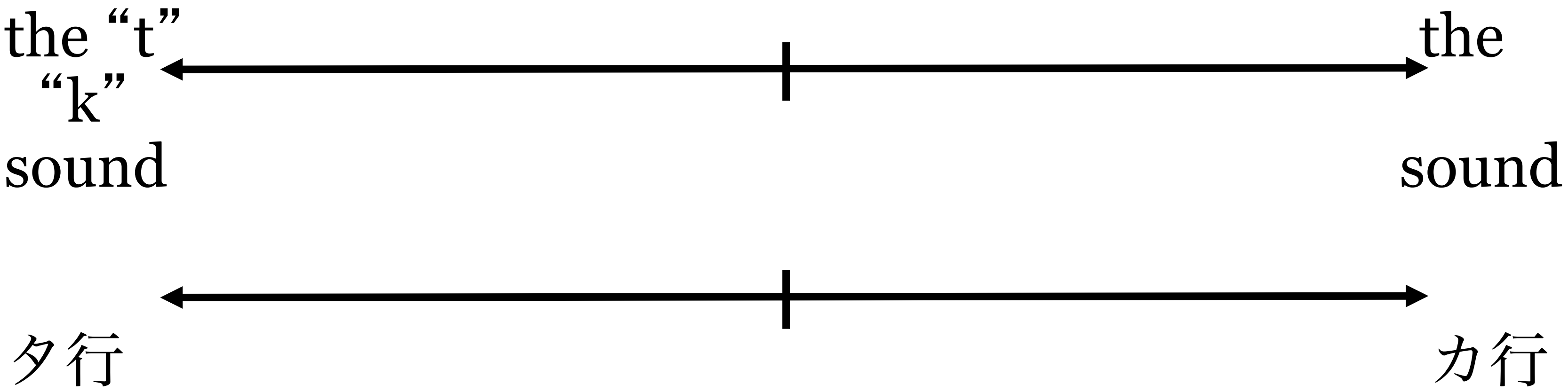
- English coronal stops are alveolar and typically apical
- Japanese coronal stops are dental and laminal

What are the effects of such differences in coronal stops?

- Cues to stop place are necessarily more distributed over the spectrum of the stop release burst and the transition into the following vowel. Which cues dominant when?
- Synthesis of stop place continua for perceptual studies is much more challenging.
- Acoustic measures of stop place are not as well developed as acoustic measures of fricative place, even for adults.

Using VAS to study stop place perception

- Munson, Yoneyama, and Edwards (October, 2012) extract CV portion of words beginning with dorsal and coronal plosives in Japanese- and English-speaking children’s production.
- English- and Japanese-speaking adults use VAS scale with end points labeled as appropriate for the two different writing systems:



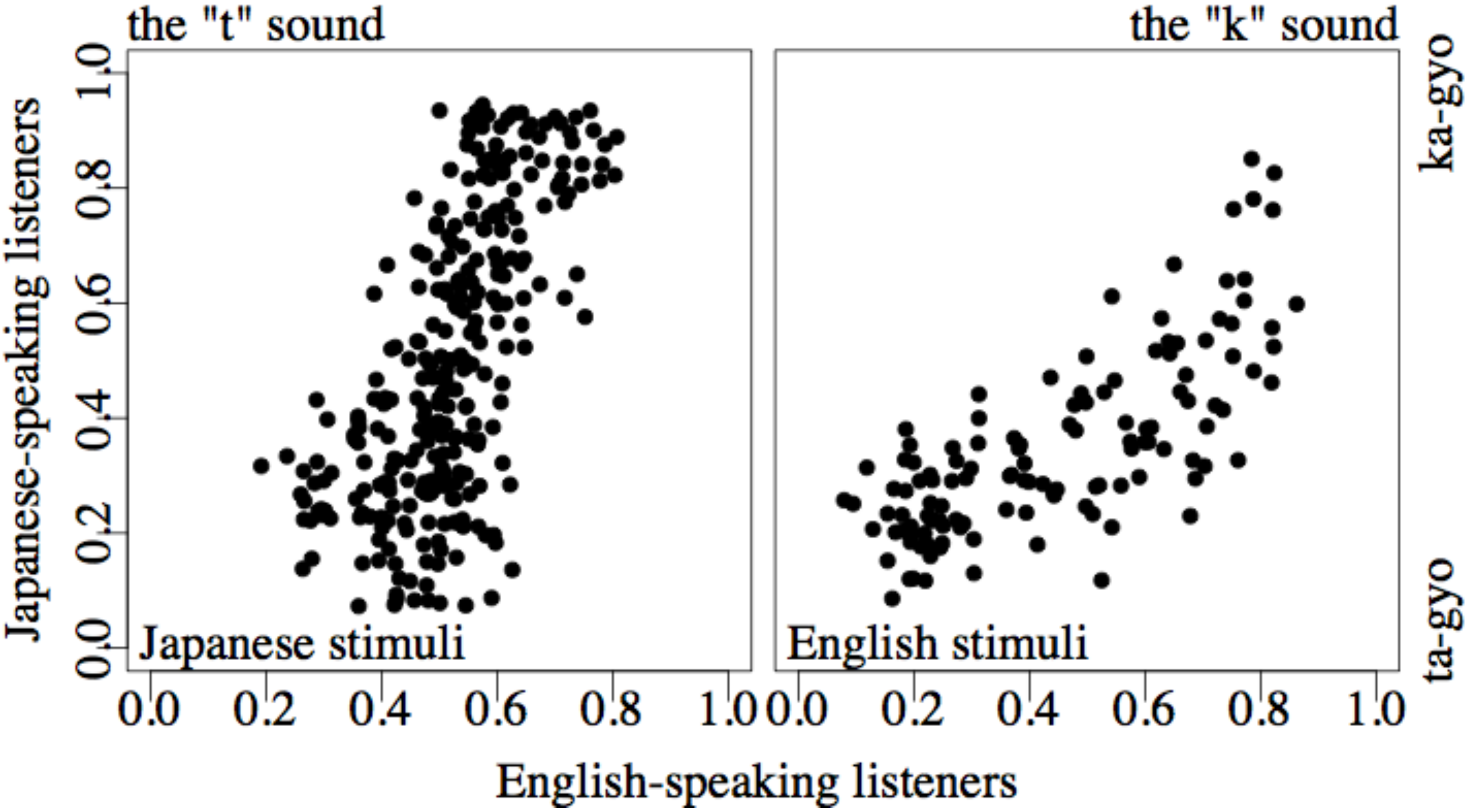
the “t” the

“k”

