Open Library of Humanities

(Non-)effects of linguistic environment on early stable consonant production: A cross cultural case study

Bram Peute, Radboud University Nijmegen, NL, Houtlaan 4, 6525 XZ Nijmegen, NL; Leiden University, Rapenburg 70, 2311 EZ Leiden, NL, aakvs@xs4all.nl

Marisa Casillas, Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, NL; University of Chicago, Rosenwald Hall, 3rd Floor, 1101 E. 58th St., Chicago, IL, USA, mcasillas@uchicago.edu

Recent evidence shows that children reach expected basic linguistic milestones in two rural Indigenous communities, Tseltal and Yélî, despite their infrequent exposure to child-directed speech from adults. However, those results were partly based on measures that are fairly robust to environmental variation, e.g. the onset of babbling. By contrast, directed speech input is typically linked to lexical development, which is environmentally sensitive. We investigate whether these children's vocal motor schemes—a motor-phonological measure of stable consonant production related to early vocabulary—show a similar "expected" developmental pattern to what has been found for typically developing children in urban "child-centered" linguistic communities. We also compare development between the two languages, whose phonological inventories differ greatly in size and complexity. Using spontaneous speech from clips sampled across children's waking days at home, we find that children's canonical babble and stable consonant production is overall comparable to previous work with typically developing children. We find no evidence for difference between the two languages.

1. Introduction

Children's acquisition of spoken language, from early vocalizations to first words and beyond, offers insight into the evolutionary origins and environmental sensitivity of the human language learning system. The trajectory of early vocal development is well documented across a diverse set of language communities: infants start producing protophones soon after birth, then around 6 to 8 months of age they begin to babble, and then some time around the first birthday their first words appear (Oller 1980; Oller et al. 1998; Oller 2000; Lee et al. 2017; Cychosz et al. 2021). Canonical babble serves as an important step along the road to first words. Canonical babble is composed of well-formed syllables, typically reduplicated consonant-vowel structures, like "mama" or "dada" (Lee et al. 2017; McGillion et al. 2017). This kind of early babble is a milestone in the child's motor development. It marks the beginning of a developmental path that ends in their ability to reliably and distinguishably produce the phones relevant to their home language(s).

The robustness of this overall developmental trajectory is underscored in recent work documenting similar developmental patterns despite cross-cultural variability in children's linguistic input. For example, while children in the Tseltal and Yélî Dnye speech communities are directly addressed at a rate similar to middle-class children in urban contexts (e.g., in the US), a substantial portion of their directed speech input (henceforth CDS; "child-directed speech") comes from other children—not adults (Bunce et al. in revision). Diverging caregiver ideologies between these two communities about how to talk to young children also mean that typical CDS input differs greatly in style, everyday activities, topics under discussion, and likely interlocutors (Brown & Casillas in press). Despite these differences, Tseltal and Yélî Dnye children's single- and multi-word lexical vocalizations showed a similar onset to each other and to children in urban, child-centric talk settings where a greater portion of CDS comes from adults (Bunce et al. in revision; Casillas et al. 2020; 2021). This result appears, at face value, to run counter to other work clearly showing that "high quality" adult CDS is strongly associated with faster-growing receptive and productive vocabularies (e.g., Hart & Risley 1995; Hoff 2003; Shneidman & Goldin-Meadow 2012; Weisleder & Fernald 2013). Critically, however, the investigations by Casillas and colleagues were not focused on vocabulary development. The authors instead focused on a few broad categories of vocalization type, several of which are known to be fairly robust to environmental variation (e.g., Oller et al. 1995; Lee et al. 2017; Cychosz et al. 2021).

A stronger test of the impact of CDS input type on early vocal development would instead be to look at patterns in production that link children's pre-lexical behavior to their eventual lexical development. If there were still no signs of delayed development with measures that relate to the lexicon, it would constitute better evidence that these children indeed have some means of gathering language information on the same timescale as children in child-centric environments,

and would justify further work investigating how children pick up on linguistic information from these environments (e.g., via sibling-provided input, third-party input, or other means; Rogoff et al. 1993; Shneidman & Goldin-Meadow 2012; see Casillas et al. 2020 for a discussion).

