This manuscript is currently under review, as of April 2023

Linking acoustic variability in the infants' input to their early word production

Federica Bulgarelli[1,2] & Elika Bergelson[2]

[1] University at Buffalo

[2] Duke University

Author Note

This work was supported by grants to EB (NIH-OD, DP5 OD019812-01) and FB (NIH-NICHD, F32 HD101216). We thank all of the research assistants at Duke University who aided with data preparation. The authors have no conflict of interest to disclose.

The data used in this manuscript, along with fully reproducible analyses and manuscript, are available on OSF: https://osf.io/3m2rf/?view_only=02b86a5b9605422fb26fc196a14f2592

The authors made the following contributions. Federica Bulgarelli: Conceptualization, Writing - Original Draft Preparation, Writing - Review & Editing; Elika Bergelson: Writing - Review & Editing, Supervision.

Correspondence concerning this article should be addressed to Federica Bulgarelli, 211 Mary Talbert Way, Buffalo NY 14260. E-mail: fbulgare@buffalo.edu

Linking acoustic variability in the infants' input to their early word production

Research highlights

- Talker variability shapes learning in the lab and is available in the real world, we ask whether variability predicts word learning in the real world
- Acoustic measurements of words in infants' input predicted when infants say those same words beyond the effects of frequency
- Speech register also predicts when infants will say words, alongside effects of acoustic variability
- Our results provide a deeper understanding of how sources of variability inherent to children's input influence their learning and development

Abstract

Talker variability shapes how learning unfolds in the lab, and similar types of variability have been shown to be available to infants in the real world. Here, we ask whether talker variability also influences age of first production for common nouns, above and beyond the effects of frequency. Then, we ask whether these effects are redundant with effects of speech register. We predicted children's month of first production using acoustic measurements for highly common nouns from a longitudinal corpus of North-American infants. In addition of frequency, variability in how words sound in 6-17mo's input predicted when children said those same words. Further, while proportion of child-directed-speech also predicts month of first production, it does so alongside measurements of acoustic variability in children's real-world input. Together, this adds to a growing body of literature showing that how children hear words influences learning both in the lab and in daily life.

Introduction

The relationship between what children hear and their language development has been of interest to researchers for decades. Much of this research has focused on the quantity and quality of input, using metrics such as types and tokens, syntactic variability, and referential transparency (e.g. Huttenlocher, Waterfall, Vasilyeva, Vevea, & Hedges, 2010; Rowe, 2012). While these properties effectively describe some aspects of the input, they generally stop short of measuring the acoustic properties of speech, and how these may influence spoken word learning. Since acoustic variability has been shown to influence word learning in the lab (Bulgarelli & Bergelson, 2022, 2023; Galle, Apfelbaum, & Mcmurray, 2015; Hoehle, Fritzsche, Meb, Philipp, & Gafos, 2020; Rost & McMurray, 2009) and to be readily and similarly available to infants in the real world (Bulgarelli, Mielke, & Bergelson, 2021), the current manuscript tests whether infant's own experiences with acoustic variability are related to their early word production. Put otherwise: can we link the acoustic variability with which infants' hear words in daily life to when they start to say those same words?

To date, research on the effects of talker variability on word learning has yielded mixed results. While initial studies with adults suggested that talker variability may be hard for learners (Mullennix, Pisoni, & Martin, 1989), studies with infants report that it can be helpful for generalization to new talkers (Bulgarelli & Bergelson, 2022) and during challenging word learning tasks (e.g. learning minimal pairs; Stager and Werker (1997); Galle et al. (2015); Hoehle et al. (2020)). At the same time, talker variability can be hard for infants under certain conditions. For example, talker variability resulted in 8-month-olds over-extending what should 'count' as an instance of a new word (Bulgarelli & Bergelson, 2022); and made learning novel dissimilar-sounding words, ('neem' and 'lof', not minimal pairs) more challenging for 14-month-olds (Bulgarelli & Bergelson, 2023). Taken together, the literature suggests that while talker variability can be helpful under specific circumstances, it can also interfere with learning.

All of the above studies were conducted in the lab and featured stimuli intended to minimize or maximize acoustic variability stemming from different talkers. In a recent corpus analysis, Bulgarelli et al. (2021) extracted tokens of highly frequent nouns from a longitudinal corpus of daylong recordings (baby, ball, book, water, dog), quantified the amount of acoustic variability infants heard, and related that to other input properties (e.g. number of tokens or talkers in the input). Results suggested that infants experienced similar acoustic variability in their day-to-day life as they do in the lab. Between-talker variability was modestly correlated with a variety of input properties; hearing more talkers overall and hearing a higher percentage of speech produced by children (relative to adults or electronics) each correlated with hearing more acoustic variability. Overall, these findings suggest that acoustic measurements of variability provide additional information about how children's input varies beyond previously considered measures. But this work leaves open how this variability may connect to early production of these words, which we tackle here.

Notably, there is some overlap in what characterizes child-directed speech (CDS hereafter) and high talker-variability (e.g. pitch range and duration variance), and CDS has been linked to improved word learning (Graf Estes & Hurley, 2013; Ma, Golinkoff, Houston, & Hirsh-Pasek, 2011). For example, Graf Estes and Hurley (2013) found that 17-month-old infants performed better on a word mapping task in the lab when hearing CDS, and properties of CDS in infant's input at 7 months are related to infant's vocabulary size at 2 years (Newman, Rowe, & Bernstein Ratner, 2016). Thus, a secondary question we consider is the separability of measures of talker variability and CDS in predicting early word production.

In sum, research to date suggests that talker-based acoustic variability (1) influences word learning in the lab (sometimes helping and other times increasing difficulty), (2) is available to infants in their real world input, and (3) is not simply redundant with other descriptive properties (e.g. number of tokens, talkers). However, this leaves open how different aspects of talker-based acoustic variability in the input may influence which words infants say and when. Further, CDS and talker-based acoustic properties are often conflated in prior work, leaving it an open question whether these represent the same source of variability in their effects on word learning. In what follows, we seek to link infants' experiences with highly frequent and early learned words to their own word production. Specifically, we first ask whether acoustic variability in infants' own input for highly common nouns (from daylong home recordings) is related to the age of first production of those same nouns. Then, using a subset of our data, we assess whether our acoustic measures of talker variability are redundant with measures of CDS, and whether CDS provides further power in predicting early noun production.

Methods

We report on three types of data, all derived from the SEEDLingS corpus (Bergelson, 2017), described below: 1) acoustic measurements of highly frequent and early learned words in infants' input; 2) ratings of whether words were produced in child- or adult-directed speech, and 3) vocabulary data regarding which words infants themselves produce by 18 months.

