Comparing Models of Unsupervised Adaptation in Speech Perception

Shaorong Yan, and T. Florian Jaeger

Brain and Cognitive Sciences, University of Rochester

UNIVERSITY of ROCHESTER

Background & Question

- Talkers differ in how they produce the same phoneme ("Lack of invariance", Liberman et al., 1967).

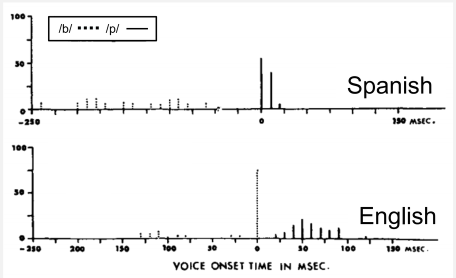


Figure 1. Differences in VOT distributions for /b/s and /p/s between English and Spanish speakers (taken from Lisker & Abramson, 1964).

- Inter-talker differences overcome in part by adaptation to individual talkers (Bradlow & Bent, 2008; Clarke & Garret, 2004; Norris et al., 2003).
- Captured by supervised learning models (ideal adapter, Kleinschmidt & Jaeger, 2015)
- Adaptation/Learning also occurs over unlabeled data (Clayards et al., 2008; Kleinschmidt et al., 2015)
- Current study:
  - Modeling **unsupervised speech adaptation**
  - What prior knowledge listeners possess

Unsupervised adaptation algorithms

Ideal adapter framework (Kleinschmidt & Jaeger, 2015)

Updated mean(μ) and variance(σ^2) of category i after observing input x at the kth trial

Probability of x being from category i

Probability of x *not* being from category i

$$\boxed{p(\mu_{c_i,k},\sigma^2_{c_i,k}\mid x)} = \boxed{p(\mu_{c_i,k},\sigma^2_{c_i,k}\mid x, C=c_i)} * \boxed{p(C=c_i\mid x)} + \boxed{p(\mu_{c_i,k-1},\sigma^2_{c_i,k-1})} * \boxed{p(C\neq c_i\mid x)}$$

Update if x is from category i

Do not update category i if x is not from category i

How do listeners adapt when there is uncertainty about the category label of the input?

- ***Baseline, No uncertainty*** **(Supervised Model)**: Only update the labeled category

$$p(C=c_i\mid x)=\begin{cases}1 & \text{if } x \text{ is labeled as category } i\\ 0 & \text{if } x \text{ is not labeled as category } i\end{cases}$$

- ***Discard uncertainty*** **(Posterior-Takes-All Model)**: Only update the more likely category

$$p(C=c_i\mid x)=\begin{cases}1 & \text{if } p(C=c_i\mid x,\mu_{c_i,k-1},\sigma^2_{c_i,k-1}) > 0.5\\ 0 & \text{if } p(C=c_i\mid x,\mu_{c_i,k-1},\sigma^2_{c_i,k-1}) < 0.5\end{cases}$$

- ***Maintain uncertainty*** **(Posterior Model):** Optimally integrate uncertainty about labels (i.e., fully Bayesian)

