

Phonetic variability minimally degrades word segmentation in infant-directed speech

Gabriel Scharff, Megha Sundara

University of California, Los Angeles

A common idealization in modeling language acquisition is that speech input is represented as a sequence of phonemes. But the speech signal, even when directed to infants [1], is gradient, with predictable allophonic variation, and less predictable rate of speech effects like deletion and assimilation. Because young infants encode phonetic distinctions [2], their speech intake includes information about such variation. We evaluated the impact of encoding phonetic distinctions on word segmentation, the critical task of breaking up the speech signal into discrete units.

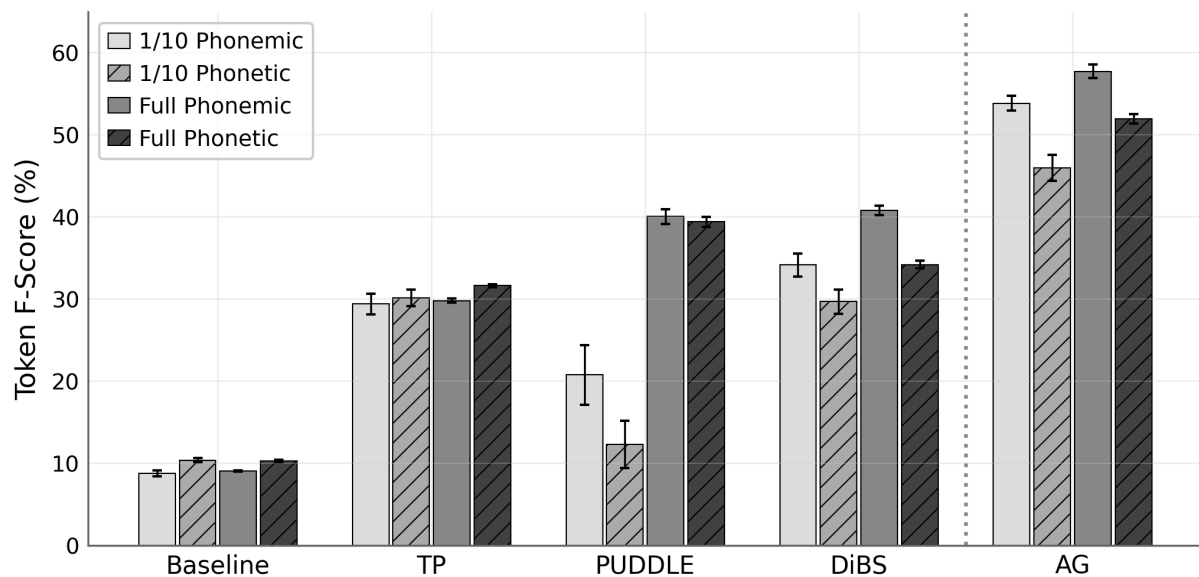
We compared the performance of unsupervised word segmentation algorithms on phonemic versus phonetically-transcribed infant-directed speech (IDS) in PhonProv [1]. The PhonProv corpus has 31,000+ words [6970 utterances] from ~10 hours of interaction with 5 infants in the Providence corpus, an upper bound for ~1 day of speech addressed to infants. Phonetic transcription in this corpus captures consonant variation (e.g., aspiration, flapping), but not vowel variation.

Four algorithms in the WordSeg package [3] were compared: PUDDLE [4], and DiBS [5] exploits segment phonotactics; TP, exploits transitional probabilities between syllables [6], and an Adaptor Grammar (AG) model [7], that uses hierarchical Bayesian inference to find words and segments. To simulate variation in the input received by individual infants, we created ten balanced replicates, each containing 90% of the original corpus. Model performance was evaluated using token F-scores, the harmonic mean of precision and recall benchmarked against the gold (correct) segmentation. A baseline model that assigned word boundaries randomly ($p=0.5$) was also included [8]. We report results from models that segmented the entire corpus, as well as ones that were fed a much smaller corpus size (1/10th) to simulate the earliest stages of word segmentation.

All models outperformed the baseline (Figure 1). Surprisingly, when trained on the full corpus of IDS, there was only a small reduction in performance when the input encoded phonetic variation (range: 0-0.07). The reduction in performance was greater (range 0.04-0.23) when phonetically transcribed adult-directed speech (ADS) in the Buckeye corpus was segmented by TP, DiBS and PUDDLE [9]. Even with the corpus size reduced to a tenth, reduction in F-scores was minimal in IDS for 3 of the 4 algorithms. What we did not replicate was an overall advantage for segmenting IDS [10, 11; see 12 for similar findings as ours]; even the best segmentation performance here was lower than that reported for ADS [9].

Conclusion: We did not find IDS to be easier to segment. However, contrary to findings for ADS [9], we found phonetic variation to be less detrimental to word segmentation in IDS, even for small corpus sizes. This could be because the shorter utterances in IDS allow for more segments at utterance edges, a prosodic position that typically licenses allophony, providing cues to word boundaries, potentially aiding segmentation. We will discuss the implications of these results for mechanisms by which infants learn to segment words.

Figure 1: model accuracy when produced phonemic or phonetically-transcribed IDS



References

- [1] Khlystova, E., Chong, A. J., & Sundara, M. (2023). Phonetic variation in English infant-directed speech: A large-scale corpus analysis. *J. of Phonetics*, 100, 101267. [2] Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development*, 7(1), 49–63. [3] Bernard M, Thiollere R, Saksida A, Loukatou GR, Larsen E, Johnson M, Fibla L, Dupoux E, Daland R, Cao XN, Cristia A. WordSeg: Standardizing unsupervised word form segmentation from text. *Behav Res Methods*. 2020 Feb;52(1):264-278. doi: 10.3758/s13428-019-01223-3. PMID: 30937845. [4] Monaghan, P., & Christiansen, M. H. (2010). Words in puddles of sound: modelling psycholinguistic effects in speech segmentation. *J. of Child Language*, 37(03), 545-564. [5] Daland, R., & Pierrehumbert, J.B. (2011). Learning diphone-based segmentation. *Cognitive science*, 35(1), 119–155 [6] Saksida, A., Langus, A., & Nespors, M. (2017). Co-occurrence statistics as a language-dependent cue for speech segmentation. *Developmental Science*, 20(3). [7] Johnson M., Griffiths, T. L., & Goldwater, S. (2007). Adaptor grammars: A framework for specifying compositional nonparametric Bayesian models. In *Advances in neural information processing systems* (pp. 641-648). [8] Lignos, C. (2012). Infant word segmentation: An incremental, integrated model. *Proceedings of WCCFL* (Vol. 30, pp. 13-15). [9] Beech, C., & Swingley, D. (2023). Consequences of phonological variation for algorithmic word segmentation. *Cognition* 235, 10.1016/j.cognition.2023.105401. [10] Batchelder, E. O. (2002). Bootstrapping the lexicon: A computational model of infant speech segmentation. *Cognition*, 83, 167–206. [11] Ludusan, B., Mazuka, R., Bernard, M., Cristia, A., & Dupoux, E. (2017). The role of prosody and speech register in word segmentation: A computational modelling perspective. In *Proceedings of the Annual Conference of the Association for Computational Linguistics (Volume 2: Short papers)* (pp. 178–183). Vancouver, Canada. [12] Cristia A, Dupoux E, Ratner NB, Soderstrom M. (2019). Segmentability Differences Between Child-Directed and Adult-Directed Speech: A Systematic Test With an Ecologically Valid Corpus. *Open Mind*. 3:13-22.