Evaluating the learnability of vowel categories from Infant-Directed Speech

Jahnavi Narkar[1], Ekaterina A. Khlystova[1], Connor J. Mayer[2], Ann Aly[3], Ji Young Kim[4], Megha Sundara[1]
[1]Department of Linguistics, UCLA; [2]Department of Language Science, UC Irvine; [3]Tech Flow, Cape Coral, FL, USA; [4]Department of Spanish and Portuguese, UCLA

BACKGROUND

- Hyper-articulation – increased distance between centroids of vowels – in infant-directed speech (IDS) is thought to facilitate acquisition (e.g., Trainor & Desjardins, 2002; Liu et al, 2005).
- But vowels in IDS are also more variable (Cristia & Seidl, 2014; Martin et al, 2015; Ludusan et al. 2021)

ALTERNATIVE APPROACH

- Evaluate ***distributional overlap***
  - By combining category ***distance*** and ***variability***
  - Measures used extensively in socio-phonetics and machine learning (e.g., Hay, Warren & Drager, 2006; Kelly & Tucker, 2020)
- Independently ***test learnability*** via previously implemented Gaussian learner (Feldman et al., 2013)

- Two predictions of a facilitation account:
  (1) Vowels in IDS have less-overlapping distributions
  (2) Extracting vowel categories from less overlapping distributions is easier

METHODS

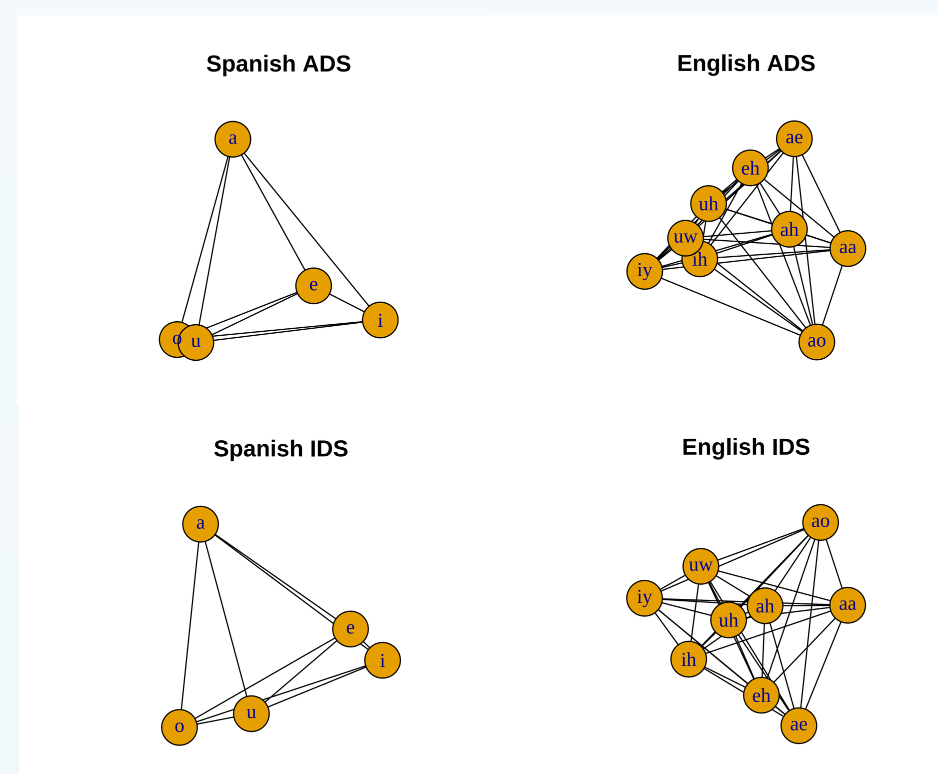
- Four connected speech corpora analyzed:
  - English IDS: Providence Corpus (Demuth et al. 2007; ~ 20K tokens)
  - English ADS: Buckeye Corpus (Pitt et al. 2007; ~20K tokens)
  - Spanish IDS: adult-child dyads recorded in lab (Sundara et al. 2020; ~5K tokens)
  - Spanish ADS: adult Spanish speakers (Kim & Repiso-Puigdelliura 2021; ~5K tokens)
- Extracted F1, F2, F3 & duration in Voicesauce (Shue et al., 2011)
- **Indexing overlap between categories**:
  (a) Pillai scores (0 = complete overlap; 1 = no overlap e.g., Hay et al. 2006)
  (b) KL divergence - machine learning statistic to measure the difference between 2 distributions (0 = complete overlap; larger number = less overlap)
- **Extracting vowel categories**: Bayesian model of distributional learning (Feldman et al., 2013)

RESULTS

Do vowel categories in IDS have less overlap than in ADS?

