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

Ruolan Li[1,2], Thomas Schatz[1], Yevgen Matusevych[3], Sharon Goldwater[3], and Naomi H. Feldman[1]
[1]University of Maryland [2]University of Rochester [3]University of Edinburgh

During the first year of life, infants become attuned to the sounds of their native language(s). For example, between 6-8 months and 10-12 months, infants learning American English get better at distinguishing American English [ɹ] and [l], as in 'rock' vs 'lock', compared to infants learning Japanese [1]. This phenomenon has been attributed to infants' acquisition of phonetic categories, i.e. the vowels and consonants of their native language [2, 3, 4]. A recent modeling study, however, challenged this interpretation [5]. It provided the first demonstration of a learning mechanism that predicts this cross-linguistic difference in [ɹ]-[l] discriminability from raw speech input. Furthermore, it did so without learning phonetic categories. However, the models in that study were trained on less than 20 hours of speech from 20 or more speakers in roughly equal proportions, which is not particularly representative of an infant's learning conditions.

In this project, we ask whether—similar to infants, who robustly acquire their language across a variety of environments [6, 7]—the learning mechanism from [5] produces stable learning outcomes across a range of ecologically plausible learning conditions. We focus on the effect of two parameters: (1) the total amount of input, and (2) the input distribution over speakers, i.e. the number of speakers and the share of input produced by each speaker. We assemble a new dataset containing more speech overall (1900 hours for American English and 190 for Japanese) and more speech per speaker (up to 900 and 60 hours per speaker for American English and Japanese, respectively). This allows us to cover a range of learning conditions that are more representative of infants' experience. Furthermore, we construct the dataset from public domain audiobooks, whereas the previous study used proprietary corpora, making the results more accessible and replicable.

References

[1] Patricia K Kuhl, Erica Stevens, Akiko Hayashi, Toshisada Deguchi, Shigeru Kiritani, and Paul Iverson. Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, 9(2):F13–F21, 2006.

[2] Patricia K Kuhl, Barbara T Conboy, Sharon Coffey-Corina, Denise Padden, Maritza Rivera-Gaxiola, and Tobey Nelson. Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493):979–1000, 2007.

[3] Janet F Werker and Suzanne Curtin. Primir: A developmental framework of infant speech processing. *Language Learning and Development*, 1(2):197–234, 2005.

[4] Catherine T Best et al. The emergence of native-language phonological influences in infants: A perceptual assimilation model. *The development of speech perception: The transition from speech sounds to spoken words*, 167(224):233–277, 1994.

[5] Thomas Schatz, Naomi H Feldman, Sharon Goldwater, Xuan-nga Cao, and Emmanuel Dupoux. Early phonetic learning without phonetic categories – Insights from machine learning. 2019.

[6] Alejandrina Cristia, Emmanuel Dupoux, Michael Gurven, and Jonathan Stieglitz. Child-directed speech is infrequent in a forager-farmer population: A time allocation study. *Child Development*, 2017.

[7] Paul Vogt, J Douglas Mastin, and Diede MA Schots. Communicative intentions of child-directed speech in three different learning environments: Observations from the Netherlands, and rural and urban Mozambique. *First Language*, 35(4-5):341–358, 2015.