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. 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 are 'filtered' through the inventory of speech-sounds they have mastered. More specifically, we find that infants with well-established speech sounds produce consonants in babble that match (i) their caregivers' speech, and (ii) objects in their environment, while infants who do not have such sounds only match caregiver's speech, but not attended objects. Our findings suggest that infants' early productions shape their path from babble to words: the consonants they produce reliably before 12 months shape their responses to what they hear and see around them.

Keywords: *infant production, phonological development, babble, prelinguistic vocalizations, input effects, language development*

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

Caregivers play a key role in infants' transition from pre-linguistic babble to meaningful words (Albert, Schwade, & Goldstein, 2018; Goldstein, King, & West, 2003; Goldstein & Schwade, 2008; Gros-Louis, West, & King, 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). Related work by Boysson-Bardies and colleagues finds that the phonological features of babble map on to properties of the ambient language, collected in home recordings (Boysson-Bardies, Halle, Sagart, & Durand, 1989; Boysson-Bardies & Vihman, 1991). Bridging the approaches of these previous lab-studies and home recordings, the present study uses ecologically-valid and naturalistic home recordings to investigate how infants' early-established phonological inventories shape the transition from babble to words. Before turning to the specifics of the present work, we ground it within the preceding literature. Specifically, we briefly review the literature on the role of early babble, and how it links to the ambient language and the early lexicon.

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 meaningless and unconnected to the language being learned. However, subsequent research has called this account into question, garnering several convergent lines of evidence for the opposite viewpoint, namely, that babble is neither wholly egocentric, nor unrelated to the ambient language.

First, 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; Vihman, Macken, Miller, Simmons, & Miller, 1985). Moreover, contingent caregiver feedback to infant vocalizations also appears tied to word learning. For instance, Goldstein, Schwade, Briesch and Syal (2010) find 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.

A second set of critiques to Jakobson's framing comes from considering the specifics of infants' pre-linguistic productions more closely. Through a typological lens, cross-linguistic data links infants' babble both to what may be 'biologically-oriented' in Jakobson's sense, but also to properties of the ambient language. For instance, Japanese, Swedish, French and English-learning infants do indeed show a high proportion of e.g. labials and dentals across the board, but also show *language-specific* distributions in their babble, e.g. across stop, fricative, and nasal production (Boysson-Bardies & Vihman, 1991).

From an interactional perspective, Goldstein and colleagues have also provided evidence against a wholly 'egocentric' view of early vocalizations. 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).

Articulatory and Perceptual Limitations to Early Productions

While on-line linguistic experience has been found to affect the phonological structure of babble, Goldstein and Schwade (2008) report that the *sounds* of infants' babble appear unrelated to the surrounding linguistic input. This conclusion was based on further analysis of the mother-infant interactions reported above, which assessed the proportion of matching phonemes between 9.5-month-olds' productions and their mothers' immediately-preceding word forms. However, this analysis did not consider infants' articulatory abilities, nor the phonological detail that we can expect English-learning infants of 9-10 months to process in the linguistic input. The current study raises the possibility that infants' consonant productions are indeed contingently linked to surrounding input, once their own speech-sound abilities are accounted for.[1]

To address this, we must first consider which features of the caregivers' input we can expect infants to perceive and/or (re)produce. Perceptually, infants are more sensitive to the phonological features of 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 between these sounds in production until almost a year later (Macken, 1980), and they rarely produce certain other types of sounds (e.g. /ɹ/ or /f/; McLeod & Crowe, 2018). With this in mind, it is necessary to

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

consider the phonological resources that are available to infants when analyzing their babble production in relation to adult speech.

To this end, we adapt the approach of McCune and Vihman (2001), who analyzed prelinguistic vocalizations to show that some consonants are stably and consistently reproduced. These stable consonants are known as 'vocal motor schemes' (hereafter **VMS**), and indicate the establishment of a phonological repertoire. The establishment of one or two VMS consonants has been consistently related to later speech-sound and word production abilities (Majorano, Vihman, & DePaolis, 2014; McCune & Vihman, 2001; McGillion et al., 2017) suggesting that this marks a turning-point in infants' linguistic development. Vihman (1993) proposes that stable consonants serve as an 'articulatory filter', which establishes a perceptuomotor link between what an infant hears and what they produce.