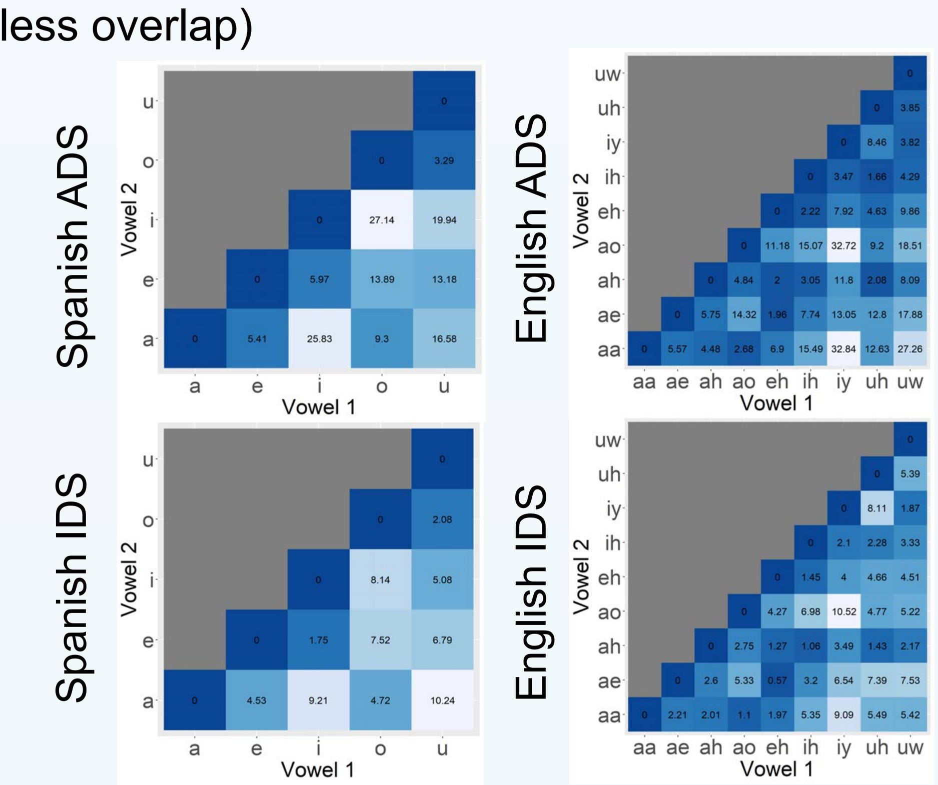
Pillai scores

- Pillai scores to generate dissimilarity metric for vowel pairs in IDS and in ADS
- 2-D Multi-Dimensional Scaling (MDS) solution to visualize dissimilarity space