$$p(C=c_i\mid x)=p(C=c_i\mid x,\mu_{c_i,k-1},\sigma^2_{c_i,k-1})$$

- ***Utmost uncertainty*** **(Half-Half Model)**: Always update both categories

$$p(C=c_i\mid x)=p(C\neq c_i\mid x)=0.5$$

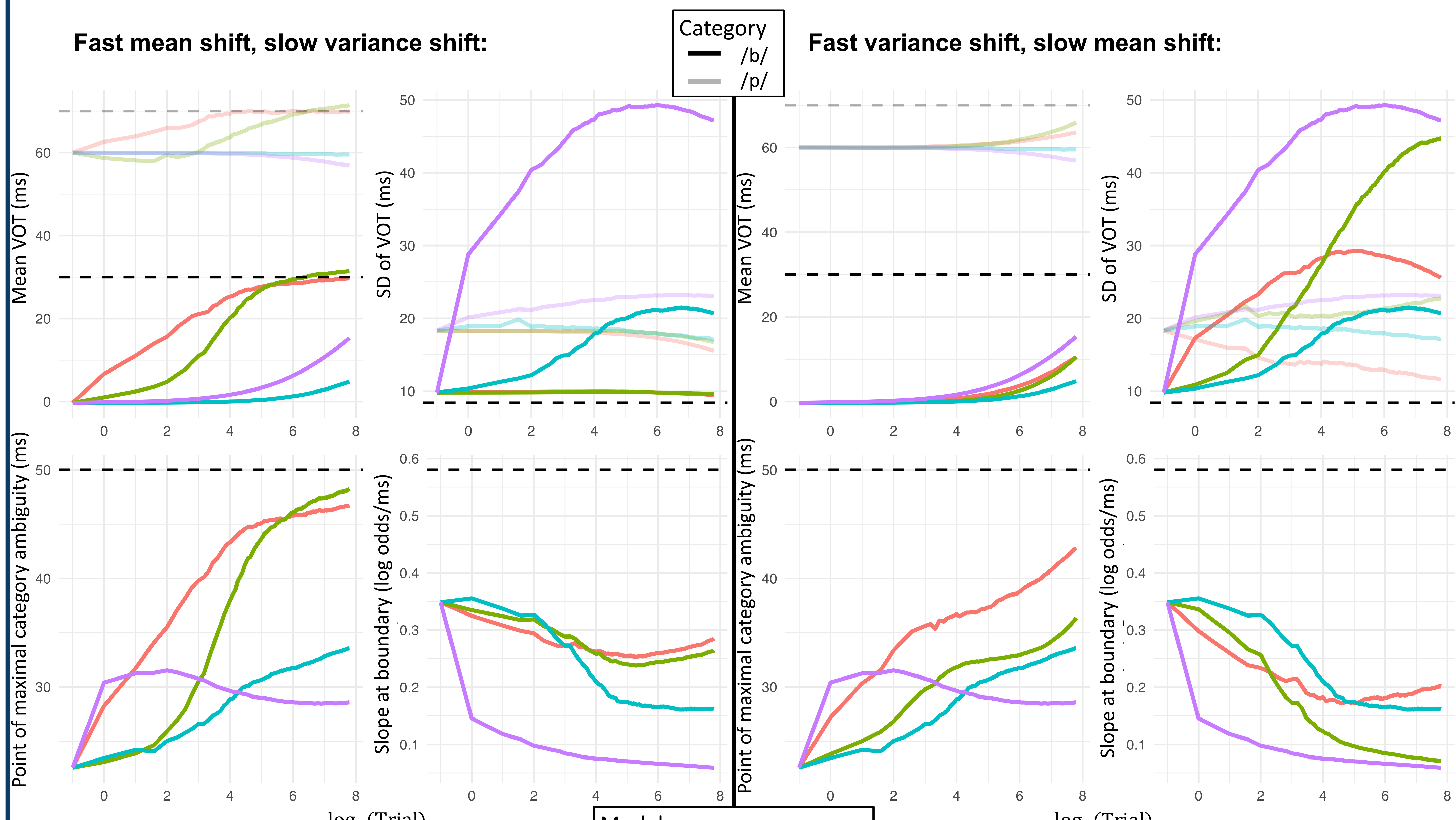
Discussion

- Our unsupervised models provide good fit to human behavior data
- Posterior model that maintains uncertainty has overall better fit than Posterior-Takes-All Model that discards uncertainty (in contrast with McMurray, Aslin & Toscano, 2009).
- But Posterior-Takes-All Model seems to have better fit to change in slope.
- **Future directions**:
  - Individual differences in prior knowledge
  - Covariance across categories (Chodroff et al., 2015)
  - Normality of distributions → sampling based update
  - Unsupervised adaptation with algorithmic models? e.g., particle filter (Kleinschmidt, 2018)

Predicted patterns of adaptation

Fast mean shift, slow variance shift:

Fast variance shift, slow mean shift:



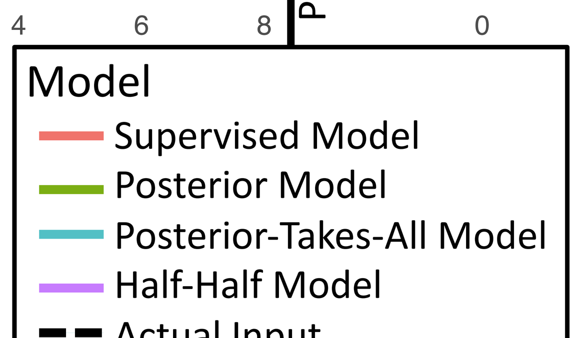
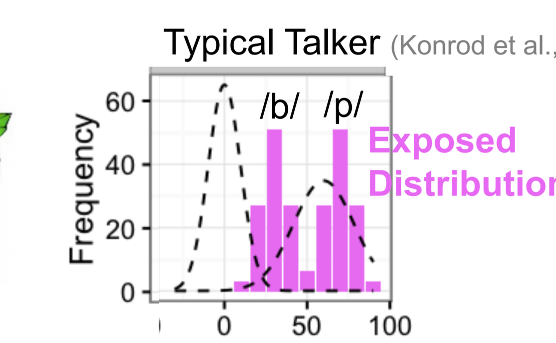


Figure 2. Predicted changes in VOT distributions and boundary (κ = 1, υ = 200).

Figure 3. Predicted changes in VOT distributions and boundary (κ = 200, υ = 1).

Behavioral dataset

Example trial



Input statistics

VOT Distribution (ms)	/b/ Mean (SD)	/p/ Mean (SD)
Typical	-0.3 (9.8)	60 (18.3)
Exposed	30 (8.4)	70 (8.4)

Model fit against human categorization

Overall model fit

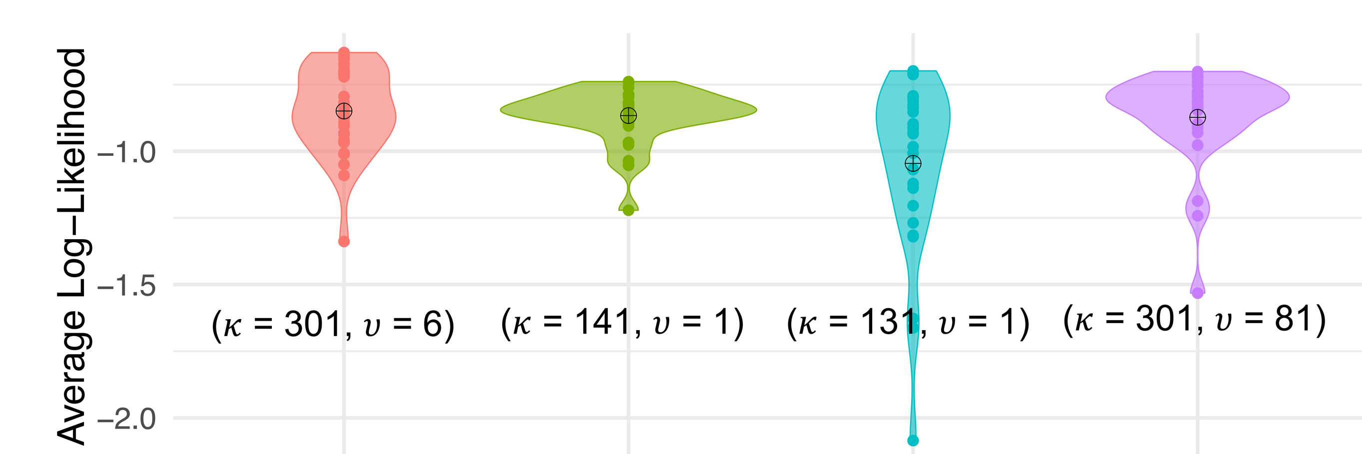


Figure 4. Model fit (each point is a subject) for best priors (combination of κ and υ).