Recent work also links infants' own *perception* to their special VMS consonants (Majorano, Bastianello, Morelli, Lavelli, & Vihman, 2019; Majorano et al., 2014), offering a possible explanation for why commonly babbled consonants occur in first words: infants' perception *and* production may be influenced by a phonological repertoire established through babble, leading to acquisition of words that match the consonants that are most stably represented in their output.

The Present Study

In what follows, we expand on findings from Goldstein and colleagues (2003, 2008) with a consideration of VMS (DePaolis, Vihman & Keren-Portnoy, 2011; McCune & Vihman, 2001), extending the articulatory filter account (Vihman, 1993). We analyze infants' pre-lexical productions in relation to their VMS to consider whether infants' pre-linguistic vocalizations are shaped towards 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 phonologically linked to 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. We consider two aspects of context – the immediately-preceding speech input from the caregiver (caregiver input) , and the objects the infants attend to during babble production (attended object). We expect VMS to affect babble in two ways:

1) We predict that infants with a stable VMS repertoire will produce more babble that matches a given context than infants with no stable VMS in their repertoire.
2) We predict that when infants have stable VMS consonants, they will be more likely to produce congruent babble when the context in question (caregiver input/attended object) interfaces with their established inventory. E.g. 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 responses with infants' individual production inventories, suggesting that babble production is indeed affected by the input, *if* the infant has a phonological repertoire to draw from.

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.

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; we enrolled and retained 44 infants 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).

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 hours) 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 Team, 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.

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; reliability measures for each stage of the annotation are reported below. Since there are many methodological details pertinent to classifying infant's 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.

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.

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 extracted the 30-minute segment in the day-long recordings where the child vocalized the most (excluding or resampling in the case of crying/feeding, see Supplementary Material, S1). We then tallied supraglottal consonants produced in that 30-minute segment.[2] In line with previous work, we used these tallies to establish whether infants had a stable consonant repertoire ('withVMS') or not ('noVMS'; see Table 1) (DePaolis et al., 2011; McGillion et al., 2017).

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

Table 1. *Glossary of acronyms and terminology used in the analysis.*

Terminology	Definition
VMS	**V**ocal **M**otor **S**chema, i.e. stable consonant in an infant's productive repertoire; 'stable' is operationalized in withVMS and noVMS below.
withVMS	Infant who produces one or more consonant types ≥50 times in a half-hour recording session at 0;10 or 0;11, *or* who produces ≥20 but <50 instances of a single consonant at 0;10 *and* 0;11.
noVMS	Infant who produces <50 of *any* consonant type at 0;10 *or* who produces ≥20 but <50 instances of a single consonant at 0;10 but <20 at 0;11.
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 15s 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. withVMS infants), a consonant production **in** the infant's phonological **rep**ertoire.
OUTREP	For infants with stable Vocal Motor Scheme (i.e. withVMS infants), a consonant production not in the infant's phonological repertoire. All consonant productions are OUTREP for noVMS infants.

First, adopting DePaolis and colleagues' (2011) approach, we classed a child as 'withVMS' if they produced one or more consonants ≥50 times in 30 minutes at 0;10. Thirteen infants were withVMS according to these criteria (across all consonant tokens: M=178.54, SD=75.99, R=86-412). 14/29 infants had low production at 0;10 and were classed as noVMS (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 %=.50, SD=.17, R=.23-.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.

Using our initial criterial (≥50 of one or more consonants in 30 minutes), 14 of the 17 infants assessed at 11 months were classified as noVMS (n=6) or withVMS (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). 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 3/4 sessions), and that these infants demonstrated 'well-practiced and longitudinally stable vocal productions' (McCune & Vihman, 2001, p.671), we classed them withVMS.[3] See Table 2, S1 and Figure S1.

This approach ensures distinctiveness between our withVMS and noVMS groups, allowing us to compare responses across two stages of phonological development, rather than age. In all cases, we analyzed 11-month video data for all infants who were re-sampled at 0;11, and 10-month video data for those who were only sampled at 0;10.

Descriptively, 24/44 infants had at least one VMS consonant (see Table 2, Figure 1, Figure S1 and Table S1 for further breakdowns by child and consonant). A naïve research assistant re-coded six infants' 30-minute segments. There was 100% agreement regarding infants' VMS group (withVMS: n=4 or noVMS: n=2) and, for the withVMS infants, which consonant(s) were classed as being in repertoire.

