VOT is necessary but not sufficient for describing the voicing contrast in Japanese

Eun Jong Kong*, Mary E. Beckman*, Jan Edwards †

(*Ohio State University, †Univ. of Wisconsin at Madison)

THE OHIO STATE UNIVERSITY

THE UNIVERSITY of WISCONSIN MADISON

Introduction

- Since the seminal work of Lisker and Abramson (1964), Voice Onset Time (VOT) has been used as the primary measure for comparing word-initial stop voicing and aspiration contrasts across languages.

e.g.,

- Spanish: /d/ vs. /t/
  lead VOT vs. short lag VOT
- Cantonese: /t/ vs. /t^h/
  short lag vs. long lag VOT.
- English: /d/ vs. /t/

  lead or short lag VOT vs. long lag VOT

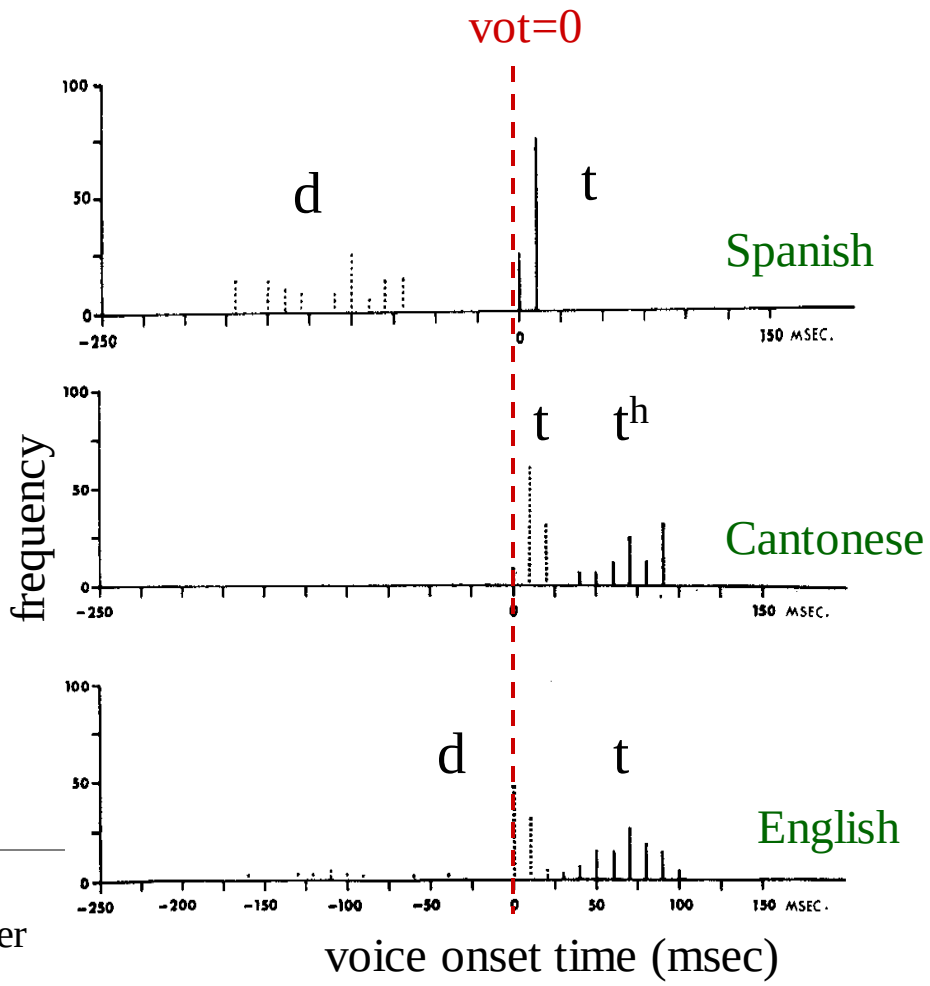


Figure.1 Voice onset time distribution of apical (dental and alveolar) stops of two-category languages. Taken from Lisker & Abramson (1964).

Introduction

- VOT has also been a useful acoustic measure for describing children's mastery of word-initial stops in languages with voicing and/or aspiration contrasts.

e.g., Thai (Gandour et al 1986)

- stops with three-way contrast

: /d/ vs. /t/ vs. /t^h/

- lead VOT mastered later than short lag VOT or long lag VOT

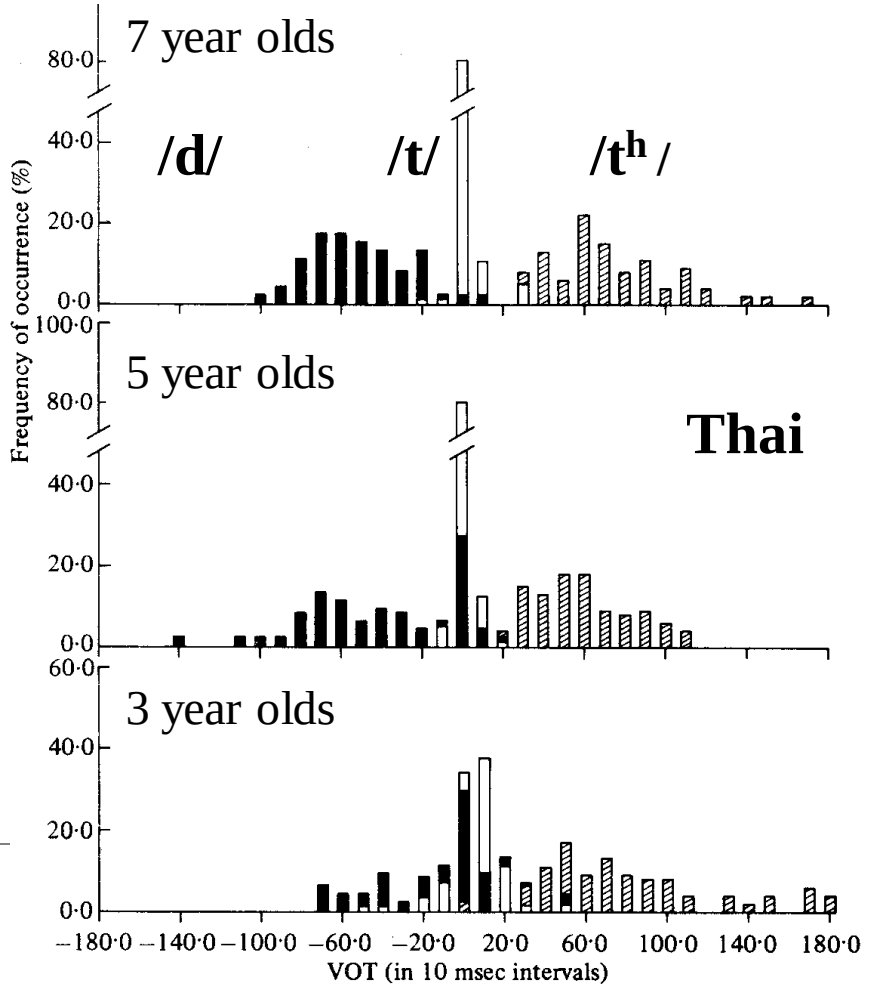


Figure.2 VOT distribution of alveolar stops in Thai. Taken from Gandour et al (1986).

Introduction

- Is VOT the whole story?
- Japanese stops and VOT
  - Two-way voicing contrast (Homma 1980, Shimizu 1989)
    - voiced stops: not only lead VOT, but also short lag VOT (Takada 2004)
    - voiceless stops: neither clearly short lag nor clearly long lag, but intermediate between the two (Riney et al 2007)
  - This results in overlap in VOT range between the two categories
  - Is there another acoustic measure that helps to disambiguate?

Goal of the study

- To evaluate whether VOT is a sufficient acoustic measure in distinguishing voiced stops from voiceless stops in Japanese, we investigate

  - how the acoustic parameter of VOT relates to native speaker/transcriber judgments of accuracy for voiced and voiceless stop consonants in English- and Japanese-acquiring children.
  - whether another acoustic parameter is also needed to predict native speaker/transcriber judgments of these productions.

Research questions

- Children's stop productions were analyzed to address the following questions.

Question 1) Are there differences between the time-courses for mastering the stop voicing contrasts in English and Japanese?

Method; judgments by trained native speaker/phoneticians, logistic regression.

