Cognitive Psychology 122 (2020) 101308

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

From babble to words: Infants' early productions match words and objects in their environment

Catherine Laing*, Elika Bergelson

Centre for Language and Communication Research, Cardiff University, UK
Department of Psychology and Neuroscience, Duke University, USA

ARTICLE INFO

Keywords:
Infant production
Phonological development
Babble
Prelinguistic vocalizations
Input effects
Language development

ABSTRACT

Infants' early babbling allows them to engage in proto-conversations with caretakers, well before clearly articulated, meaningful words are part of their productive lexicon. Moreover, the well-rehearsed sounds from babble serve as a perceptual 'filter', drawing infants' attention towards words that match the sounds they can reliably produce. Using naturalistic home recordings of 44 10–11-month-olds (an age with high variability in early speech sound production), this study tests whether infants' early consonant productions match words and objects in their environment. We find that infants' babble matches the consonants produced in their caregivers' speech. Infants with a well-established consonant repertoire also match their babble to objects in their environment. Our findings show that infants' early consonant productions are shaped by their input: by 10 months, the sounds of babble match what infants see and hear.

1. Introduction

In order for infants to begin producing their first words, they must first begin matching the speech sounds they can produce with the words they understand. Previous studies have established the caregiver's role in this process, as they initiate a 'feedback loop' between the sounds of babble and the sounds of speech (Albert et al., 2018; Goldstein et al., 2003; Goldstein & Schwade, 2008; Gros-Louis et al., 2014; Wu & Gros-Louis, 2014). For instance, well-timed caregiver feedback has been found to increase the 'speechlikeness' of early babble in lab studies (Goldstein et al., 2003; Goldstein & Schwade, 2008). This highlights infants' attunement to their input, and points to an interface between infants' perception of the ambient language and their own babble production. However, it does not address the shift that infants must make from the supposedly random syllables of babble to the memorized and planned sounds of their early words. Building on these previous lab-based studies, the present study uses naturalistic home recordings to investigate how infants' early-established consonant inventories shape their babble production on-line, in relation to caregiver speech and objects in their point of focus. This may have important implications for our understanding of how infants transition from babble to words.

1.1. The role of early babble

Jakobson (1940/1968, p. 24) famously referred to babble as a "purposeless egocentric soliloquy…[a] biologically oriented 'tongue delirium'." According to his account, the sounds produced in babble are phonetically meaningless and unconnected to the

* Corresponding author at: Centre for Language and Communication Research, Cardiff University, Cardiff CF10 3EU, UK.
E-mail address: laingc@cardiff.ac.uk (C. Laing).

https://doi.org/10.1016/j.cogpsych.2020.101308
Received 23 January 2020; Received in revised form 30 April 2020; Accepted 15 May 2020

language being learned (see Oller, 2000). On one hand, some support for this view can be seen in the universality of babble cross-linguistically. Namely, infants generally reach defined 'stages' of phonological development at roughly equal rates, and the canonical syllables produced in babble tend to be phonetically similar across languages (Oller, 2000). On the other hand, influence from the input has also been found to affect babble production, and subsequent research has established several convergent lines of evidence to show that babble is neither wholly egocentric, nor unrelated to the ambient language.

Indeed, rather than babble occurring in a vacuum, research has found clear continuity between babble and first words, with early babble becoming increasingly language-like and speech-like (Menn & Vihman, 2011; Oller, Wieman, Doyle, & Ross, 1976; Vihman, Macken, Miller, Simmons, & Miller, 1985). For example, in an analysis of home-recorded data taken weekly between age 0;9 to 1;2, Vihman et al. (1985) show how babble production becomes more 'speech-like' over time, both in interactive and solitary situations (i.e. with and without direct input from a caregiver). Furthermore, they found that babble production did not cease once the infants began to produce words, though the authors do report a change in the phonological properties of babble after the onset of word production. The phonetic structure of infants' earliest words shows clear similarities to the properties of their babble (Oller et al., 1976), suggesting that infants draw upon what is articulatorily accessible to them when acquiring and producing their first words. That is, infants capitalize on the well-rehearsed sounds and structures from babble when establishing their early lexicon.

Importantly, babble has been identified as a predictor of later word production. Wu and Gros-Louis (2014) show that the number of babble vocalizations produced at 10 months predicts language outcomes five months later, while McGillion et al. (2017) find age of babble onset to be a significant predictor of age of first word production. This may be because babble has a direct role in supporting word learning: Goldstein, Schwade, Briesch, and Syal (2010) found that infants were able to learn word-object associations more readily when adults responded contingently to their object-directed vocalizations with object labels. The authors posit that early vocalizations in *social contexts* in particular support language development.

Relatedly, input has been established as playing a key role in the transition from babble to words. Boysson-Bardies and colleagues find that the phonetic features of babble map broadly onto properties of the ambient language (de Boysson-Bardies, Halle, Sagart, & Durand, 1989; de Boysson-Bardies & Vihman, 1991): language-specific place and manner categories were found to distinguish the consonants of babble across infants acquiring French, English, Japanese and Swedish (de Boysson-Bardies & Vihman, 1991), and similar findings were identified in the babbled vowels of infants acquiring English, French, Cantonese and Algerian Arabic (de Boysson-Bardies et al., 1989). As well as an ambient language influence, contingent caregiver feedback to infant vocalizations also affects the phonological properties of babble. In two studies, Goldstein, King and West (2003) and Goldstein and Schwade (2008) show that vowel quality and consonant-vowel transitions become more speech-like following contingent caregiver feedback. By manipulating mothers' responses to their infants' vocalizations, they showed that only *contingent* responses generated more speech-like productions in infants' babble. These findings support a role for early social interactions in phonological development (cf. Warlaumont, Richards, Gilkerson, & Oller, 2014), and contradict a wholly 'egocentric' view of early vocalizations.

Thus, taken together, although there are certainly some universal aspects in the timing and content of babble development (Oller, 2000), there are also many indications that early babble is influenced by the ambient phonetic and social environment, and tied to subsequent – and necessarily language-specific – lexical learning.

1.2. Articulatory and perceptual limitations to early productions

The importance of babble production on a child's overall language development is well-established, showing a vital role for linguistic input in shaping the transition from babble to words. Although even deaf infants babble, the timing, prevalence, and types of vocalizations they make vary from hearing infants, likely because of the difference in direct auditory feedback that their babbling engenders (Oller & Eilers, 1988). Vihman (1993, 2017) explores the role of input in shaping early vocalization in more detail, proposing that infants draw on input from *their own productions* in early vocal development. That is, through attending to the sounds produced most often in their own babble, infants' perception of the input is shaped by what they are most easily able to produce themselves.

Expanding on this idea, McCune and Vihman (2001) analyzed prelinguistic vocalizations to show that some consonants are stably and consistently reproduced in babble. These stable consonants are known as 'vocal motor schemes' (hereafter **VMS**), and indicate the establishment of a consonant repertoire. The establishment of such a repertoire may signify a crucial turning-point in phonological development: Vihman (1993) proposes that stable VMS consonants serve as an 'articulatory filter', which establishes a perceptuo-motor link between what an infant hears and what they produce. That is, an infant's perception of the language in their input is filtered through what they know from production, so that words containing the sounds they can most easily retrieve and reproduce are most salient to the child.

Recent work also links infants' own *perception* to their special VMS consonants (Majorano, Vihman, & DePaolis, 2014; Majorano, Bastianello, Morelli, Lavelli, & Vihman, 2019), offering a possible explanation for why an infant's most commonly babbled consonants occur in their first words: infants' perception *and* production may be influenced by a consonant repertoire established through babble, leading to acquisition of words that match the consonants that are most stably represented in their output. This has been shown by Majorano et al. (2014; see also DePaolis, Vihman, & Keren-Portnoy, 2011), who tested Italian infants' perception of novel words containing consonants that either matched or did not match their own VMS consonants. Infants with one established VMS consonant attended longer to words that contained their VMS: infants with [t] in their VMS repertoire attended longer to novel words such as *dite*, and infants with [p] in repertoire attended longer to words such as *pabo*. This raises the question of how the articulatory filter might shape the transition from babble to speech: if infants' attention is drawn towards words that contain the sounds they can most easily produce, does this in turn make them more likely to attempt to reproduce these sounds in their babble?

This could explain why infants' first words are so similar to their babble in both phonetic and phonological content (McCune & Vihman, 2001; Oller et al., 1976).

Work by Goldstein and Schwade (2008) addresses this question to some degree, though they did not take into account the infants' VMS repertoire. Alongside their analysis of on-line linguistic experience on the phonological structure of babble, they tested whether the *sounds* of infants' babble matched the consonants in preceding caregiver speech. They found no relation between the two. This conclusion was based on further analysis of the mother-infant interactions reported above, which assessed the proportion of matching consonants between 9.5-month-olds' productions and their mothers' immediately-preceding utterances.[1] However, this analysis failed to consider infants' own production abilities, which may be crucial to uncovering specific links between the sounds in babble and caregiver productions.

That is, before ruling out the possibility of a match between infants' babble and preceding caregiver speech, it may be important to consider both the articulatory filter and infants' articulatory and perceptual abilities. Goldstein and Schwade (2008) analyzed all of the consonants and vowels in mothers' preceding utterances across all word positions, compared with all of the consonants and vowels of infants' responses. We propose that this approach may not have captured the nuances of early phonological acquisition. First, it is important to consider which features of the caregivers' input we can expect infants to perceive and/or (re)produce.[2] Perceptually, infants are more sensitive to word onsets than offsets, particularly in stressed syllables (Swingley, 2005; Vihman, Nakai, Depaolis, & Hallé, 2004). They also have difficulty well into year two in distinguishing voicing contrasts, e.g. [p] vs. [b], particularly at the ends of words (Zamuner, 2006). Notably, when it comes to production, not all consonants are equal: in a meta-analysis of consonant acquisition across 27 languages, McLeod and Crowe (2018) show that infants begin to produce stop consonants (e.g. [p]) early, but approximants and fricatives (e.g. [ɹ], [f]) as late as the third year. Furthermore, while many 9.5-month-olds can produce a sound like [b] or [p], they don't reliably distinguish voicing contrasts in their own production until almost a year later (Macken, 1980), and they rarely produce certain other types of sounds. With this in mind, it is necessary to consider the phonological resources that are available to infants when analyzing their babble production in relation to adult speech.