[3] Combining production at 0;10 and 0;11, all three of these infants produced >50 tokens of their VMS consonants. Moreover, all reported results were consistent when we removed these infants from the analysis. 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 2. *Number of infants by VMS (Vocal Motor Scheme) groups according to age (0;10 and 0;11 months) and sex (number of females in brackets). withVMS is broken down into number of VMS. See text and SI for details regarding VMS classification and age.*

Group	**0;10**	**0;11**
withVMS	13(5)	11(8)
One VMS	6(2)	7(5)
Two VMS	5(2)	4(3)
Three VMS	2(1)	0
noVMS	14(5)	6(3)
	27(10)	**17(11)**

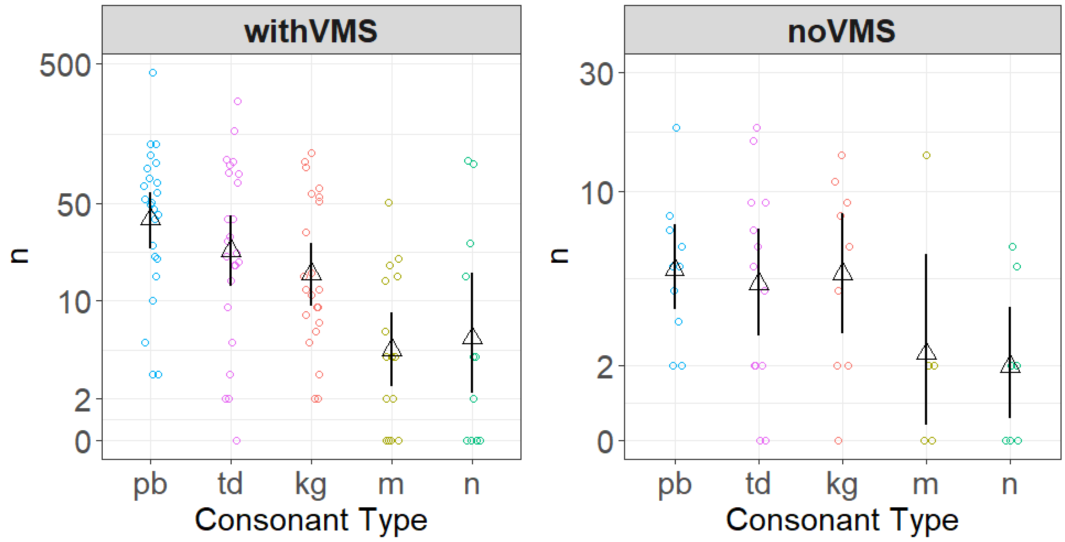

Fig. 1. *Number of each consonant type produced by withVMS and noVMS infants in the audio recordings. Colored circles represent individual infants' consonant productions; points are jittered horizontally for clarity. Black triangles represent the mean number of productions of each consonant type, with 95% bootstrapped confidence intervals.*

Annotating video data. The video data – taken within a week of the audio recording but on a different day (see Table 3) – was then annotated for infant consonant productions (CPs), caregiver input, and objects attended to during infant production. See Table 1 and S2. 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=.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. Once each of the 1,916 CPs across the 44 infants' data was annotated, the data was tagged for caregiver input and attended object.

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

Data	**All**	**withVMS**	**noVMS**
Audio	320.25	310.48	317.2
Video	318.59	308.63	315.85

Caregiver input. Input speech in the 15 seconds preceding each CP, which included any live speaker captured in the video (usually mother or father) was analyzed. 51% (n=976, per infant: M=22.18, SD=19.03) 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 228 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=.68, z=35.5).[4]

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

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=25.58, SD=23.37) 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=.9, z=69.1) for attended object tags.

Next, the phonetic properties of these two context measures (i.e. caretaker input and attended object label) were compared to the infant's CP. In the vast majority of cases (98%), the first consonant in the caretaker input 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 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*), since stressed syllables are particularly perceptually-salient to young infants (Vihman et al., 2004). As above, voicing contrasts were collapsed ('bat' or 'pig' were tagged in the /b,p/ category); similarly, word-initial fricatives 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. For both caregiver input 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 caregiver prompts=40; n attended objects=58).