Participants

Participants were from the SEEDLingS dataset (Bergelson, 2017), a corpus of 44 infants recruited for a year-long study of word learning, who were recorded monthly from 6-17 months of age (23 males, 21 females). All infants were born full term (40 +/- 3 weeks), had no known hearing or vision problems, and were reported to hear at least 75% English. Forty-two of the infants were White, two were multiracial. Maternal education ranged from high school degree to advanced degree (high school degree: n=1; some college: n=3; associate or bachelor's degree: n=18; advanced degree: n=22). This sample includes one pair of dizygotic twins; both are included. This was a convenience sample.

Corpus Recording and Initial Annotation Procedure

Starting at 6 months and continuing for a year, families were audio-recorded once a month for a full day (up to 16h, using LENAs), and video recorded once a month for an hour (using head mounted cameras and a tripod); on separate days (see Bergelson, 2017; Bergelson, Amatuni, Dailey, Koorathota, & Tor, 2018 for data and description).

Approximately 54 audio recorded hours and 12 video recorded hours for each child were annotated for instances of concrete nouns (based on the broader goals of this project). Each imageable concrete noun said directly to or near the target child was manually tagged by annotators, along with individual talker labels (see Bergelson et al. (2018); Bulgarelli and Bergelson (2019) report reliability for speaker tags was high, kappa = 0.93). Addressee was not initially coded.

Dataset

We identified 13 of the most frequent nouns across the entire corpus, which were "baby", "ball", "book", "water", "dog(gy)", "hand(s)", "car", "hat", "kitty", "milk", "nose", "head", and "mouth". The initial dataset (before the exclusions described below) was 44669 tokens of 13 words across 44 infants. Bulgarelli et al. (2021) report acoustic analyses for five of these (baby, ball, book, water, and dog(gy)). We extracted an audio-clip for each annotated noun instance based on its timestamp and a 0.5s buffer on each side. Research assistants transcribed these segments using Praat, and then we aligned the transcribed textgrids to the audio wav files using the Montreal Forced Aligner (McAuliffe, Socolof, Mihuc, Wagner, & Sonderegger, 2017). All force-aligned files were reviewed and alignment of the target words was adjusted as necessary. Lastly, we extracted the wav files containing the bare target words for each token of each target word for each participant.

Acoustic measurements. We measured acoustic properties that are not lexically contrastive in English. These measurements included *mean pitch, median pitch, max pitch,* *mean pitch slope, duration, and harmonics-to-noise ratio.* Each of these measurements was conducted on the whole word using an automated approach in PraatR (Albin (2014); see script on OSF for details about how each measurement was calculated), see Bulgarelli et al. (2021) for additional details.

Excluding unusable tokens. Following previous research (Bulgarelli et al., 2021), we excluded tokens that would incorrectly effect our measurements of variability. While we don't want to exclude all extreme values (e.g. ones that might be considered outliers) because we are interested in measuring the variability infants hear, we excluded 6424 tokens that had consecutive pitch measurements that differed by more than an octave (double or half the previous pitch), as such jumps in pitch are classic signatures of pitch-tracking errors and are unlikely to occur in natural speech.

We also excluded 3281 tokens that included sounds in addition to the target word, such as background noise from other speakers, animals, or toys (among others, Bulgarelli et al. (2021) report kappa = 0.65 for this exclusion criteria). Next, we excluded 2026 tokens with a harmonics-to-noise ratio <1. While this cutoff is not intrinsically meaningful, this excludes the small tail of tokens with a relatively high ratio of aperiodic noise relative to periodic speech. Lastly, we excluded 461 tokens for which acoustic measurements could not be measured; e.g. when pitch information was missing. After all exclusions, the current dataset includes 32477 tokens, see Supplemental Table for breakdown by word.

Data aggregation. Our variability analyses combined tokens heard by all speakers for each infant, using the **standard deviations** for each acoustic measurement for each word. See Supplementals for means for all words.

Ratings of Child-directed speech. As the SEEDLingS corpus was not annotated for likely addressee of each noun instance, we took a citizen science approach to gathering ratings of CDS. Of the 44 participants, 32 gave permission for short clips of their recordings to be used on public-facing platforms. For these, we submitted the audio clips (including context) to a web-based citizen science platform called Zooniverse. For each clip, annotators on Zooniverse were notified of the target word they were listening for (one of the 13 listed above), and asked to classify it as: a) adult-directed speech, b) child-directed speech, c) utterances containing more than one instance of the target word, and d) junk (noise, baby sounds, not containing the target word). For clips that were marked as containing more than one instance of the target word, annotators then rated each instance of the target word as a) adult-directed, or b) child-directed. Most of the instances of a word were rated 7 times (mean = 6.99), but we included ratings for any that were rated at least 5 times. Generally, a given instance was tagged by a set of unique annotators, however since participation was anonymous we cannot verify that none were tagged by the same person twice.

Raters on Zooniverse rated 34280 tokens of these 13 words from 32 subjects. After annotations were complete, we considered a token of a word as being produced in CDS or ADS if it was rated as such by >70% of annotators. Instances for which there wasn't this strong level of agreement were excluded from the CDS calculations. 62% of instances reached this threshold and were included.

Vocabulary data. In addition to the audio and video recordings, caregivers were asked to fill out monthly MacArthur-Bates Communicative Development Inventories (MCDIs) from 6 to 18 months, providing parent-reported vocabulary data for each child every month. For each of our target words, we computed the age at which each child was first reported to produce that word, and used that as our age-of-acquisition measure, hereafter called MonthFirstProduction, see Table 1.