Furthermore, if infants attend preferentially to words that contain their well-established, stable consonants (i.e. their VMS) in an experimental setting (e.g. Majorano et al., 2014), then it is possible that they do so in an ecologically more realistic setting as well. If this is the case, then they may be more likely to reproduce the sounds that match their VMS consonants. In turn, this may aid earlier acquisition of the forms that match infants' VMS, as has been noted in previous studies (McCune & Vihman, 2001). Infants would therefore also be more likely to begin pairing their VMS consonants with objects they know in their environment, initiating the link between what an infant knows about the world (i.e. object names) and the sounds they can most easily produce (their VMS). The present study tests these proposals by considering the intermediate step between babble and word production: infants' pairing of their VMS consonants to words and objects in their environment.

1.3. The present study

In what follows, we expand on findings from Goldstein et al. (2003) and Goldstein and Schwade (2008) with a consideration of VMS (DePaolis et al., 2011; McCune & Vihman, 2001), extending the articulatory filter account (Vihman, 1993). We determine whether infants' babbled responses to their input may be influenced by their own production experience, rather than being shaped directly by the preceding caregiver utterance (as in Goldstein & Schwade, 2008). We analyze infants' pre-lexical productions in relation to their VMS to consider whether pre-linguistic vocalizations are related to immediately-preceding input speech. We also extend this analysis to consider whether babble is related to objects attended to by the infant at the time of vocalization; that is, to observe whether infants' own object-directed vocalizations may be phonetically linked to the names of the objects they attend to. This will allow us to test whether babbled consonants are related to stimuli in the infant's immediate environment, filtered through each infant's own articulatory repertoire. More concretely: given an infant's babble, we consider two aspects of the context in which it occurred; i) the immediately-preceding speech input from the caregiver (caregiver prompt),[3] and ii) the objects the infants attended to during babble production (attended object). We expect established consonants (i.e. their VMS) to affect babble in two ways:

1) We predict that infants with a stable VMS repertoire will produce more babble that matches a caregiver prompt or attended object context **overall** than infants with no stable VMS in their repertoire.
2) We predict that infants with stable VMS consonants will be more likely to produce **their particular VMS consonants** in a matching caregiver prompt or attended object context. For example, an infant with only [t] in VMS will be more likely to respond with [ta] to a /t/-initial word such as *teddy* than with [ba] to a /b/-initial word such as *baby*. In line with (1), we do not expect to see any relation between babble and input for infants who lack VMS consonants.

Support for these predictions would suggest that the transition to first words is guided by the articulatory filter, as infants tune in to input that matches their stable consonant inventory (i.e. their VMS). This would constitute the first evidence linking input-contingent productions with infants' individual production inventories, suggesting that babble production is indeed affected by the

[1] The authors do not specify whether these were single- or multi-word utterances.

[2] While infants' babbles (even canonical ones, which have adult-like consonant-to-vowel transitions) can often contain recognizable speech sounds from the adult phonology, they are of course not as fully developed as true adult phonemes (Ramsdell, Oller, Buder, Ethington, & Chorna, 2012).

[3] We use 'prompt' for expositional ease, rather than to imply these were deliberate, intentional prompts.

input, *if* the infant has a consonant repertoire to draw from.

2. Methodology

The current work utilizes observational data collected as part of a yearlong study. The full study included home recordings, lab experiments and development questionnaires taken on a monthly basis between the ages of 6 and 18 months (see Bergelson, Amatuni, Dailey, Koorathota, & Tor, 2019 for full details of home recordings). The present analyses focus solely on the home-recorded data at 10–11 months, when VMS are first established (DePaolis et al., 2011; McGillion et al., 2017). We also consider family demographics data taken at the start of data collection, when the infants were 6 months old. Demographics data for each participant, along with all other supplementary materials, are available on the GitHub and Open Science Framework repositories for this study at https://github.com/cathelaing/Laing-Bergelson-CongruentBabble and https://osf.io/xbf2h/.

2.1. Participants

Forty-four infants took part in the yearlong study. The targeted sample size for the lab-and-home sample was 48 infants during an eight-month enrollment window for the study; 44 infants were enrolled and retained over this window. This n was chosen due to sub-studies that split the sample into three groups, for which 16 was the standardly accepted minimum sample size (see Bergelson et al., 2019). The final sample included one pair of dizygotic twins; a further two families dropped out in the early stages of data collection. The infants (21 females) were growing up in largely white, middle-class households in upstate New York. Thirty-three of the mothers reported having a BA or higher. Forty-two infants were reported as White, two as biracial. All infants were full-term with no reported speech or hearing problems. Data collection was carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki; written informed consent was obtained from a parent/guardian for each child prior to data collection. All procedures were approved by the IRB at the University of Rochester (where the data were initially collected) and Duke University (where they continue to be analyzed).

2.2. Data

One day-long audio recording using a LENA recorder (LENA Research Foundation, 2018), and one hour-long session recorded by video were obtained on two separate days during each month of data collection. These recordings generally occurred within one week of infants turning one month older (e.g. 10mo., 11mo., etc.). During the audio recordings, infants wore a waistcoat holding a LENA recorder in a small chest pocket, which captured all speech in the infant's surroundings. Caregivers were instructed to switch the recorder on when the infant awoke in the morning, and to leave it running until the battery ran out (~16 h) or until the infant's bedtime. Researchers later collected the recorders from the families' homes. For the video recordings, the infants wore two Looxcie video cameras attached to a hat or headband – one pointing slightly upwards, one slightly downwards. This provided a view of the scene from the infant's perspective. If the infant seemed likely to remove the camera during the recording, the caregiver also wore a head-mounted Looxcie camera. A camcorder (Panasonic HC-V100 or Sony HDR-CX240) was also set up in the home. Research assistants set up the equipment and left it running for one hour, then returned to collect the cameras. The video-recordings were merged and annotated in DataVyu (DataVyu, 2014), while the audio-recordings were processed by LENA's proprietary algorithm and then manually annotated as described below. See Bergelson et al. (2019) for further details of the original study and its annotation pipeline.

2.3. Procedure

Our procedure consisted of two stages: 1) determining infants' consonant repertoires from the daylong audio recordings, and 2) annotating infants' consonant productions in the video recordings in relation to our two types of context: objects in the infants' environment, and caretakers' language. Both stages of annotation were carried out by the first author, a trained phonetician.[4] See S1 for full details of the transcription process. Reliability measures for each stage of the annotation are reported below. Since there are many methodological details pertinent to classifying infants' consonants and the contexts in which they occurred that must precede the subsequent statistical analyses in the results, we provide intermediary descriptives, tables, and figures within the methods section.

2.3.1. Establishing consonant repertoire from day-long audio recordings

We first established infants' consonant repertoires (VMS). To our knowledge, we are the first to use day-long recordings for this purpose, which required relying on previously established approaches and adapting them to our data, as we sought to determine whether infants had stable VMS.

[4] We adopt phonetic notation to denote infant productions, reserving phonemic notation when referring to adults' production (brackets and slashes, respectively). While adults generally treat infant sounds as if they were phonemic, early babble bears a complex relationship to adult phoneme, consonant, and vowel categories (Oller, 2000). As used here, for instance, [ba] refers to a syllable with an initial segment resembling a bilabial stop consonant at onset, but does not assume consistent close correspondence to the target phoneme /b/, though for simplicity we use the term 'consonant' to describe it.

Given previous results establishing that infants generally acquire their first VMS at around 10 months (DePaolis et al., 2011; McGillion et al., 2017), we began by analyzing all 44 audio recordings taken at 10 months. Using LENA's automatically-generated child-vocalization counts, we were able to leverage the entire daylong recording in order to extract the 30-minute segment in the day where the LENA algorithm determined that the child vocalized the most. We excluded or resampled in the case of crying/feeding (see Supplementary Material, S1). We then tallied supra-glottal consonants[5] produced in canonical syllables during that 30-minute segment.[6] Following McCune and Vihman (2001), we analyze only supra-glottal consonants with a complete closure in the vocal tract (i.e. not glides) as these are known to be relevant to the transition to first word production.

In line with previous work, we used these tallies to establish whether infants had a stable consonant repertoire ('with-VMS') or not ('no-VMS'; see Table 1) (DePaolis et al., 2011; McGillion et al., 2017).

First, adopting DePaolis et al.'s (2011) approach, we classed a child as 'with-VMS' if they produced one or more consonants ≥50 times in 30 min at 0;10. 13 infants were with-VMS according to these criteria (across all consonant tokens: M = 178.54, SD = 75.99, R = 86–412). 14 of the remaining 31 infants had low production at 0;10 and were classed as no-VMS (M = 8.5, SD = 8.46, R = 0–28).

The remaining 17 infants showed signs of a developing VMS consonant at 0;10 that merited further consideration. That is, a given consonant was dominant in the session (defined here as ≥20 tokens of the same consonant accounting for > 20% of overall consonant production, cf. Athari, Wang, Day, & Rvachew, 2019), but did not reach our initial criteria of ≥50 tokens (M = 29.5, SD = 8.64, R = 20–48; Mean % = 0.50, SD = 0.17, R = 0.23–0.9). As this suggests that these infants were on the cusp of establishing a VMS, we re-assessed VMS for this subset at 0;11. We only re-assessed infants who did not clearly meet our VMS criterion at 0;10, and so our final sample includes infants aged 10 (n = 27) and 11 (n = 17) months.

