5-month-olds are sensitive to phonotactic patterns in their native language

Megha Sundara & Canaan Breiss
UCLA Department of Linguistics

A prevalent view is that infants learn language-specific phonotactic distributions from words. Supporting evidence comes from studies showing infants can segment words before they are sensitive to native language phonotactics. Thus, 5-month-old English-learning infants can segment words (e.g., Johnson & Tyler, 2010), but they are sensitive to language-specific phonotactic distributions only at 9 months (Jusczyk et al., 1994; Mattys et al., 1999; Mattys & Jusczyk, 2001).

We present data from two infant experiments and one adult experiment that challenge this developmental trajectory. In Jusczyk et al.'s classic experiments on phonotactics, infants were presented lists of words that differed simultaneously on two metrics, positional probability of a segment within a word (*unigram*) and co-occurrence probabilities of two phonemes within a word (*bigram*). In Experiment 1, we asked adults to assign numerical ratings to nonce sequences based on the likelihood that they could be adopted as a new word of English (using magnitude estimation, cf. Hayes & White, 2013). Using the 480 stimuli from Jusczyk et al., as well as novel stimuli, we tested adults in the lab (n = 85) and then replicated the results in a larger, online Amazon Mechanical Turk experiment (n = 168). Our results (Figure 1) confirmed that low values on each of these metrics is associated with decreased phonotactic acceptability in adults (Daland et al., 2011, Albright, 2009). Crucially, we found that unigram and bigram probabilities independently predict adult native English speakers phonotactic judgments with the effects of each metric being clearest among items with low-probability values on the other (Figure 1).

Based on the adult experiments, we identified 4 sets of CVC sequences: (a) two had items with low bigram probabilities that differed only on unigram probabilities, and (b) two had items with low unigram probabilities that differed only on bigram probabilities. In Experiment 2 (n = 30), using an infant-controlled version of the Headturn Preference Procedure (Jusczyk et al., 1994) we compared monolingual English learning 5-month-olds's listening times to 7 lists of words each with high and low *unigram* probabilities (14 trials). Listening times were log-transformed then analyzed using a maximally-specified Bayesian hierarchical linear regression model. Results (raw listening times in Figure 2, left panel) indicated that 5-month-olds listened significantly longer to lists with high compared to low unigram probability items. We are currently testing monolingual English learning 5-month-olds (n = 23) using the same methods on lists with high and low *bigram* probability items (Figure 2, right panel). Of the 23 infants, 15 listened longer to lists with high compared to low bigram probability items.

Together, our results show that we can isolate the contribution of individual lexical statistics to phonotactic acceptability. Once we do so, we see that infants are sensitive to language-specific unigram (and possibly bigram) probabilities at the earliest age at which they have been shown to segment words. Our results cannot be reconciled with accounts where sensitivity to language-specific phonotactics is learned from words. Instead, our results argue for models in which infants use language-specific phonotactic probabilities from the unsegmented speech stream (e.g., Adriaans & Kager, 2010; Brent & Cartwright, 1996; Daland & Pierrehumbert, 2011).

Figure 1 Mean ratings by adult native English speakers.

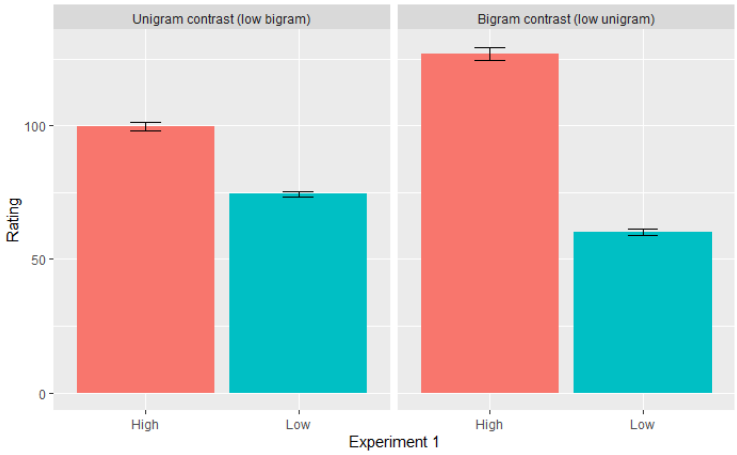
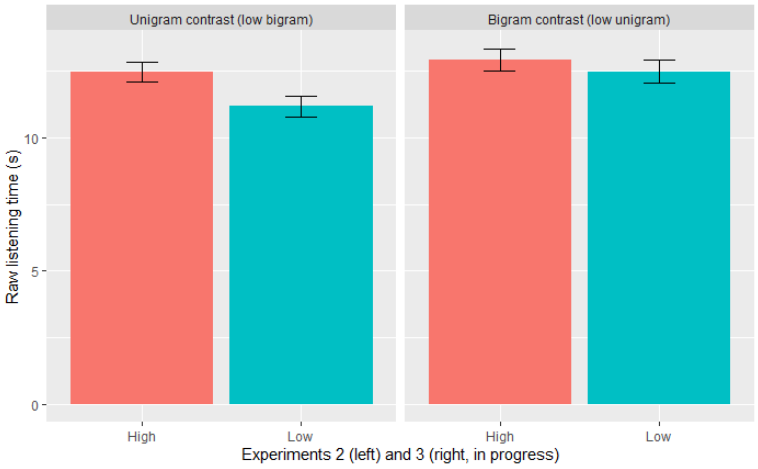


Figure 2 Raw listening times from monolingual 5-month-olds tested in Experiment 2 and 3.



References

Albright, A. (2009). Feature-based generalization as a source of gradient acceptability. *Phonology*, *26*(1), 9-41.

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, *61*(1-2), 93-125.

Daland, R., & Pierrehumbert, J. B. (2011). Learning diphone-based segmentation. *Cognitive Science*, *35*(1), 119-155.

Daland, R., Hayes, B., White, J., Garellek, M., Davis, A., & Norrmann, I. (2011). Explaining sonority projection effects. *Phonology*, *28*(2), 197-234.

Hayes, B., & White, J. (2013). Phonological naturalness and phonotactic learning. *Linguistic Inquiry*, *44*(1), 45-75.

Johnson, E.K. & Tyler, M. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science,* 13, 339-345.

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, *33*(5), 630.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, *38*(4), 465-494.