Variation in the acquisition of English stop voicing, revisited

Mary E. Beckman (Ohio State University, Jan Edwards (University of Maryland), Benjamin Munson (University of Minnesota, Minneapolis)

Introduction

- Recognizing the "need to develop new methodologies to quantify robustness of knowledge at multiple levels of abstraction and in multiple sensory domains" [1], we compare three production-based measures and a perception-based measure of the robustness of the English voicing contrast in a database of word-initial stops produced by 147 children, aged 28-39 months.
- Following Macken & Barton's landmark study of "covert contrast" [2], as well as subsequent work on acquisition of phonation contrasts [e.g., 3, 4, 5], we use voice onset time (VOT) as the acoustic cue for the production-based measures.
- We use judgments of voicing by a panel of four phonetically-trained student clinicians as the perception-based measure.

Eliciting the word productions

- Real word productions were elicited using a picture-prompted word repetition task [6] (schematized in Fig. 1).
- Stimuli were photographs and dialect-appropriate recordings of words likely to be known by a 30-month old (see Table 1).
- Two dialect-appropriate versions of the audio prompts were produced in a child-directed voice by adult female speakers of (1) a Madison (WI) area African-American English dialect and (2) a Minneapolis (MN) area Northern Cities dialect.

Figure 1. Child sees picture and hears audio prompt, and then repeats.

Table 1. Stop-initial words. These were elicited in 2 (or *4) trials per word type.

/d/	/t/	/g/	/k/
dish	tickle	give*	kitchen
dinner	teddy bear	get*	kitty
daddy	table	gum	cake
dance	tape	garbage	cat
dog*	tongue	go	candy
duck	toast	good	cup
door	tummy		car
	tooth		coat
	~~tooth~~		cookie

VOT values and production-based measures of robustness of contrast

- Recordings were tagged to demarcate an interval for the elicited production for each trial. A phonetically-trained research assistant then listened to the trials for stop-initial words and categorized each as a "stop", "affricate", or "other" manner.
- The assistant then tagged the release and the onset of periodicity for each of the (about 9500) tokens that were identified as plosives ("stop" or "affricate"), using criteria in [4] for these two landmarks – roughly, time of peak amplitude of burst and time of zero crossing in first clear glottal pulse – and VOT was then defined as the time difference (in ms) between the tags.
- Fig. 2 shows the distribution of VOT values in a representative subset of the children. (Note that only positive values are shown even though onset of periodicity was tagged as occurring before release in 1% of tokens. Since the sociophonetic status of such "prevoicing" in the two target dialects is not known, these tokens were set aside for future analysis.)