Using the same initial criterion applied at 10 months (≥50 of one or more consonants in 30 min), 14 of the 17 infants assessed at 11 months could be clearly classified as no-VMS (n = 6) or with-VMS (n = 8). The final three infants were consistent in producing the same consonant stably at 0;11 just as they were at 0;10 (≥20 productions of a consonant type, accounting for > 20% of all consonant tokens), but still short of the ≥50 mark; across both months, all three infants did produce > 50 tokens of the same consonant. Given that our criteria were more conservative than previous approaches (e.g. McCune & Vihman, 2001, required infants to reach ≥10 of a given consonants across 3 of 4 sessions), and that these infants demonstrated 'well-practiced and longitudinally stable vocal productions' (McCune & Vihman, 2001, p.671), we classed them as with-VMS. That said, all reported results were consistent when we removed these three infants from the analysis altogether.[7] See Table 2, S1 and Fig. S1. It is worth noting that given our goal of rigorously evaluating links between infants' babble and environmental context, and given that babble doesn't develop at the same age for all infants, our approach prioritized the distinctiveness of infants' stage of phonological development (i.e. with-VMS vs. no-VMS), rather than age.

Descriptively, 24/44 infants had at least one VMS consonant, and these all mapped broadly onto plosive and nasal phonetic categories (see Table 2, Fig. 1, Fig. S1 and Table S1 for further breakdowns by child and consonant); however, infants produced a wider variety of consonant types overall, including sounds that mapped onto adult liquid and fricative categories. A naïve research assistant re-coded six infants' 30-minute segments. There was 100% agreement regarding infants' VMS group (with-VMS: n = 4 or no-VMS: n = 2) and, for the with-VMS infants, which consonant(s) were classed as being in repertoire.

2.3.2. Annotating video data

The video data – taken within a week of the audio recording but on a different day (see Table 3) – was then phonetically transcribed for infant consonant productions (**CP**s) within canonical syllables. We analyzed 11-month video data for all infants whose audio-recordings were re-sampled at 0;11, and 10-month video data for those whose audio-recordings were only sampled at 0;10. Every consonant that the infant produced in their hour-long video was transcribed, as it was for audio. Following the initial annotation, a research assistant trained in phonetic transcription re-transcribed 10% of the original annotations. Coder agreement was 77% (Cohen's kappa = 0.72, z = 24.2), which is typical of the analysis of infant production (e.g. McGillion et al., 2017); disagreements were resolved through re-listening and discussion. After all 44 infants' data had been transcribed, we coded each consonant production for our two context variables (caregiver prompt and attended object) for each infant production. See Tables 1 and S2.

Caregiver prompt. Input speech in the 15 s preceding each CP, which included any live speaker captured in the video (usually mother or father; usually child-directed speech) was analyzed. We chose this timeframe based on previous studies showing turns between mother and infant (i.e. two conversational turns) to last around 10.5 s (Jaffe et al., 2001), combined with studies of parent-child interactions that show a child's early tendency to repeat material from earlier turns (Casillas, Bobb, & Clark, 2016; Dunn & Shatz, 1989). We tagged the most salient word in this 15 s segment as the 'caregiver prompt' and compared the infant's babble production in relation to this. Average duration between caregiver prompt and infant CP was 6.07 s (SD = 3.8); results were consistent when we reanalyzed the data with only prompts from < 10 s before the infants' CP.

[5] We focus on consonants, given wide variability in early vowels, and the difficulty of mapping them onto adult categories (Shriberg, Austin, Lewis, Mcsweeny, & Wilson, 1997).

[6] Following DePaolis et al. (2011), we collapsed voicing contrasts (e.g. [ba] vs. [pa]) as infants don't produce them distinctively until ~18 m (Macken, 1980); [babapaba] was tagged as four [p,b] instances.

[7] While our VMS criteria stem from our best effort to adapt previous approaches to the nature of our daylong audio-recordings, the data will be shared to allow others to implement other criteria as they see fit. See https://github.com/cathelaing/Laing-Bergelson-CongruentBabble.

Table 1
Glossary of acronyms and terminology as used in our analyses. See text for details.

Terminology	Definition
VMS	**V**ocal **M**otor **S**cheme, i.e. stable consonant in an infant's productive repertoire
with-VMS	Infant who produces one or more VMS consonants in the 30 min of their daylong recording with the most infant vocalizations
no-VMS	Infant who does not produce a VMS consonant in the 30 min of their daylong recording with the most infant vocalizations
CP	**C**onsonant **P**roduction, i.e. one consonant + vowel syllable (e.g. [ba]).
Caregiver Input	The concrete noun or otherwise stressed/salient word produced by the caregiver in the 15 s preceding the infant's vocalization (e.g. 'ball').
Attended Object	The object an infant attended to during their vocalization (e.g. 'cup').
Congruent CP	Consonant Production that matches the segmental properties of the context (e.g. infant produces [ba] after hearing 'ball' in the input, or while attending to a ball).
Incongruent CP	Consonant Production that does not match the segmental properties of the context (e.g. infant produces [ba] after hearing 'dog' in the input, or while attending to a dog).
INREP	For infants with stable Vocal Motor Schema (i.e. with-VMS infants), a consonant production **in** the infant's **rep**ertoire.
OUTREP	For infants with stable Vocal Motor Schema (i.e. with-VMS infants), a consonant production not in the infant's repertoire. All consonant productions are OUTREP for no-VMS infants.

Table 2
Number of infants by VMS (Vocal Motor Scheme) group according to age (0;10 and 0;11 months) and sex (number of females in parentheses). With-VMS is broken down into number of VMS. See text (Section 2.3.1) and SI for details regarding VMS classification and age.

Group	0;10	0;11
with-VMS	13(5)	11(8)
One VMS	6(2)	7(5)
Two VMS	5(2)	4(3)
Three VMS	2(1)	0
no-VMS	14(5)	6(3)
	27(10)	17(11)

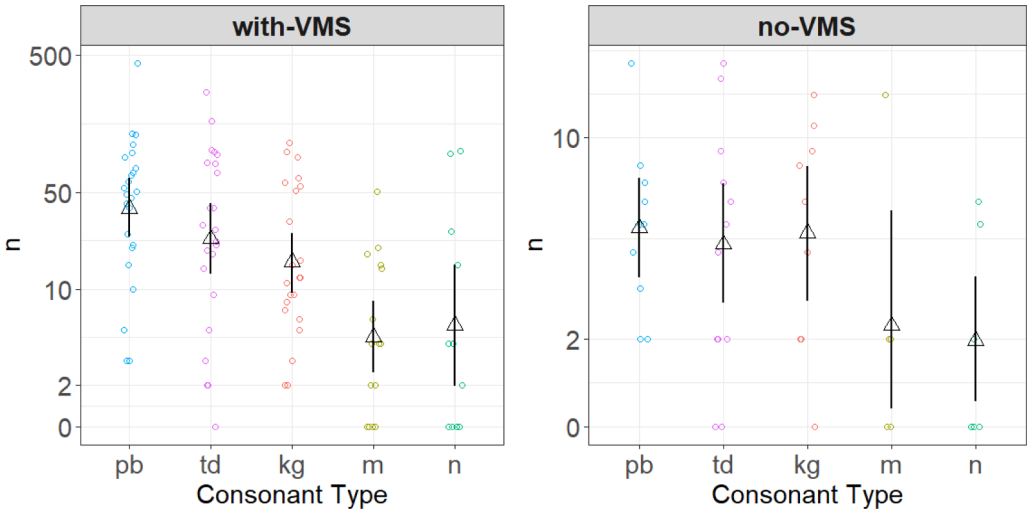

Fig. 1. Number of each consonant type produced in canonical babble by with-VMS and no-VMS infants in the audio recordings. Colored circles represent individual infants' consonant productions; points are jittered horizontally to clarify overlap. Black triangles represent the mean number of productions of each consonant type, with 95% bootstrapped confidence intervals. Y-axis utilizes log-transformed vertical spacing for visual clarity.

Table 3
Mean (SD) infant age in days for each data type, overall and across VMS groups.

Data type	All infants	with-VMS	no-VMS
Audio data	320.25 (14.99)	310.48 (15.35)	317.2 (14.32)
Video data	318.59 (14.87)	308.63 (15.25)	315.85 (14.06)

50% (n = 951, per infant: M = 21.6, SD = 18.6) of CPs were preceded by audible caregiver speech in this timeframe. The majority of these (74%) included a concrete noun, which we tagged as the caregiver prompt for that CP. For example, the utterance "let's make another pattern with the boxes" includes two nouns, one concrete (*boxes*) and one abstract (*pattern*); in this case we coded *boxes*. If two concrete nouns occurred in the segment, we selected the noun that was the most salient; that is, the noun that stood out more due to repetition, loudness, or modifications in pitch or duration (cf. Adriaans & Swingley, 2017). Of the 243 segments that did not include a noun, we coded the word that was impressionistically the most salient in the segment. See S2. A trained research assistant, naïve to the purpose of the study, coded 10% of the data in line with the procedure described here, showing 85% agreement with the original annotations (Cohen's k = 0.68, z = 35.5).[8]

Attended object. We also annotated whether or not the infant was attending to an object at the point of consonant production, based on whether a) it was clear from the video that they were looking directly at an object (i.e. the infant's face was visible on one of the camera feeds), b) they were pointing to an object while looking in its direction, or c) they were holding an object while looking in its direction. When available, we used the caretaker's label for a given object (e.g. an infant attended to a toy that the mother referred to as "Oscar", so we labelled this object as *Oscar*; see S2). For book-reading, we used the picture as the 'object', when relevant (e.g. if an infant was looking at a page in a book with a picture of a cow, we annotated *cow*, not *book*). Attended objects accounted for 62% of our CP data (n = 1,179 CPs). After removing instances when the infant was attending to an object but its identity or label was unclear (n = 78), we were left with 1,101 data points (M = 26.8, SD = 24.03) for the attended object analysis. Again, a trained research assistant re-annotated 10% of the data (including CPs with no attended object, n = 24), revealing an inter-coder agreement of 91.2% (k = 0.9, z = 69.1) for attended object tags.

