Supervised and unsupervised learning in phonetic adaptation

Dave F. Kleinschmidt, Rajeev Raizada, and T. Florian Jaeger, University of Rochester
dkleinschmidt@bcs.rochester.edu, raizada@bcs.rochester.edu, fjaeger@bcs.rochester.edu

Our question:

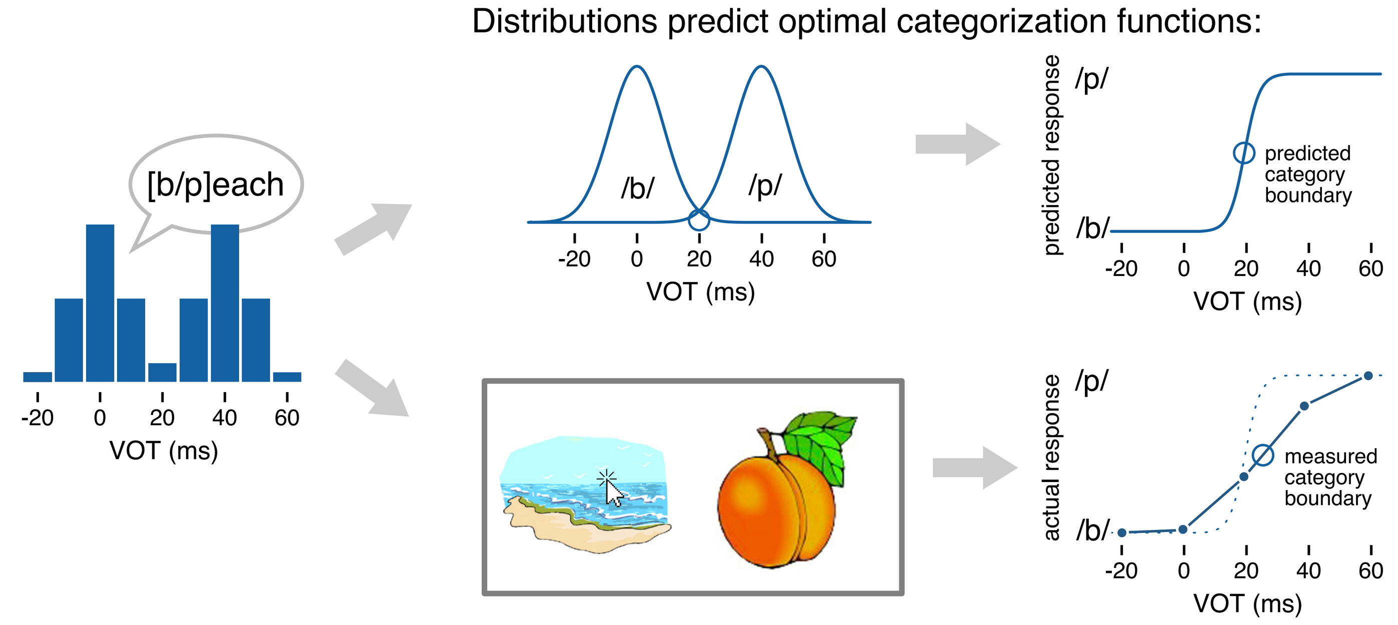
Do people use **category labels** during adaptation?

Language learning doesn't stop once you reach adulthood: talkers use linguistic cues to realize their intentions in different ways. To **adapt** to a new talker, you have to learn the way they use cues. If you know their intented meaning, this learning should be a lot easier. Learning with known category labels is called **supervised learning**, and learning from cues only is called **unsupervised learning**.

Distributional learning

Listeners hear different distributions of VOTs in /b/-/p/ minimal pair words

Distributions predict optimal categorization functions:



Listeners classify each word by clicking on a matching picture.

Learning is measured by comparing listeners' actual categorizations with the optimal categorization

With and without labels

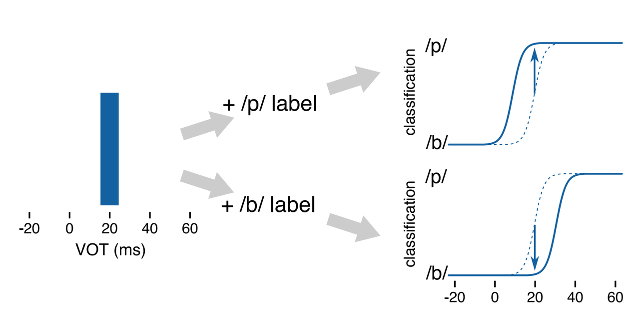
Unsupervised learning: all trials are **unlabeled**

(Semi-)supervised learning: some trials are **labeled**

Unlabeled: /b/ and /p/ response options are minimal pair. VOT is ambiguous between /b/ and /p/

Labeled: /b/ and /p/ response options are non-minimal pair. Rest of word labels VOT as /b/ or /p/.