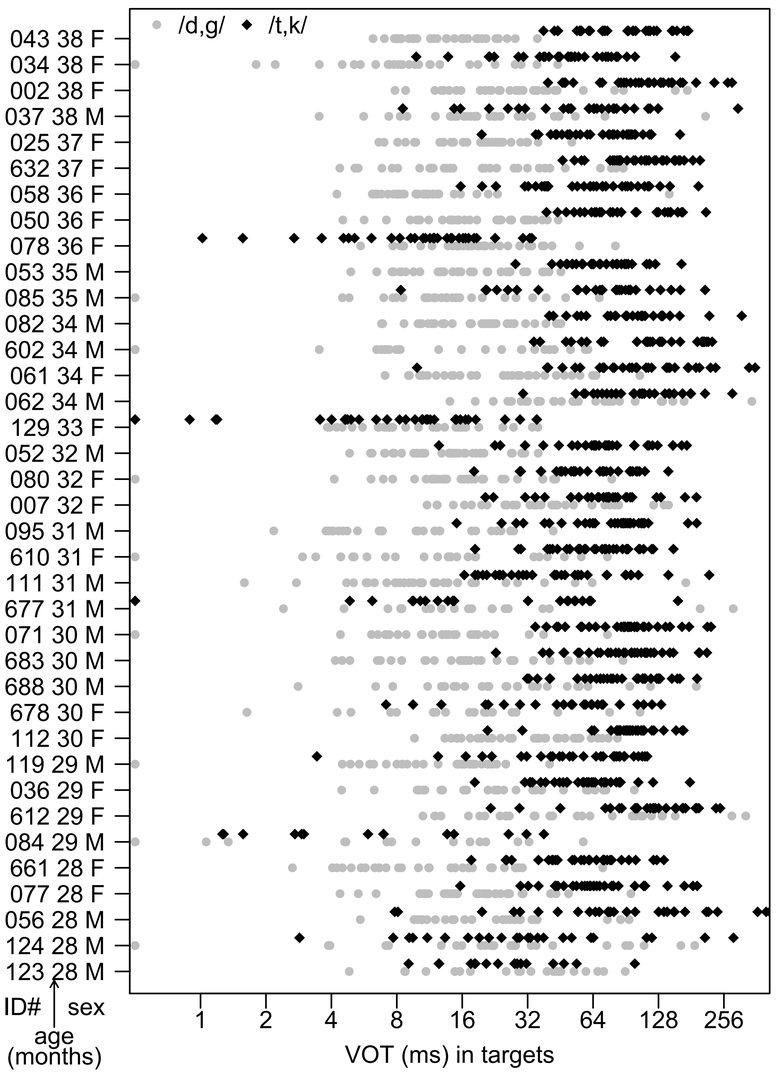

Figure 2. Distributions of VOT values for voiced and voiceless stops for a representative 25% the children, arranged by age (youngest at bottom) and within age, by the first robustness of contrast measure (Fig. 3, top).

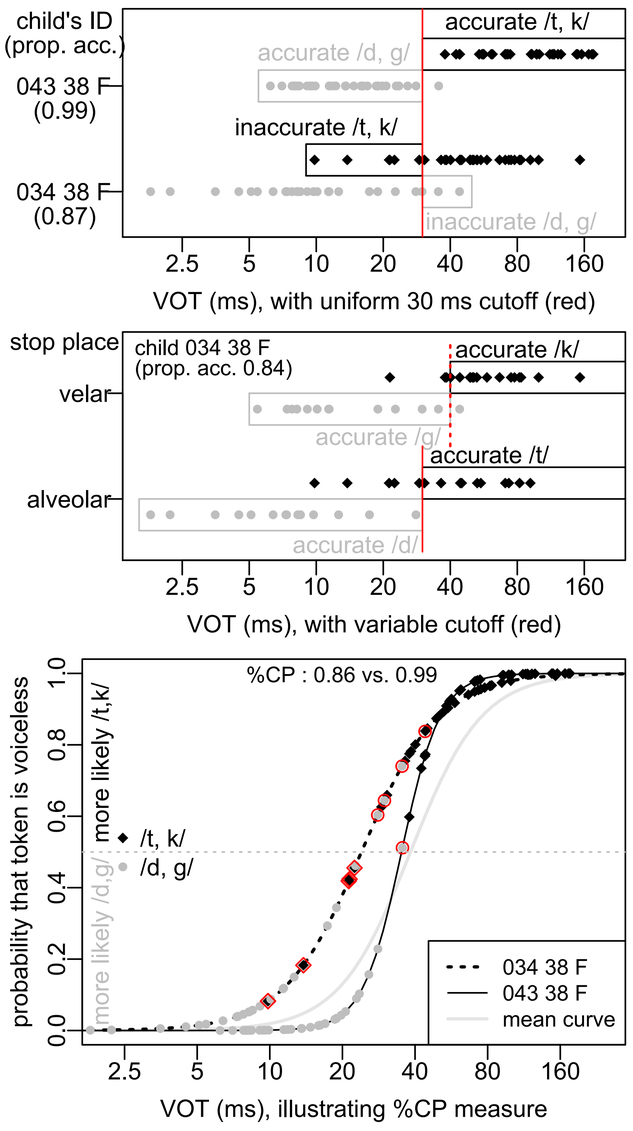

Figure 3. Calculation of the production-based measures.

- The VOT values for each child were used to calculate three measures of the robustness of the voicing contrast.
- Two of the measures (Fig. 3, top and middle panels) are modeled on the Hitchcock & Koenig's "Accuracy" measure [5].
- The other measure (Fig. 3, bottom) is modeled on the Holliday et al. "%CP" measure [7] – i.e. the proportion of tokens for which the child-specific BLUPs from a logistic mixed effects model (Eq. 1) correctly predicted the target voicing value.

