

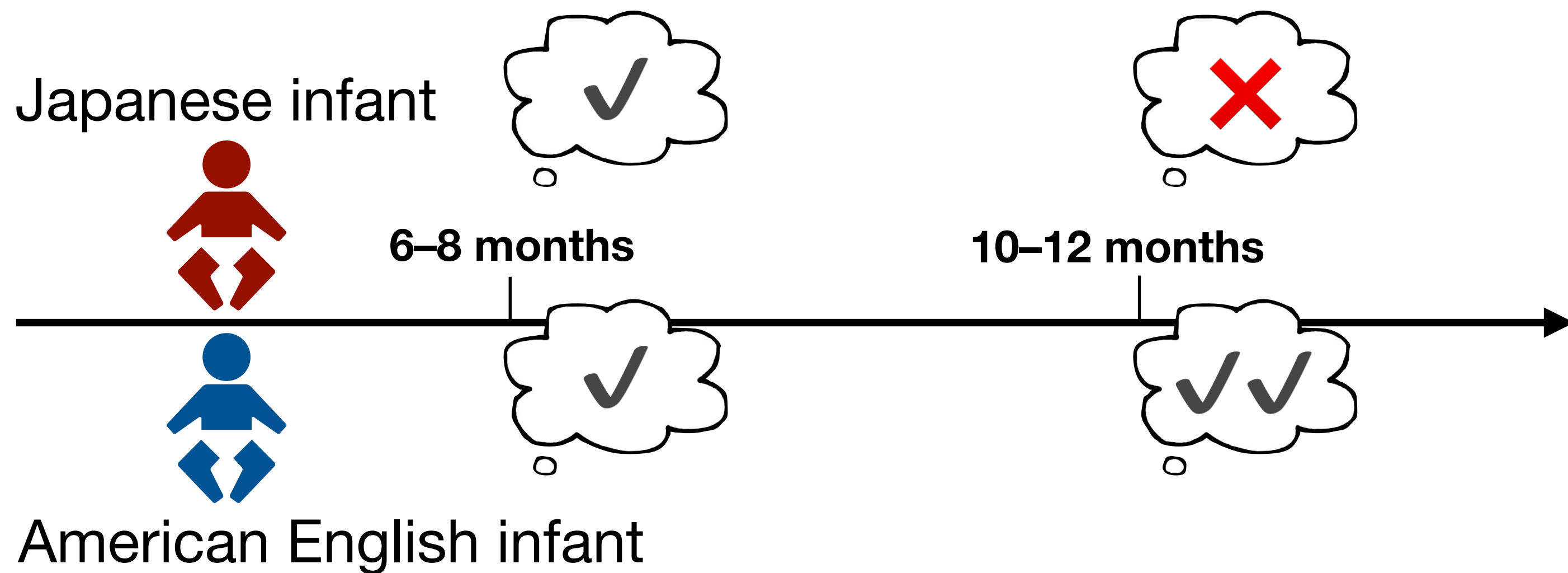
# Modeling early phonetic learning: The effect of input size and speaker distribution



Ruolan Li<sup>1,2</sup>, Thomas Schatz<sup>1</sup>, Yevgen Matuselych<sup>3</sup>, Sharon Goldwater<sup>3</sup>, and Naomi H. Feldman<sup>1</sup>  
<sup>1</sup> University of Maryland    <sup>2</sup> University of Rochester    <sup>3</sup> University of Edinburgh