While past work on early vocal development has been linguistically diverse (de Boysson-Bardies & Vihman 1991; Kunnari 2003; Fikkert & Levelt 2008; Lee et al. 2010; Shneidman & Goldin-Meadow 2012; Weisleder & Fernald 2013; Lee et al. 2017), links between early phonological development, caregiving practices, and everyday language use have yet to be drawn for any rural, Indigenous context. Especially considering that many languages spoken in rural and traditional or Indigenous contexts have typological features underrepresented in developmental research (Kidd & Garcia 2022), exploring the predictive relationship between productions at the pre-lexical and early lexical stages of development in these communities is a key next step for research on language development.

Caregiving practices and everyday activities certainly shape CDS input differently in the two language communities studied here. While Tseltal children are socialized to become expert observer-participants in the interactions around them, Yélî children are more often placed at the center of social interactions (Brown 2011; 2014). CDS in both languages maintains some lexical and prosodic features that distinguish it from adult-directed speech, but Yélî Dnye CDS has been noted to be—relative to Tseltal—more overtly affect-laden and is more often heard from men and boys compared to the Tseltal context (Brown & Casillas in press; Bunce et al. in revision). Infants in the Yélî context are more often passed between multiple caregivers throughout the day and, once they are walking, often spend their afternoons in larger child playgroups than what is typically found in the Tseltal community (Brown & Casillas in press). In brief, while both communities have a large amount of CDS input from other children (e.g., compared to the US), Yélî Dnye-speaking children hear much more CDS from other children and from males compared to Tseltal-speaking children.

In addition to these differences in CDS, the two language phonologies differ greatly, with Yélî Dnye using three times the number of contrastive consonants as Tseltal, and within relatively few phonological parameters (e.g., no contrastive voicing). Further, little is known about the acquisition of phonemes in each language (e.g., Tseltal: ejectives, Yélî Dnye: doubly articulated consonants), some of which are rare across the world's languages. For these reasons, similar developmental trajectories between the two languages would suggest substantial robustness in early vocal development and would fall in line with other work showing that (a) similar early consonants arise in children's first productions across a range of languages and that (b) cross-language differences in prelexical babble are relatively subtle and lead to similar phonetic forms in children's first words (i.e., counter to Jakobson 1941/1968; see, e.g., Oller et al. 1976; Stoel-Gammon & Cooper 1984; Vihman et al. 1985; see also Lee et al. 2017).

In addition to canonical babble, which was investigated in prior work with these communities, we here investigate children's "vocal motor schemes" (henceforth VMS), which are motor plans for phone production developed through babbling. VMSs provide a bridge between pre-lexical babble and early productive vocabulary. They provide the child with the means of producing "auditory approximations" to the phonemes present in the target words produced by adults, but without the child having to deduce the exact phonemes involved (Vihman 1993). This capacity for consistent phonetic patterning is theorized to help prepare the child for producing stable consonants, described as "a generalized action plan that generates consistent phonetic forms [...] a formalized pattern of motor activity that does not require heavy cognitive resources to enact" (McCune & Vihman 2001: 152). McCune and Vihman (2001) operationalize VMS in their longitudinal dataset of at-home play sessions as follows: If a child produces 10 realizations of a given consonant in 3 or 4 contiguous monthly sessions, the child is said to have acquired the VMS for that consonant (see **Figure 1** for an illustration of our similar counting process). On average, the typically developing American English-learning children in McCune and Vihman's (2001) study acquired two VMS consonants at 12 months (SD = 1.78, range = 9–14). A larger dataset of at-home play sessions with typically developing British English-learning children by DePaolis and colleagues (2011) showed a slightly earlier average achievement of 2 VMS: 10 months (SD = 1.35, range = 9–15). Though we note that this dataset used an additional criterion for counting a VMS (50 uses of a consonant within 30 minutes). While these datasets do not provide sufficient evidence for a clear linguistic milestone (e.g., as we have for the onset of canonical babble; Oller 2000; Lee et al. 2017; Cychosz et al. 2021), they do set a preliminary range of typical early VMS development: 2 VMS between 9 and 15 months, reached, on average, by 10–12 months.