Eq. 1. glmer(targetVfactor ~ logVOT.c + (1 | ID) + (0 + logVOT.c | ID), dat, family="binomial")

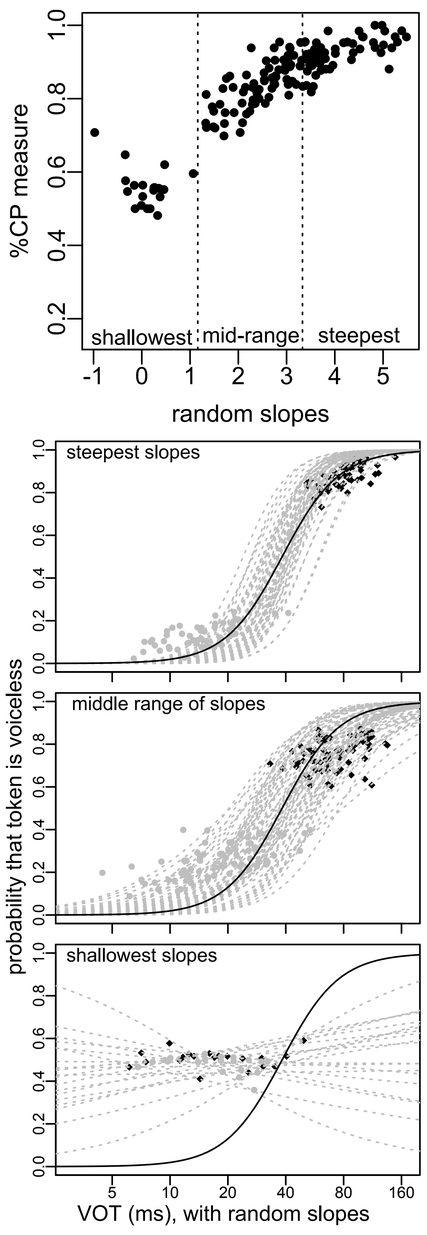

Figure 4. (top panel) Relationship of %CP to child-specific slope (see other 3 panels).

- The top panel of Fig. 4 shows how this %CP measure relates to the child-specific random slope for the VOT effect (Fig. 4, other three panels). This slope is another measure of robustness of contrast that Holliday and colleagues had explored in earlier pilot work.

Relationships among the two types of production-based measure

- We explored the differences among the production-based measures first by examining the relationships between pairs of measures (Fig. 5).

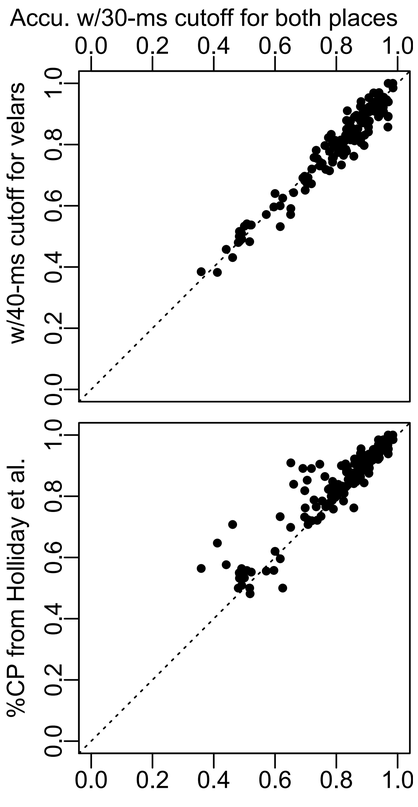

Figure 5. (top) Relationship of the Accuracy measure with a variable cutoff value to the Accuracy measure with a uniform 30-ms cutoff value. (bottom) Relationship of the Holliday et al. %CP measure to the Accuracy measure with a uniform 30-ms cutoff value .

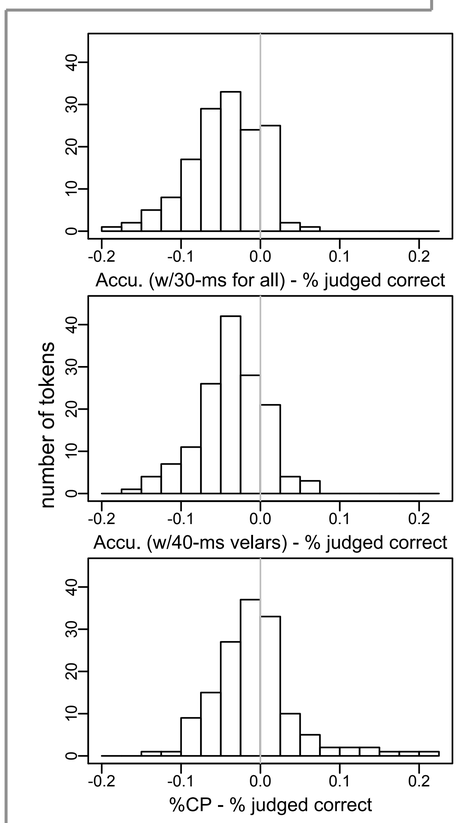

Figure 8. Distribution of differences between each of the production-based measures and the perception-based measure of robustness of contrast (see red and blue arrows in Fig. 7, right).