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 at https://github.com/cathelaing/Laing-Bergelson-CongruentBabble. We calculated several outcomes. (1) *For all infants*, in the cases where there was a relevant context (either a caregiver input or an attended object, as described above), we determined whether their 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. Finally, 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.

In order to establish 'chance' performance, following Goldstein and Schwade's (2008) analysis, we scrambled data across infants to determine what we might expect if infants' CPs were random. 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 scrambled the context variable data (separately for the caretaker input 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 below.

Results

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 input and attended objects).

To be clear: our first hypothesis is that having stable consonants in repertoire, i.e. being a withVMS infant, will lead to babble that is more congruent with the context (caregiver input and/or attended object). We test this by comparing the proportion of congruent consonant productions (CPs) by VMS group (withVMS vs. noVMS) 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 withVMS group. Finally, we will compare the proportion of congruent OUTREP CPs produced by both withVMS and noVMS infants to consider the extent to which babble production is influenced by caregiver input/attended objects when we factor out the influence of VMS.

All figures show non-transformed data unless otherwise specified. None of the infants were outliers in our dataset (defined as CP types or token ±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 analysis. See S3 for further details.

Infant production: Audio recordings

On average, infants produced 118 consonant tokens in the half-hour segment of the daylong audio-recordings we analyzed. Variability was high (Range=0-702; see Table 4); three infants (2 females) produced zero consonants. The most prominent consonant was /p,b/ (n=2,013), followed by /t,d/ (n=1,641) and /k, g/ (n=947; Figure 1).

Table 4. *Descriptive Statistics (mean(SD)) for* ***C****onsonant* ***P****roductions in video and audio recordings across all infants, and by VMS group. Tokens refer to individual CPs, types refer to distinct phonemes. Columns 4 and 5 show mean(SD) for proportion of congruent tokens for the two context variables, with group size 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(117)	43.55(34.13)	.48(.23)	.38(.27)
CP tokens	withVMS	194.29(126.21)	60.13(34.59)	.55(.17)	.49(.27)
CP tokens	noVMS	27.15(34.14)	23.65(20.59)	.41(.28)	.25(.22)
CP types	All	3.18(1.42)	4.34(1.27)		
CP types	withVMS	4(.88)	4.67(1.31)		
CP types	noVMS	2.2(1.32)	3.95(1.15)		

We began by testing whether there was any quantitative difference in babble production between groups. As expected, withVMS infants produced more babble in the audio recordings than noVMS infants, and this difference was significant for both types and tokens (Types: Est.Diff.=1.99, $\delta=.74$, $p<.001$, 95% CI=[0.45, .89]; Tokens: Est.Diff=132, $\delta=.99$, $p<.001$, 95% CI=[.95, .99]; by Wilcoxon Rank Sum test; see Table 4 & Figure 1).

Infant production: Video recordings

On average, infants produced 44 CP tokens in the hour-long video recordings and 4.34 CP types. See Table 4 and Figure 2. Once again, withVMS infants produced significantly more tokens than noVMS infants (60.13 vs. 23.65), while the number of CP types was marginally higher for withVMS infants (4.67 vs. 3.95; see Table 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 assessible for infants who had such a VMS repertoire (i.e. withVMS infants).

Indeed, withVMS 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 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=.63, p<.01, 95% CI=[4.08, 24.12]; paired Wilcoxon Signed-Rank test). This demonstrates continuity between withVMS infants' production in the audio and video recordings: As well as producing their INREP consonants more consistently in the video recordings, withVMS infants also produced more babble overall and a wider variety of consonants than the noVMS group. This thereby validates our initial VMS classifications.

