Computational modeling of infant word-finding from the phonetic ground up

Daniel Swingley[1] & Robin Algayres[2]

[1]University of Pennsylvania, [2]Ecole Normale Superieure & Paris Sciences et Lettres & Inria

Since the 1990s, researchers have tested how infants detect words in continuous speech, evaluating infants' use of phonetic and distributional information. Computational modeling of this process has examined distributional information, but mostly ignores phonetics. Virtually all models presuppose that (a) parental speech realizes every phone in words' canonical pronunciations, and that (b) infants reliably recognize these phones in real-world contexts. Under these assumptions, word segmentation (over text, essentially) is only moderately difficult and can be solved to infant-plausible levels of success using diverse statistical heuristics (Bernard et al., 2020). Because everyone nevertheless knows that (a) is wrong and (b) is doubtful, computational models capable of working from the speech signal itself are needed. But models operating over acoustics are hard to build and hard to evaluate. We present an advance on both problems. A new, state-of-the-art self-supervised computational parsing model, DP-Parse (Algayres et al., 2022), is evaluated over actual natural-speech experimental stimuli (Pelucchi, Hay, & Saffran, 2009a, 2009b). The model does not compute transitional probabilities, but finds the same wordforms infants do, favoring words with high syllabic transitional probabilities despite equal occurrence frequency. We believe this is the first time infants and a model have been tested on the same acoustic word-segmentation stimuli. DP-Parse takes audio as input (here, the Italian-language sentences of Pelucchi et al.), converts the audio into a computational embedding starting from wav2vec2 (Baevski et al., 2020), trained without supervision on ~8 months' worth of speech (Bergelson et al., 2019), estimates pseudofrequencies of potential words, and computes the most likely lexical parse, using a generalization of Goldwater et al. (2009). Like infants, the model finds high-transitional-probability words more often than low-transitional-probability words. It is now possible to benchmark models against infant outcomes using the same materials, and we propose iterative testing of infants and model systems.