Consonant production (CP). Next, the phonetic properties of these two context measures (i.e. caregiver prompt and attended object label) were compared to the infant's CP. In the vast majority of cases (98%), the first consonant in the caregiver prompt word or object label was annotated (e.g. if the prompt/object was *dog*, we coded this as the word-initial consonant /d/), except when multisyllabic words began with vowels, liquids or glottal fricatives (e.g. /a/, /ɹ/ or /h/); in these cases, the initial consonant of the stressed syllable was coded (e.g. /k/ in *avocado*, /m/ in *remote*), since stressed syllables are perceptually more salient to young infants than unstressed syllables (Vihman et al., 2004).[9] As above, voicing contrasts in the target forms were collapsed ('bat' or 'pig' were tagged in the /b,p/ category); similarly, word-initial fricatives in the adult word were collapsed to /s/, the only fricative produced by infants in our data. Monosyllables that did not contain a supra-glottal consonant in word-initial position (e.g. *whisk*, *ring*, *house*) were omitted from the analysis, given that we only considered infant productions that contained supra-glottal consonants, as these represent a shift towards speech-like vocalizations (McCune & Vihman, 2001). For both caregiver prompt and attended object, in around 5% of cases this meant there was no codable consonant for a word (e.g. neither *wow* nor *whisk* have a supra-glottal consonant with full closure in syllable-initial position. These instances were omitted from the analysis ($n_{\text{caregiver prompts}}$ = 40; $n_{\text{attended objects}}$ = 58).

2.4. Data aggregation and analysis

Data across all infants was aggregated and analyzed in R (R Core Team, 2018). All scripts and tabular data are available on GitHub. We calculated several outcomes. (1) *For all infants*, in the cases where there was a relevant context (either a caregiver prompt or an attended object, as described above), we determined whether their consonant production (CP) matched the context or not (CONGRUENT vs. INCONGRUENT). For instance, if a child's [ta] was preceded by mom's 'ball,' this would be an INCONGRUENT token; if preceded by mom's 'tub' it would be a CONGRUENT token. We then calculated the proportion of all CPs with a relevant context that were CONGRUENT for each child. (2) *For infants with a stable VMS repertoire*, we determined whether each CP in the video was in their individual VMS inventory as established in the audio-recording (INREP vs. OUTREP; see Table 1). For instance, if an infant only has [p,b] in their VMS, a 'pa' counts as an INREP token, while 'ta' would count as an OUTREP token. We then calculated the proportion of all CPs that were INREP for each of these children. For each context (caregiver prompt and attended object), we calculated the percent of CPs that were both in infants' VMS repertoire *and* congruent with the context (i.e. % of INREP *and* CONGRUENT CPs) for each of the context variables.

Critically, establishing 'chance' performance for matches between an infant's CP and their context is crucial here. For example, if infants' high production of [t] was driven by a universal articulatory bias towards [t] and/or a predominance of /t/ in caregivers' speech or objects in the environment, then our results might reflect general linguistic tendencies and not on-line influences from individual infants' surrounding environments. We opted to follow the approach of Goldstein and Schwade (2008) to circumvent this concern. Namely, we scrambled data across infants to determine what we might expect if infants' CPs were random. This approach controls for individual variability in the data, as well as taking into account consonant frequency in English, and so is more appropriate than a set value (e.g. 20% to represent the 5 VMS consonant categories), which would assume all consonant types are equally likely to occur. We scrambled the context variable data (separately for the caretaker prompt and attended object analyses) to compare infants' CONGRUENT and INCONGRUENT CPs to chance, pairing each CP with the onset consonant of a word/object from a randomly chosen CP in the data. We then compared results in the original vs. scrambled data, as reported in (3) below.

[8] While these results include all word categories, analyses including only nouns were consistent (S4, Table S2).

[9] Note also that Vihman et al. (2004) found that changes to word stress did not block recognition, while changes to the segmental properties of the stressed syllable onset did, suggesting that infants in our study will still be sensitive to the word-initial consonant of a multisyllabic word with second-syllable stress, e.g. *pyjamas*.

Table 4
Descriptive Statistics (mean(SD)) for **C**onsonant **P**roductions in video and audio recordings overall, and by VMS group. Tokens refer to individual CPs, types refer to distinct consonants. Columns 4 and 5 show mean(SD) for proportion of congruent tokens in the videos for the two context variables (group n in column header; one infant was not included in the Attended Object match data due to having 0 attended objects in their video recording).

	Group	Audio	Video	Caregiver prompt match (n = 44)	Attended object match (n = 43)
CP tokens	All	96.6(1 1 7)	43.55(34.13)	0.48(0.23)	0.38(0.27)
CP tokens	with-VMS	194.29(126.21)	60.13(34.59)	0.55(0.17)	0.49(0.27)
CP tokens	no-VMS	27.15(34.14)	23.65(20.59)	0.41(0.28)	0.25(0.22)
CP types	All	3.18(1.42)	4.34(1.27)		
CP types	with-VMS	4(0.88)	4.67(1.31)		
CP types	no-VMS	2.2(1.32)	3.95(1.15)		

2.5. Data analysis plan

We begin by considering overall babble production in the audio and video recordings to determine differences in quantity (consonant *tokens*) and variability (consonant *types*) between VMS groups. We then test our two hypotheses by analyzing infants' production in the video recordings, for each of our two context variables (caregiver prompts and attended objects).

To be clear: our first hypothesis is that having stable consonants in repertoire, i.e. being a with-VMS infant, will lead to babble that is more congruent with the context overall (caregiver prompt and/or attended object). We test this by comparing the proportion of congruent consonant productions (CPs) by VMS group (with-VMS vs. no-VMS) to each other, and to chance (i.e. the scrambled data).

Next we will test our second hypothesis: that among children with stable consonants in their repertoire, congruent consonant productions are more likely to be the *specific sounds* in their repertoire (i.e. INREP CPS) as opposed to sounds they cannot yet reliably and stably produce (i.e. OUTREP CPs). We test this by comparing the proportion of congruent INREP CPs in relation to the proportion of congruent OUTREP CPs. For instance, if an infant produced 10 CPs, two of which were both INREP *and* congruent, and three that were OUTREP *and* congruent, we would compare 20% INREP vs. 30% OUTREP; this analysis is only possible for infants with VMS consonants, i.e. the with-VMS group. Finally, we will compare the proportion of congruent OUTREP CPs produced by both with-VMS and no-VMS infants to consider the extent to which babble production is influenced by caregiver prompts/attended objects when we factor out the influence of VMS. See Illustration 1a–c.

All figures show non-transformed data unless otherwise specified. None of the infants were outliers in our dataset (defined as CP types or tokens ± 3 standard deviations from the mean). Since many aspects of the data we analyze are not normally distributed, we use Wilcoxon tests for two-sample and paired-comparisons throughout for ease of readability; in cases where the subset of data being analyzed did not differ from a normal distribution, Wilcoxon and t-tests rendered the same pattern of results. Reported effect sizes were generated using Cliff's Delta (δ) calculations for independent samples and regression coefficients (*r*) for paired samples. We include 95% bootstrapped confidence intervals (CI) for mean dependent variables. An initial analysis of infant age, sex, and maternal education showed that these variables did not account for significant variance in the number of CP types or tokens, in audio-recordings or in videos, and thus data are collapsed across these dimensions in all further analyses. See S3 for further details.

3. Results

3.1. Infant production: Audio recordings

On average, infants produced 96 consonant tokens in the half-hour segment of the daylong audio-recordings we analyzed. Variability was high (Range = 0–610; see Table 4); three infants (2 females) produced zero consonants. The most prominent consonant was [p,b] (n = 1,718), followed by [t,d] (n = 1,320) and [k,g] (n = 742; Fig. 1).

We began by testing whether there was any quantitative difference in babble production between groups. As expected, with-VMS infants produced more babble in the audio recordings than no-VMS infants, and this difference was significant for both types and tokens (Types: Est.Diff. = 1.99, $\delta = 0.74$, $p < .001$, 95% CI = [0.45, 0.89]; Tokens: Est.Diff = 132, $\delta = 0.99$, $p < .001$, 95% CI = [0.95, 0.99]; by Wilcoxon Rank Sum test; see Table 4 & Fig. 1).

3.2. Infant production: Video recordings

On average, infants produced 43.55 CP tokens in the hour-long video recordings and 4.34 CP types. See Table 4 and Fig. 2. Once again, with-VMS infants produced significantly more tokens than no-VMS infants (60.13 vs. 23.65), and the number of CP types was marginally higher for with-VMS infants (4.67 vs. 3.95; see Tables 4 and 5). We next considered whether infants' consonant productions during the videos matched their stable VMS repertoire as determined by the audio-recordings; this was only assessable for infants who had such a VMS repertoire (i.e. with-VMS infants).

Indeed, with-VMS infants produced significantly more consonants that were in their stable VMS repertoire (INREP consonants) relative to other consonants (OUTREP consonants), as assessed by comparing (total number of INREP tokens)/(total number of VMS

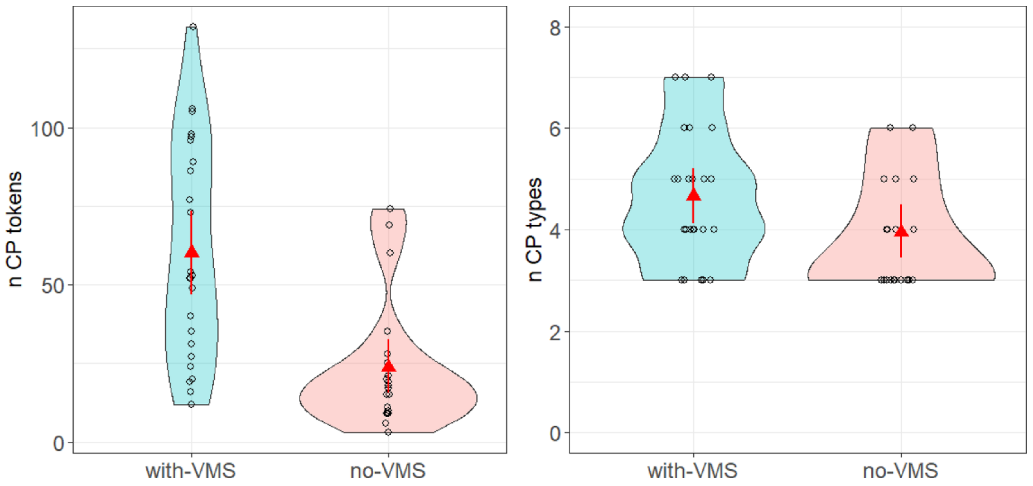

Fig. 2. Number of consonants (tokens: left panel, types; right panel) produced by with-VMS and no-VMS infants in the video recordings. Circles (jittered horizontally for clarity) represent individual infants' consonant productions; triangles represent the mean number of productions across groups, with 95% bootstrapped confidence intervals.