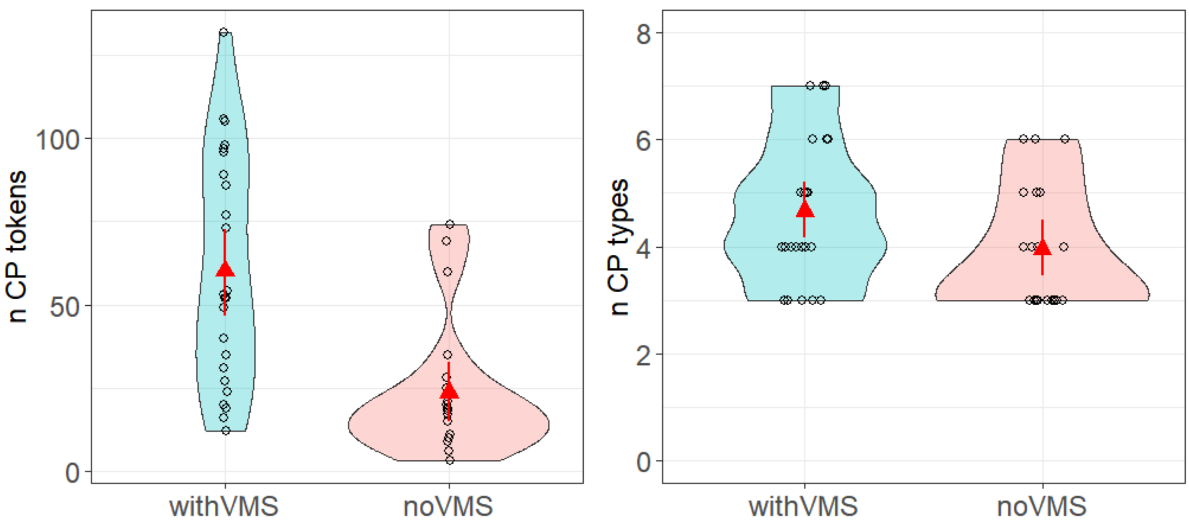

Fig. 2. *Number of consonants (tokens: left panel, types; right panel) produced by withVMS and noVMS infants in the video recordings. Circles (jittered horizontally for clarity) represent individual infants' consonant productions; red triangles represent the mean number of productions across groups, 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	δ=.68	<.001***	[15, 57.99]
CP types ~ VMS group		.99	δ=.32	.060	[<.001, 1]
2) Caregiver prompt	936				
% congruent CPs~VMS group		.11	δ=.28	.111	[-.03, .30]
withVMS group vs. chance		.33	δ=.90	<.001***	[.23, .43]
noVMS group vs. chance		.24	δ=.45	.014*	[.04, .37]
INREP vs. OUTREP CPs (withVMS only)		.35	*r*=.67	.001**	[.15, .50]
withVMS vs. **noVMS** (OUTREP only)		-.02	δ=-.06	.741	[-.20, .14]
3) Attended Object	1101				
% congruent CPs~VMS group	(n=43)	.28	δ=.46	.003**	[.09, .43]
withVMS group vs. chance		.28	δ=.61	<.001***	[.15, .40]
noVMS group vs. chance	(n=19)	<.1	δ<.01	1	[-.17, .12]
INREP vs. OUTREP CPs (withVMS only)		.32	*r*=.42	.056	[-.01, .52]
withVMS vs. noVMS (OUTREP only)	(n=43)	.08	δ=.13	.318	[-.07, .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 withVMS group productions in all comparisons ~ VMS group; bolded term had higher values for all other comparisons.[5] Effect sizes were calculated using Cliff's Delta for independent samples and regression coefficients for paired samples. *$p<.05$; **$p<.01$; ***$p<.001$

Caregiver prompts

Having established qualitative and quantitative differences in the babble production of infants with and without a VMS consonant, we now turn to our two context variables to

[5] Wilcoxon Estimated-Differences are similar but not identical to mean differences (Hollander & Wolfe, 1999).

address our main hypotheses. Overall, 1,451 consonant productions (CPs) in the data were produced alongside *either* a caregiver prompt *or* an attended object; 527 CPs were produced alongside *both*. Here we consider all instances of caregiver input, including the 527 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 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, 51% of the infants' CPs were preceded by caregiver speech (SD=.18), and 48% of these matched the salient word in this preceding segment (e.g. mother says 'ball' and baby says 'ba'; SD=.23). We ran a Wilcoxon Rank Sum test to determine whether the proportion of each infant's congruent CPs differed by VMS group. This difference was not significant (Est.Diff.=.11, δ=.28, p=.111, 95% CI=[-.03, .3]; see Table 5); withVMS infants' CPs matched caregiver prompts 14% more often than noVMS infants' did (.55(.17) vs. .41(.28)).