Results

Predicting age of first production based on word-specific input

Our first set of analyses tests whether the MonthFirstProduction of a specific word is related to children's own experiences with that word. Given the well-documented effects of

Table 1

Word-level properties in the corpus. Columns with sd refer to average standard deviations, which serve as our measure of variability; hnr = harmonics-to-noise ratio, meanpitch = mean pitch, slope = pitch slope. cds column reports the percent of word tokens identified as child-directed-speech for a subset of 32/44 participants; the 'all' row averages across all words.

word	%produce	MonthFirstProd	frequency	sd.hnr	sd.meanpitch	sd.duration	sd.slope	%CDS
baby	40.91	15.28	86.91	5.06	86.67	232.45	403.06	80.44
ball	79.55	14.20	60.50	4.35	86.15	166.90	432.24	91.09
book	59.09	14.96	73.66	3.88	97.89	126.40	711.79	87.43
car	40.91	15.72	41.30	4.35	85.14	166.52	406.44	51.26
dog	68.18	13.63	72.45	4.31	87.01	175.32	407.36	83.55
hand	15.91	16.43	82.66	4.93	81.47	157.76	356.01	84.07
hat	31.82	16.14	41.16	4.10	92.94	126.74	514.95	86.43
head	18.18	16.12	51.50	4.63	84.09	196.81	476.08	63.66
kitty	38.89	15.43	39.19	4.40	79.26	161.39	480.48	92.83
milk	46.51	15.55	41.98	4.49	81.63	132.45	462.20	73.71
mouth	22.73	16.40	51.36	4.79	80.88	134.58	396.37	87.36
nose	43.18	15.89	42.25	5.60	84.45	211.71	368.86	90.51
water	40.91	15.44	61.23	4.27	70.55	153.27	330.57	72.64
all	42.10	15.16	57.68	4.55	84.55	164.90	441.45	80.19

frequency on language development (e.g. Ambridge, 2015), we start with a baseline model that predicts MonthFirstProduction based on the (log-transformed) frequency with which that word was heard by that child over the course of the sparsely sampled year:

$$MonthFirstProduction \sim LogFrequency + (1|subj) + (1|word)$$

Fixed effects in the baseline model accounted for 6.8% of the variance (with random effects, the model accounted for 51% of variance), and included a significant effect of frequency ($t(229.37) = -4.65$, $p < .001$, d = -0.61), such that hearing a word more often resulted in saying it earlier.

Next, we add our acoustic variability metrics (standard deviations of mean pitch, max pitch, median, duration, pitch slope, and harmonics-to-noise ratio), and their interactions with frequency to the model. We also include a set of descriptive properties (how many talkers produced the word, and proportion of tokens from electronics and other children), and word length properties (number of phonemes and number of syllables), which could predict how easy a word is to say in the first place. Since many of these are highly correlated with each other (e.g. mean pitch and median pitch), we conduct backwards stepwise model selection with AIC (e.g. Yamashita, Yamashita, & Kamimura, 2007) to determine the best model for the data. Using this approach, the best fit model is:

$$MonthFirstProduction \sim LogFrequency + MeanpitchVariability + MaxpitchVariability +$$

$$DurationVariability + LogFrequency\mathbf{x}MeanpitchVariability +$$

$$LogFrequency\mathbf{x}DurationVaribility + (1|subj) + (1|word))$$

The fixed effects in this model accounted for 12.4% of the variance and this model was a significantly better fit for the data than the baseline model (p = 0.01), see Table 2 and Figure 1 for model comparison.

There was a significant effect of frequency ($t(213.34) = -2.20$, $p = .029$, d = -0.30), such that hearing a word more often led to an earlier month of first production, as well as a significant effect of max pitch variability ($t(215.53) = -2.09$, $p = .038$, d = -0.28), such that hearing a word more variably in max pitch (holding all other things constant) resulted in an earlier month of first production. There was also a significant interaction between frequency and mean pitch variability, ($t(215.65) = 2.48$, $p = .014$, d = 0.34) such that higher frequency words that infants heard with less variable mean pitch were produced by them earlier. For instance words with a log frequency of 2.5 that were said 1SD *less* variably in mean pitch

Table 2

Model comparison table showing (1) baseline model with just frequency, (2) best model based on backward model selection. The fixed effects in Model 1 account for 6.8 perfect of the variance in the baseline AoA model, in Model 2 they account for 12.4 perfect

	Dependent variable:	
	MonthFirstProduction	
	Baseline model	Best fit model
Frequency	-1.8^{***} (0.4)	-3.0^{**} (1.4)
Meanpitch Variability		-0.04 (0.03)
Maxpitch Variability		-0.02^{**} (0.01)
Duration Variability		0.01^{*} (0.01)
Frequency × Meanpitch Variability		0.04^{**} (0.02)
Frequency × Duration Variability		-0.01^{**} (0.01)
Constant	18.7^{***} (0.7)	21.8^{***} (2.2)
Observations	237	237
Note:		*p<0.1; **p<0.05; ***p<0.01