Question 2) How well does the single acoustic dimension of VOT predict the native speaker/transcriber's judgments of voiced vs. voiceless stops produced by English- and Japanese-acquiring children?

Question 3) Is there another acoustic dimension that improves the prediction of the native speaker/transcriber's judgments of the voicing contrast in stops produced by these children?

Method; acoustic analysis, logistic regression

Data collection

1) Production data come from παιδολογος project

- cross-language investigation of phonological development

www.ling.ohio-state.edu/~edwards/

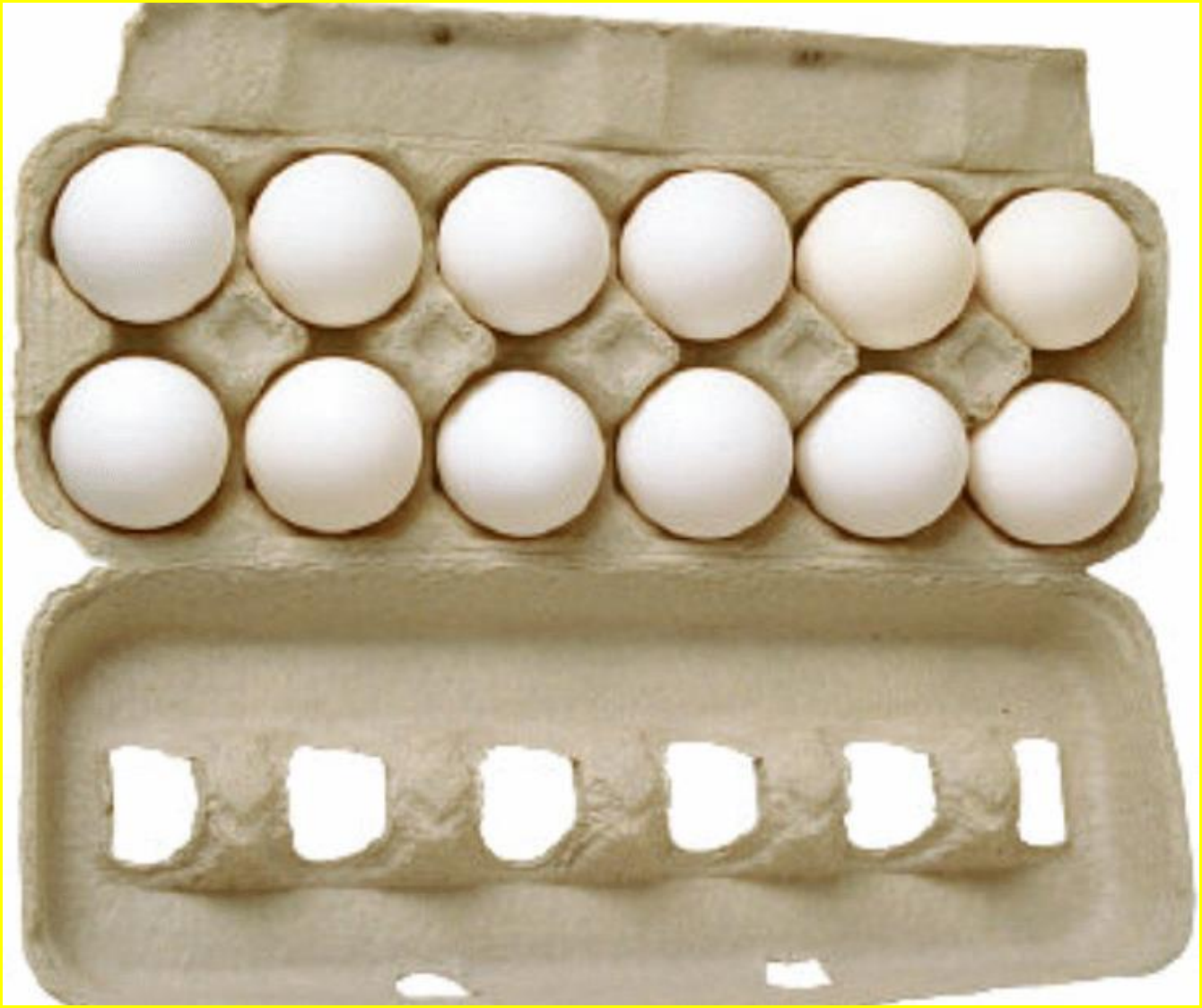
2) Subjects

- 51 children (2;0-6;0) , 20 adults (18;0-30;0) recorded in Tokyo
- 50 children (2;0-6;0) , 15 adults (18;0-30;0), recorded in Ohio

3) Materials: word-initial pre-vocalic lingual stops — e.g.,

- Japanese /d/ *daikon* 'radish' vs. /t/ *tamago* 'egg'
- English /d/ *dove* vs. /t/ *tongue*

  (velar stops were also recorded but not discussed here)



tamago ‘egg’

daikon ‘radish’

Correct Voicing

Voicing Error

Correct Voicing

Voicing Error

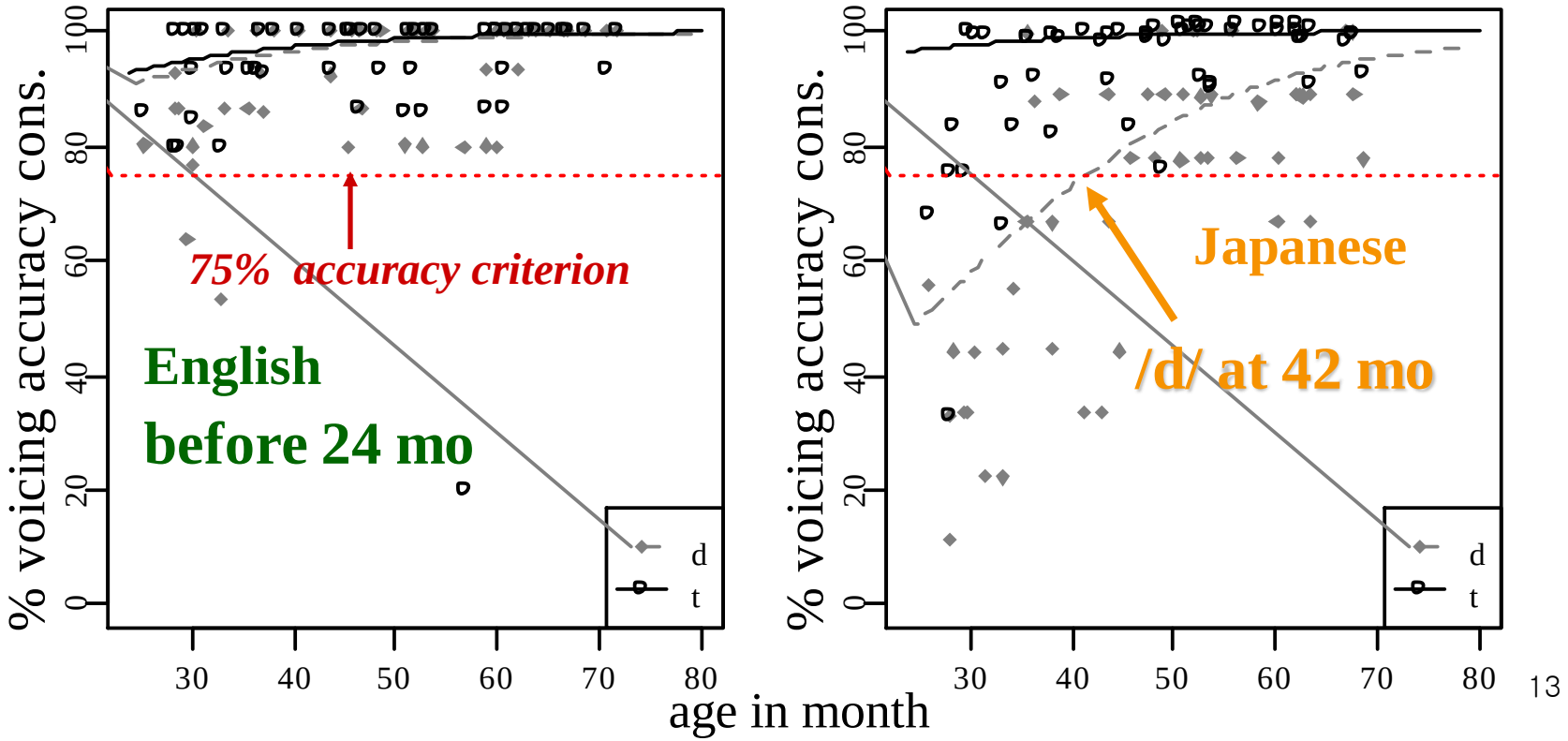
Analysis 1: Transcription

Question 1) Are there differences between the time-courses for mastering the stop voicing contrasts in English and Japanese?