Background: Phonetic adaptation has been observed in both forms



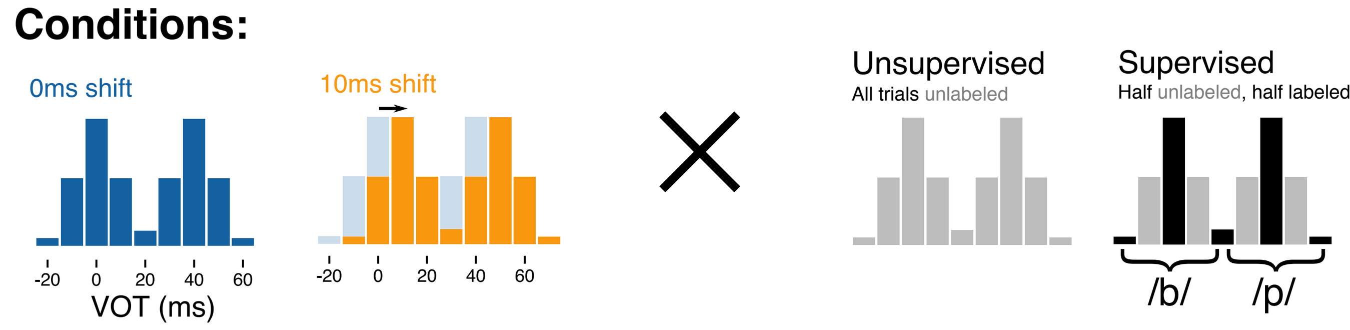
Supervised: Recalibration/perceptual learning [Bertelson et al. 2003, Norris et al., 2003, Kraljic & Samuel, 2005]. Ambiguous /b/-/p/ with visual or lexical information that consistently labels it. If labeled as a /b/, later classify more of a /b/-/p/ continuum as /b/, and vice-versa

Unsupervised: Distributional learning [Clayards et al., 2008; Munson, 2011]. Hear /b/-/p/ minimal pair words randomly drawn from bimodal distribution on /b/-/p/ continuum. Classification of continuum changes to reflect clusters in distribution.

Real life adaptation is generally a mix, some labeled data and some not. **Can listeners use some labeled data to improve learning from unlabeled data?**

Experiment 1 Basic test of semi-supervised distributional learning

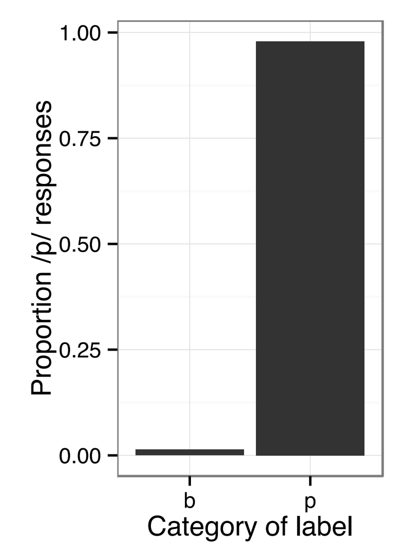
Conditions:

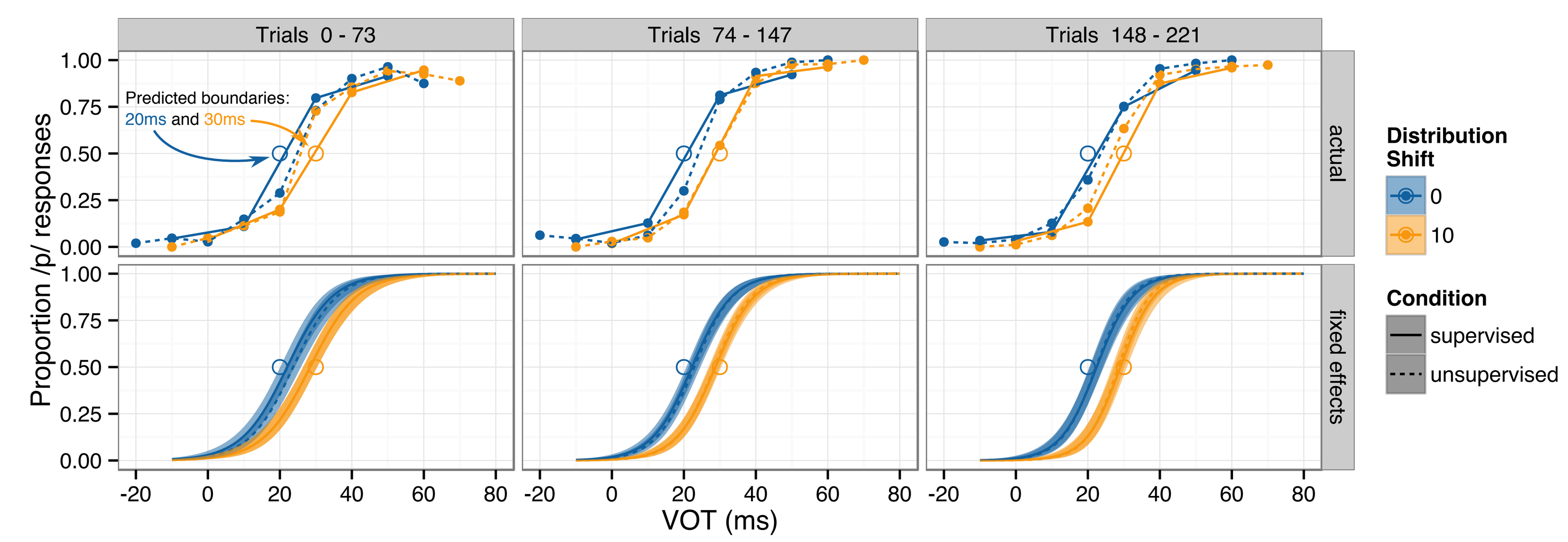


Labeled trials:
Listeners do **use labels** to guide responses on labeled trials

Unlabeled trials:
Listeners **adapt**: classification depends on distribution (color), is close to optimal (open circles), and boundary gets steeper with exposure.

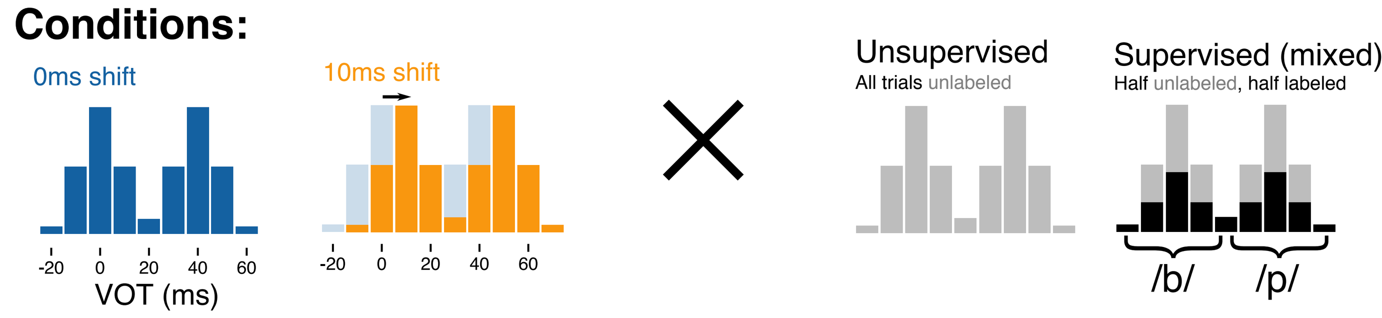
But supervised learning is no different than unsupervised (dashed vs. solid lines)





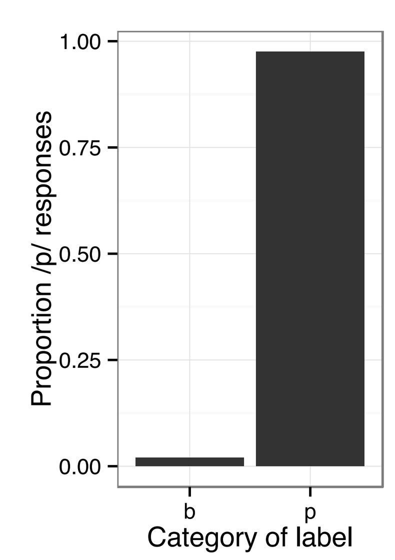
Experiment 2 Spread labeled trials out more evenly