Variation in VMS development is meaningful for later lexical development. McCune and Vihman (2001) first linked VMS to children's shift to referential word use, defined in their study as use of a word across two or more contexts. Most children achieved 2 VMS just before making this referential transition, and most of these stable early word forms included the child's acquired VMS consonants. More recent work also supports the idea that individual variation in VMS development predicts multiple aspects of lexical development and language processing, including: the subsequent appearance of first words (McGillion et al. 2017), early expressive vocabulary size (Majorano et al. 2014), and attention to auditory presentations of familiar and unfamiliar words, or words that do and don't include the child's acquired VMS consonants (DePaolis et al. 2011; Majorano et al. 2014). Although VMS consonants do not directly reflect frequency of use in caregiver input (DePaolis et al. 2011; Majorano et al. 2014), the established link between variance in VMS acquisition and variance in lexical development—the latter of which is known to be very sensitive to environmental variation in CDS (Hart & Risley 1995; Hoff 2003; Shneidman & Goldin-Meadow 2012; Weisleder & Fernald 2013)—suggests the possibility that broader characteristics of VMS development may vary across diverse CDS input contexts.

For example, more adult CDS may be associated with average earlier VMS acquisition in *addition* to average larger vocabulary size later on. If so, we might expect that children in language communities with less adult-produced CDS, such as the Tseltal and Yélî groups studied here, show somewhat later VMS development than children in environments with more frequent adult CDS. Similarly, we might expect that children learning Yélî Dnye, who hear more of their directed input from other children, would show somewhat later VMS development than Tseltal children. If not, it would lend support to the ideas that (a) VMS development (and early words, which use VMS consonants) are primarily motor-driven processes and that, (b) while individual differences in VMS development predict aspects of early referential word use, VMS development itself is not subject to the same predictors that have been recognized for later lexical growth (e.g., quantity of quality directed speech from adults).

1.1 The present study

The present study focuses on the acquisition of consonants (both in onset and coda position) by children born into two speech communities: Tenejapan Tseltal and Yélî Dnye. Tseltal is a Mayan language spoken in the highlands of Chiapas in Southern Mexico. This language, which has five main dialects, is in vigorous use with estimates of over 400,000 to 500,000 speakers who are bilinguals in Spanish, and an estimated 40,000 to 50,000 monolingual speakers (Polian 2013; Eberhard et al. 2020). Yélî Dnye is an isolate spoken on Rossel Island, in the Milne Bay Province of Papua New Guinea. The language, which has two main dialects, is in vigorous use as well, with a loosely estimated 5000–7000 native speakers, most of whom are at least partly bilingual in English, Tok Pisin, and/or other languages of Papua New Guinea (Levinson 2022; Eberhard et al. 2020). Many aspects of daily life, as well as the overall rate of CDS, are similar between these two patrilocally organized, swidden horticulturalist communities (Casillas et al. 2020; 2021). Before age 3;0, children in both communities are directly addressed around 3–3.5 minutes per hour during a day at home (Casillas et al. 2020; 2021). While the quantity of directed linguistic input is relatively similar, differences in early caregiving and ideas about speaking to young children create differences in who speaks to young children and how, with more input from men and children in the Yélî community compared to Tseltal. Additionally, the consonant inventories of the two languages are highly distinct from each other. Yélî Dnye has 60 contrastive consonants compared to Tseltal's 20, with the additional consonants including both single- and doubly-articulated stops with nasal, labial, and/or palatal releases and pre-nasalization (Levinson 2022; Polian 2013). The fact that these two language communities use similar rates of directed input but have highly divergent phonological inventories makes the comparison of children's early consonant development interesting to explore. While we would predict similar early consonants to appear across languages on the basis of past work (e.g., Oller et al. 1976; Stoel-Gammon & Cooper 1984; Vihman et al. 1985), the understudied nature of these two very different languages offers an opportunity to test this assumption.

If Tseltal and Yélî Dnye-acquiring children show similar VMS development to each other, and to same-aged English-acquiring children, it would suggest that even lexically related early phonological development is cross-culturally robust to variation in CDS use, and that children's linguistic growth in this domain is spurred on by other factors, including their own motor development, and potentially other cues from their linguistic environment to spur their early linguistic growth. For example, many scholars focusing on language development in non-urban contexts have highlighted the role of observable third-party behavior for learning (e.g., Rogoff et al. 1993; Chavajay & Rogoff 1999; Gaskins & Paradise 2010). In addition to VMS, we also attempt to replicate the finding that canonical babble onset shows no delayed development in these two communities, this time using a larger sample and independent annotations from the original study data (Casillas et al. 2020; 2021).