## Introduction

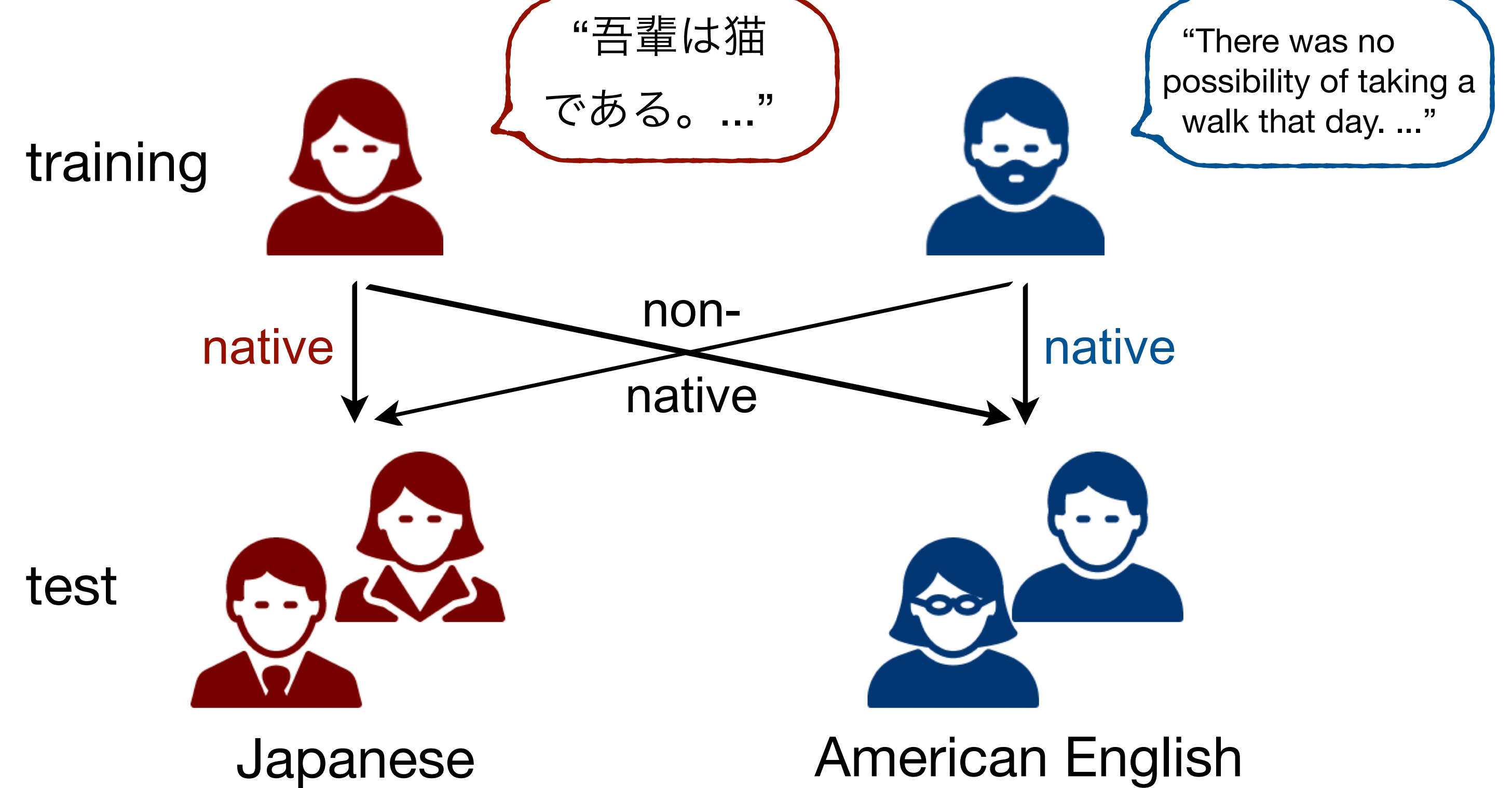
**Early phonetic learning:** discrimination of [l]-[ɹ]<sup>1</sup>