Relationship between VOT and transcription

- We evaluated the perception-based measure of robustness first by using a mixed effects logistic model in which the dependent variable was the transcriber's judgment of whether the consonant in the excised CV snippet was voiceless, with the VOT and the consonant place as fixed effects, and random intercepts for children and for transcribers (Eq. 2).

Eq. 2. glmer(judgment ~ 1 + logVOT.c + targetPlace + (1 | ID) + (1 | judge), dat, family="binomial")

- There was a small but significant effect of place of articulation on the perceived voicing, as expected from widely attested differences in VOT between velars and more anterior stops (Fig. 6, top panel).
- However the place effect was smaller than the differences among the four transcribers, as reflected in the random intercepts for this random effect (Fig. 6, bottom panel).

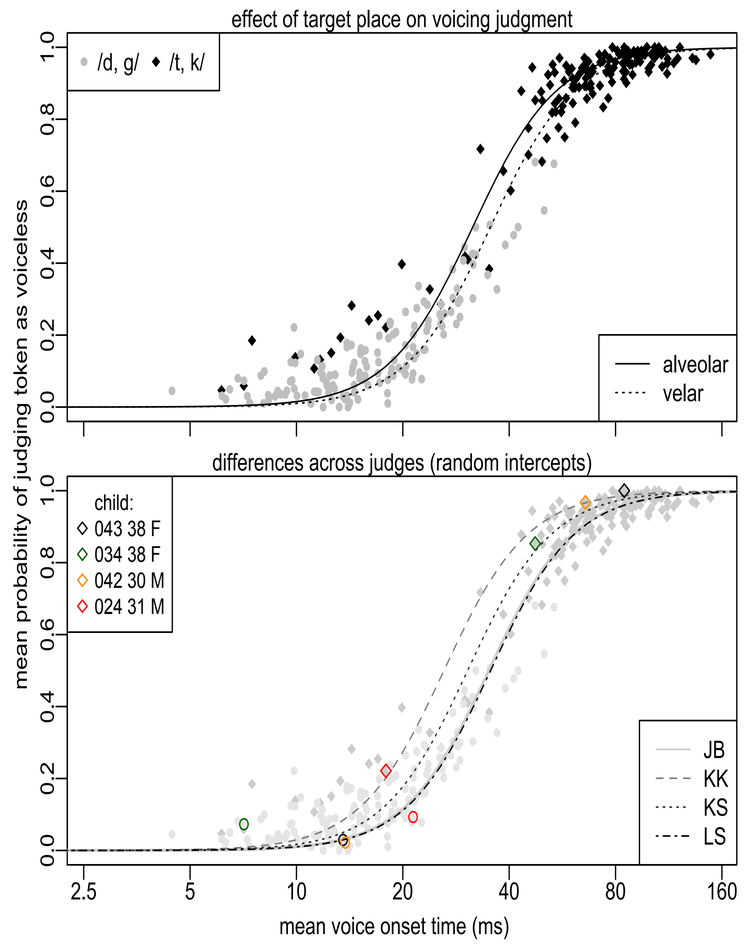

Figure 6. The circles and diamonds plot the relationship between the proportion of judgments as voiceless and the mean VOT, calculated separately by target voicing for each child participant, with overlaid curves for: (top) the fixed effects of VOT and of target place from the mixed effects logistic regression in Eq. 1, (bottom) the random intercepts for the 4 transcribers from the mixed effects logistic regression in Eq. 2. Data points in color (bottom) are for 2 older girls and 2 younger boys whose VOT values (and transcription judgments) represent a range of robustness of contrast.

Relationship between production- and perception-based measures

- We used the transcriptions to calculate a perception-based measure of robustness of the voicing contrast – namely, the proportion of judgments for each child's tokens where the tokens' voicing was transcribed as matching the target voicing.
- We also evaluated the production-based measures by examining their relationships to this perception-based measure of robustness of the voicing contrast (Figs. 7 & 8).