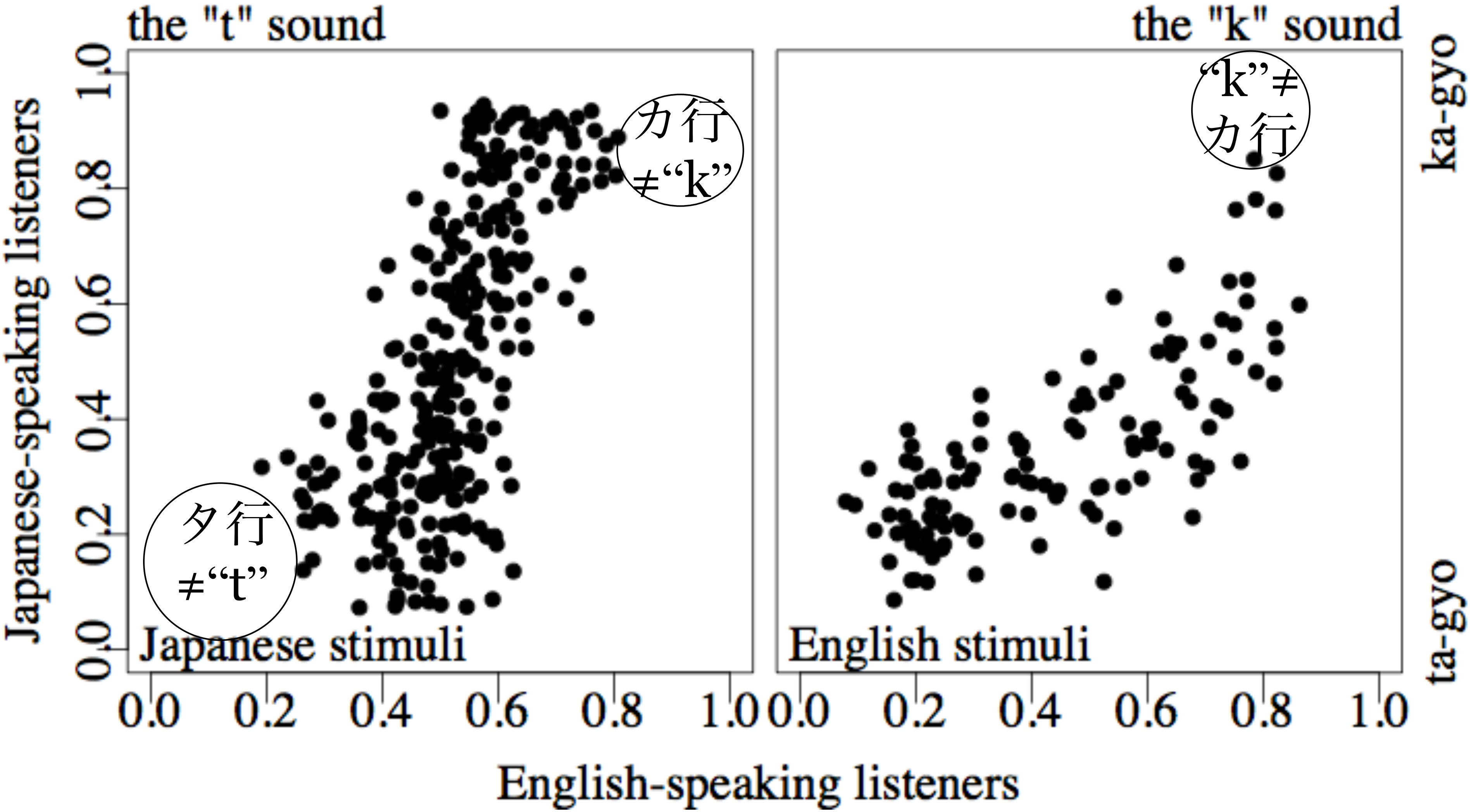
sound sound

タ行 カ行

Preliminary results



Preliminary results



Summary 3: Measuring stop place

- Acoustic measurement of stop place contrasts is extremely challenging.
- However, methods used to examine the perception of fricative place contrasts can be applied to stimuli that vary in stop place.

☞Cross-language differences in perceptual responses to dorsal and coronal stops tell us how to focus our efforts to develop better acoustic measures.

Ask me about:

Case study 4: The listen-rate-say task

Watch this place for further developments:

http://learningtotalk.org

Overall summary

- Cross-linguistic comparisons uncover many apparent contradictions in “markedness” based accounts.

These apparent contradictions prompt us to …

1. develop finer-grained measures of adults’ and children’s productions of difficult sounds
2. develop more sensitive methods for measuring adults’ perception of children’s productions
3. apply these methods to focus our examination of contrasts that are phonetically especially challenging
4. and in examining the interaction between adult perception and productions in response to children

Acknowledgements to

- 日本音声学会 for inviting me to give this talk
- my collaborators, especially those listed on the title slide
- funding sources listed on the title slide
- the children who lent us their voices & their caretakers
- and you for your kind attention

Ευχαριστώ πολύ

감사합니다

ありがとう

謝謝

Thank you

Study 4. Adult perception affects production

The Listen-Rate-Say (LRS) task (Munson & Julien 2011)

Stimuli

- 200 consonant-vowel sequences excised from the initial position of words produced by 2- and 3-year-old monolingual English-acquiring children in a picture-prompted repetition task.
- All of the target words had initial /s/ and /ʃ/ targets.
- Children's productions had been transcribed as being either correct or a substitution of /ʃ/→[s] or /s/→[ʃ].

Participants

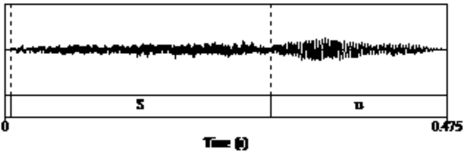
- 22 adult native speakers of English with no specialized training in speech and language.

The Listen-Rate-Say task

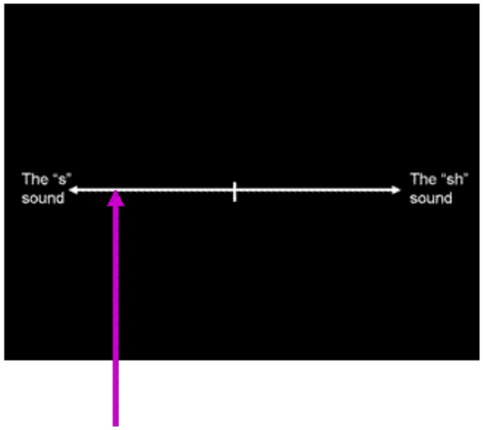
Listen to the initial CV of a child's attempt to say an /s/- or /ʃ/-initial word while looking at the picture the child was naming

Rate the child's production using a visual analog scale, (as in Urberg-Carlson et al., 2008)

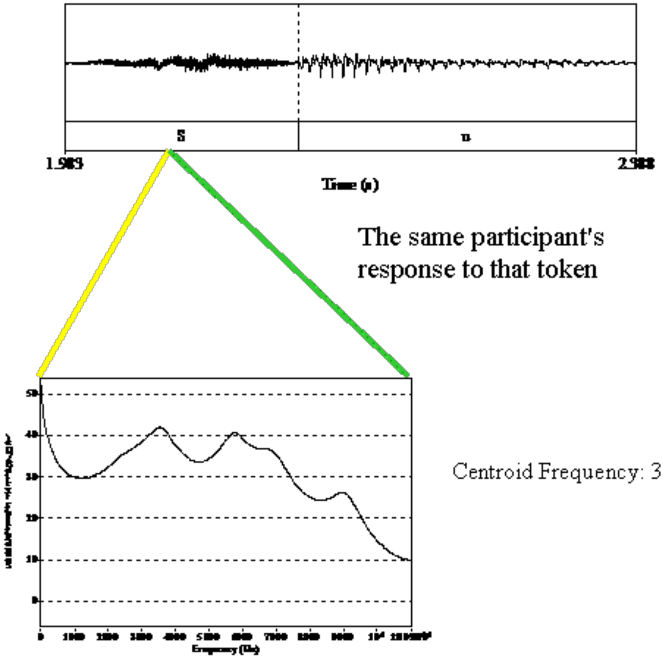
Say the word that the child was attempting, 'as if you were responding to the child whose speech you just rated'