### Research Question

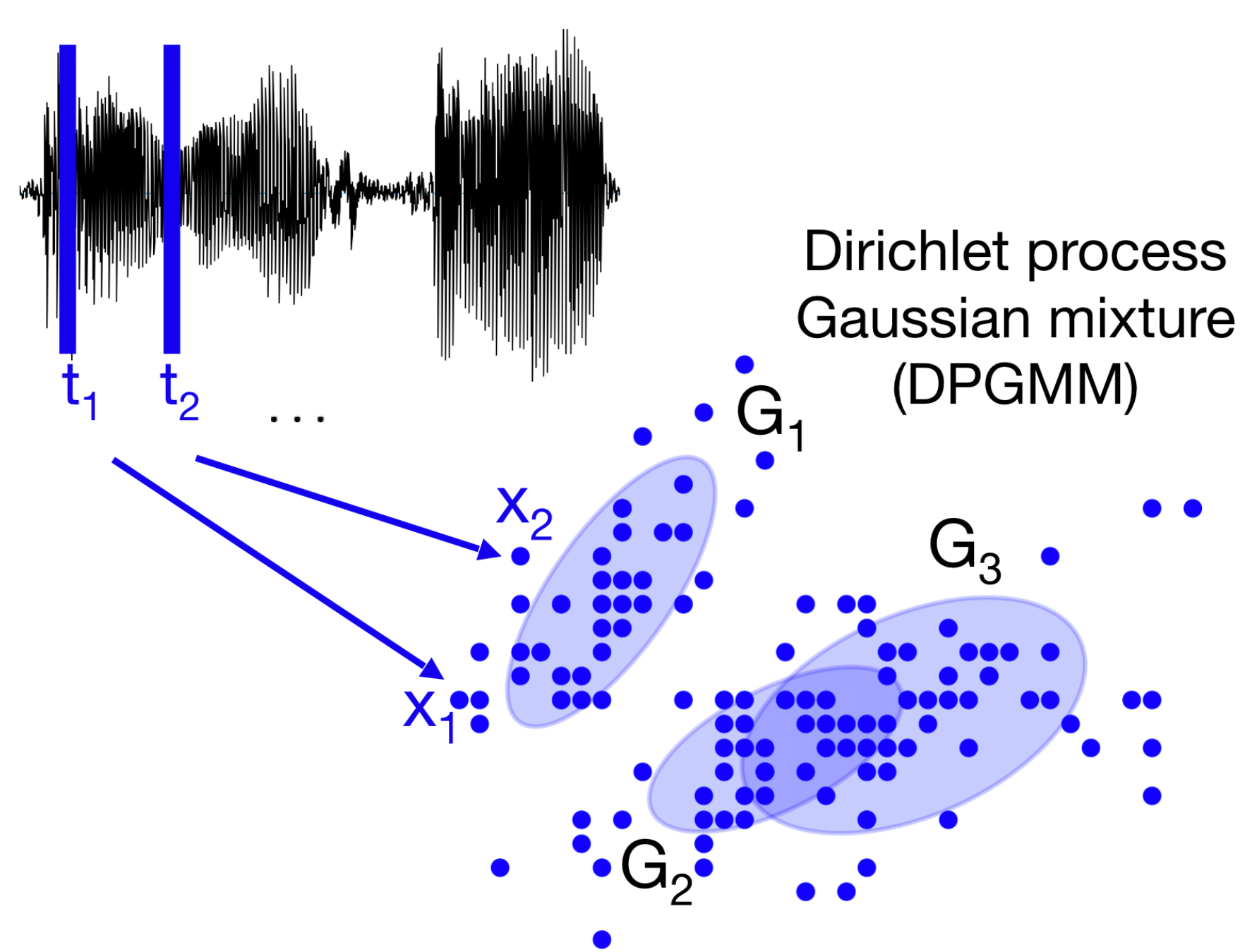
How does early phonetic learning interact with various possible ecological conditions?