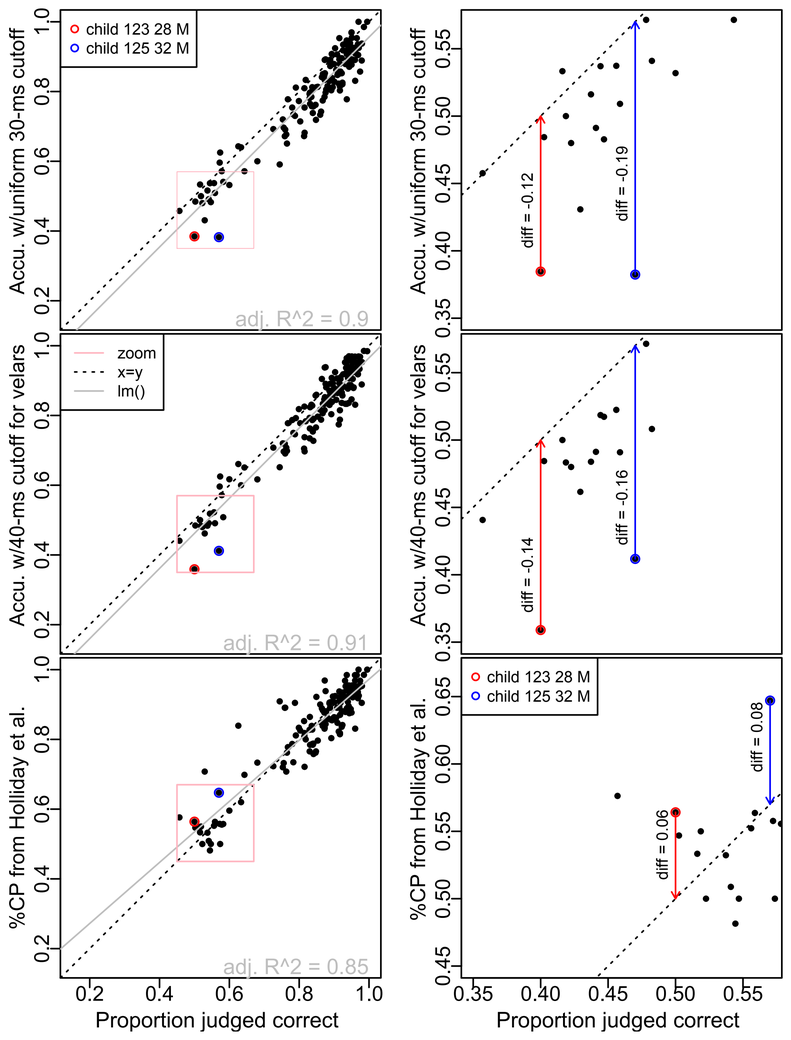

Figure 7. (left) Scatterplots showing the relationship between each of the production-based measures of robustness of contrast and the transcription-based measure. The gray lines and the gray text R^2 values in the lower right are results of linear regression models. (right) The same three scatterplots, but zooming in on the regions outlined by the red squares in the scatterplots on the left.

References

[1] Edwards, J., Munson, B., & Beckman, M. E. 2011. Frequency effects in phonological acquisition. *Journal of Child Language*, 38: 35–40.

[2] Macken, M., & Barton, D. 1980. The acquisition of the voicing contrast in English: a study of voice onset time in word-initial stop consonants. *Journal of Child Language,* 7: 41–74.

[3] Scobbie, J. M., Gibbon, F., Hardcastle, W. J., & Fletcher, P. 2000. Covert contrast as a stage in the acquisition of phonetics and phonology. In Broe, M. B., & Pierrehumbert, J. B. (Eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*. Cambridge: Cambridge Uni. Press, 194–207.

[4] Kong, E. J., Beckman, M. E., & Edwards, J. (2012). Voice onset time is necessary but not always sufficient to describe acquisition of voiced stops: The cases of Greek and Japanese. *Journal of Phonetics*, 40: 725-744.

[5] Hitchcock, E., & Koenig, L. 2013. The effects of data reduction in determining the schedule of voicing acquisition in young children. *Journal of Speech, Language, and Hearing Research*, 56: 41–457.

[6] Edwards, J., & Beckman, M. E. (2008). Methodological questions in studying consonant acquisition. *Clinical Linguistics and Phonetics*, 22(12), 937-956.

[7] Holliday, J. J., Reidy, P. F., Beckman, M. E. & Edwards, J. 2015. Quantifying the robustness of the English sibilant fricative contrast in children. *Journal of Speech, Language, and Hearing Research*, 58, 622-637.

author urls: (1,3) http://learningtotalk.org, (2) http://learningtotalk.umd.edu/
author emails: (1) beckman.2@osu.edu, (2) edwards@umd.edu, (3) munso005@umn.edu

Work supported by NIDCD grant R01 02932 to Jan Edwards, Mary E. Beckman, & Benjamin Munson.