- Measure: voicing accuracy from transcriptions by a trained phonetician native speaker of English/Japanese.
  - voicing correct: /t/ → [t], /d/ → [d], /d/ → [g], /t/ → [k]
  - voicing error: /t/ → [d], /d/ → [t], /t/ → [n]
- Criterion for mastery: 75% voicing accuracy (adapted from criteria used in norming studies such as Smit et al., 1990).

Transcription: results

- Mixed effects logistic regression.
  - Dependent variable: token by token voicing accuracy (correct / incorrect)
  - Independent variable: age of child and target voicing (fixed effect) + subject (random effect)



Analysis.1: interim conclusion

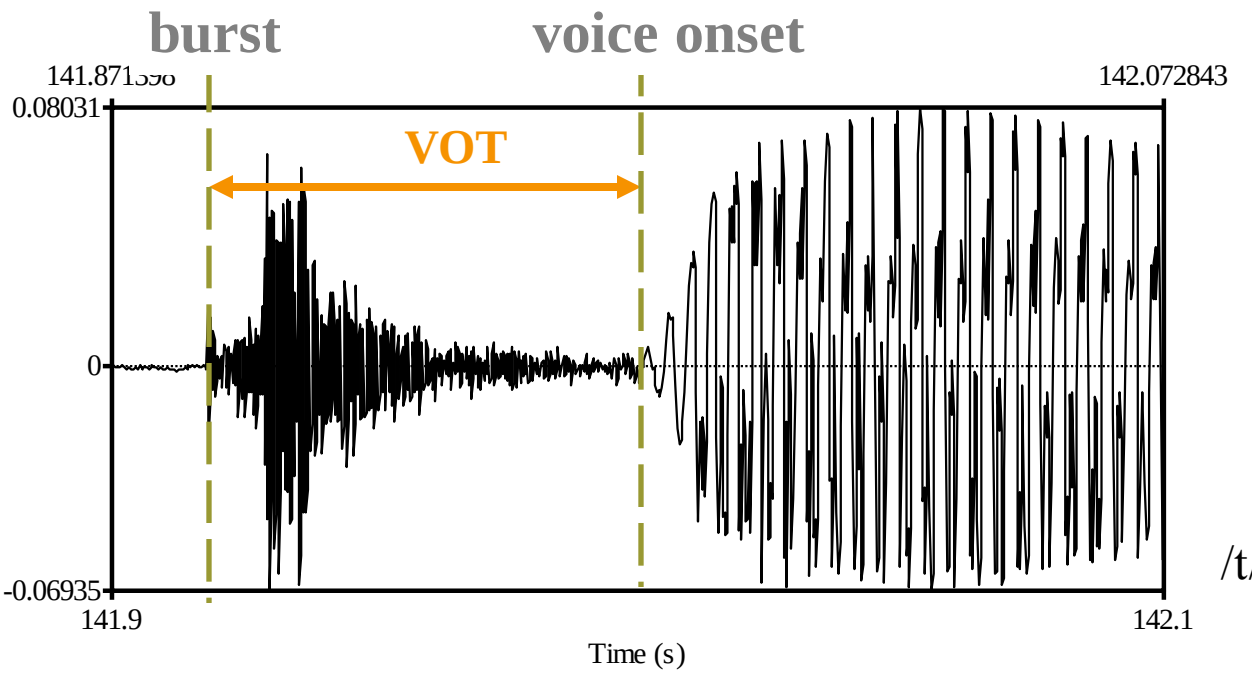
Transcription Analysis

- The voicing contrast is mastered later by Japanese-speaking children, as compared to English-speaking children.

Analysis 2: VOT

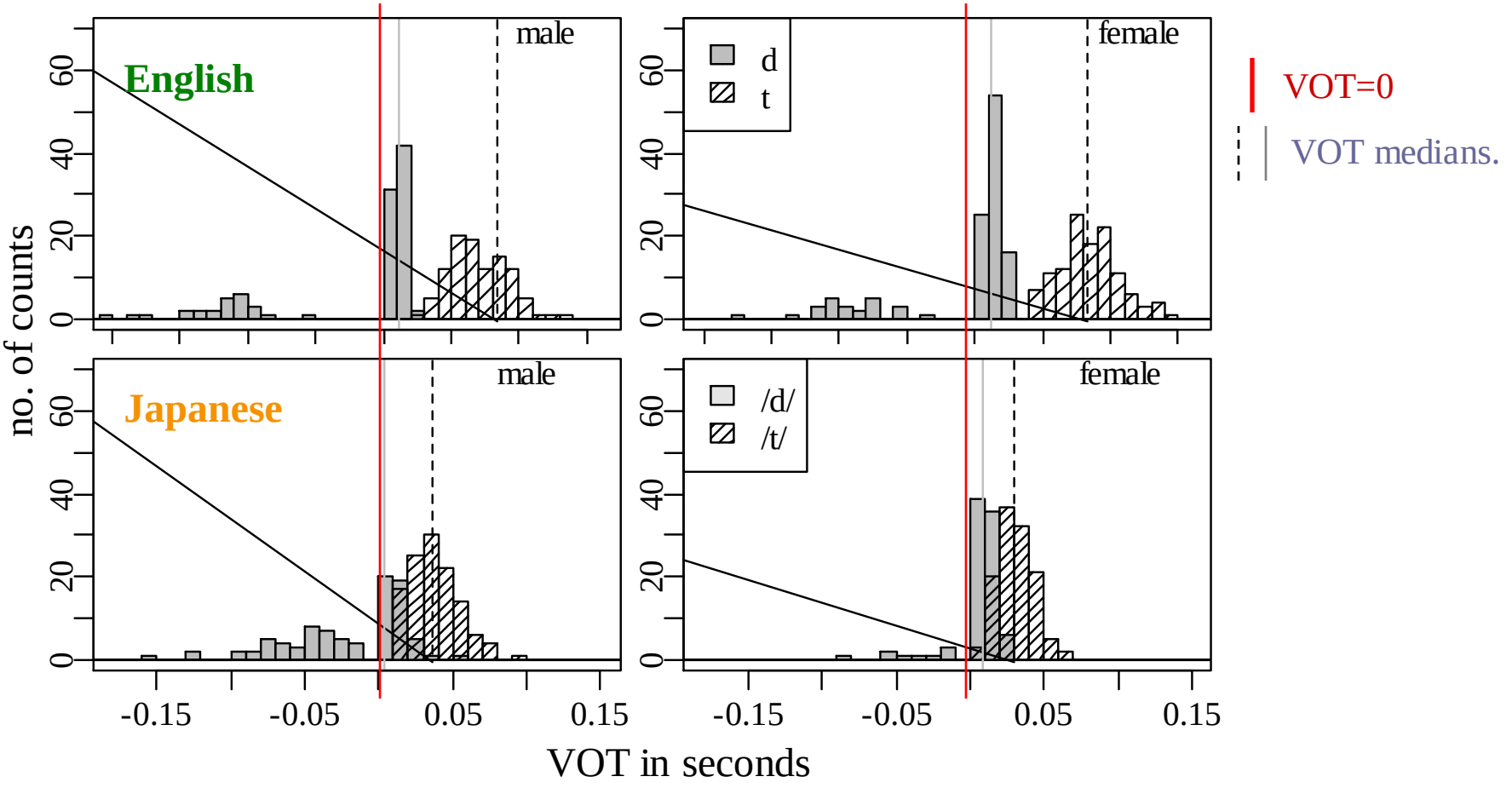
Question 2) How well does the single acoustic dimension of VOT predict the native speaker/transcriber's judgments of voiced vs. voiceless stops produced by English- and Japanese-acquiring children?

- VOT: the latency between the burst and the voicing onset.