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 Figure 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). WithVMS infants' CPs matched the consonantal properties of their caregivers' speech significantly more in the real data set than in the scrambled one (.55(.17) vs. .22(.14), Est.Diff.=.33, δ=.90, $p<.001$, 95% CI=[.23, .43], by Wilcoxon test), as did noVMS infants' CPs (.41(.28) vs. .2(.16), Est.Diff.=.24, δ=.45, $p=.014$, 95% CI=[.04, .37]). Thus, in contrast with our first hypothesis, we find that infants' babble production matches their caregiver's input more than we would expect by chance, even when they don't yet have a stable VMS in their repertoire. See Figure 3.

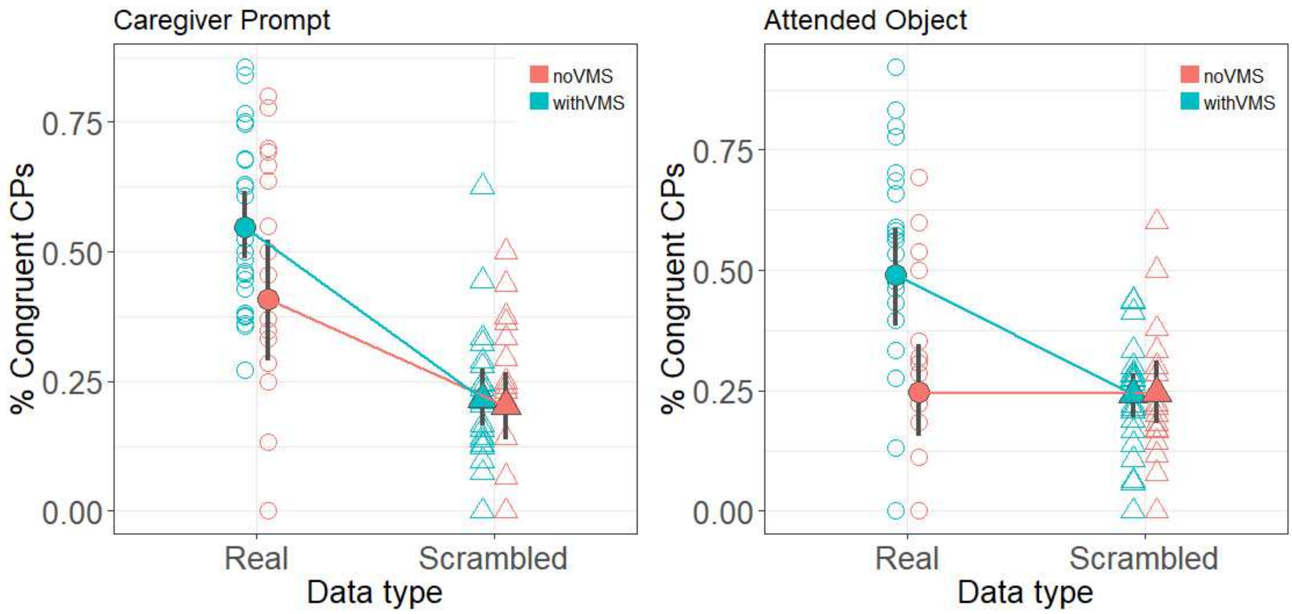

Fig. 3. *Proportion of infants'* ***C****onsonant* ***P****roductions that match a) caregiver prompts and b) attended objects compared with scrambled data (Real vs. Scrambled data). Filled points show the means and 95% bootstrapped CIs for withVMS and noVMS infants (pink and blue points, respectively). Colored lines link mean group values across Real and Scrambled data (circles, and triangles, respectively).*

We next tested whether infants with a stable VMS repertoire (i.e. withVMS infants) *selectively* responded to caregiver input that matched the infants' own consonant repertoire (as opposed to the preceding analysis, which looked at matches between infants' CPs and caregiver input 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 withVMS infants produced significantly more CPs to match caregiver prompts when the prompt was INREP (M=.71, SD=.24) than when it was OUTREP (M=.38, SD=.23; Est.Diff.=.35, *r*=.67, *p*=.001, 95% CI=[.15, .50]). See Figure 4. 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).