**Method:** train DPGMM model<sup>4</sup> on audiobook speech

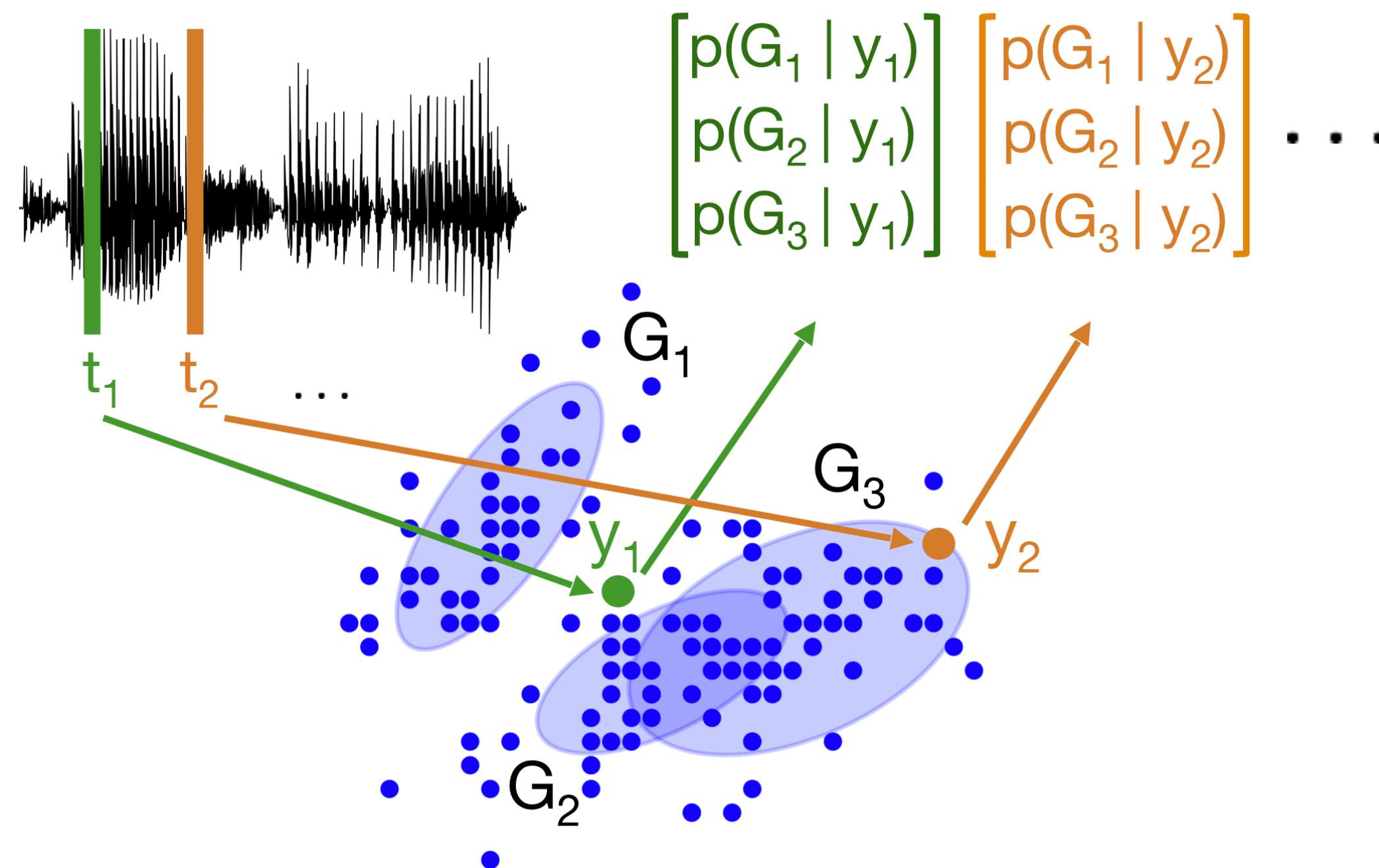


## Model specification<sup>4</sup>

**Training:** Maximize  $p(G_1 G_2 G_3 | x_1 \dots x_T)$

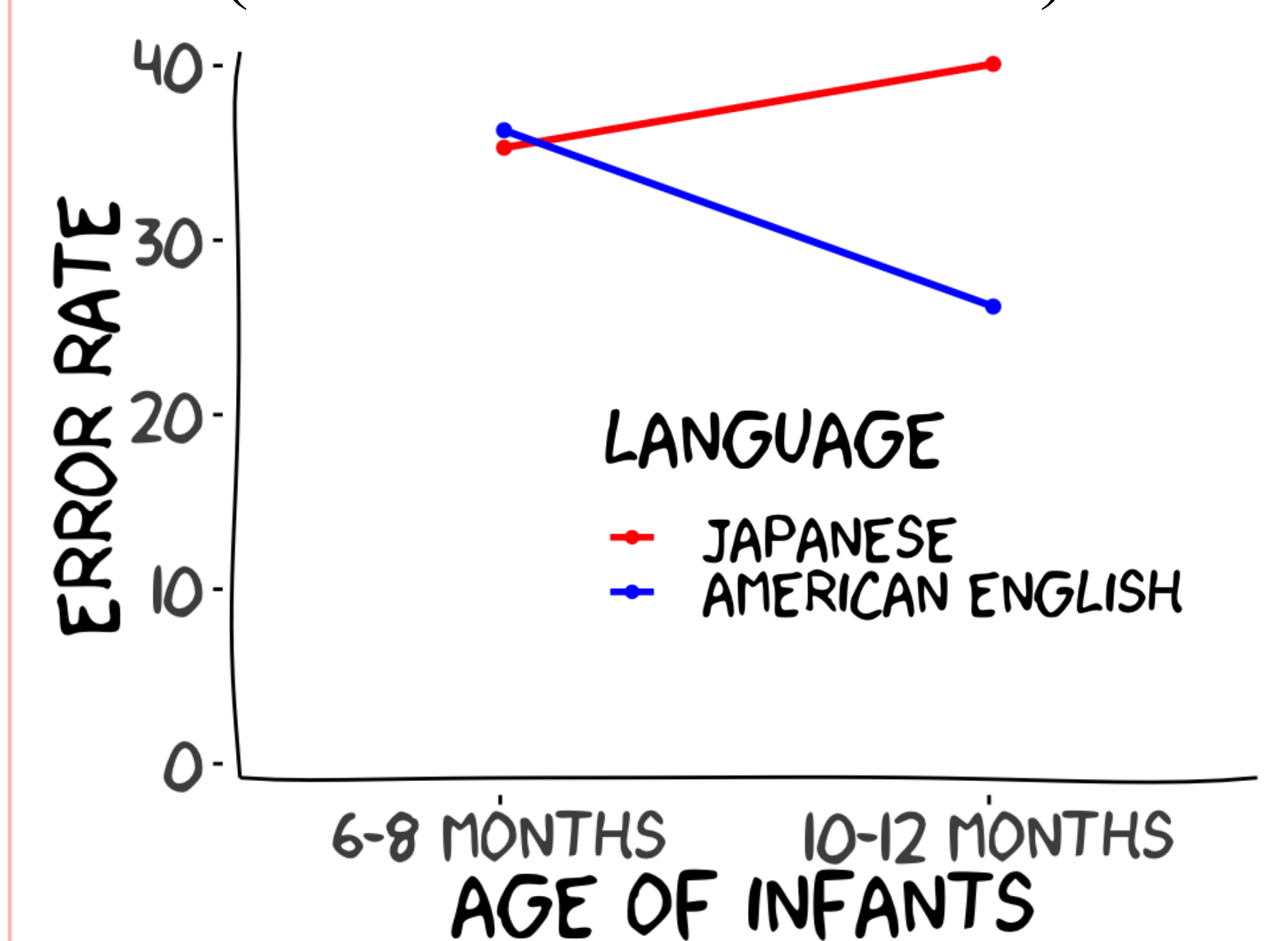


**Test:** Extract DPGMM posteriors



**Prediction**

(based on human data<sup>1</sup>)