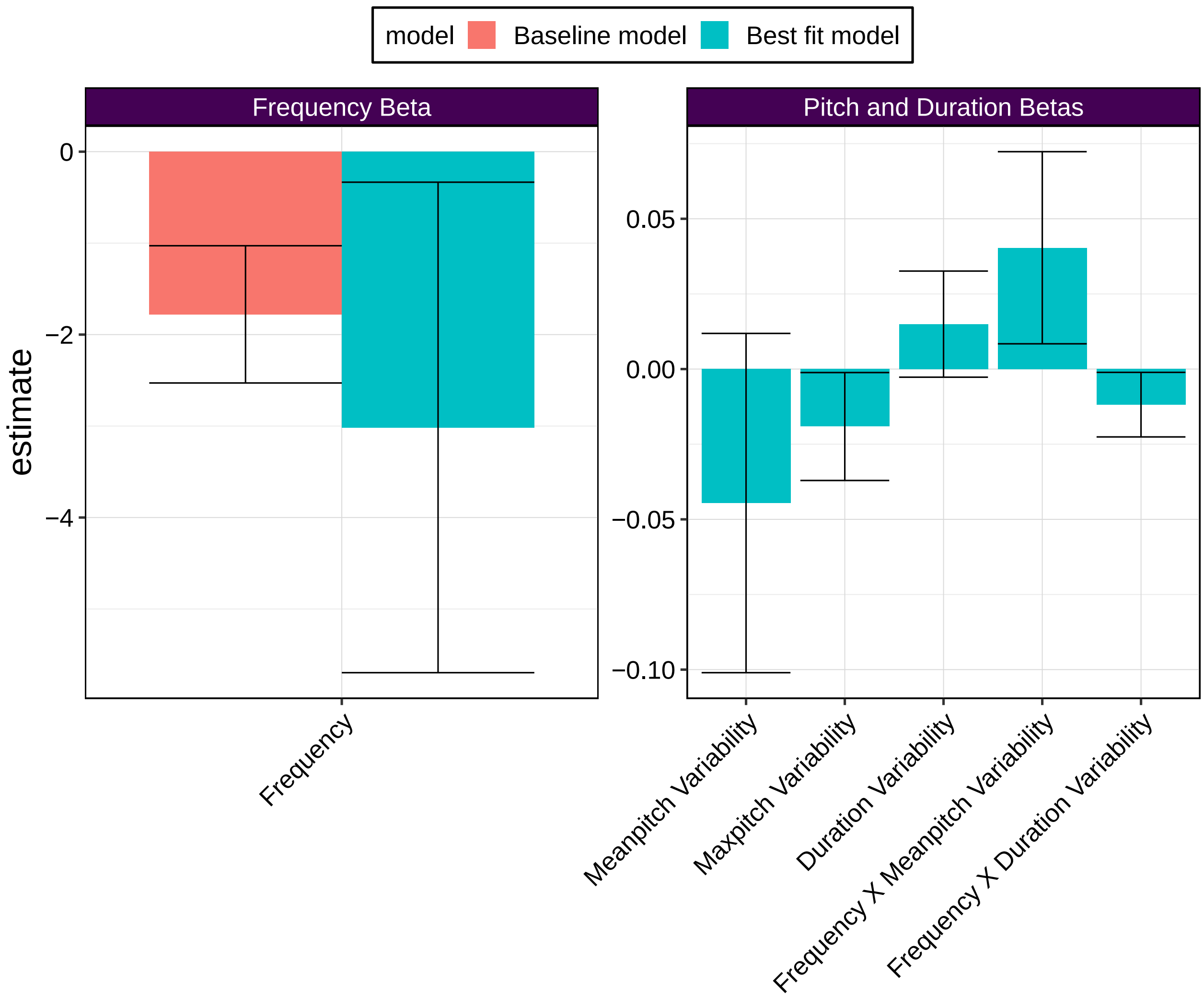

Figure 1. Beta estimates (change in months) and standard errors for frequency in both baseline model and best fit model (left; baseline model in pink, best fit in blue) and additional predictors in best fit model (right) for the full dataset (44/44 Ss). Frequency, Maxpitch Variability, Frequency × Meanpitch Variability, and Frequency × Duration Variablity were individually significant predictors. Higher negative betas for these predictors indicate earlier MonthFirstProduction. N.B. y-axes differ across facets. The fixed effects account for 6.8% of the variance in the baseline AoA model and 12.4% in the best fitting model.

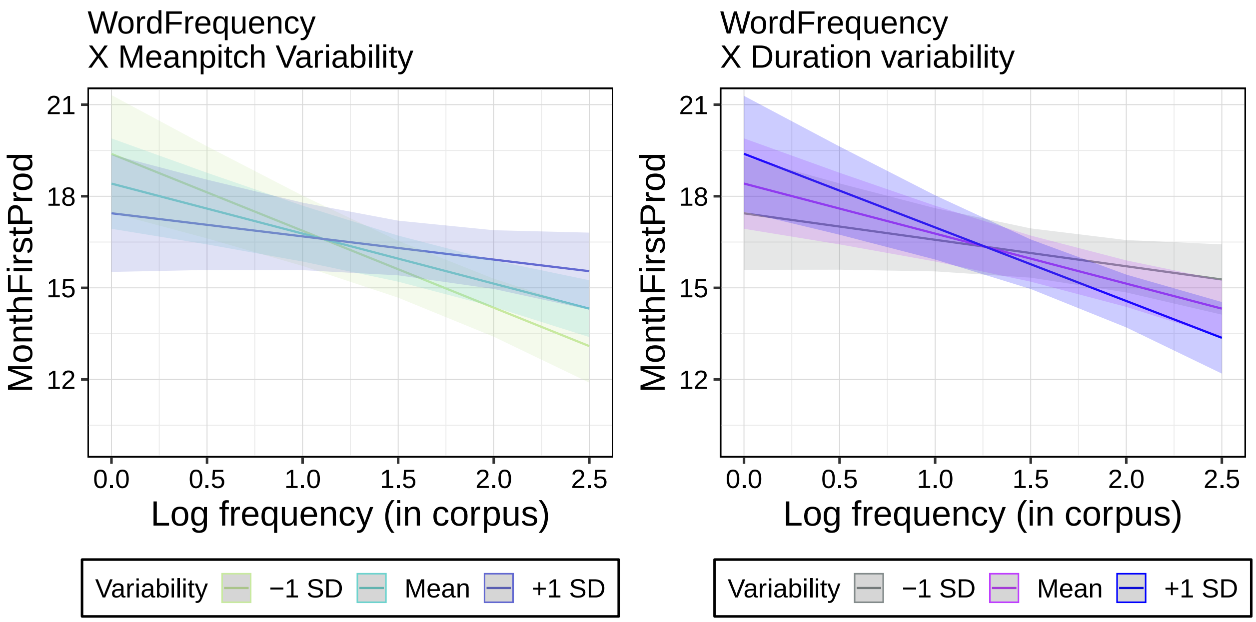

Figure 2. Visualization of predicted values for significant interactions. Left: Interaction between frequency and meanpitch variability. Right: Interaction between frequency and duration variability. Lines represent the predicted data for the mean variability value, +/- 1SD.

were predicted to be produced 2.45 months sooner than more frequent words heard 1SD *more* variably in mean pitch. In contrast, words with a log frequency of 0.5 that were said 1SD *more* variably in mean pitch were predicted to be produced 1.06 months sooner than lower frequency words heard 1SD *less* variably in mean pitch, see Figure 2.

This model also included an interaction between frequency and duration variability ($t(208.73) = -2.16$, $p = .032$, d = -0.30) such that more frequent words infants heard said more variably in duration were produced by them earlier and vice versa. For instance, more frequent words that were said 1SD *more* variably in duration were predicted to be produced 1.91 months sooner than more frequent words heard *less* variably in duration. In contrast, less frequent words that were said 1SD *less* variably in duration were predicted to be produced 1.18 months sooner than lower frequency words heard 1SD *more* variably in duration, see Figure 2.

The main effects of mean pitch variability and duration variability were not significant on their own , all *p*s $>$ .05, but were retained based on our model selection approach (see above).

This first set of analyses suggests the variability with which infants hear words is related to *when* they produce those words in the real world. We found that hearing a word more, and hearing it more variably in mean pitch, max pitch and duration predicted when a word was first produced. Children who heard words more said them earlier, and broadly put, hearing words more variably also resulted in earlier production. The exact pattern of the interaction between frequency and these acoustic variables differed slightly as a function of whether the word is higher or lower frequency (among our already high frequency words).

The role of child-directed-speech

One possibility is that the effects reported above are largely due to speech register, as CDS is often characterized by higher mean and max pitch and changes in word duration. As these are the variables that were found to be significantly related to word production, we next ask whether including the proportion of CDS for a given word as a predictor in the models would explain further variance, or better account for variance otherwise explained by our acoustic variability metrics.