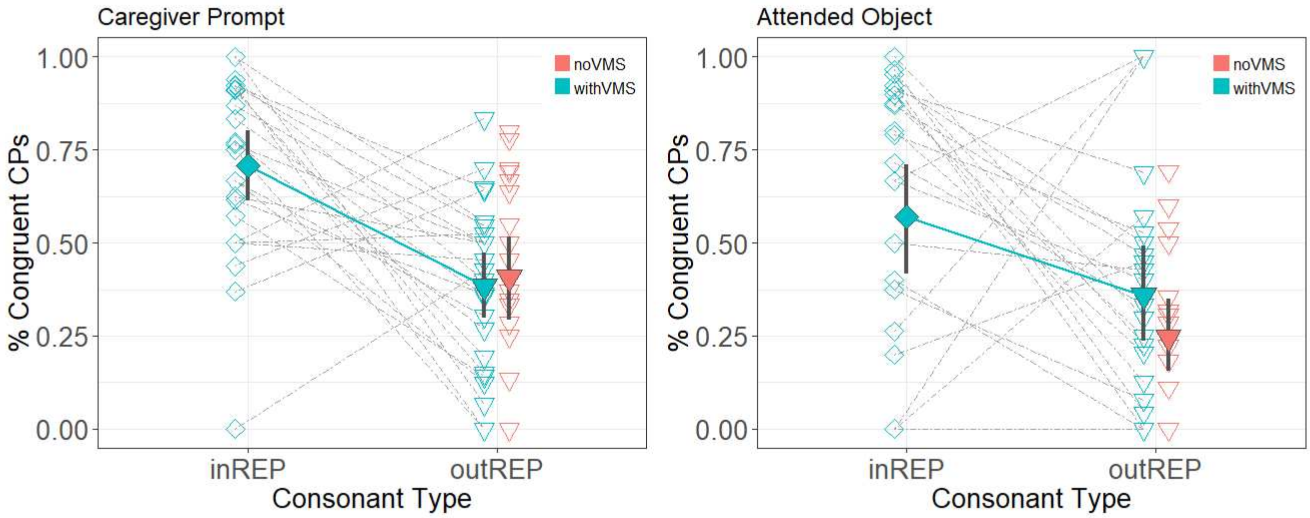

Fig. 4. *Proportion of infants'* ***C****onsonant* ***P****roductions that match a) caregiver prompts and b) attended objects in relation to each infants' phonetic repertoire (i.e. INREP and OUTREP consonants). Filled points show the means and 95% bootstrapped CIs for withVMS and noVMS infants (blue and pink points, respectively). Grey lines link withVMS infants' INREP and OUTREP consonants (diamonds and triangles, respectively). N.B.: all consonants are OUTREP for infants in the noVMS group, by definition.*

Finally, we compared infants' consonant productions following caregiver prompts that were not in their repertoire (i.e. outREP consonants for withVMS infants, and all consonants for noVMS infants). Overall, 39% of OUTREP consonants matched a caregiver prompt (SD=.25). A Wilcoxon Rank Sum test revealed no difference between withVMS (M=.38, SD=.23) and noVMS infants' CPs (M=.41, SD=.28) when the CP was not INREP

(Est.Diff.=-.02, p=.741, δ=-.06, 95% CI=[-.2, .14]). See Table 5. That is, in response to a caregiver prompt outside of their inventory, withVMS and noVMS infants responded with matching consonants equivalently (see Figure 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=.22) were produced while attending to an object, and 38% showed a phonological match with the attended object (SD=.27). 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. WithVMS infants had more matching CPs than noVMS infants (Est.Diff.=.28, δ=.46, p=.003, 95% CI=[.09, .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 (i.e. 51% vs. 25%, respectively; see Table 4 & Figure 3).

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 (.49(.27) vs. .24(.11), Est.Diff.=.28, δ=.61, p<.001, 95% CI=[.15, .40]; by Wilcoxon Rank Sum test), while infants who lacked such a repertoire did not (.25(.22) vs. .24(.16), Est.Diff.<.1, δ<.01, p=1, 95% CI=[-.17, .12]). See Figure 3. Thus here we find support for our first hypothesis, i.e. that infants with stable inventories will babble more congruently with their input than noVMS infants, but only within the context of an 'attended object.'