Table 5
Results from statistical tests comparing VMS groups with CP production, including 1) Wilcoxon Rank Sum tests comparing total number of CP types and tokens produced by infants in each VMS group. 2) Wilcoxon Rank Sum and paired Wilcoxon Signed Rank tests for caregiver prompt data and 3) Wilcoxon Rank Sum and paired Wilcoxon Signed Rank tests for attended object data. Effect sizes and 95% bootstrapped CIs are shown for each test.

		Test result			
	n CPs	Est.Diff.	effect size	*p*-value	95% CI
1) All data	1916				
CP tokens ~ VMS group		33	$\delta = 0.68$	< 0.001***	[15, 57.99]
CP types ~ VMS group		0.99	$\delta = 0.32$	0.060	[< 0.001, 1]
2) Caregiver prompt	911				
% congruent CPs ~ VMS group		0.11	$\delta = 0.28$	0.12	[−0.09, 0.59]
with-VMS group vs. chance		0.33	$\delta = 0.90$	< 0.001***	[0.23, 0.43]
no-VMS group vs. chance		0.25	$\delta = 0.47$	0.01*	[0.09, 0.74]
INREP vs. OUTREP (with-VMS only)		0.35	$r = 0.64$	0.001**	[0.15, 0.50]
with-VMS vs. **no-VMS** (OUTREP only)		−0.02	$\delta = -0.06$	0.72	[−0.4, 0.29]
3) Attended Object	1042				
% congruent CPs ~ VMS group	(n = 43)	0.28	$\delta = 0.46$	0.003**	[0.09, 0.43]
with-VMS group vs. chance		0.28	$\delta = 0.61$	< 0.001***	[0.15, 0.40]
no-VMS group vs. chance	(n = 19)	< 0.1	$\delta < 0.01$	1	[−0.17, 0.12]
INREP vs. OUTREP (with-VMS only)		0.32	$r = 0.42$	0.056	[−0.01, 0.52]
with-VMS vs. no-VMS (OUTREP only)	(n = 43)	0.08	$\delta = 0.13$	0.318	[−0.07, 0.25]

Note: All comparisons where infants were excluded note n parenthetically; otherwise all infants are included in analysis. Estimated Difference between groups are from Wilcoxon tests: positive values indicate higher with-VMS group productions in all comparisons of VMS group differences; bolded term had higher values for all other comparisons. (Wilcoxon Estimated-Differences are similar but not identical to mean differences (Hollander & Wolfe, 1999).) Effect sizes were calculated using Cliff's Delta for independent samples and regression coefficients for paired samples. *$p < .05$; **$p < .01$; ***$p < .001$.

consonant types) vs. (total number of OUTREP tokens)/(total number of non-VMS consonant types) produced in the video data (INREP: mean tokens = 23.3; OUTREP: mean tokens = 8.7; Est.Diff. = 9.6, $r = 0.63$, $p < .01$, 95% CI = [4.08, 24.12]; paired Wilcoxon Signed-Rank test). This demonstrates continuity between with-VMS infants' production in the audio and video recordings. As well as producing their INREP consonants more consistently in the video recordings, with-VMS infants also produced more babble overall and a wider variety of consonants than the no-VMS group. This thereby validates our initial VMS classifications.

3.3. Caregiver prompts

Having established quantitative and qualitative differences in the babble production of infants with and without an established VMS repertoire, we now turn to our two context variables to address our main hypotheses. Overall, 1,441 consonant productions (CPs) in the data were produced alongside *either* a caregiver prompt *or* an attended object; 512 CPs were produced alongside *both*. Here we consider all instances of caregiver prompt, including the 512 with an attended object. Since we might expect that infants would be more likely to produce congruent babble if the caregiver labelled an object that the child was attending to (i.e. caregiver

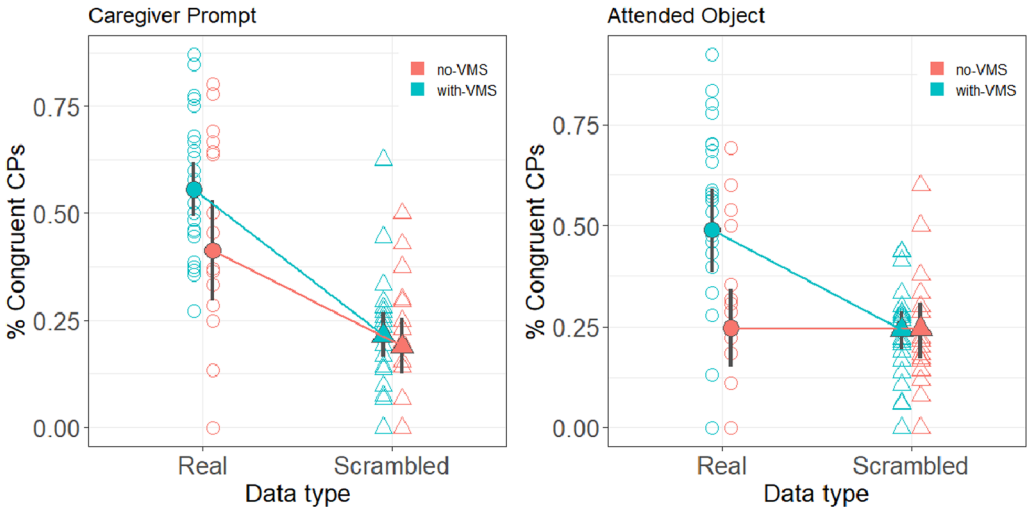

Fig. 3. Proportion of infants' **C**onsonant **P**roductions that match caregiver prompts (left-hand panel) and attended objects (right-hand panel) compared with scrambled data (Real vs. Scrambled data). Filled points show the means and 95% bootstrapped CIs for with-VMS and no-VMS infants (blue and pink points, respectively). Colored lines link mean group values across Real and Scrambled data (circles, and triangles, respectively). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

says *bottle* while infant attends to a bottle), we also ran all analyses with these matching prompt-object instances removed; this generated results consistent with those reported below. See S4 and Table S3 for details.

We begin by observing infants' babble in relation to caregiver prompts occurring in the 15-second segment prior to the CP in question. On average, 50% of the infants' CPs were preceded by caregiver speech (SD = 0.18), and 48% of these matched the salient word in this preceding segment (e.g. mother says 'ball' and baby says 'ba'; SD = 0.23). With-VMS infants' CPs matched caregiver prompts more often than no-VMS infants' did (0.55(0.17) vs. 0.41(0.28)), but a Wilcoxon Rank Sum test comparing the proportion of each infant's congruent CPs showed no difference between VMS groups (Est.Diff. = 0.11, δ = 0.28, p = .12, 95% CI = [−0.09, 0.59]; see Table 5).

As mentioned above, it is hard to know what 'true chance' of matching caregiver prompts would be in this type of comparison since not all consonants are equally likely to occur in infants' or caregivers' productions (see Fig. 1). To approximate chance given the data distribution, we compared the proportion of infants' CPs that matched caregiver prompts with the proportion of such matches in a scrambled dataset (i.e. where caregiver prompts were scrambled such that each CP was paired with a randomly selected caregiver prompt from the dataset). With-VMS infants' CPs matched the consonantal properties of their caregivers' speech significantly more in the real data set than in the scrambled one (0.55(0.17) vs. 0.22(0.14), Est.Diff. = 0.33, δ = 0.90, $p < .001$, 95% CI = [0.23, 0.43], by Wilcoxon test), as did no-VMS infants' CPs (0.41(0.28) vs. 0.19(0.15), Est.Diff. = 0.25, δ = 0.47, p = .01, 95% CI = [0.09, 0.74]). Thus, in contrast with our first hypothesis, we find that infants' babble production matches their caregiver's prompt more than we would expect by chance, even when they don't yet have a stable VMS in their repertoire. See Fig. 3, left-hand panel.

We next tested whether infants with a stable VMS repertoire (i.e. with-VMS infants) *selectively* responded to caregiver prompts that matched the infants' own consonant repertoire (as opposed to the preceding analysis, which looked at matches between infants' CPs and caregiver prompts regardless of whether those particular CPs were in infants' VMS repertoire, i.e. INREP). To do so, we determined how many CPs matched *both* the caregiver prompt *and* the infant's VMS. A paired Wilcoxon Signed-Rank test showed that with-VMS infants produced significantly more CPs to match caregiver prompts when the prompt was INREP (M = 0.71, SD = 0.25) than when it was OUTREP (M = 0.39, SD = 0.23; Est.Diff. = 0.35, r = 0.64, p = .001, 95% CI = [0.15, 0.50]). See Fig. 4, left-hand panel. That is, infants with stable consonant repertoires matched input speech more often when it contained sounds that were specifically part of their own consonant inventory (VMS), rather than being equally likely to respond with a congruent CP regardless of whether or not the caregiver prompt featured a consontant in their VMS repertoire.

Finally, we compared infants' consonant productions following caregiver prompts that were not in their repertoire (i.e. outREP consonants for with-VMS infants, and all consonants for no-VMS infants). Overall, 40% of OUTREP consonants matched a caregiver prompt (SD = 0.25). A Wilcoxon Rank Sum test revealed no difference between with-VMS (M = 0.39, SD = 0.23) and no-VMS infants' CPs (M = 0.41, SD = 0.28) when the CP was not INREP (Est.Diff. = −0.02, p = .72, δ = −0.06, 95% CI = [−0.4, 0.29]). See Table 5. That is, in response to a caregiver prompt outside of their inventory, with-VMS and no-VMS infants responded with matching consonants equivalently (see Fig. 4, left-hand panel).