This second set of analyses is conducted on a subset of the original dataset (on which we first rerun our original models before considering CDS, see below), due to parental permissions. Ratings revealed that 80% of the tokens were produced in CDS, ranging from 0%-100% of tokens of any given word for any participant.

Correlations between acoustic measurements and child-directed-speech. We first ask whether proportion of CDS was related to the variability with which the words were said. That is, did children who heard more CDS also hear more variability in e.g. mean pitch? Correlations between proportion of CDS and each of our acoustic variables are in

Table 3

Correlation (Kendall's tau) between proportion of child-directed-speech and each acoustic variability metric.

	Acoustic measurements					
correlation	meanpitch	maxpitch	median	slope	hnr	duration
child-directed-speech	0.03	0	0.05	0.05	0.13**	0.01

[a] **significant after Bonferroni-correction for multiple comparisons (n=6, new p threshold = .008), *p<.05.

Table 3. Only the correlation between proportion of CDS and harmonics-to-noise ratio variability withstood correction for multiple comparisons ($\tau = .13$, $z = 3.92$, $p < .001$), such that hearing a higher proportion of CDS also resulted in hearing more variability in harmonics-to-noise ratio. Nonetheless, this correlation was small in magnitude, suggesting that the proportion of CDS is not a simple redescription of how acoustically variable the speech sounds.

Non-contrastive acoustic measurements and child-directed speech. We now ask whether the proportion of CDS for a word in the input helps predict when a child starts saying that word in this dataset. We first conduct our model selection process again to find the best fit model with our acoustic variability metrics on this subset of the data, then we add CDS as a predictor the model could choose and see whether that changes the best fit model.

The best fitting model for this subset of the data was identical to the model identified for the full dataset:

$$MonthFirstProduction \sim LogFrequency + MeanpitchVariability + MaxpitchVariability +$$

$$DurationVariability + LogFrequency\mathbf{x}MeanpitchVariability+$$
$$LogFrequency\mathbf{x}DurationVaribility + (1|subj) + (1|word))$$

The fixed effects in this model accounted for 12.1% of the variance and all effects went in the same direction as the model with all participants described above. This model was a significantly better fit for the data relative to a model on the subset of the data with just frequency (p = .002). Model estimates can be found in Table 4.

We next add proportion of CDS and its interaction with frequency to the model selection process. The best fit model was as follows:

$$MonthProduction \sim LogFrequency + DurationVariability + HnRVariability+$$
$$PropCDS + LogFrequency\mathbf{x}PropCDS + LogFrequency\mathbf{x}DurationVariability+$$
$$LogFrequency\mathbf{x}HnRVariability + (1|subj) + (1|word))$$

The fixed effects in this model accounted for 13.2% of the variance and this model was a significantly better fit for the data relative to the baseline model (p = .001), see Table 4 and Figure 3.

This model included a significant effect of harmonics-to-noise ratio (HnR) variability, ($t(163.58) = -1.99$, $p = .048$, d = -0.31) such that hearing a word less variably in harmonics-to-noise ratio resulted in an earlier month of first production. There was also a significant interaction between frequency and duration variability ($t(164.44) = -2.09$, $p = .038$, d = -0.33). Consistent with the model on the full dataset, more frequent words (e.g. ones with a log frequency of 2.5) that were heard 1SD *more* variably in duration were predicted to be produced 2.44 months sooner than more frequent words heard 1SD *less* variably in duration. Less frequent words (e.g. ones with a log frequency of 0.5) that were

