

Infants construct phonemes from early phonetic categories: Evidence from the acquisition of phonotactics, morphology and word recognition

Megha Sundara

Department of Linguistics, University of California

Segments are not always produced the same way across contexts, even in speech addressed to infants. Such gradient phonetic detail is thought to pose a challenge for infants acquiring their native language because of the lack of perceptual constancy. In computational modeling of acquisition is it ignored by using idealized dictionary pronunciations of the input. An assumption common to corpus, behavioral and computational investigations of acquisition is that phonetic variation is noise that needs to be filtered out. Or that infants acquire knowledge about phonetic variation later in life as a result of language experience.

I will first introduce a new corpus of phonetically-transcribed infant-directed speech, PhonProv [1], and a visualization tool, IDSViz, to study variation [now available for download]. This corpus includes 6500 utterances from the Providence Corpus [2] that are forced aligned, transcribed and hand checked by phonetically-trained annotators. Throughout the talk I will show the extent of variation in infant-directed speech using PhonProv.

Then, based on experiments on morphological acquisition and word recognition, I will show that it is phonemes that infants gradually construct as a result of language experience. In the final set of experiments, I will present evidence from the acquisition of phonotactics that infants' earliest categories are not phonemic, but phonetic.

Infants construct phonemic categories in the second year of life

Evidence from morphology: In American English, words ending in [t] and [d] are obligatorily produced as taps, when suffixed with *-ing*. By 8-months, English-learning infants relate sequences like *babbling* to *bab*, demonstrating their discovery of the suffix *-ing* [3]. However, suffix discovery is blocked if the stem changes; infants fail to relate *patting* [pætɪŋ] with a stem-final tap and *pad* at 8-months; they only succeed at 12-months [4]. At the same age, English-learning infants fail to relate *patting* [pætɪŋ] with a stem-final tap and *pat*. That is, only by 12-months do English-learning infants establish the equivalence between taps and [d]. They still fail to relate taps and [t].

Evidence from word recognition: Other pronunciation variation is optional. Word-final [t] in American English is most frequently produced as a glottal stop. We tested 14-month-olds on their preference for [t]-final familiar words produced with either a glottal stop or an unaspirated, released [t] [5, 6]. In both experiments, these familiar words were contrasted with minimally different coda mispronunciations. American English-learning 14-month-olds listened longer to familiar words, regardless of whether they were produced with glottal stops or a canonical variant of word-final [t]. That is, by 14-months, English-learning infants establish the equivalence between glottal stops and [t].

Infants' earliest categories are not phonemic

Evidence from phonotactics: To some extent segments vary in their pronunciation across positions in all languages. In American English, for instance, initial [t]s are typically aspirated, medial [t]s are often unaspirated or tapped and final [t]s are overwhelmingly produced as glottal stops or deleted altogether. Such variation provides critical information about sequencing restrictions – or phonotactics. For instance, aspirated [t] is most likely to be preceded by an utterance or word boundary, whereas a glottal stop is most likely to be followed by one. By corollary, it is extremely unlikely that an aspirated stop in American English is followed by an utterance or word boundary, or that a glottal stop (underlying [t]) is preceded by one. If infants' earliest categories are phonemic, that is aspirated stops, unaspirated stops, glottal stops and taps (or any subset) are treated as equivalent, they would simply not be able to learn phonotactic restrictions.

We show exactly that based on 3 experiments on American English-learning 5-month-olds [7, 8]. First, I will show that even 5-month-olds can distinguish 3 sets of high versus low probability English nonce sequences. These sequences differ in segment and segment pair frequencies. Next, using *k-fold* cross validation, a popular method in machine learning, I will show that these 3 sets of nonce sequences, that are successfully distinguished by 5-month-olds, cannot be distinguished based on segment and segment pair frequencies derived from a phonemically transcribed infant-direct speech corpus.

In sum, based on the extent of pronunciation variation in speech addressed to infants, behavioral experiments on American English-learning 5-, 8-, 12- and 14-month-olds combined with corpus analysis using methods from machine learning, I will argue that infants build phoneme categories incrementally from early phonetic categories.

References

- [1] Khlystova, E., Chong, A.J., & Sundara, M. (2023). Phonetic variation in English infant-directed speech: A large-scale corpus analysis. *Journal of Phonetics*, 100, 101267. <https://doi.org/10.1016/j.wocn.2023.101267>
- [2] Demuth, K., Culbertson, J., & Alter, J. (2006). Word-minimality, epenthesis and coda licensing in the early acquisition of English. *Language and Speech*, 49(2), 137–174. <https://doi.org/10.1177/00238309060490020201>
- [3] Kim, Y. J., & Sundara, M. (2021). 6-month-olds are sensitive to English morphology. *Developmental Science*, 24, 4, e13089. <https://doi.org/10.1111/desc.13089>
- [4] Sundara, M., White, J., Kim, Y.J. & Chong, A.J. (2021) Stem similarity modulates infants' acquisition of phonological alternations. *Cognition*, 209, 104573. <https://doi.org/10.1016/j.cognition.2020.104573>
- [5] Liang, K. Khlystova, E., & Sundara, M. (2026). 14-month-old English-learning infants encode positional variants of word-final /t/. Oral presentation at *International Congress of Infant Studies*, Panama City
- [6] Khlystova, E. & Sundara, M. (2023). The lexical representation of coda /t/ by 14-month-olds. 185th Meeting of the *Acoustic Society of America*, Sydney
- [7] Sundara, M., & Breiss, C. (2026). How infants learn segmental phonotactics in their native language: Integrating insights from machine learning and adult and infant experiments. *Cognition*, 273, 106525. <https://doi.org/10.1016/j.cognition.2026.106525>
- [8] Sundara, M., Breiss, C.M., Dickson, N., & Mayer, C. What's in a 5-month-old's proto-lexicon? Under revision