3.4. Attended objects

We next considered infants' CPs in relation to the objects they were attending to at the time of production. On average, 56% of CPs (SD = 0.22) were produced while attending to an object, and 38% showed a phonetic match with the attended object (SD = 0.27).

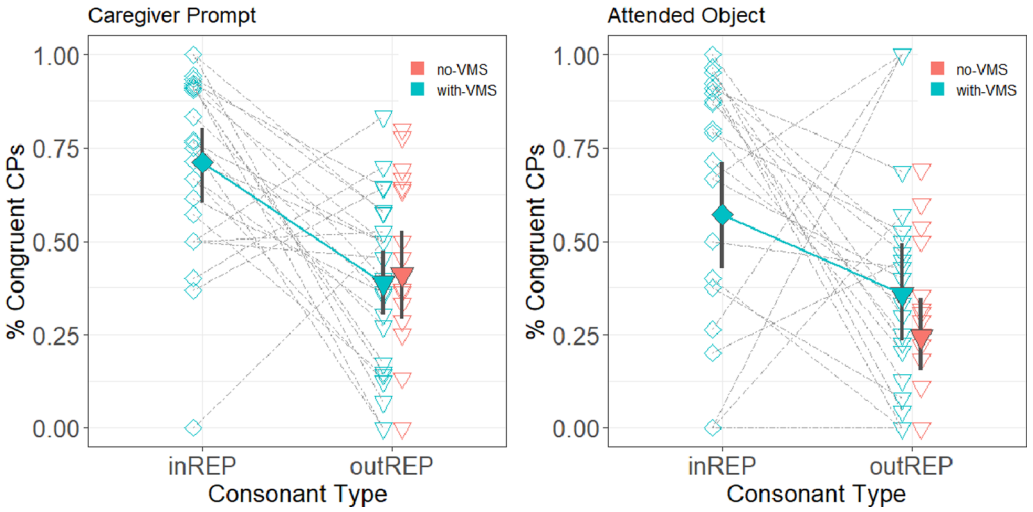

Fig. 4. Proportion of infants' **C**onsonant **P**roductions that match caregiver prompts (left-hand panel) and attended objects (right-hand panel) in relation to each infants' phonetic repertoire (i.e. INREP and OUTREP consonants). Filled points show the means and 95% bootstrapped CIs for with-VMS and no-VMS infants (blue and pink points, respectively). Grey lines link with-VMS infants' INREP and OUTREP consonants (diamonds and triangles, respectively). N.B.: all consonants are OUTREP for infants in the no-VMS group, by definition. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

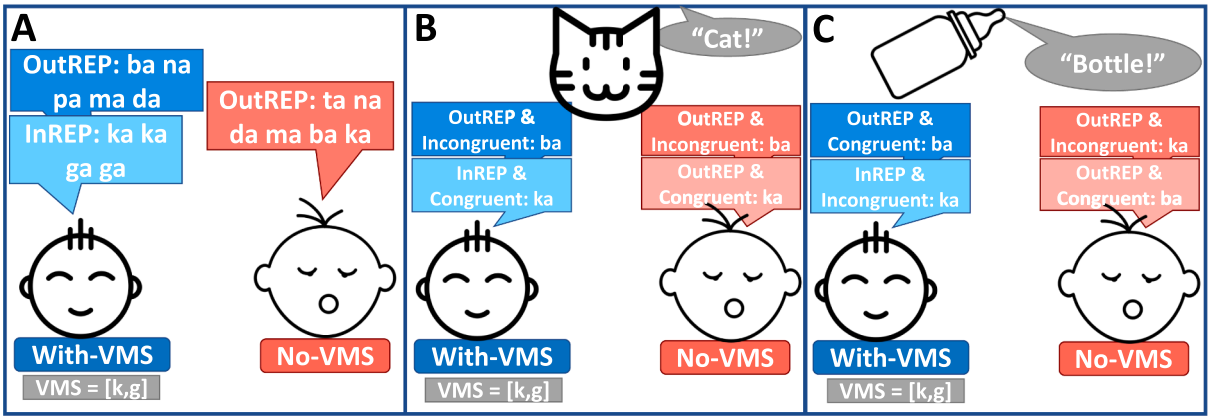

Illustration 1. (a–c) Possible infant CP responses to input across both VMS groups. With-VMS infant shown on left, no-VMS infant on right in all three panels. Illustration 1a shows example INREP/OUTREP CPs for each group (all consonants are OUTREP for no-VMS infants). (b)–(c) show examples of congruent and incongruent responses to *cat* and *bottle*, respectively. Illustrations by Alex Podolsky, Dewi Imaniyah, Nibras@design & rivercon from Noun Project (https://thenounproject.com/).

To test whether the proportion of CPs that matched attended objects differed by VMS group, we ran a Wilcoxon Rank Sum test, comparing the proportion of matching CPs by VMS group. With-VMS infants had more matching CPs than no-VMS infants (Est.Diff. = 0.28, δ = 0.46, p = .003, 95% CI = [0.09, 0.43]; Table 5). That is, infants with a stable consonant repertoire matched their consonants productions to an attended object significantly more often than infants who lack such a repertoire (49% vs. 25%, respectively; see Table 4 & Fig. 3, right-hand panel).

Here too we further compared the proportion of infants' CPs that matched attended objects with scrambled attended object data (i.e. attended objects were scrambled so each CP was paired with a randomly selected attended object from the dataset) to allow for a comparison against chance. We found that infants with a stable consonant repertoire matched their CPs to an attended object significantly more in the real dataset than in the scrambled one (0.49(0.27) vs. 0.24(0.11), Est.Diff. = 0.28, δ = 0.61, p < .001, 95% CI = [0.15, 0.40]; by Wilcoxon Rank Sum test), while infants who lacked such a repertoire did not (0.25(0.22) vs. 0.24(0.16), Est.Diff. < 0.1, δ < 0.01, p = 1, 95% CI = [−0.17, 0.12]). See Fig. 3, right-hand panel. Thus here we find support for our first hypothesis, i.e. that infants with stable inventories will babble more congruently with their input than no-VMS infants, but only within the context of an attended object.

Next, as we did for caregiver prompts, we determined whether or not with-VMS infants' object-matching CPs were influenced by their particular consonant inventory. Using a paired Wilcoxon Signed Rank test, we compared the proportion of infants' CPs that matched both the attended object *and* were part of their established VMS inventory (INREP; 0.57(0.37)) with the proportion of CPs that

matched the attended object but were not in the infant's VMS inventory (OUTREP, 0.36(0.32)). Similar to the caregiver prompts, infants produced more object-matching CPs when the object's label provided a phonetic match with their VMS inventory, though this difference was only marginally significant (Est.Diff. = 0.32, r = 0.42, p = .056, 95% CI = [−0.01, 0.52]). This result extends (cautious) support for our second hypothesis that infants with stable inventories will congruently babble more specifically when the context (in this case the label of the attended object) features a consonant in their repertoire. See Fig. 4, right-hand panel and Table 5.

Finally, we examined infants' consonant productions when the attended object's label did not match their VMS repertoire (i.e. OUTREP consonants for with-VMS infants, and all consonants for no-VMS infants). Overall, 31% of OUTREP CPs matched an attended object (SD = 0.29), and this did not differ between VMS groups (Est.Diff. = 0.08, δ = 0.13, p = .318, 95% CI = [−0.07, 0.25]; with-VMS: M = 0.36, SD = 0.32; no-VMS: M = 0.25, SD = 0.22). That is, given objects whose labels were outside of their inventory, with-VMS and no-VMS infants responded with matching consonants at low and equivalent rates.

4. Discussion

We considered 10–11-month-olds' pre-linguistic vocalizations in relation to their established consonant repertoire (operationalized via Vocal Motor Schema, i.e. VMS). We aimed to discover whether babbled speech sounds are random – as suggested by findings from Goldstein and Schwade (2008) – or congruent with the input. To do this, we invoked Vihman's (1993) articulatory filter framework, to test whether having the beginnings of an established phonological system supports infants' ability to link the consonants produced in their babble with the words and objects in their environment.

Our analysis yielded two key findings that highlight a variable role for stable production inventories (VMS) as a function of input context. First, regardless of whether infants had a stable consonant inventory or not (i.e. whether they were with-VMS or no-VMS infants), their babble matched salient words in their caregiver's speech. That is, both with-VMS and no-VMS infants' consonant productions in the videos were congruent with the caregiver's preceding speech roughly 50% of the time, relative to the ~25% established as 'chance' from the scrambled data (see Fig. 3, left-hand panel). In contrast, only with-VMS infants' babble matched the objects they attended to; no-VMS infants' babble did not. That is, while with-VMS infants' consonant productions matched the attended object ~50% of the time, no-VMS infants' consonant productions only did so ~25% of the time, with chance again around 25% based on the scrambled data (see Fig. 3, right-hand panel). This offers partial support for our first hypothesis: having stable VMS consonants did predict whether infants' babble matched the input context, but only for attended objects. In contrast, both with-VMS and no-VMS infants' babble matched the caregiver prompt context.

Second, the presence of a VMS repertoire also affected infants' specific consonant productions relative to their input. That is, infants with a stable consonant inventory (with-VMS infants) were more likely to produce babble that matched the context (especially for the caregiver input context) when the *particular stimulus* was within the infant's productive repertoire, i.e. contained their own VMS consonants. With-VMS infants' consonant productions matched the input context *and* were part of their VMS inventory (i.e. INREP CPs) ~70% of the time for caregiver input, and ~60% of the time for attended objects; their consonant productions matched the context but were *not* part of their VMS inventory (OUTREP CPs) ~40% of the time for each context. Intriguingly, the CPs produced by no-VMS infants (all of which are OUTREP by definition) matched the input at a statistically equivalent (but slightly lower) rate as the with-VMS children's OUTREP CPs, ~35% of the time for caregiver input and ~25% of the time for attended objects. It is worth noting that while these results generally support our second hypothesis i.e. that with-VMS infants would be more likely to match the particular sounds they produce when they occur in their input context, the data were stronger in the caregiver input context: the INREP vs. OUTREP difference was only marginally significant in the context of attended objects; we return to this below.