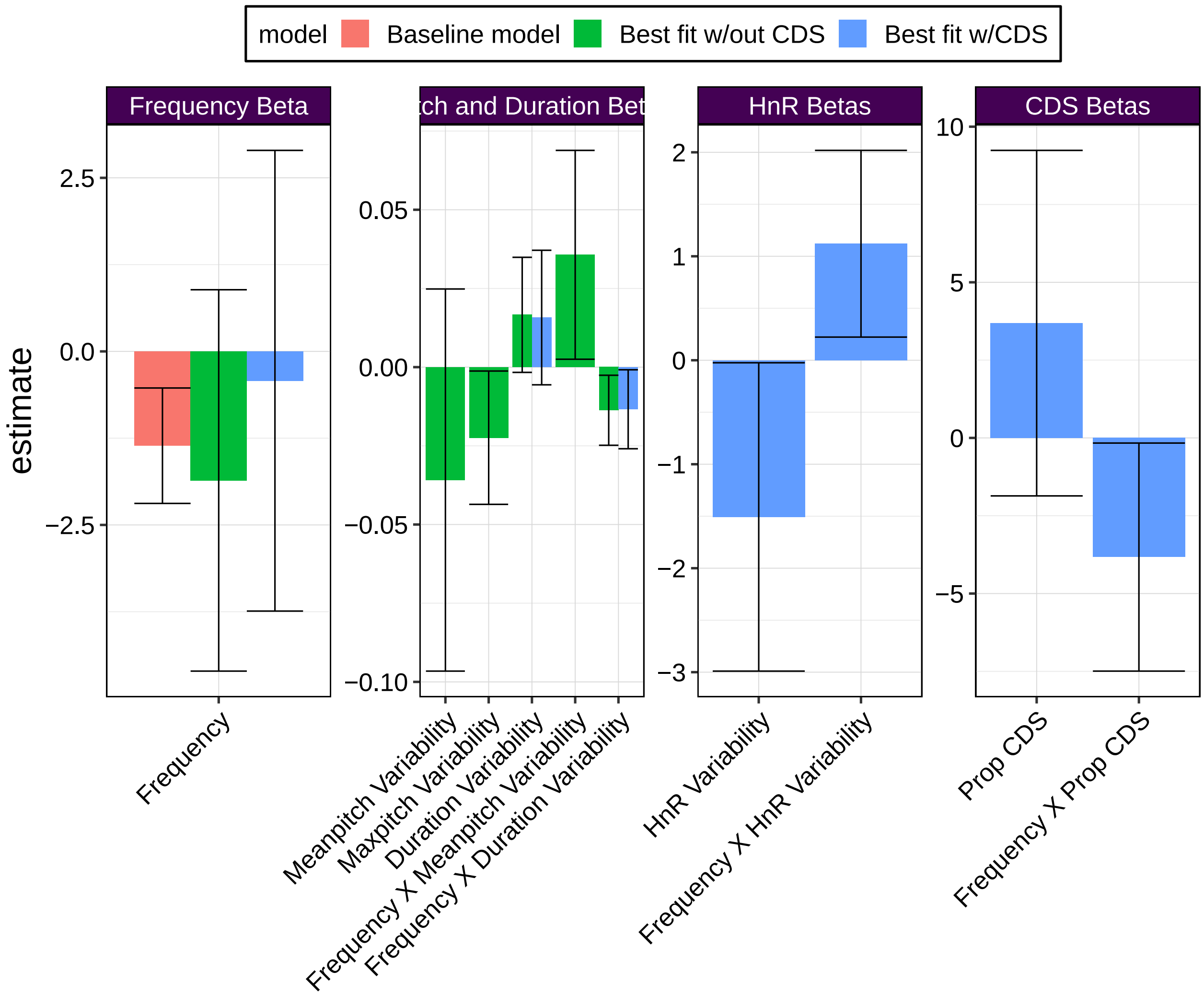

Figure 3. Beta estimates (change in months) and standard errors for models of the subset of data that includes CDS measures (32/44 infants). Left to right each panel shows frequency, pitch and duration, harmonics-to-noise (HnR), and CDS estimates, respectively. Color indicates which model the terms occurred in (pink = baseline, green = best fitting model without CDS, blue= best fitting model with CDS). Higher negative betas indicate earlier MonthFirstProduction. N.B. y-axes differ across facets. The fixed effects in the baseline AoA model for the CDS subset accounts for 4.8%, the fixed effects in the best fit model without CDS account for 12.1% of the variance; and for the best fit model with CDS account for 13.2% of the variance.

Table 4

Model comparison table showing, for the subset of data with CDS ratings (1) baseline model with just frequency, (2) best model based on backward model selection, and (3) best model with proportion of CDS. Model 1 accounts for 4 percent of the variance, Model 2 accounts for 12.1 percent, and Model 3 accounts for 13.2 percent

	Dependent variable:		
	MonthFirstProduction		
	Baseline model	Best fit w/o CDS	Best fit w/CDS
Frequency	-1.4^{***} (0.4)	-1.9 (1.4)	-0.4 (1.7)
Duration Variability		0.02^{*} (0.01)	0.02 (0.01)
Maxpitch Variability		-0.02^{**} (0.01)	
Meanpitch Variability		-0.04 (0.03)	
HnR Variability			-1.5^{**} (0.8)
Proportion CDS			3.7 (2.8)
Frequency × Duration Variability		-0.01^{**} (0.01)	-0.01^{**} (0.01)
Frequency × Meanpitch Variability		0.04^{**} (0.02)	
Frequency × HnR Variability			1.1^{**} (0.5)
Frequency × CDS			-3.8^{**} (1.9)
Constant	18.0^{***} (0.8)	20.5^{***} (2.3)	18.0^{***} (2.5)
Observations	187	187	187
Note:			*p<0.1; **p<0.05; ***p<0.01

said 1SD *less* variably in duration were predicted to be produced 1.25 month sooner than lower frequency words heard 1SD *more* variably in mean pitch, see Figure 4.

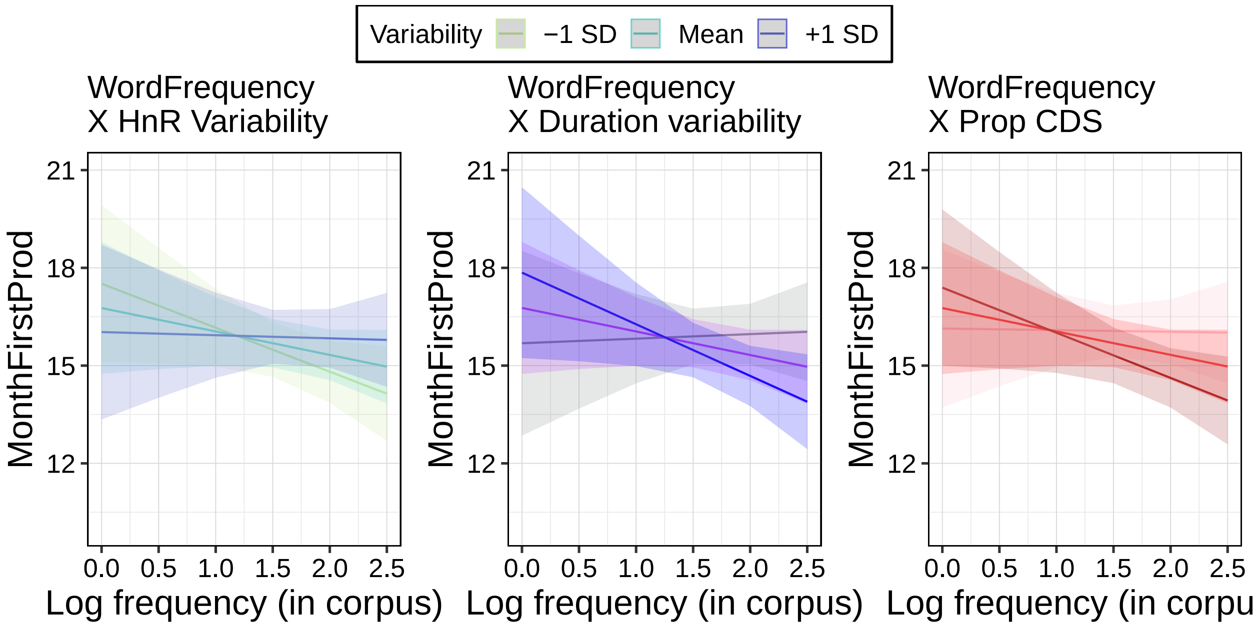

Figure 4. Visualization of predicted values for significant interactions. Left: Interaction between frequency and meanpitch variability. Right: Interaction between frequency and duration variability. Lines represent the predicted data for the mean variability value, +/- 1SD.

The model also included a significant interaction between frequency and harmonics-to-noise variability ($t(163.12) = 2.45$, $p = .015$, d = 0.38). That is, more frequent words that were said 1SD *less* variably in harmonics-to-noise ratio were predicted to be produced 2.40 months sooner than frequent words produced 1SD *more* variably in harmonics-to-noise ratio. Less frequent words that were produced 1SD *more* variably in harmonics-to-noise ratio were predicted to be produced 1.75 months sooner than lower frequency words produced 1SD *less* variably in harmonics-to-noise ratio. Lastly, there was a significant interaction between frequency and proportion of CDS ($t(150.03) = -2.05$, $p = .042$, d = -0.33). In this case, more frequent words that were produced in CDS more often were predicted to be produced 2.12 months sooner than those produced in CDS 1SD less often, while lower frequency words that were produced in CDS 1SD less often were predicted to be produced 0.64 months sooner than those produced in CDS more often.