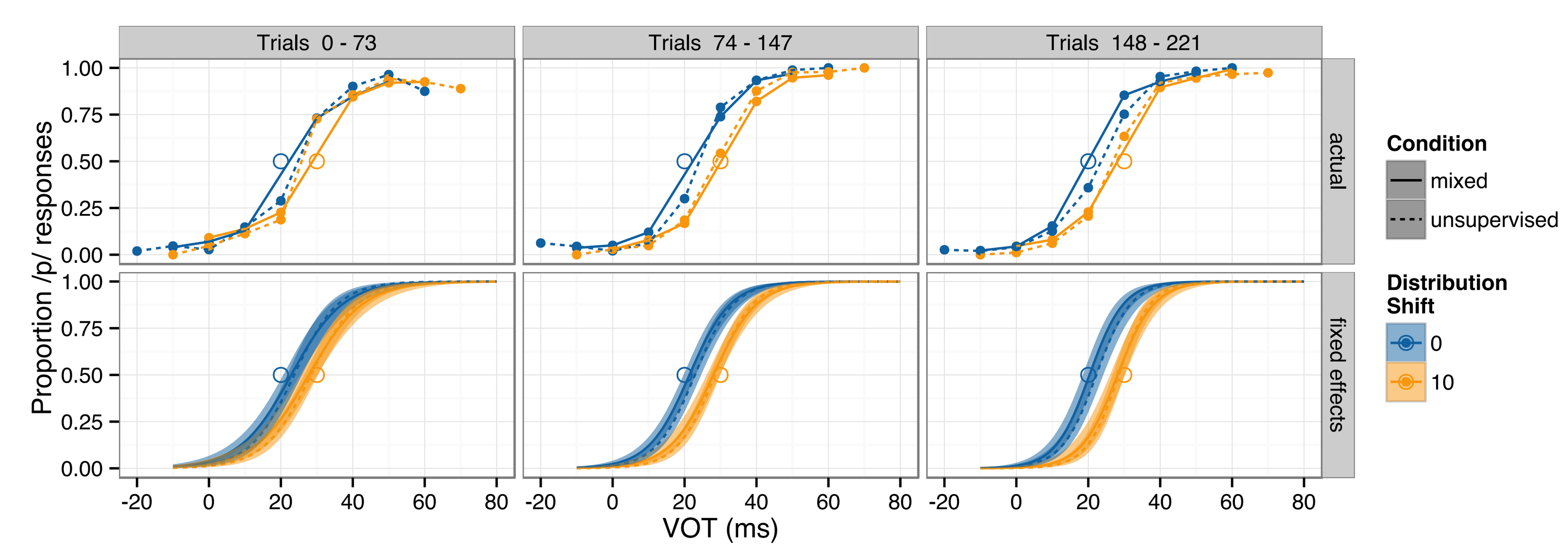
Conditions:



Labeled trials:

Unlabeled trials: Still no advantage for supervised learning

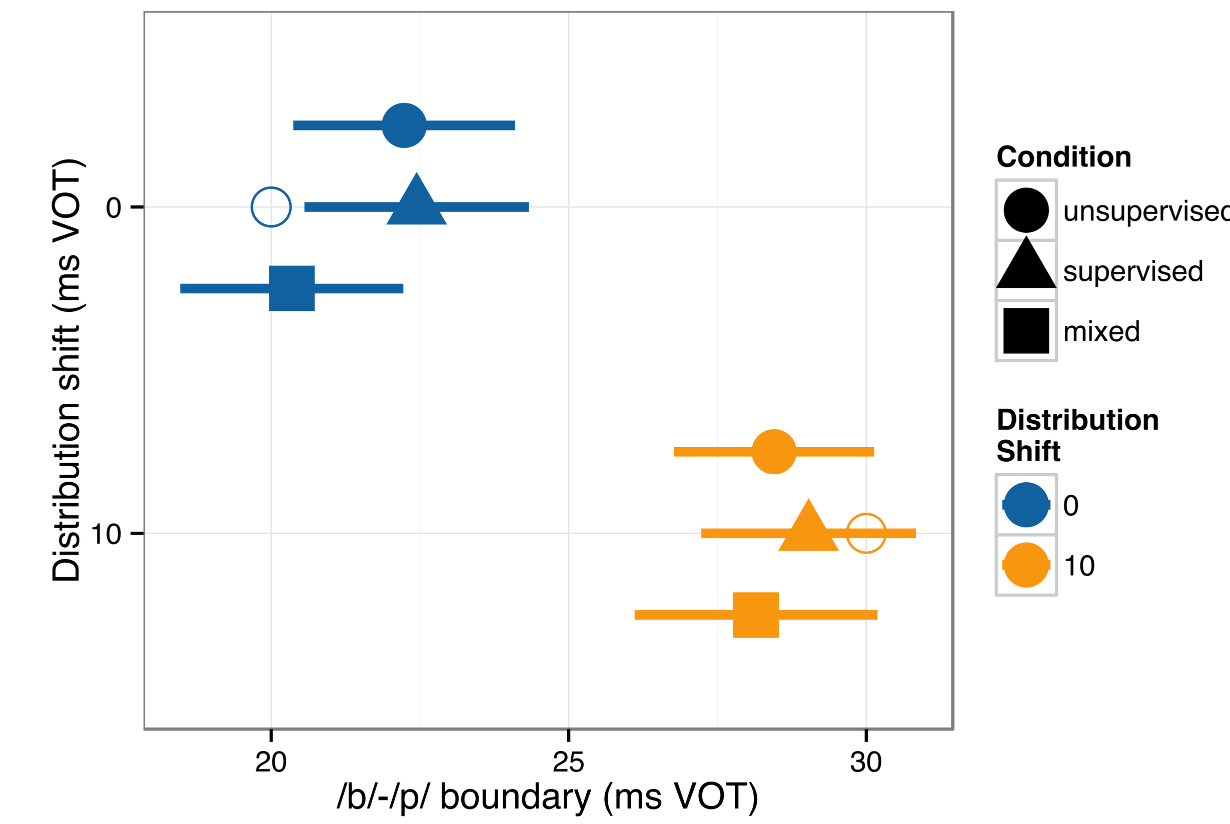




Methods details: Analyzed data from 172 subjects recruited on Mechanical Turk. Each subject was randomly assigned to a supervision condition (unsupervised, supervised, or mixed) and a distribution condition (0ms and 10ms). There were an average of 29 per cell (26 to 31). Each subject got 222 trials drawn from the appropriate distribution, with three minimal pairs (beach/peach, bees/peas, beak/peak).

Fit a logistic GLMM with fixed effects of trial, VOT, condition (unsupervised, supervised, or mixed), and distribution (0ms or 10ms shift), and the maximal random effects structure (random intercepts and slopes for trial and VOT by subject). Predictors were appropriately centered and scaled or sum-coded before fitting. Estimated category boundaries from the fixed effects coefficients, and for visualization computed their standard errors based on the fixed effects variance-covariance matrix (not taking into account random effects).

Summary: Category boundaries



Listeners' category boundaries **reflect the distributions** they heard. But they **don't differ between unsupervised and semi-supervised learning**.

Conclusions

Surprisingly, **category labels did not make adaptation faster or better**, even though they were used in classification.

Two possible reasons why:

1) Other studies use **intrinsic labels** (lexical or audio-visual cues). Labels that aren't part of the speech signal might not be available for adaptation.

2) **Informativity** of labels. Unlabeled trials contain a lot of distributional information and listeners have lots of prior experience. Labels might not add that much more.

Acknowledgements:
Thanks to members of the HLP Lab for feedback on this work. This work was supported by an NSF GRFP to DFK and NICHD R01 HD075797 to TFJ.

References:
Bertelson, P., Vroomen, J., & de Gelder, B. (2003). Visual recalibration of auditory speech identification: a McGurk aftereffect. Psychological Science, 14(6), 592–597.
Clayards, M. A., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. Cognition, 108(3), 804–9.
Munson, C. M. (2011). *Perceptual learning in speech reveals pathways of processing*. (Unpublished doctoral dissertation). University of Iowa.
Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? Cognitive Psychology, 51(2), 141–78.
Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. Cognitive Psychology, 47(2), 204–238