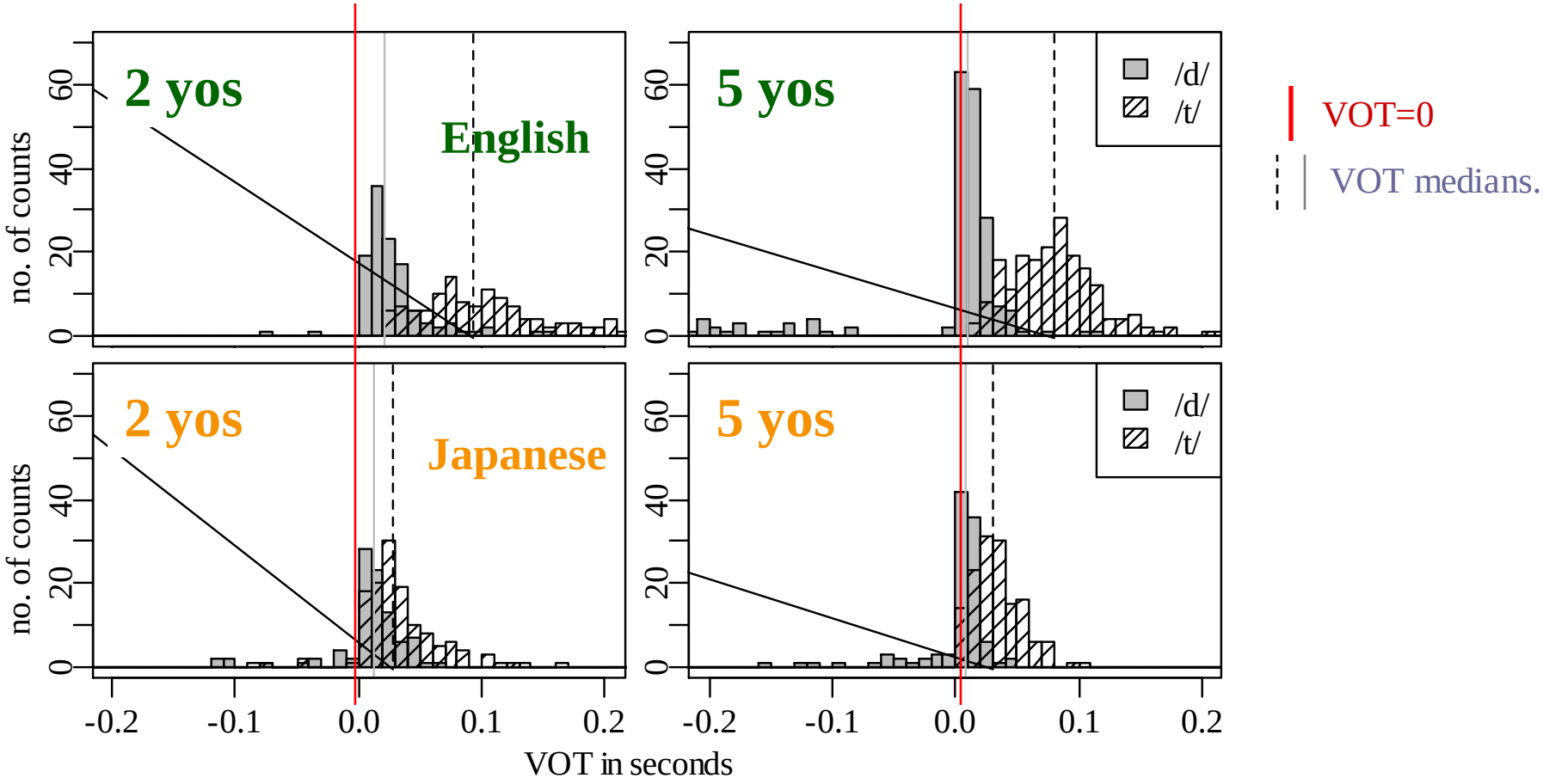
/t/ in "torn"

VOT: results (adults)



- English: clear separation between short lag (/d/) vs. long lag (/t/)
- Japanese: lead or short lag (/d/) vs. intermediate lag (/t/), with much overlap.

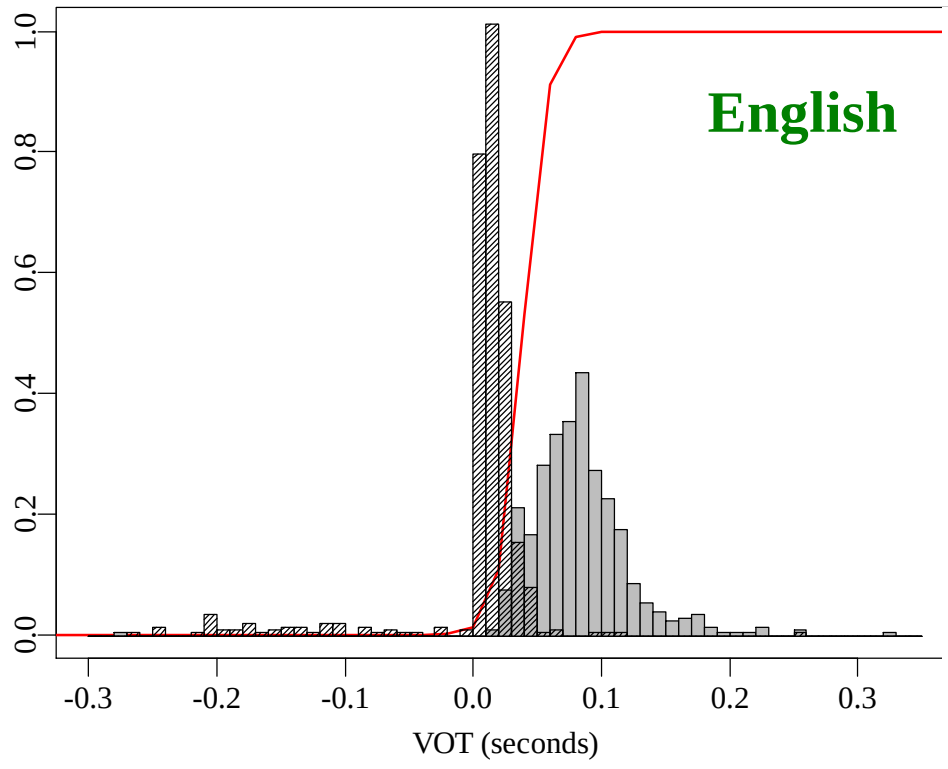
VOT: results (children)

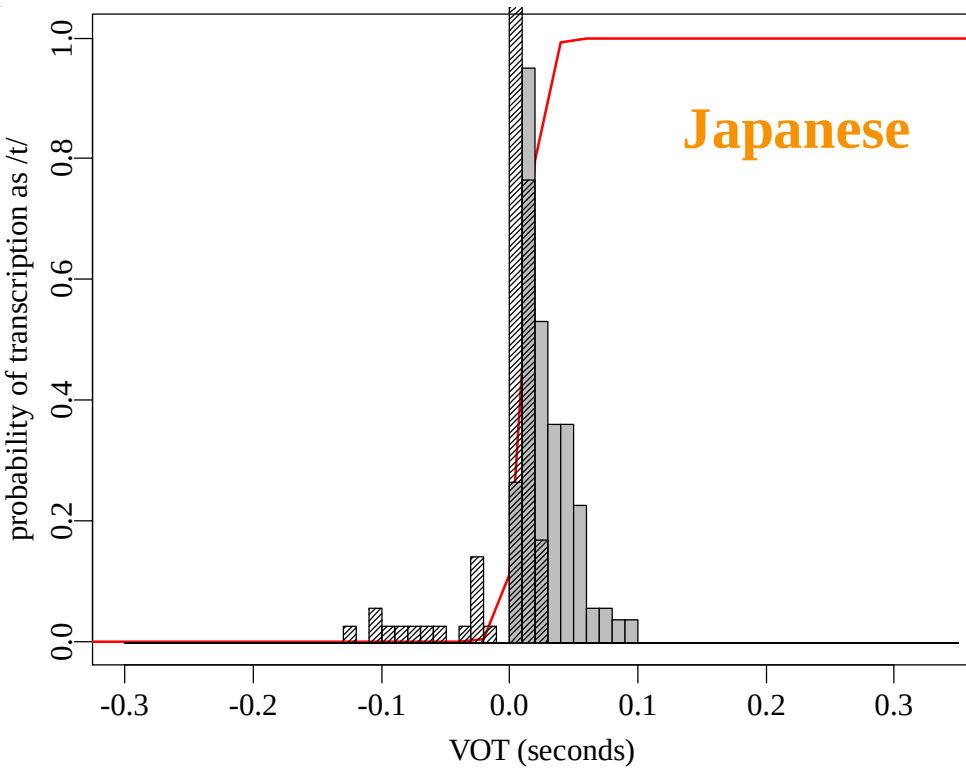


- Language specific VOT distributions in children's stops
  - English: clearly separated peaks.
  - Japanese: intermediate values for /t/ with even more overlap with /d/ than in adults.