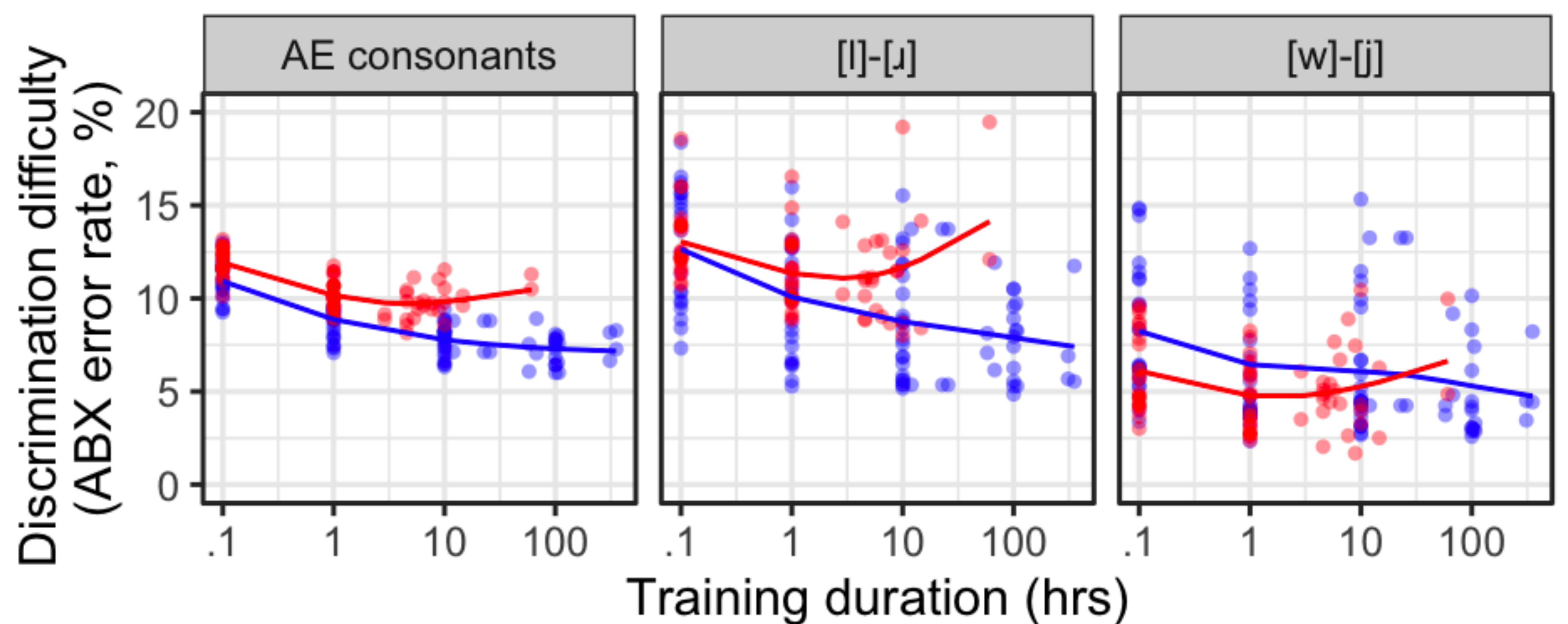
## Results

**ABX error rate vs. training duration:**

- Japanese models *rebound* after initial decrease in error rate, similar to observation in infants.<sup>1</sup>