We next compared the best fit model with and without CDS (which are not nested) via AIC. The AIC value for the model without CDS is 822.62, while for the model with CDS is 796.72, suggesting the addition of CDS improves model fit overall.

Discussion

The current study tested whether the acoustic variability with which children hear words in everyday life is related to their productions of those same words. We found that it was: hearing words more variably and in child-directed-speech influenced when those words were first produced.

Our analyses were built on a baseline model that include frequency for each word for each child, from a yearlong corpus. As predicted, hearing a word more resulted in producing that word earlier; frequency accounted for ~7% of the variance in month of first production. This is consistent with a well established effect of frequency, wherein earlier learned words tend to also be more frequent (see Ambridge, 2015; Frank et al., 2020; Goodman, Dale, & Li, 2008). These previous analyses typically calculate word frequency on a more global level - e.g. how frequent is the word 'baby' in speech to children in English? We build on this, showing that this holds on an individual level: how often a specific child heard the word 'baby' is predictive of when that same child produces it (see also Swingley & Humphrey, 2018).

Across analyses, we also found that the effects of variability varied, sometimes predicting an earlier and sometimes a later month of first production. Across all analyses, hearing *more* variability in duration resulted in producing the word earlier. The patterns for pitch-based measurements were less consistent. While more variability in max pitch resulted in earlier productions, variability in mean pitch interacted with frequency such that for higher frequency words, less mean pitch variability predicted earlier learning. In our analysis including proportion of CDS, we also found effects of CDS as well as Harmonics-to-noise ratio. In this case, higher frequency words produced with more variable harmonics-to-noise ratio were predicted to be produced later. On the other hand, higher frequency words that were produced in CDS more often were predicted to be produced sooner. This effect of CDS is consistent with research showing that acoustic properties of mother's speech in CDS predict vocabulary growth between 18 and 24 months (Han, De Jong, & Kager, 2023), and suggests that in addition to drawing infants' attention to speech (as evidenced by infant's overall preference for CDS; Cooper and Aslin (1990); Consortium (2020)), CDS also shapes children's word learning on a word-by-word basis (see also Jones, Cabiddu, Barrett, Castro, and Lee (2023)).

Why would these effects depend on frequency, and why would more variability only predict earlier production sometimes? First, we highlight that all the words used here are incredibly high frequency, within our corpus and generally in spoken English (see Perry, Perlman, & Lupyan, 2015). We cannot speak to how this would play out with truly low frequency words (which infants produce extremely rarely). Speculating based on the variable frequency of words in our corpus and across our participants, higher frequency may, for instance, give children more opportunities to make inferences about words on various levels of linguistic representation (how they sound, what they mean, etc.).

It may also be easier for infants to abstract across some acoustic properties, relative to others, in order to learn the bounds of how a word should be said. For example, in some contexts, infants do not recognize words that are presented in a different pitch (Singh, White, & Morgan, 2008), suggesting that they attend to pitch information as a cue to word identity. Thus, high pitch variability may be salient for infants, particularly for frequent words that are also more likely to be produced across talkers and contexts. Similarly, more variability in harmonics-to-noise ratio, which captures aspects of voice quality, may overlap with differences in affect, which have also been shown to influence word recognition (Singh, 2008).

In contrast, our results are compatible with the idea that unlike pitch, duration may be less salient for infants, or easier for them to abstract across tokens. While variation in duration can mark lexical stress and therefore carry meaning (PERfect vs. perFECT), it does so less consistently. If infants are sensitive to this, they may factor it in as part of e.g. a cue-weighing process (as proposed by e.g. Apfelbaum & Mcmurray, 2011; Hoehle et al., 2020), determining relevant parameters with increased exposure. Thus, we suggest that more experience may be required for abstracting across variability in pitch and harmonics-to-noise ratio relative to variability in duration. We look forward to future research directly testing this possibility. Either way, the current study suggests that not only do infants overcome a possible challenge posed by variability in duration, but they harness it during the word learning process.

While our models including acoustic variability and CDS accounted for twice as much variance as frequency alone, the vast majority of variance predicting when infants would produce these high frequency nouns remained unexplained. What else may contribute to when a word is first produced? Frank et al. (2020) found that, cross-linguistically, words are more likely to be learned if they are higher in concreteness (e.g. dog vs. happy), if they appear in shorter sentences, or in isolation. Roy, Frank, DeCamp, Miller, and Roy (2015) find similar results for a single child followed longitudinally - more frequent words, shorter words and words heard in shorter sentences tended to be produced earlier. More recent research has found that wordform variability for the same lemma (e.g. dog, doggy) also contributes to word learning (Moore & Bergelson, 2021), and differently so for higher and lower frequency words. Meaning and topic also certainly plays a role in what words children produce. For instance, across 15 languages, the first 10 words produced by children consist primarily of important family members, routines, or sounds (Frank et al., 2020). Incorporating these factors alongside acoustic variability is an exciting future direction for this work.

Our findings highlight that the acoustic variability infants hear in their input, on an *individual* level, is an important aspect to consider in our theories of language development and word production in particular. Of course, our findings focus on speech input in monolingual English-speaking homes, with typically developing infants. The extent of acoustic variability children hear is likely to vary cross-linguistically, and across contexts with more speakers of different ages. Future research will need to explore to what extent these findings generalize across linguistic communities. Nonetheless, our acoustic variability metrics combined accounted for almost as much variance as frequency alone in predicting when infants would produce specific words. Furthermore, when measurements of CDS are included, word learning was best explained by both the speech register and acoustic variability with which that word was heard. While it is perhaps unsurprising that we are unable to factor in all the sound-, meaning-, and individual-specific-properties that may predict the production of a given word, it is all the more meaningful that relatively low-level acoustic properties sampled from ~70 hours of each infant's input across a year have a measurable effect on when a given child produced specific words. While the exact mechanism by which different sources of variability shape learning remains an open question, acoustic variability may shape infant's expectations about how a word can sound, which in turn may drive their earliest efforts to produce these words themselves.

References

Albin, A. (2014). An architecture for controlling the phonetics software "Praat" with the R programming language. *Journal of the Acoustical Society of America*, *135*(4), 2198.