VOT: results (children)

- Mixed effects logistic regression
  - Dependent variable: token by token voicing judgment (**/t/ or not /t/**)
  - Independent variable: **VOT**





VOT: results (children)

- Evaluation of predictive value
  - Baseline prediction accuracy with no independent variable

    i.e., calculate the proportion of tokens where the transcriber transcribed a voiceless consonant:

    'Baseline': 49.7% and 63.3%

  - Model's prediction accuracy with VOT as an independent variable

    i.e., calculate proportion of tokens where the odds of transcribing /t/ are greater than 50% and the transcriber actually transcribed /t/:

    'VOT model': 94% and 80%

Analysis 2: interim conclusion

Transcription Analysis

- The voicing contrast is mastered later for Japanese-speaking children, as compared to English-speaking children.

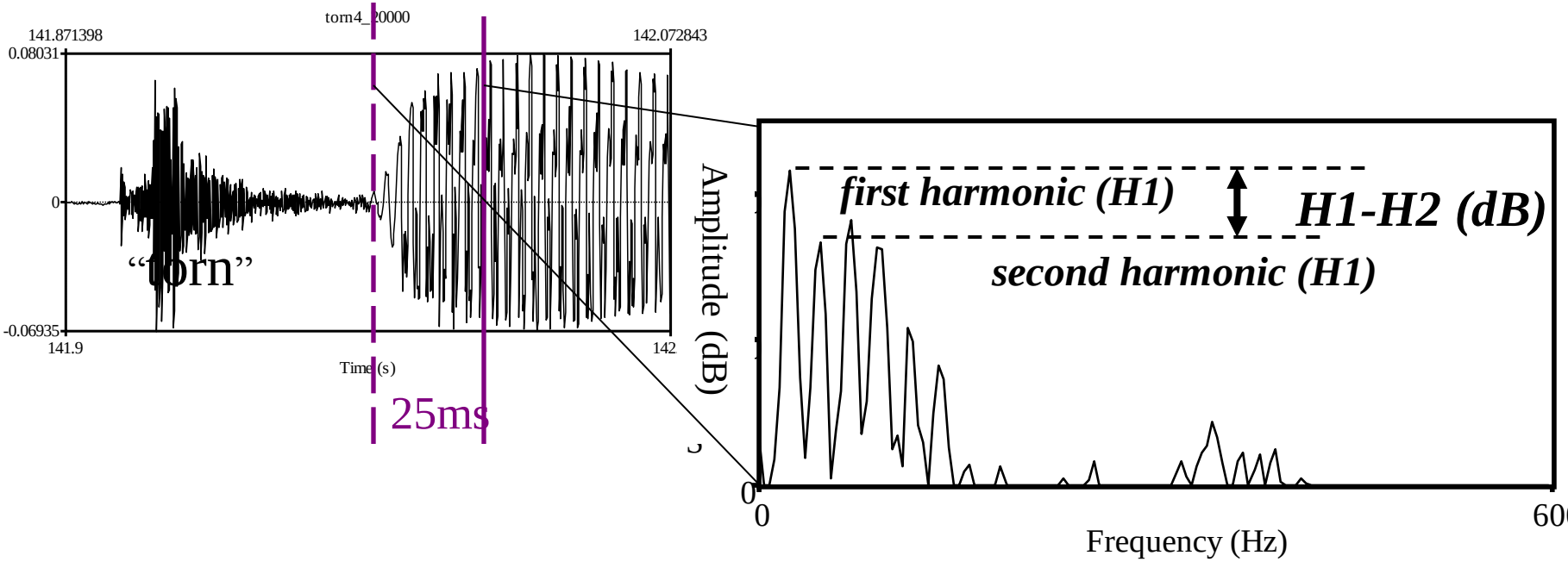
VOT

- The single acoustic dimension of VOT predicts the transcribed voicing for English productions 94% of the time.
- Accuracy of prediction for Japanese productions is much lower.

Analysis 3: H1-H2 by VOT

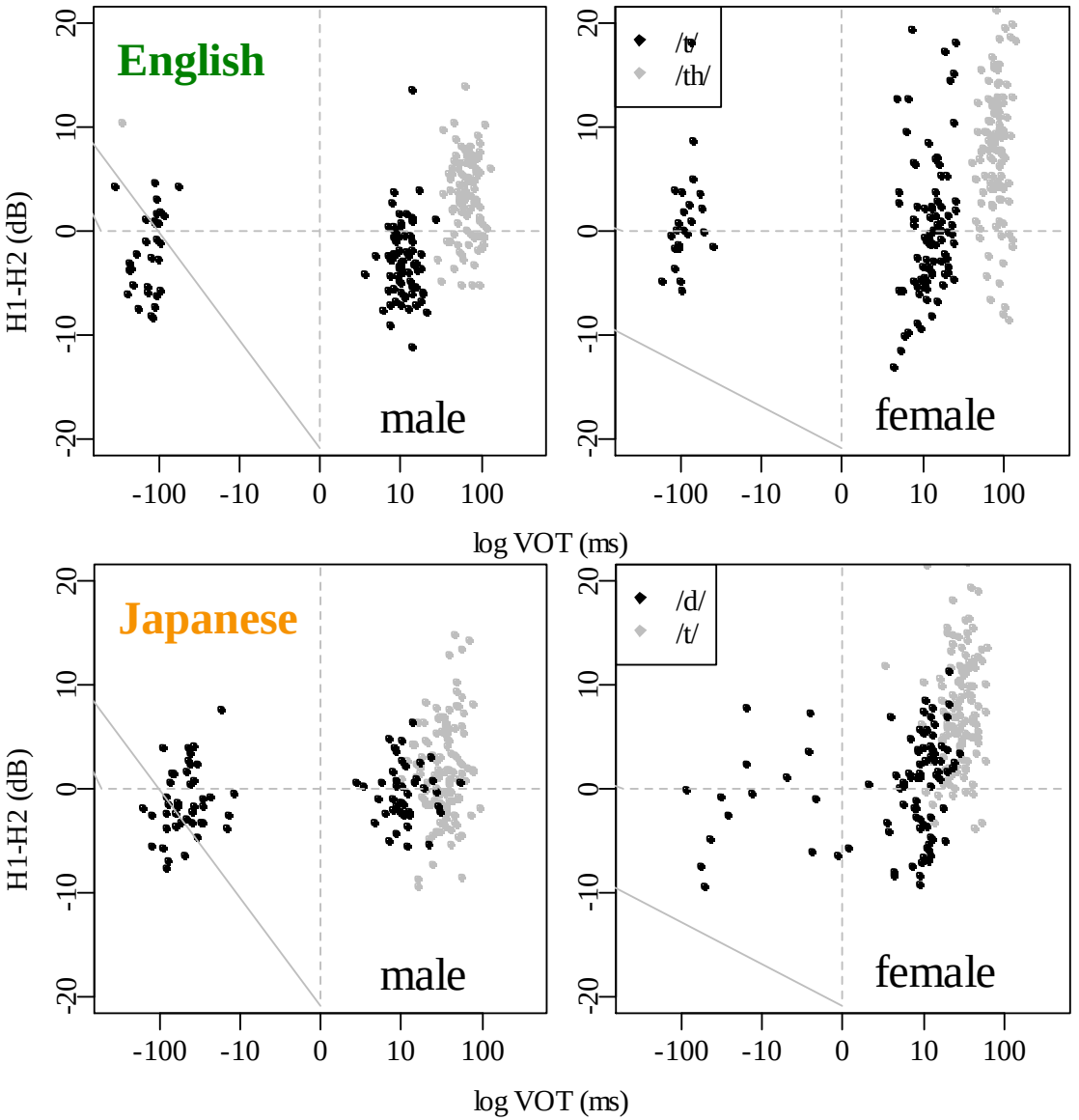
Question 3) Is there another acoustic dimension that improves the prediction of the native speaker/transcriber's judgments of the voicing contrast in stops produced by these children?

- H1-H2
  - A type of breathiness measure.
  - Amplitude difference between the first harmonic and the second harmonic.