Next, as we did for caregiver prompts, we determined whether or not withVMS 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; .57(.37)) with the proportion of CPs that matched the attended object but were not in the infant's VMS inventory (OUTREP, .36(.32)). Similar to the caregiver prompts, infants produced more object-matching CPs when the object's label provided a phonological match with their VMS inventory; this difference was marginally significant (Est.Diff.=.32, r=.42, p=.056, 95% CI=[-.01, .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 phoneme in their repertoire. See Figure 4 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 withVMS infants, and all consonants for noVMS infants). Overall, 31% of OUTREP CPs matched an attended object (SD=.29), and this did not differ between VMS groups (Est.Diff.=.08, δ=.13, p=.318, 95% CI=[-.07, .25]; withVMS: M=.36, SD=.32; noVMS: M=.25, SD=.22). That is, given objects whose labels were outside of their inventory, withVMS and noVMS infants responded with matching consonants at low and equivalent rates.

Discussion

We considered 10–11-month-olds' pre-linguistic vocalizations in relation to their established consonant repertoire. 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, analyzing whether having the beginnings of an established phonological system revealed links between an infant's babble and the words and objects in their environment.

Our analysis yielded two key findings: First, in contrast with Goldstein and Schwade (2008), our results showed that infants matched salient words in their input, and this was independent of whether they had a stable consonant repertoire (i.e. a VMS inventory). However, only infants with a stable VMS repertoire produced consonants that matched objects they were attending to. Second, infants with a stable consonant inventory were more likely to produce congruent babble when the *particular stimulus* matched that child's productive repertoire, especially in the caregiver prompt context. These findings are the first to show that the phonological properties of infants' babble are shaped to their on-line experiences of the world. While previous research focuses on the importance of early input (Baumwell, Tamis-LeMonda, & Bornstein, 1997; Gros-Louis et al., 2014) these findings demonstrate an integral role for the production experience gained through babble in early development.

By definition, infants with stable VMS exhibited consistent production of the consonants in their inventory; they also produced more consonants overall. However, contrary to expectations laid out in hypothesis 1, withVMS infants did not respond differently to caregiver prompts than noVMS infants. They were, however, significantly more likely to respond congruently to attended objects. This suggests qualitative differences between the linguistic demands of responding congruently to caregiver prompts vs. attended objects. Both groups could reproduce a salient phonological segment they'd just heard, but only infants with an established phonological 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.

WithVMS infants were better-able to reproduce consonants from input speech, and tended to produce congruent matches to objects if those consonants were stably represented in their own productive repertoires. When we considered consonants that were not in repertoire, they did not differ from noVMS infants in either condition. Together, in line with our predictions, this suggests that a VMS repertoire enhances infants' perceptual attunement to their input, making them more likely to match their own productions to salient words in the speech stream. These results also support the articulatory filter hypothesis (Vihman, 1993): infants' phonological repertoire (VMS) may act as a filter on their perception. Input that matches infants' repertoire appears to pique their interest, perhaps prompting production of their VMS consonant in response. We propose that this input-to-output 'matching' may support more speech-like consonant production, and word learning.

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 in the transition from babble to words. 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* this transition might take place, as infants bridge the gap between babble and words through the articulatory matching of input and output.

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 supports increasingly speech-like early productions; we view our perspectives as complementary. In a preliminary analysis of our home recordings, however, caregivers rarely responded to infants' babble (~25%), leading us to forego such an analysis here. This suggests 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'* capacity to match what they can produce with what they perceive. Of course, the directionality is 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. 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. 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.

Finally, 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, Dailey, Garrison, Amatuni, & Bergelson, 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). 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.

Conclusion

Our results show that infants' early babble production is shaped by their linguistic and visual input, filtered through the established consonants in their phonological repertoire.

When babble is considered through the lens of early phonological development, the sounds of early babble are not random. Rather, our results suggest infants' own production abilities help shape their transition from babble to words.

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].

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. Paper presented at the *International Child Phonology Conference*. Montreal, Canada, 14th June 2019.

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

Boysson-Bardies, B. de, 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

Boysson-Bardies, B. de, & 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

DataVyu: A video coding tool. (2014). New York University: Databrary Project. Retrieved from http://datavyu.org

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

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.). New York, NY: Wiley.

Jakobson, R. (1968). *Child Language, Aphasia and Phonological Universals*. The Hague: Mouton. https://doi.org/10.3109/13682826909011489

LENA Research Foundation. (2018). Retrieved from https://www.lena.org/

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). New York, NY: 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). Amsterdam: 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

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

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/.

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

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., 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

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