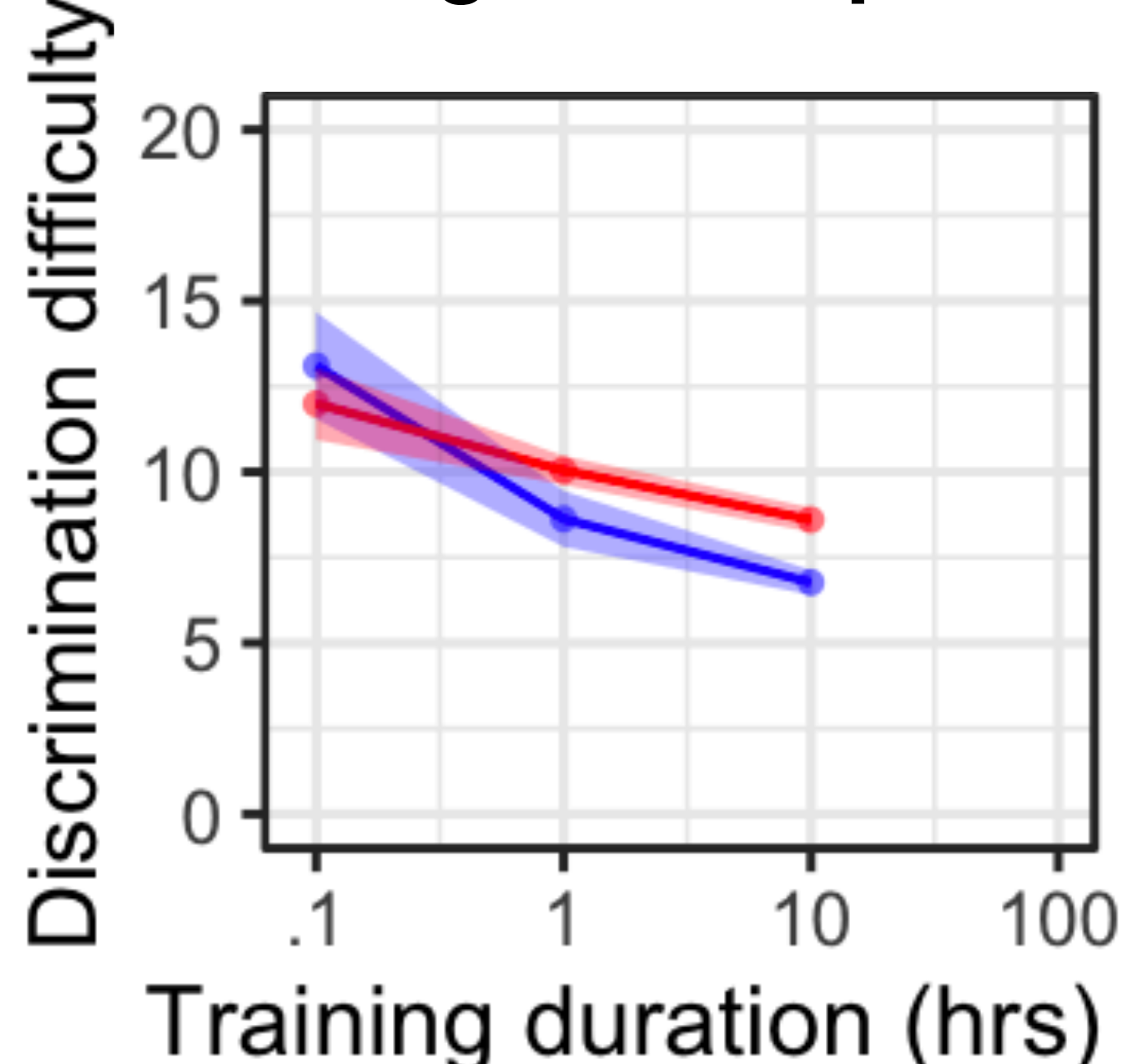
Training language:  
● Japanese  
● American English

Note: each point represents the average of 1–10 models trained on separate recordings from one single speaker (depending on the data available for that speaker).