H1-H2 by VOT: adults

- Adults
  - English
    - Higher H1-H2 and longer VOT for /t/.
    - No overlap between VOT ranges
  - Japanese
    - Higher H1-H2 and longer VOT for /t/.
    - Overlap between VOT ranges



H1-H2 by VOT: children

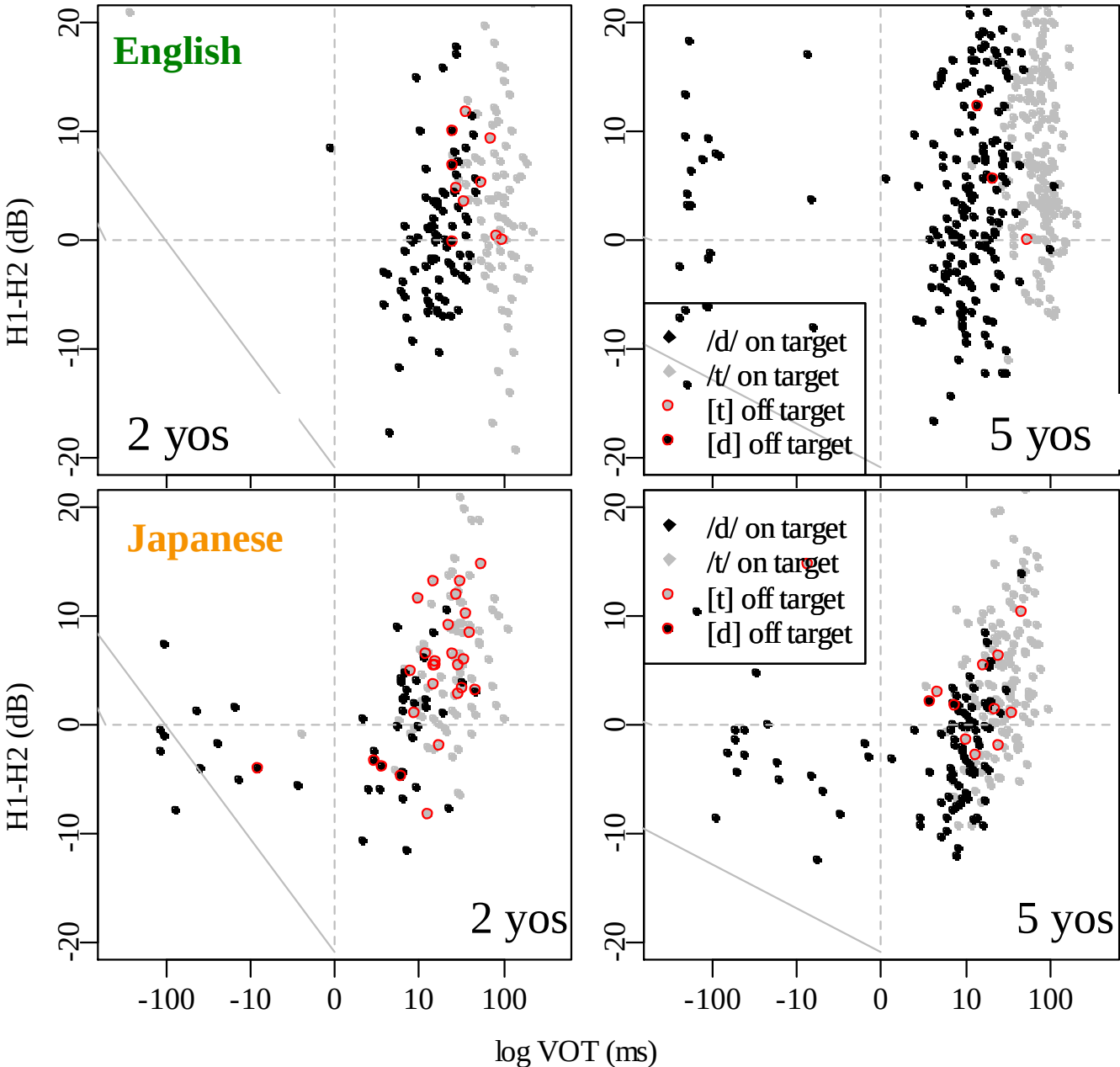
- Perceived /t/ and /d/ by transcriber.

English /t/

: longer lag VOT

Japanese /t/

: longer lag VOT,
higher H1-H2

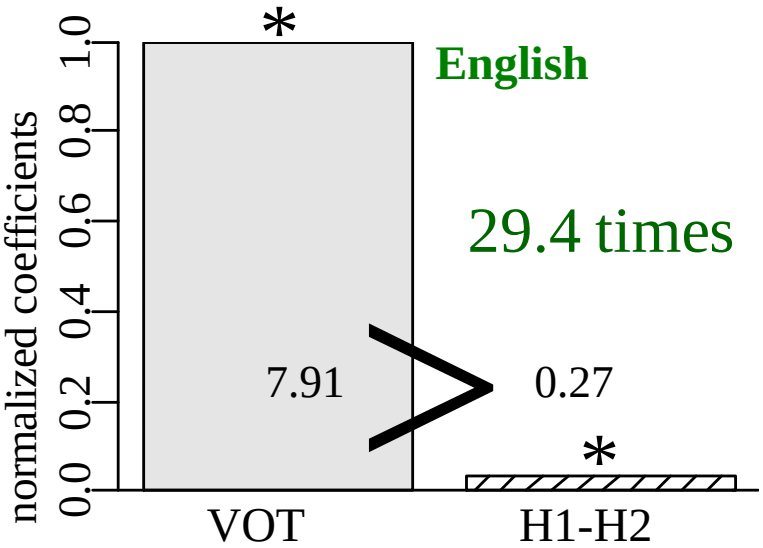


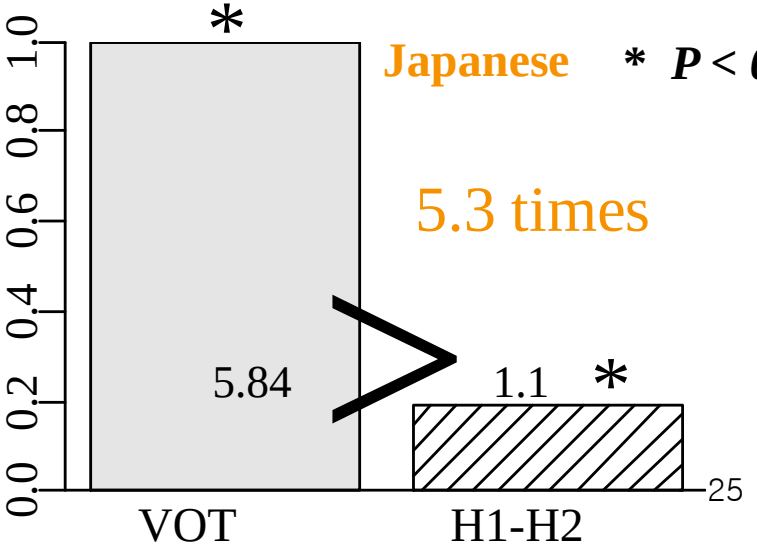
VOT: results (children)

- Mixed effects logistic regression
  - Dependent variable: token by token voicing judgment (**/t/ or not /t/**)
  - Independent variables: **VOT+ H1H2**

VOT and H1-H2: results (children)

- Evaluation of predictive value
  - Baseline prediction accuracy with no independent variable
    i.e., calculate the proportion of tokens where the transcriber transcribed a voiceless consonant: 49.7% and 63.3%
  - Model's prediction accuracy with VOT as an independent variable: 94% and 80%
  - Model's prediction accuracy with VOT and H1-H2 as independent variables: 94% and 83%





Analysis 3: interim conclusion

Transcription Analysis

- The voicing contrast is acquired later for Japanese-speaking children, as compared to English-speaking children.

VOT

- The single acoustic dimension of VOT is adequate to characterize the transcription results for English.
- However, VOT alone does not adequately characterize the transcription results for Japanese.

H1-H2 by VOT

- In Japanese, the additional acoustic parameter of H1-H2 improves the prediction of the transcription results.
  - The effects of VOT relative to H1-H2 was greater in English than in Japanese

Summary and conclusion

- Japanese-speaking children showed mastery of the voicing contrast at a later age than English speaking children.
  - However, the VOT ranges for the productions of Japanese-speaking children were similar to those of adults.
- When VOT alone was used to predict the judgments of a trained native speaker/transcriber, it was only 80% successful in Japanese, whereas it was 94% successful in English.
- Adding the acoustic parameter of H1-H2 improved the prediction of the native speaker/transcriber judgments for the productions of the Japanese-speaking children, but not for those of the English-speaking children.

