

# A Bayesian model of voice-onset time production

Mark VanDam<sup>1</sup> & Noah H. Silbert<sup>2</sup>

<sup>1</sup>Boys Town National Research Hospital  
 mark.vandam@boystown.org www.vandamMark.com  
<sup>2</sup>Indiana University, Department of Psychological and Brain Sciences  
 159th Acoustical Society of America 2pSC13  
 Baltimore, MD 20 April 2010



## Main Point

VOT produced during a listen-and-repeat task was longer after training and after exposures to over-long tokens—but only in the voiceless series.

## Background

Evidence suggests pervasive influence of linguistic experience in language storage, speech production, and speech perception. Experiences affect linguistic memory representations or lexical storage (Pierrehumbert, 2001, Goldinger, 1996, 1998, Shockey, Sabadini, & Fowler, 2004, Bybee 2001, 2006)

Perceptual learning has shown recent exposure to specific words can affect representations (Norris, McQueen, & Cutler, 2003; Kraljic & Samuel, 2005).  
 Talkers' speech production is affected by ambient language (Sancier & Fowler, 1997) and can be affected by familiarity with specific words in conversational context (Pardo, 2006).

## Method

**Dataset:** 56 target words with alveolar and velar stops in initial position, balanced by a lexical frequency: high, low

	low freq [k]	high freq [k]
low freq [g]	call - gwik	cut - gort
high freq [k]	call - gill	cut - gort
high freq [g]	car - gar	came - game

b. context: impoverished  
 could  
 rich  
 could very clearly  
 rich

c. training condition: long, short (initially lengthened or (lightly) shortened VOT  
 VOT was 80% or 80% of natural production for long or short, respectively

d. test phase: pretest, posttest

**Task:** Produce target words from orthographic cue.  
**Participants:** 2 men, 4 women, 27.4 years old, monolingual American English talkers, normal hearing/vision, no speech or language deficits, white, right-handed.

**DE:** Each of 6 participants produced about 700 forms, yielding about 4,520 total VOTs.

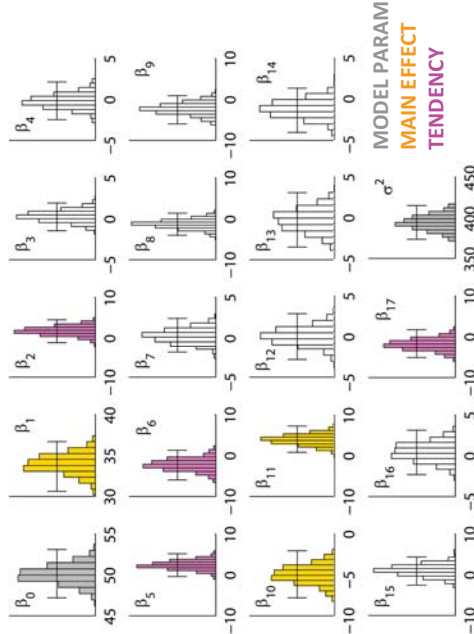


Fig. 1. Posterior distributions of model parameters. A central tendency different from zero indicates main effect and direction. Clear 'main effects' are shown in yellow and 'tendencies' are shown in pink. Model Parameters are described in Table 1.

Table 1: Model parameters, experimental factors, and effect coding scheme

Parameter	Purpose/Factor	Effect Code Values
$\sigma^2$	variance	
$\beta_0$	intercept	-1 = voiced; 1 = voiceless
$\beta_1$	voicing	-1 = low; 1 = high
$\beta_2$	lexical frequency	-1 = impoverished; 1 = rich
$\beta_3$	test phase	-1 = pre; 1 = post
$\beta_4$	training condition	-1 = short; 1 = long
$\beta_5$	voicing × frequency	1 = voiceless & high
$\beta_6$	voicing × context	1 = voiceless & rich
$\beta_7$	frequency × context	1 = high & rich
$\beta_8$	voicing × frequency × context	1 = voiceless & high & rich
$\beta_9$	voicing × phase	1 = voiceless & post
$\beta_{10}$	voicing × condition	1 = voiceless & long
$\beta_{11}$	phase × condition	1 = post & long
$\beta_{12}$	voicing × phase × condition	1 = voiceless & post & long
$\beta_{13}$	frequency × phase	1 = high & post
$\beta_{14}$	voicing × frequency × phase	1 = voiceless & high & post
$\beta_{15}$	frequency × condition	1 = high & long
$\beta_{16}$	voicing × frequency × condition	1 = voiceless & high & long
$\beta_{17}$	voicing × frequency × context	1 = voiceless & high & long

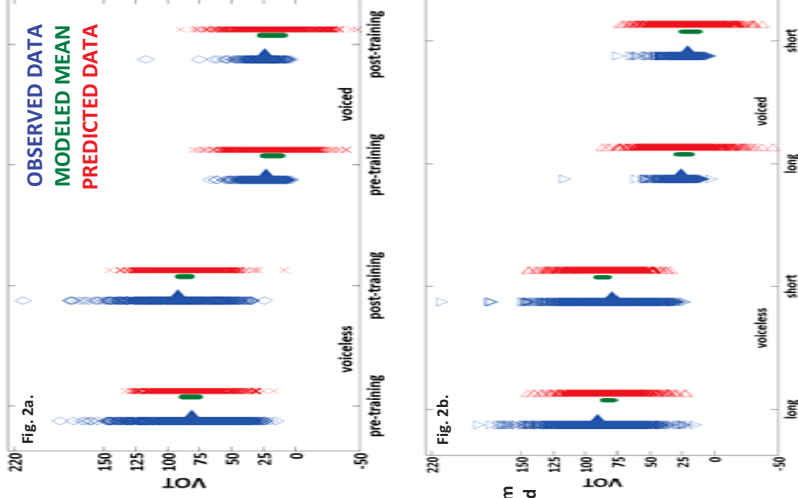


Fig. 2. Distributions and means of VOT production data for observed (blue) and modeled (green and red) data. The top panel shows voiceless VOTs were longer after training, but no change for voiced VOTs. The bottom panel shows shorter overall VOTs for those subjects trained on short exemplars.

## Conclusions

1. Voiceless stops were longer after training, but voiced stops were not (Fig. 2a).  
 → The model did a good job here.
2. Subjects trained on short VOT exemplars had shorter VOT productions (Fig. 2b).  
 → The model did not do a good job here.
3. Tendencies lean toward possible
  - a. longer VOT for high-frequency words ( $\beta_7$ ),
  - b. longer VOT when trained on long stops ( $\beta_5$ )
  - c. shorter VOT for voiceless, high-freq words ( $\beta_6$ )
  - d. shorter VOT for voiceless, high-freq, long-trained stops ( $\beta_{17}$ )

4. Asymmetry and complexity is in the fine detail of stop production, especially after specific exposure. Supports previous work on fine-grained VOT production (Miller, Green, & Reeves, 1986; Volaitis & Miller, 1992) and perception (VanDam, 2007; VanDam & Port, 2009; VanDam, Silbert, & Port, 2009).

5. The Bayesian model suggests perhaps less complexity than other data modeling such as by traditional descriptive statistics (VanDam & Port 2009) or by resampling techniques (VanDam, Silbert, & Port 2009).

Funding: NIH/NIDCD T32-DC00012, T32 DC00013-30.