One 3-year-old child's production of shoe transcribed to have an [s]-for-/ʃ/ error



One adult's accuracy rating for that token

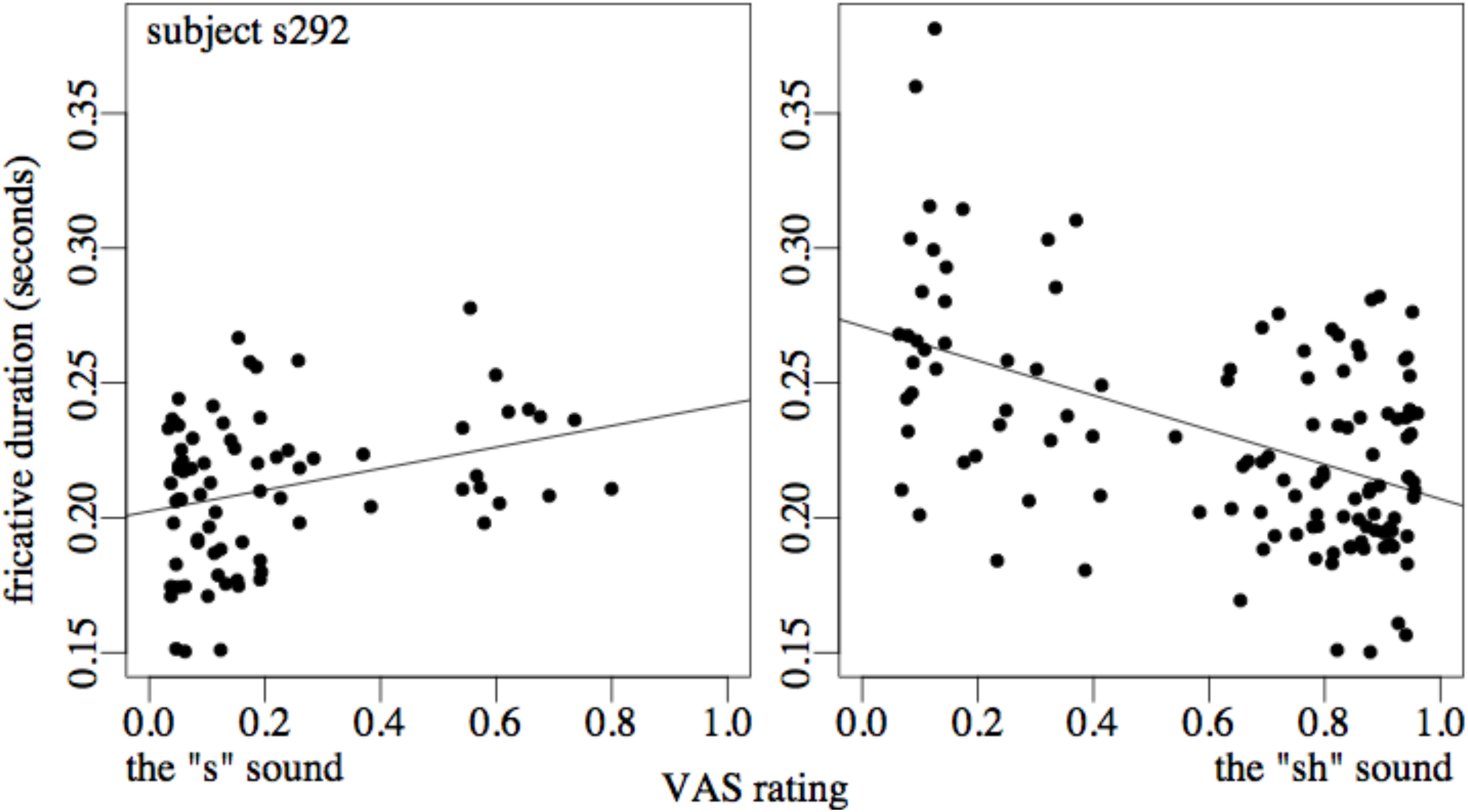


The same participant's response to that token

Centroid Frequency: 3386

Do adults adjust their productions?

For 14 of 22 talkers, duration related to VAS for [s] and/or [ʃ].


subject s292
fricative duration (seconds)
the "s" sound
VAS rating
the "sh" sound