KL divergence

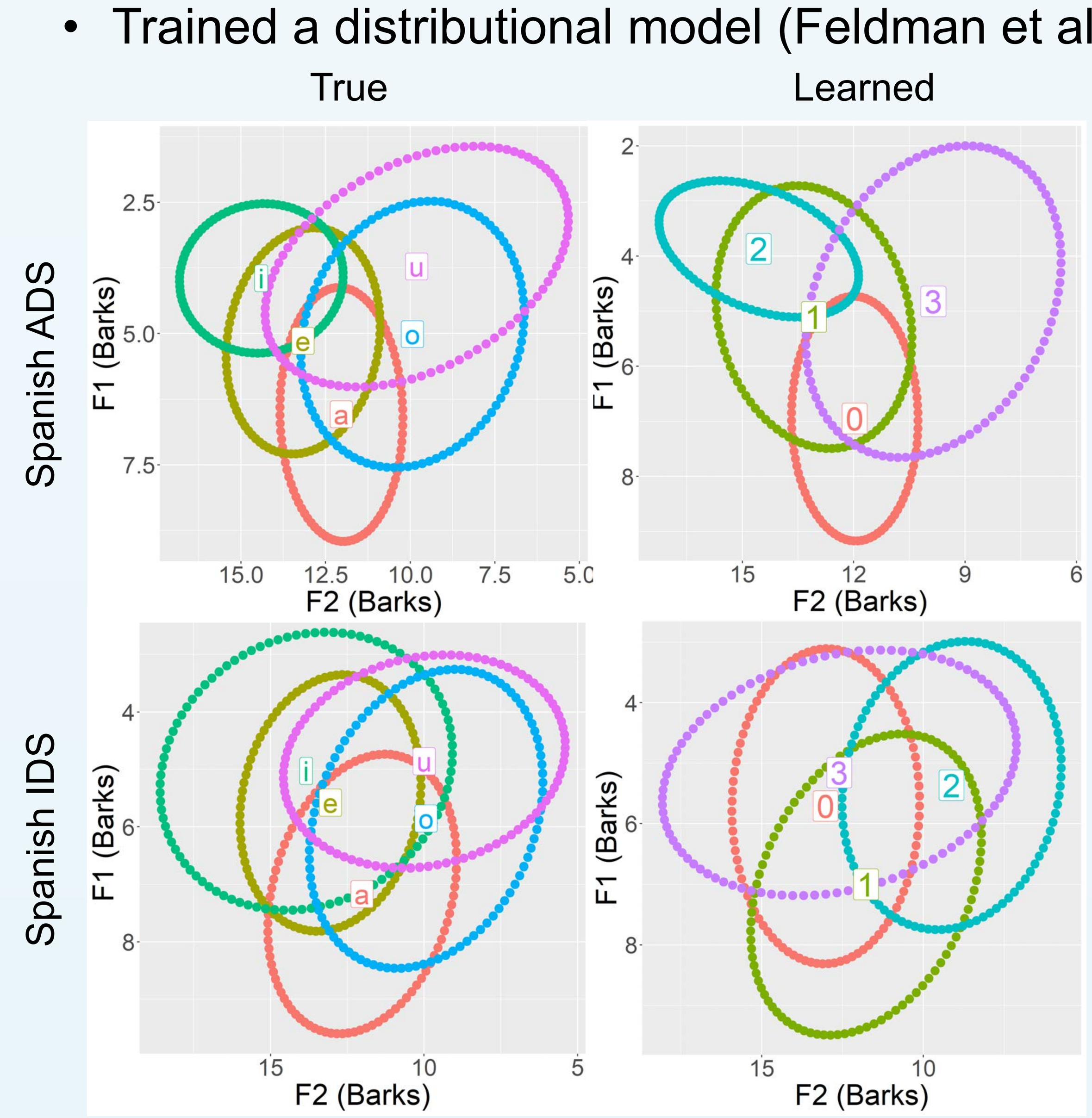
- Calculated (symmetric) KL divergence for vowel pairs in IDS and ADS
  - Greater absolute value of divergence (less overlap) in ADS
  - But relatively more pairs in IDS with greater divergence (less overlap)

