Supplementary File III for Bursty, irregular speech input to children predicts vocabulary size

Margaret Cychosz[1], Rachel R. Romeo[2,3], Jan R. Edwards[2], and Rochelle Newman[2]

[1]Department of Linguistics, University of California, Los Angeles

[2]Department of Hearing and Speech Sciences, University of Maryland, College Park

[3]Department of Human Development and Quantitative Methodology, University of Maryland-College Park

1 Overall speech input and vocabulary scores

Here we define input using the adult word counts from LENA's linguistic unit algorithm. That is, we do not distinguish between speech directed to the child versus someone else (child-directed speech versus other-directed speech). We examine whether burstiness in *overall speech input* (directed and overheard) also relates to children's receptive vocabulary sizes.

1.1 Data pre-processing

We wrote an ad-hoc script to mine the .its file associated with each LENA recording and provide us with the number of words from adults within each minute of the recording by selecting all speech segments labeled as "FAN" or "MAN," and filtering out segments >10s as anecdotal work in our lab has shown that MAN and FAN tags >10s tend to be mislabeled. We applied the same sleep classifier as in the manuscript to identify sleep portions of the recording and remove them. The word count was estimated from LENA's linguistic unit algorithm, which has been reported on widely elsewhere in the literature (Cristia et al., 2020; Gilkerson et al., 2017). We imputed "empty" minutes (epochs without any adult words) for each recording, to reflect the times when the child was not exposed to any words. In this supplementary analysis, we limited the computation to burst amplitude, since we didn't see large differences between burst amplitude and duration in the child-directed speech analysis in the full manuscript so we didn't think that overall speech input burst amplitude and duration would differ that much from one another. The average amplitude of bursty overall speech within the recordings was 49.36 words (SD=25.86, range=13.26-150.86). (Note that in the manuscript we performed additional interpolation and actually looked at 30-second increments, not 60-second as we do here.)

1.2 Model fitting

We fit a linear mixed effects model to predict receptive vocabulary scores (Peabody Picture Vocabulary Test-5 growth scale values) and included random intercepts of Child. We additionally

included fixed effects of Child Age (in months, centered and scaled), Maternal Education (binned, as described in the manuscript, and centered and scaled), and Child Gender (contrast coded). The variable of interest was the average amplitude (number of adult words) within bursty minutes. This variable did not predict vocabulary scores (β=0.04, p=0.112; model summary in Table 1). This analysis supports the idea that bursty overall *child-directed speech* input is positively related to concurrent receptive vocabulary scores. However, this same effect on vocabulary may not be present for the overall speech that children are exposed to (other-directed, overheard, etc.).

Table 1

The relationship between bursty overall speech input and concurrent receptive vocabulary size

Intercept	126.63***
	(123.12, 130.15)
	t = 70.58
	p = 0.00
Adult Word Amplitude	0.04
	(−0.01, 0.10)
	t = 1.59
	p = 0.12
Age (mos)	1.47***
	(1.38, 1.56)
	t = 33.69
	p = 0.00
Gender:Female	2.02
	(−1.44, 5.48)
	t = 1.15
	p = 0.26
Mat. Ed.	5.56***
	(4.49, 6.63)
	t = 10.22
	p = 0.00
Avg. Hourly Words	0.59
	(−2.16, 3.35)
	t = 0.42
	p = 0.68
Observations	544
Log Likelihood	−2,089.17

2 References

Cristia, A., Lavechin, M., Scaff, C., Soderstrom, M., Rowland, C., Räsänen, O., ... Bergelson, E. (2020). A thorough evaluation of the Language Environment Analysis (LENA) system. *Behavior Research Methods*.

Gilkerson, J., Richards, J. A., Warren, S. F., Montgomery, J. K., Greenwood, C. R., Oller, D. K., ... Paul, T. D. (2017). Mapping the Early Language Environment Using All-Day Recordings and Automated Analysis. *American Journal of Speech-Language Pathology*, *26*(2), 248–265. doi: 10.1044/2016_AJSLP-15-0169