Priors

- Means and variances: based on production data (Kronrod et al., 2012)
- Strength of prior for mean (κ) and variance (υ) determined through grid search

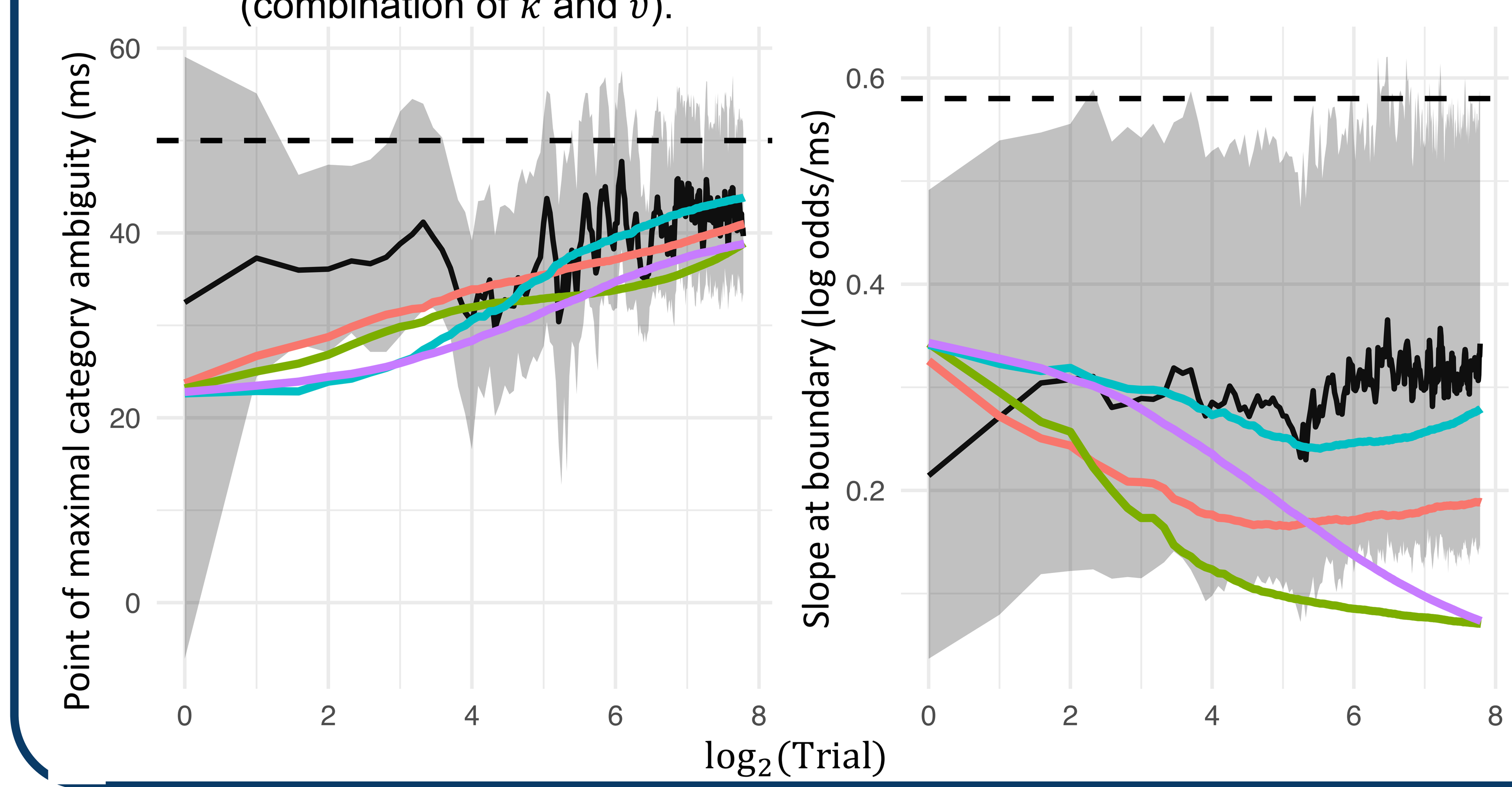


Figure 5. Predicted change in boundary for best fitting combinations of κ and υ.