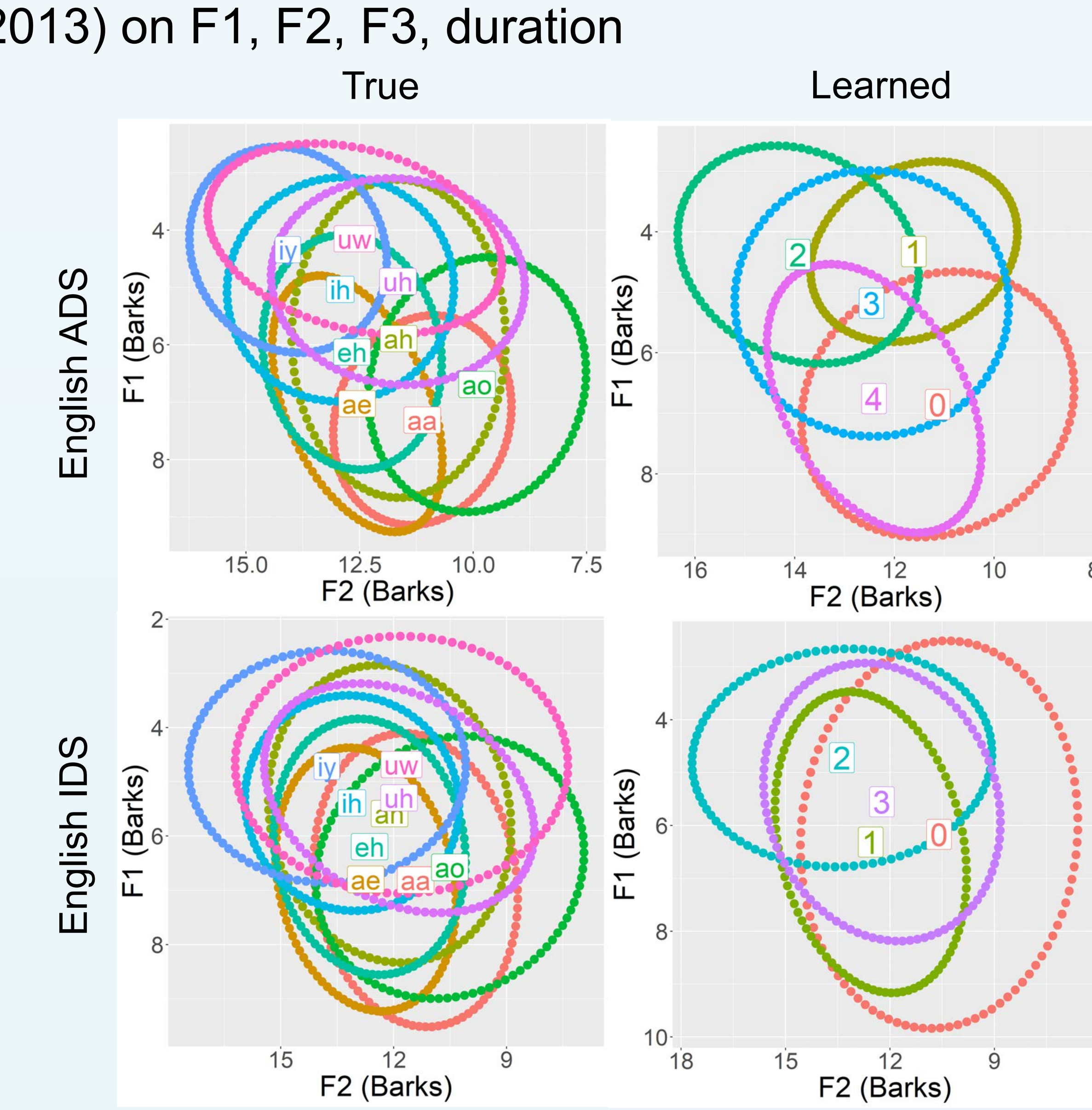


In both Spanish and English, some evidence that IDS vowels have less overlap

Extracting vowel categories via a Gaussian learner

- Trained a distributional model (Feldman et al. 2013) on F1, F2, F3, duration





- Spanish (trained on 5,000 samples):
  - Best performance on F1, F2 and duration
  - Learns 3, 4 or 5 out of 5 categories in IDS (ask us!)
  - Learns 4 out of 5 categories in ADS

- English (trained on 10,000 samples):
  - Best performance on F1, F2, F3 and duration
  - Learns 4 out of 9 categories in IDS
  - Learns 5 out of 9 categories in ADS

CONCLUSIONS

- Mixed findings in IDS
  - Pillai score for the vowel system somewhat more dispersed
  - Relatively more vowel pairs in IDS have greater KL divergence
- However, Bayesian distributional learner has lot of difficulty with connected speech
  - Worst on English 9-vowel system, though better in ADS
  - In some conditions it extracts 5 vowels, but only in Spanish IDS
- Overall, no clear evidence for facilitation in IDS

FUTURE DIRECTIONS

- Improvement needed in distributional learners to handle variation in naturalistic speech
- Perhaps IDS plays a different role in category learning
  - Could the greater spread in IDS be helpful to identify relevant acoustic cues for vowel categories?

ACKNOWLEDGMENTS

This work was supported in part by NSF BCS-2028034 to MS. We also thank members of the UCLA Phonetics Lab for their feedback on this work.

REFERENCES

Cristia, A., & Seidl, A. (2014). *Journal of Child Language*, *41*(4), 913-934. **Demuth**, K., Culbertson, J. & Alter, J. (2006). *Language & Speech*, 49, 137-174. **Pitt**, M.A. Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E. and Fosler-Lussier, E. (2007) [www.buckeyecorpus.osu.edu] Columbus, OH: Department of Psychology, Ohio State University (Distributor). **Feldman**, N. H., Griffiths, T. L., Goldwater, S., & Morgan, J. L. (2013). *Psychological review*, *120*(4), 751. **Kelley**, M. C., & Tucker, B. V. (2020). *The Journal of the Acoustical Society of America*, *147*(1), 137-145. **Kim,** J.-Y., & Repiso-Puigdelliura, G. (2021). Languages, 6 (1), 13. **Liu**, H. M., Kuhl, P. K., & Tsao, F. M. (2003). *Developmental science*, *6*(3), F1-F10. **Ludusan**, B., Mazuka, R., & Dupoux, E. (2021). *Cognitive science*, *45*(5), e12946. **Martin**, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., & Cristia, A. (2015). *Psychological science*, *26*(3), 341-347. **Shue**, Y.-L., P. Keating , C. Vicenik, K. Yu (2011) VoiceSauce: A program for voice analysis, Proceedings of the ICPhS XVII, 1846-1849. **Sundara**, M., Ward, N., Conboy, B., & Kuhl, P. K. (2020). *Bilingualism: Language and Cognition*, *23*(5), 978-991. **Trainor**, L. J., & Desjardins, R. N. (2002). *Psychonomic bulletin & review*, *9*(2), 335-340.