Taken together, these findings are the first to show that the phonetic properties of infants' babble are shaped to their on-line experiences of the world. These results contrast with previous findings from Goldstein and Schwade (2008), who found no link between caregiver speech and the sounds of infants' babble. In their discussion, the authors propose that infants' responses may have mapped onto broader categories than those adopted in their analysis. By augmenting their approach to take infants' production abilities at 9–11 months into account – i.e. collapsing voicing contrasts, and including all fricative-like segments in a single category – we find support for this proposal. Additionally, homing in on VMS in our analysis meant that we were able to take into account individual differences in each infant's *own* productive experience. Moreover, by testing only the word-initial consonant, we observed aspects of the input that we know infants both perceive and produce most easily in early development (cf. Swingley & Aslin, 2000; Vihman, DePaolis, & Davis, 1998), and were not expecting infants to undertake the articulatory mastery required to produce more extended sequences of sounds. This approach let us corroborate previous research showing the importance of early input in shaping the development from babble to words (Baumwell, Tamis-LeMonda, & Bornstein, 1997; Gros-Louis et al., 2014), and extend this account to demonstrate an integral role for the production experience gained through babble in early development.

The differences we find for the role of VMS between the caregiver input and attended object contexts merit further discussion. By definition, infants with stable VMS exhibited consistent production of the consonants in their inventory; they also produced more consonants overall. However, as noted above, with-VMS infants were not more likely to respond congruently to salient words in the input (caregiver prompts) than no-VMS infants: both groups responded congruently in equal measure. With-VMS infants were, however, more likely to respond congruently to caregiver prompts that matched their particular VMS consonant(s). This suggests that with-VMS infants may be especially attuned to words in the input that contained the sounds they can most easily produce, which may have led them to produce these sounds in response. We return to this point below.

In the attended object condition, with-VMS infants were significantly more likely to respond congruently than no-VMS infants (see Fig. 3, right-hand panel). This suggests qualitative differences between the linguistic demands of recognizing and reproducing sounds in caregiver prompts vs. considering the sounds that are used to label attended objects. Both groups could reproduce a salient segment

they'd just heard, but only infants with an established consonant repertoire – a consistent marker of early phonological advancement (e.g. Majorano et al., 2014; McGillion et al., 2017) – retrieved consonants to match their attended object.

In the attended object case, with-VMS infants were only marginally more likely to respond with the *particular* consonants that were in their VMS repertoires (unlike in the caregiver prompt case, where they more clearly did so; see Fig. 4). While further research is needed to better understand this pattern of results, if indeed infants are less responsive with the particular consonants of their VMS in an object context, it may suggest that the specifics of an infants' VMS repertoire (rather than just having such a repertoire) are less relevant for sound-object associations, or indeed, word-learning.

Our findings showing equivalent responses to caregiver prompts regardless of VMS group suggest that infants begin to establish associations between the sounds of their babble and the sounds they hear in the input even before they have a stable productive repertoire to draw upon. Given that infants know some concrete nouns by 6–9 months (Bergelson & Swingley, 2012; Parise & Csibra, 2012; Tincoff & Jusczyk, 1999, 2012), our results with 10–11-month-olds may reveal the beginnings of infants' ability to pair this knowledge with the sounds they can most readily produce. This form-meaning pairing may be crucial for *word learning* in earlier development (i.e. before VMS are established), before supporting the transition from babble to *word production* a couple of months down the line. Along these lines, recent work from Majorano et al. (2019) shows that Italian 10–11-month-olds with an established VMS repertoire were better than no-VMS infants at learning novel words, but only for words that contained the consonants commonly found in Italian infants' VMS. The authors propose that this reflects the "beginnings of lexical representations for…[infants with] greater production experience…[for whom] consonants could presumably be mapped onto…their own vocal production" (Majorano et al., 2019, p.9). That is, VMS may draw infants' attention to words they *could* produce in the input, making them more memorable and therefore good candidates for the early lexicon. This is consistent with prior work showing that the most frequent consonants in infants' babble were also in their early words (McCune & Vihman, 2001).

Our results provide first-step evidence for *how* the transition to word production might take place, as infants bridge the gap between babble and words through the articulatory matching of input and output. Indeed, at face value, our results suggest that VMS consonants may prompt articulatory 'matching' between what an infant *hears* in the input and their own vocalizations, independent of any lexical knowledge of the word in question. We propose that this may support more speech-like consonant production, as infants move from the canonical syllables of babble to the establishment of an inventory of speech sounds.

In contrast to previous work (Albert et al., 2018; Goldstein et al., 2003; Goldstein & Schwade, 2008), we looked at *infants'* responsiveness to the input rather than *caregivers'* responsiveness to infants. This previous work highlights that caregivers' selective responsiveness may support increasingly speech-like early productions; we view our perspectives as complementary. That said, in a preliminary analysis of our home recordings, caregivers rarely responded to infants' babble (~25%), leading us to forego such an analysis here.

Indeed, it is worth noting that infants' early vocal exploration is not necessarily linked to caregivers' responsiveness or presence at all. Recent work by Oller and colleagues finds that over the first year of life, in both pre- and full-term infants, early vocalization (specifically "protophone") rates are relatively stable, and show little difference between contexts where infants are alone or in the context of adult speech input, especially around the age we test here (Oller et al., 2019). The authors interpret this as "a strong indication of the endogenous nature of infant vocal exploration" (Oller et al., 2019, p. 5).

Taken together, ours and others' results suggest that the shift to more speech-like babble, and, eventually, word production, does not depend *exclusively* on caregivers' contingent responses to babble production, but may also be led by *infants'* internal drive to explore, alongside their capacity to match what they can produce with what they perceive. Of course, in the context of caregiver interaction, any sense of directionality remains unclear, as caregiver responsiveness may support a 'feedback loop' (Warlaumont et al., 2014); i.e. if the mother responded to 'ba' with 'yes, a ball!', this may have prompted further 'ba' production on the infant's part. Based on the present results, we propose that *infant responsiveness to the input* – as well as caregiver responsiveness to their output – may be central in the transition to early word production.

Furthermore, it is possible that caregivers are more responsive to the consonants that are most common in their infants' production (i.e. their VMS) and thereby promote further production of these consonants via more frequent contingent responses to them. Indeed, caregivers may be aware of their infants' VMS repertoires, or even of the words that the infant is beginning to acquire in their first lexicon; this may lead to an implicit prompting of the sounds and word-like syllables that they know their infants can produce. As babble becomes more speech-like, caregivers may increasingly provide scaffolding towards phonological acquisition, supporting the shift from proto-phonological to phonological categorization (Vygotsky, 1986; see also Oller, 2000). Similarly, if the caregiver is aware that their infant is in the process of acquiring the word *ball*, they may be more likely to use this word in their infant-directed speech. They may also be more likely to draw their infant's attention to objects that match their most common babbled sounds, thereby prompting a feedback loop between the infant's VMS, the caregiver input, and an object in their environment. This account fits within an evolutionary-developmental framework of language acquisition (e.g. Locke, 2006, 2009), whereby language development interacts with social practices in a way that allows the infant's language to develop in tandem with their caregiver's understanding of their communicative needs. We account for this in part in the supplementary materials (see S4), where we analyzed our data without these instances of matching prompt and object. However, as there were only 144 such instances in the data across all infants, we are not in a position to establish strong evidence for this sort of feedback loop here.

Future analysis of the caregiver prompt data – for example, to test whether the caregivers responded differentially to INREP vs. OUTREP consonants – would reveal more about how our results fit within the evolutionary-developmental framework. We would predict that caregivers might be more likely to label an attended object when it matches their infant's VMS repertoire, leading to an even stronger association between the child's production and the surrounding context, given the three-way 'boost' between visual and auditory stimuli and the child's VMS. We were unable to assess this empirically in our data, as there are only 60 such instances.

However, infants do indeed produce a congruent CP in the majority of these (n = 51). This provides a preliminary suggestion towards caregivers' attunement to infants' VMS consonants, which may provide a particularly supportive environment for early word learning and vocal development.

One may question the line between babble and rudimentary words. While none of the CPs included in our data were interpreted as words by our trained child phonology experts, it is possible that some CPs reflected early word production. In separate work assessing early milestones in this sample of infants (Moore et al., 2019), we find few infants produce their first word before 10 months (4 via observational data, 12 via parent report; parent vs. researcher criteria likely vary, especially for infants' very first word). However, even if infants were producing proto-words, rather than established babble consonants, our analysis relates this once again to VMS, since these vocalizations (whether babble or rudimentary words) were more often consistent when object labels contained consonants in infants' VMS repertoire.

Finally, we consider the representativeness of our VMS categorization, which necessarily differs from approaches taken in the previous literature (e.g. DePaolis et al., 2011; McCune & Vihman, 2001). As noted above, to our knowledge, this is the first analysis of infants' VMS derived from daylong recordings. Encouragingly, we found convergence between our analysis of infants' consonant productions in the hour-long video recordings and the top 30 minutes of child vocalizations from the daylong audio recordings (as determined by LENA's proprietary algorithm). Notably, infants babbled more in their half hour of audio than in their hour of video (see Table 4), highlighting the promise of combined automated and manual methods to spotlight particularly relevant portions of long-form naturalistic data in analyzing language development.

5. Conclusion

Our results show that infants' early babble production is shaped by their linguistic and visual input, filtered through the established consonants in their repertoire. When babble is considered through the lens of early phonological development, the sounds of babble are not random. Rather, our results suggest that infants' own production abilities help to shape their perception of the world around them, and their transition from babble to words.

CRediT authorship contribution statement

Catherine Laing: Conceptualization, Methodology, Formal analysis, Investigation, Visualization. **Elika Bergelson:** Formal analysis, Resources, Data curation, Visualization, Supervision, Funding acquisition.

Acknowledgements

Special thanks to Andrei Amatuni, as well as research assistants at University of Rochester, Duke and Cardiff.