Summary and conclusion

- English and Japanese encode their stop voicing contrast in the acoustic dimensions in language-specific ways.
  - English: exclusively along VOT dimension
  - Japanese: more than VOT dimension

- Unlike English, VOT is not a sufficient acoustic measure of stop voicing contrast in Japanese.
  - It was necessary to examine other relevant acoustic dimensions such as breathiness to correctly characterize Japanese stop voicing contrast.

Acknowledgement

- This work was supported by by NIDCD grant 02932 to Jan Edwards.
- We thank the children who participated in the task, the parents who gave their consent, and the principals and teachers at the schools at which the data were collected.

Thank you for your attention!

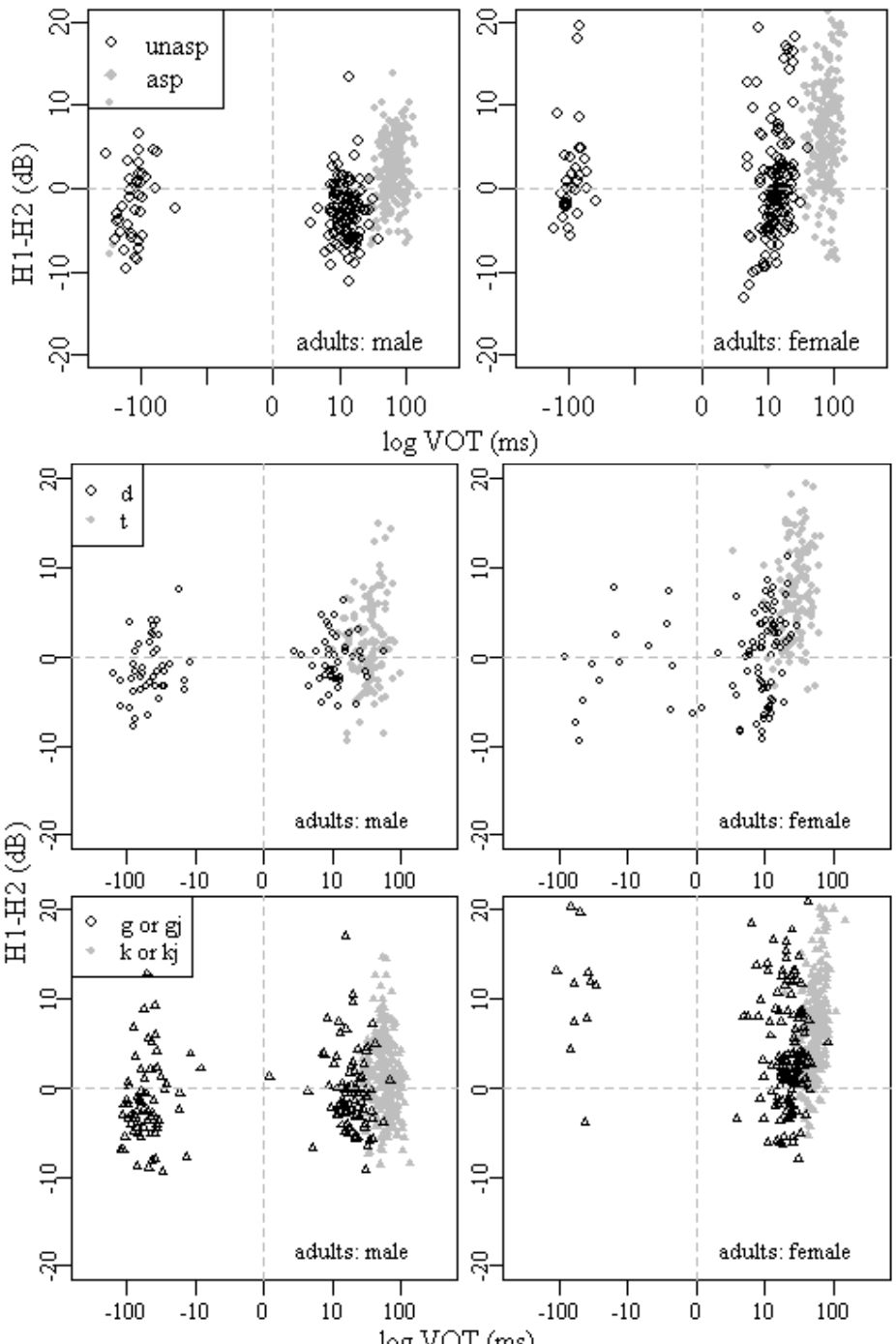
reference

- Lisker, L. and A. Abramson. 1964. A cross-language study of voicing in initial stops: acoustical measurements. Words, 20.
- Riney, T., N. Takagi, K. Otaa, and Y. Uchida. 2007. The intermediate degree of vot in japanese initial voiceless stops. Journal of Phonetics, 35.
- Smit, A.B., L. Hand, J. Freilinger, J renthal, and A Bird. 1990. The iowa articulation norms project and its nebraska replication. Journal of Speech and Hearing Disorders, 55.
- Gandour, H. S. H., J., R. Petty, S. Dardarananda, Dechongkit, and S. Mukongoen. 1986. The acquisition of the voicing contrast in thai: A study of voice onset time in word-initial stop consonants. Journal of Child Language, 13.
- Takada, M. 2004. VOT tendency in the initial voiced alveolar plosive /d/ in Japanese and the speakers' age. Journal of the Phonetic Society of Japan, 8(3), 57-66.
- Homma, Y. (1980). Voice onset time in Japanese stops. Onseigakkai Kaihoo, 163, 7-9.
- Sander, E.1972. When are speech sounds learned? Journal of Speech and Hearing Disorders, 37: 55-63.

Extra I: Velars adults scatterplts

English adults: coronals + velars

Japanese adults: coronals (top) + velars (bottom)



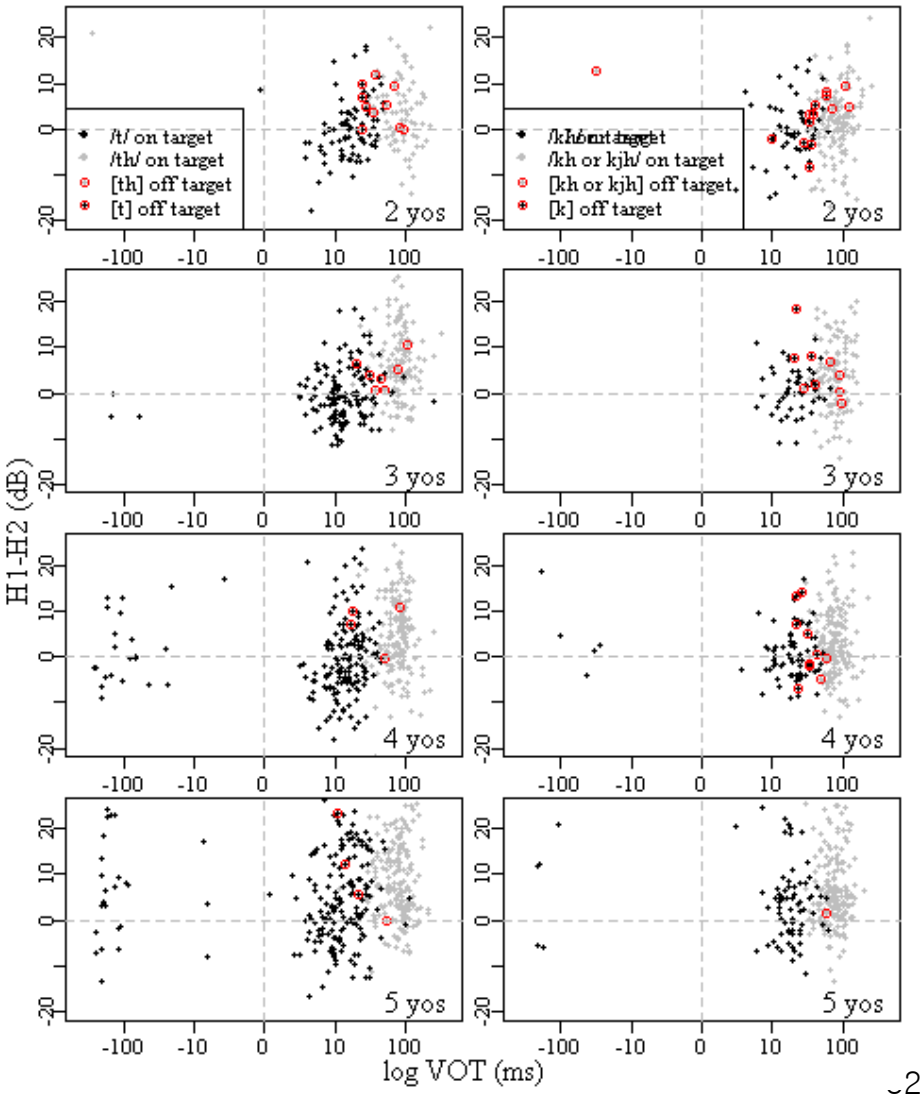
Extra I: Velars children scatterplots

- English children

(alv: left, velar: right)

- VOT only model: 93%

- VOT&H1-H2 model: no improvement. VOT was the only effective parameter.