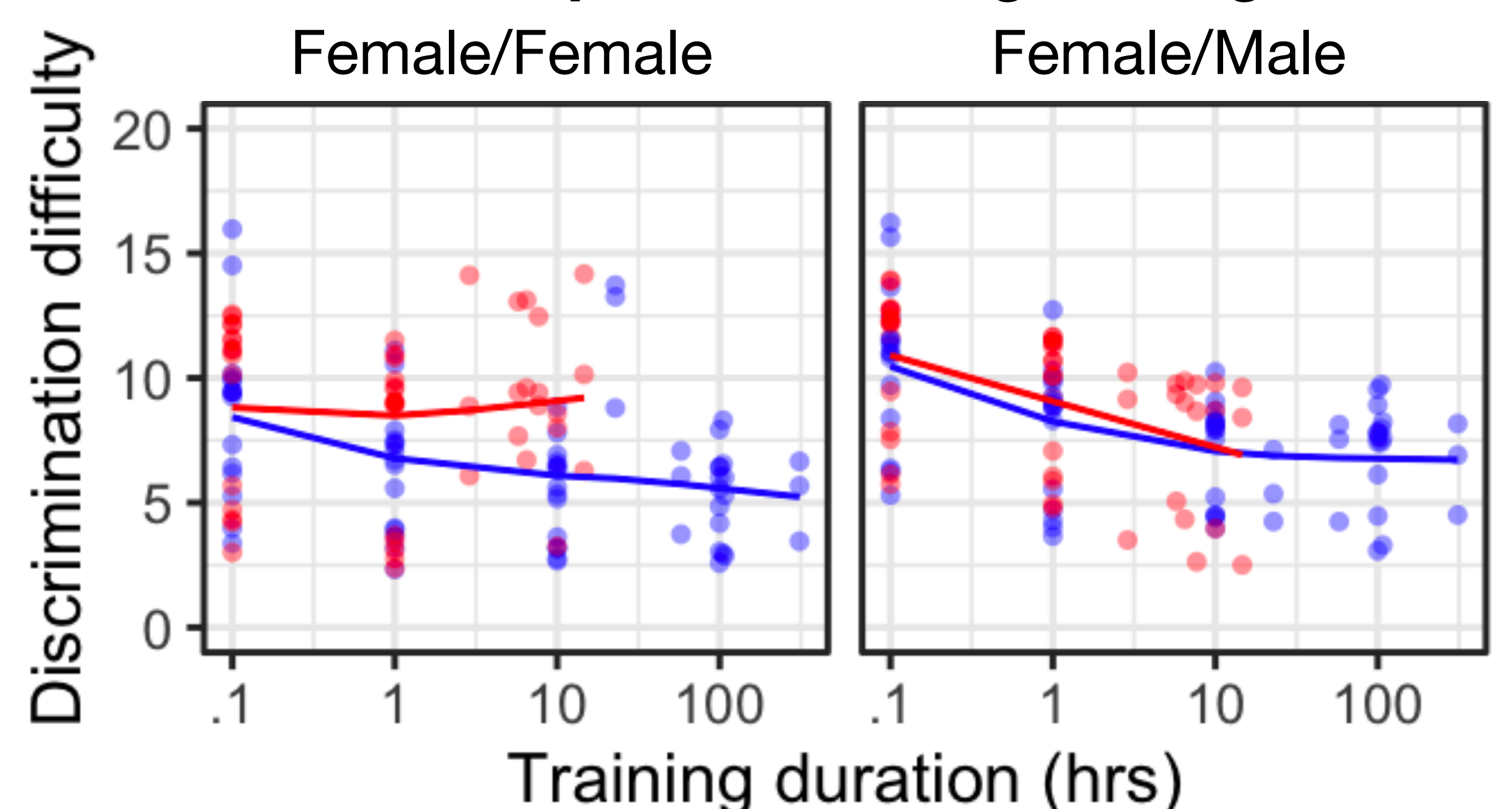
## Additional findings on [l]-[ɹ]

**Training: Multi-speaker**



- Multi-speaker training results: between matching- and mismatch- gender results.
- Matching gender replicates empirical findings & is closer to what infants hear<sup>1,5</sup>

**Gender-specific training/testing:**



## References & Acknowledgement

[1] Patricia K. Kuhl, Erica Stevens, Akiko Hayashi, Toshisada Deguchi, Shigeru Kiritani, & Paul Iverson. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, 9, 13–21.

[2] Schatz, T., Peddinti, V., Bach, F., Jansen, A., Hermansky, H., & Dupoux, E. (2013). Evaluating speech features with the minimal-pair ABX task: Analysis of the classical MFC/PLP pipeline. *Proceedings of the*

Annual Conference of the International Speech Communication Association, INTERSPEECH, 1781–1785.

[3] Thomas Schatz. ABX-Discriminability Measures and Applications. *Cognitive science*. Université Paris 6 (UPMC), 2016. English. f1tel-01407461

[4] Schatz, T., Feldman, N. H., Goldwater, S., Cao, X., & Dupoux, E. (2019). Early phonetic learning without phonetic categories – Insights from machine learning. <https://doi.org/10.31234/OSF.IO/FC4WH>

[5] Bergelson, E., Casillas, M., Soderstrom, M., Seidl, A., Warlaumont, A. S., & Amaturi, A. (2019). What Do North American Babies Hear? A large-scale cross-corpus analysis. *Developmental Science*, 22(1), 1–12. <https://doi.org/10.1111/desc.12724>

This research is supported by NSF BCS-1734245, ESRC ES/R006660/1.