Funding

This work was funded by the National Institutes of Health [grant number DP5-OD019812].

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cogpsych.2020.101308.

References

Adriaans, F., & Swingley, D. (2017). Prosodic exaggeration within infant-directed speech: Consequences for vowel learnability. *Journal of the Acoustical Society of America, 141*(5), 3070–3078. https://doi.org/10.1121/1.4982246.

Albert, R. R., Schwade, J. A., & Goldstein, M. H. (2018). The social functions of babbling: Acoustic and contextual characteristics that facilitate maternal responsiveness. *Developmental Science, 21*(5), 1–11. https://doi.org/10.1111/desc.12641.

Athari, P., Wang, C. H., Day, R., & Rvachew, S. (2019). An exploration of mother-infant verbal interaction at the transition to canonical babbling. In International child phonology conference.

Baumwell, L., Tamis-LeMonda, C. S., & Bornstein, M. H. (1997). Maternal verbal sensitivity and child language comprehension. 20(2), 247–258.

Bergelson, E., Amatuni, A., Dailey, S., Koorathota, S., & Tor, S. (2019). Day by day, hour by hour: Naturalistic language input to infants. *Developmental Science, 22*(1), e12715. https://doi.org/10.1111/desc.12715.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences, 109*(9), 3253–3258. https://doi.org/10.1073/pnas.1113380109.

de Boysson-Bardies, B., Halle, P., Sagart, L., & Durand, C. (1989). A crosslinguistic investigation of vowel formants in babbling. *Journal of Child Language, 16*(1), 1–17. https://doi.org/10.1017/S0305000900013404.

de Boysson-Bardies, B., & Vihman, M. M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language, 67*(2), 297–319. https://doi.org/10.1353/lan.1991.0045.

Casillas, M., Bobb, S. C., & Clark, E. V. (2016). Turn-taking, timing, and planning in early language acquisition. *Journal of Child Language, 43*(6), 1310–1337. https://doi.org/10.1017/S0305000915000689.

DataVyu: A video coding tool (2014). Databrary Project.

DePaolis, R. A., Vihman, M. M., & Keren-Portnoy, T. (2011). Do production patterns influence the processing of speech in prelinguistic infants? *Infant Behavior and Development, 34*(4), 590–601. https://doi.org/10.1016/j.infbeh.2011.06.005.

Dunn, J., & Shatz, M. (1989). Becoming a conversationalist despite (or because of) having an older sibling. *Child Development, 60*(2), 399–410.

Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National*

Academy of Sciences of the United States of America, 100(13), 8030–8035. https://doi.org/10.1073/pnas.1332441100.
Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science, 19*(5), 515–523. https://doi.org/10.1111/j.1467-9280.2008.02117.x.
Goldstein, M. H., Schwade, J., Briesch, J., & Syal, S. (2010). Learning while babbling: Prelinguistic object-directed vocalizations indicate a readiness to learn. *Infancy, 15*(4), 362–391. https://doi.org/10.1111/j.1532-7078.2009.00020.x.
Gros-Louis, J., West, M. J., & King, A. P. (2014). Maternal responsiveness and the development of directed vocalizing in social interactions. *Infancy, 19*(4), 385–408. https://doi.org/10.1111/infa.12054.
Hollander, M., & Wolfe, D. A. (1999). Nonparametric statistical methods (2nd ed.). Wiley.
Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., Jasnow, M. D., Rochat, P., & Stern, D. N. (2001). Rhythms of dialogue in infancy: Coordinated timing in development. *Monographs of the Society for Research in Child Development, 66*(2), https://doi.org/10.4135/9781452286143.n496.
Jakobson, R. (1940). Child language, aphasia and phonological universals. Mouton. https://doi.org/10.3109/13682826909011489.
LENA Research Foundation (2018). https://www.lena.org/.
Locke, J. L. (2006). Parental selection of vocal behavior. *Human Nature, 17*(2), 155–168. https://doi.org/10.1007/s12110-006-1015-x.
Locke, J. L. (2009). Evolutionary developmental linguistics: Naturalization of the faculty of language. *Language Sciences, 31*(1), 33–59. https://doi.org/10.1016/j.langsci.2007.09.008.
Macken, M. A. (1980). Aspects of the acquisition of stop systems: A cross-linguistic perspective. In G. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), Child phonology, Vol. 1: Production (pp. 143–168). Academic Press.
Majorano, M., Bastianello, T., Morelli, M., Lavelli, M., & Vihman, M. M. (2019). Vocal production and novel word learning in the first year. *Journal of Child Language, 1–11*. https://doi.org/10.1017/S0305000918000521.
Majorano, M., Vihman, M. M., & DePaolis, R. A. (2014). The relationship between infants' production experience and their processing of speech. *Language Learning and Development, 10*(2), 179–204. https://doi.org/10.1080/15475441.2013.829740.
McCune, L., & Vihman, M. M. (2001). Early phonetic and lexical development: A productivity approach. *Journal of Speech, Language and Hearing Research, 44*(3), 670–684. https://doi.org/10.1044/1092-4388(2001/054).
McGillion, M., Herbert, J. S., Pine, J., Vihman, M. M., DePaolis, R., Keren-Portnoy, T., & Matthews, D. (2017). What paves the way to conventional language? The predictive value of babble, pointing, and socioeconomic status. *Child Development, 88*(1), 156–166. https://doi.org/10.1111/cdev.12671.
McLeod, S., & Crowe, K. (2018). Children's consonant acquisition in 27 languages: A cross-linguistic review. *American Journal of Speech-Language Pathology, 1–26*. https://doi.org/10.1044/2018_AJSLP-17-0100.
Menn, L., & Vihman, M. M. (2011). Features in child phonology: Inherent, emergent, or artefacts of analysis? In N. Clements & R. Ridouane (Eds.), Where do phonological features come from? Cognitive, physical and developmental bases of distinctive speech categories (Language F, pp. 261–301). John Benjamins.
Moore, C., Dailey, S., Garrison, H., Amatuni, A., & Bergelson, E. (2019). Point, walk, talk: Links between three early milestones, from observation and parental report. *Developmental Psychology*. https://doi.org/10.1037/dev0000738.
Oller, D. K. (2000). *The emergence of the speech capacity*. Lawrence Erlbaum Associates.
Oller, D. K., & Eilers, R. (1988). The role of audition in infant babbling. *Child Development, 59*(2), 441–449. https://doi.org/10.2307/1130323.
Oller, D. K., Caskey, M., Yoo, H., Bene, E. R., Jhang, Y., Lee, C., ... Vohr, B. (2019). Preterm and full term infant vocalization and the origin of language. *Scientific Reports,* 1–10. https://doi.org/10.1038/s41598-019-51352-0.
Oller, D. K., Wieman, L. A., Doyle, W. J., & Ross, C. (1976). Infant babbling and speech. *Journal of Child Language, 3*(1), 1–11. https://doi.org/10.1017/S0305000900001276.
Parise, E., & Csibra, G. (2012). Electrophysiological evidence for the understanding of maternal speech by 9-month-old infants. *Psychological Science, 23*(7), 728–733. https://doi.org/10.1177/0956797612438734.
Ramsdell, H. L., Oller, D. K., Buder, E. H., Ethington, C. A., & Chorna, L. (2012). Identification of prelinguistic phonological categories. *Journal of Speech, Language, and Hearing Research, 55*(6), 1626–1639. https://doi.org/10.1044/1092-4388(2012/11-0250).
R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
Shriberg, L. D., Austin, D., Lewis, B. A., Mcsweeny, J. L., & Wilson, D. L. (1997). The Percentage of Consonants Correct (PCC) metric: Extensions and reliability data. *Journal of Speech, Language and Hearing Research, 40*, 708–722.
Swingley, D. (2005). 11-Month-olds' knowledge of how familiar words sound. *Developmental Science, 8*(5), 432–443. https://doi.org/10.1111/j.1467-7687.2005.00432.x.
Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition, 76*(2), 147–166. https://doi.org/10.1016/S0010-0277(00)00081-0.
Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science, 10*(2), 172–175. https://doi.org/10.1111/1467-9280.00127.
Tincoff, R., & Jusczyk, P. W. (2012). Six-month-olds comprehend words that refer to parts of the body. *Infancy, 17*(4), 432–444. https://doi.org/10.1111/j.1532-7078.2011.00084.x.
Vihman, M. M. (1993). Variable paths to early word production. *Journal of Phonetics, 21*, 61–82.
Vihman, M. M. (2017). Learning words and learning sounds: Advances in language development. *British Journal of Psychology, 108*, 1–27. https://doi.org/10.1111/bjop.12207.
Vihman, M. M., Macken, M. A., Miller, R., Simmons, H., & Miller, J. (1985). From babbling to speech: A re-assessment of the continuity issue. *Language, 61*(2), 397–445.
Vihman, M. M., Nakai, S., Depaolis, R. A., & Hallé, P. (2004). The role of accentual pattern in early lexical representation. *Journal of Memory and Language, 50*, 336–353. https://doi.org/10.1016/j.jml.2003.11.004.
Vihman, M. M., DePaolis, R. A., & Davis, B. L. (1998). Is there a "trochaic bias" in early word learning? Evidence from infant production in English and French. *Child Development, 69*(4), 935–949. https://doi.org/10.1111/j.1467-8624.1998.tb06152.x.
Vygotsky, L. (1986). *Thought and Language (Translation by Alex Kozulin)*. MIT Press.
Warlaumont, A. S., Richards, J. A., Gilkerson, J., & Oller, D. K. (2014). A social feedback loop for speech development and its reduction in autism. *Psychological Science, 25*(7), 1314–1324. https://doi.org/10.1177/0956797614531023.
Wu, Z., & Gros-Louis, J. (2014). Infants' prelinguistic communicative acts and maternal responses: Relations to linguistic development. *First Language, 34*(1), 72–90. https://doi.org/10.1177/0142723714521925.
Zamuner, T. S. (2006). Sensitivity to word-final phonotactics in 9- to 16-month-old infants. *Infancy, 10*(1), 77–95. https://doi.org/10.1207/s15327078in1001_5.