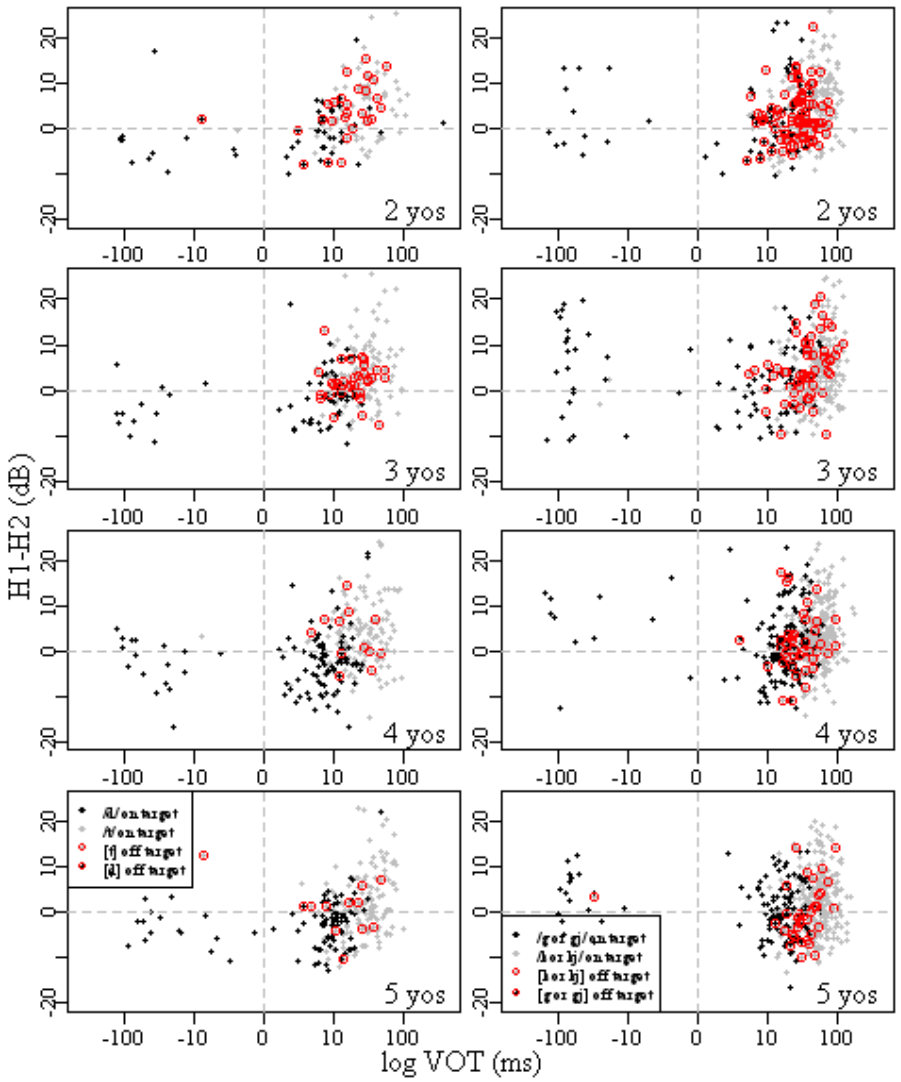
Extra I: Velars children scatterplots

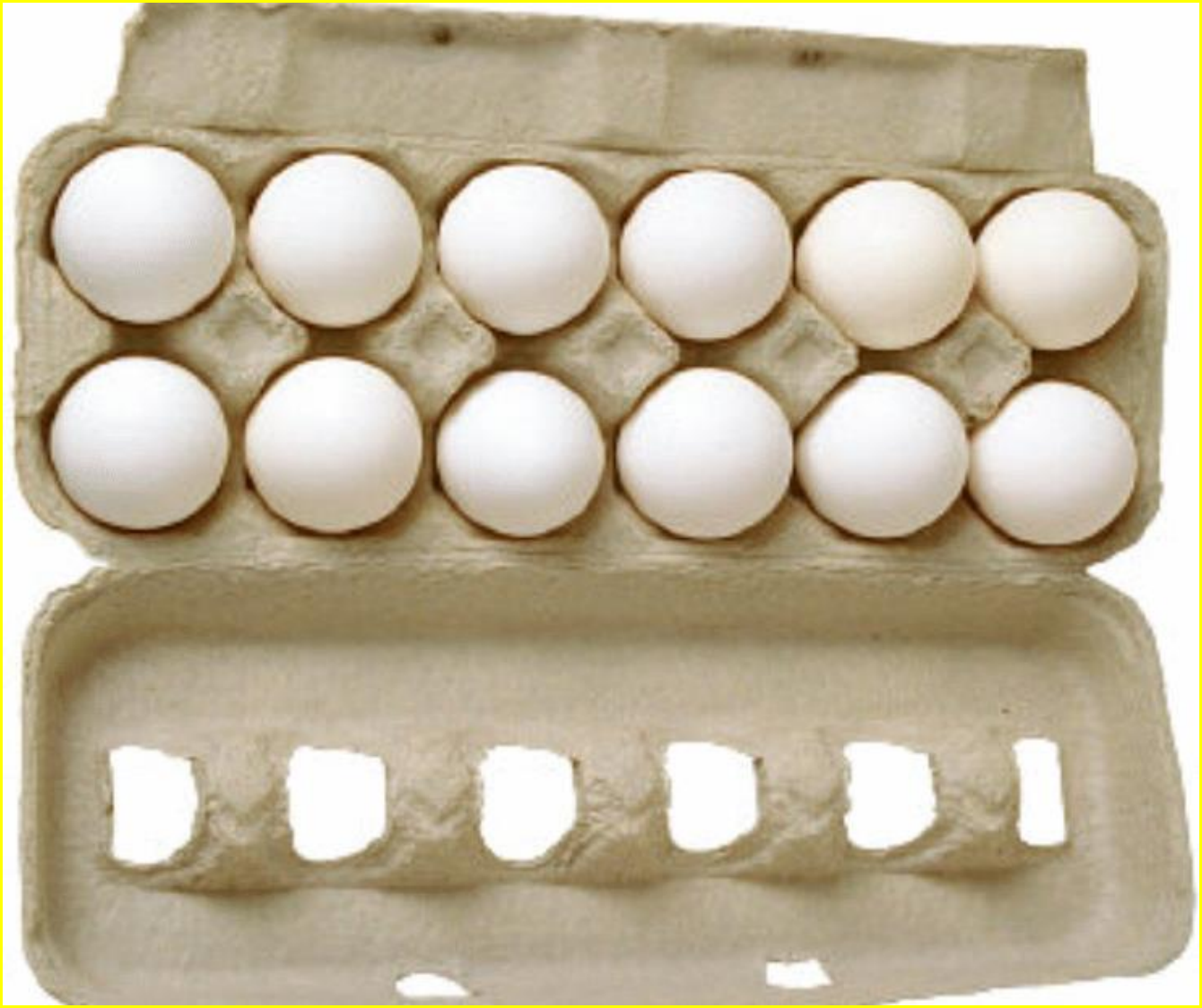
- Japanese children

(alv: left, velar: right)

- VOT only model: 87%

- VOT&H1-H2 model: no improvement. VOT was the only effective parameter.





Correct Voicing

Voicing Error