Ambridge, B. (2015). *The ubiquity of frequency effects in first language acquisition.*

Apfelbaum, K. S., & Mcmurray, B. (2011). Using variability to guide dimensional weighting: Associative mechanisms in early word learning. *Cognitive Science*, *35*(6), 1105–1138. https://doi.org/10.1111/j.1551-6709.2011.01181.x

Bergelson, E. (2017). *Bergelson Seedlings HomeBank Corpus.* https://doi.org/10.21415/T5PK6D

Bergelson, E., Amatuni, A., Dailey, S., Koorathota, S., & Tor, S. (2018). Day by day, hour by hour: Naturalistic language input to infants. *Developmental Science*, e12715. https://doi.org/10.1111/desc.12715

Bulgarelli, F., & Bergelson, E. (2019). Look who's talking: A comparison of automated and human-generated speaker tags in naturalistic day-long recordings. *Behavioral Research Methods*, *1–13*. https://doi.org/10.3758/s13428-019-01265-7

Bulgarelli, F., & Bergelson, E. (2022). Talker variability shapes early word representations in English-learning 8-month-olds. *Infancy*, 1–28. https://doi.org/10.1111/infa.12452

Bulgarelli, F., & Bergelson, E. (2023). Talker variability is not always the right noise: 14 month olds struggle to learn dissimilar word-object pairs under talker variability conditions. *Journal of Experimental Child Psychology*, *227*, 105575. https://doi.org/10.1016/j.jecp.2022.105575

Bulgarelli, F., Mielke, J., & Bergelson, E. (2021). Quantifying Talker Variability in North-American Infants' Daily Input. *Cognitive Science*, *46*(1), e13075. https://doi.org/10.1111/cogs.13075

Consortium, M. (2020). Quantifying sources of variability in infancy research using the infant-directed-speech preference. *Advances in Methods and Practices in Psychological Science*, *3*(1), 24–52.

Cooper, R. P., & Aslin, R. N. (1990). Preference for Infant-directed Speech in the First Month after Birth. *Child Development*, *61*(5), 1584–1595. https://doi.org/10.1111/j.1467-8624.1990.tb02885.x

Frank, M. C., Alcock, K. J., Arias-Trejo, N., Aschersleben, G., Baldwin, D., Barbu, S., … Soderstrom, M. (2020). Quantifying Sources of Variability in Infancy Research Using the Infant-Directed-Speech Preference. *Advances in Methods and Practices in Psychological Science*, *3*(1), 24–52. https://doi.org/10.1177/2515245919900809

Galle, M. E., Apfelbaum, K. S., & Mcmurray, B. (2015). *The Role of Single Talker Acoustic Variation in Early Word Learning The Role of Single Talker Acoustic Variation in Early Word Learning*. https://doi.org/10.1080/15475441.2014.895249

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, *35*(3), 515–531. https://doi.org/10.1017/S0305000907008641

Graf Estes, K., & Hurley, K. (2013). Infant-directed prosody helps infants map sounds to meanings. *Infancy*, *18*(5), 797–824. https://doi.org/10.1111/infa.12006

Han, M., De Jong, N. H., & Kager, R. (2023). Relating the prosody of infant-directed speech to children's vocabulary size. *Journal of Child Language*, 1–17. https://doi.org/10.1017/S0305000923000041

Hoehle, B., Fritzsche, T., Meb, K., Philipp, M., & Gafos, A. (2020). Only the right noise? Effects of phonetic and visual input variability on 14-month-olds' minimal pair word learning. *Developmental Science*, 0–2. https://doi.org/10.1111/desc.12950

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). *Sources of variability in children ' s language growth*. *61*, 343–365. https://doi.org/10.1016/j.cogpsych.2010.08.002

Jones, G., Cabiddu, F., Barrett, D. J. K., Castro, A., & Lee, B. (2023). How the characteristics of words in child-directed speech differ from adult-directed speech to influence children's productive vocabularies. *First Language*, 01427237221150070.

https://doi.org/10.1177/01427237221150070

Ma, W., Golinkoff, R. M., Houston, D. M., & Hirsh-Pasek, K. (2011). Word Learning in Infant- and Adult-Directed Speech. *Language Learning and Development*, *7*(3), 185–201. https://doi.org/10.1080/15475441.2011.579839

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal forced aligner: Trainable text-speech alignment using kaldi. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *2017-Augus*, 498–502. https://doi.org/10.21437/Interspeech.2017-1386

Moore, C., & Bergelson, E. (2021). *Wordform variability in infants' language environment and its effects on early word learning*. OSF Preprints. https://doi.org/10.31219/osf.io/n3phk

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *The Journal of the Acoustical Society of America*, *85*(1), 365–378. https://doi.org/10.1121/1.397688

Newman, R. S., Rowe, M. L., & Bernstein Ratner, N. (2016). Input and uptake at 7 months predicts toddler vocabulary: The role of child-directed speech and infant processing skills in language development. *Journal of Child Language*, *43*(5), 1158–1173. https://doi.org/10.1017/S0305000915000446

Perry, L. K., Perlman, M., & Lupyan, G. (2015). Iconicity in English and Spanish and Its Relation to Lexical Category and Age of Acquisition. *PLOS ONE*, *10*(9), e0137147. https://doi.org/10.1371/journal.pone.0137147

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*(2), 339–349. https://doi.org/10.1111/j.1467-7687.2008.00786.x.Speaker

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech vocabulary development. *Child Development*, *83*(5), 1762–1774. https://doi.org/10.1111/j.1467-8624.2012.01805.x

Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., & Roy, D. (2015). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(41), 12663–12668. https://doi.org/10.1073/pnas.1419773112

Singh, L. (2008). Influences of high and low variability on infant word recognition. *Cognition*, *106*(2), 833–870. https://doi.org/10.1016/j.cognition.2007.05.002

Singh, L., White, K. S., & Morgan, J. L. (2008). Building a Word-Form Lexicon in the Face of Variable Input: Influences of Pitch and Amplitude on Early Spoken Word Recognition. *Language Learning and Development*, *4*(2), 157–178. https://doi.org/10.1080/15475440801922131

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Letters to Nature*, 381–383. https://doi.org/10.1038/41102

Swingley, D., & Humphrey, C. (2018). Quantitative Linguistic Predictors of Infants' Learning of Specific English Words. *Child Development*, *89*(4), 1247–1267. https://doi.org/10.1111/cdev.12731

Yamashita, T., Yamashita, K., & Kamimura, R. (2007). A Stepwise AIC Method for Variable Selection in Linear Regression. *Communications in Statistics - Theory and Methods*, *36*(13), 2395–2403. https://doi.org/10.1080/03610920701215639