1.2 Predictions

Based on prior findings, we hypothesize that the onset of canonical babble takes place between 6 and 8 months in both datasets (Oller 1980; Oller et al. 1998; Oller 2000; Lee et al. 2017; Cychosz et al. 2021). We also predict that VMS acquisition by Yélî and Tseltal children will be comparable to that of Western children; that is, children will reach 2 VMS consonants between 9 and 15 months, typically between 10 and 12 months (McCune & Vihman 2001; DePaolis et al. 2011). We make this prediction in line with the prior finding that the onset of single- and multi-word utterances is not delayed in these communities (Casillas et al. 2020; 2021), and despite the fact that VMS is related to early lexical development, which is known to later be sensitive to ambient adult CDS rates. Finally, given the differences in phonological inventories between Tseltal and Yélî Dnye, we also examine whether the onset of VMS consonants is slightly later for children learning a language with a larger phonological inventory. This finding is not predicted if children's initial consonants are drawn from core types that appear across languages (Oller et al. 1976; Stoel-Gammon & Cooper 1984; Vihman et al. 1985) but regardless we examine this possibility in the present study given that we are working with two understudied languages and a recording type that is relatively new for studies of VMS. As such, we predict that, if there *were* a difference, Yélî children would acquire their VMS consonants slightly later than Tseltal children because the Yélî consonant inventory is large, complex, and fits many contrasts into a relatively small acoustic space (Levinson 2022; see also Cristia & Casillas 2022). That said, differences based on inventory effects may not emerge with children this young (Jakobson 1968; Vihman & de Boysson-Bardies 1994; Monnin & Lœvenbruck 2010).

2. Methods

2.1 The communities

We analyzed a total of 15 hours and 45 minutes of Tseltal and 4 hours and 30 minutes of Yélî Dnye audio recordings for the phonological content of children's spontaneous vocalizations. The data consist of annotated segments from one daylong recording per child. From each recording we

extract a single estimate for canonical proportion (i.e., the proportion of vocalizations containing at least some canonical babble) and a single estimate for VMS acquisition.

This approach differs from most other studies on VMS and early babble, where multiple recordings or longitudinal data are used to establish stable evidence of phonological development (e.g., Vihman & de Boysson-Bardies 1994; McCune & Vihman 2001; DePaolis et al. 2011; Majorano et al. 2014; Oller et al. 1995; Lee et al. 2017). Laing and Bergelson (2020) established VMS counts for their participants (mostly) at single timepoints by adopting DePaolis and colleagues' (2011) criterion of 50 productions of a consonant within 30 minutes. They then used automated output from the LENA software accompanying their daylong recordings to identify the top 30-minute clips of infant vocalization from the long (up to 16 hours) recording, manually choosing among these clips and sometimes recombining subparts of them to estimate VMS from 30 minutes of high-quality infant vocalization.

In the remote fieldwork context in which the present data were collected, recordings could only be made during a few weeks' visit every 1–2 years, rendering longitudinal data collection impractical. The original dataset aimed to capture one recording from as many children as possible, leaving little time to make multiple recordings with individual children. Lacking longitudinal data, repeated recordings, or automated annotation output, we are forced to adapt the VMS measure to our present dataset. We do, however, use a secondary measure to establish that the resulting estimates reflect stable consonant production (see below). That said, the lack of multiple samples per child, longitudinal or not, increases the likelihood that some of the resulting estimates reflect production patterns that aren't representative of the child's true abilities at that time; thus the patterns we find should be considered at the population level, not at the individual level.

The recordings were collected in 2015 (Tseltal) and 2016 (Yélî Dnye) and can be accessed via the Casillas HomeBank corpus (Casillas et al. 2017). Participant consent processes and data collection were conducted in accordance with ethical guidelines approved by the Radboud University Social Sciences Ethics Committee. We focused on children between 5 and 20 months because children are expected to begin canonical babble production around 6 months, sometimes earlier, and by 20 months nearly all children would be expected to have started producing recognizable words (Oller et al. 1998; McGillion et al. 2017; Casillas et al. 2020).

The Tseltal-speaking children come from a farming community in the highlands of Chiapas in Southern Mexico, where they are typically raised in patrilineally organized, multigenerational households. During the day, infants are carried on their mother's back while she goes about her business, or they are left at home with other family members while the mother works elsewhere (e.g., in the field). The majority of children in this community grow up monolingually until they go to school (Casillas et al. 2020), and their linguistic environments have been characterized as non-child-centered and non-object-centered (Brown 1998; 2011; 2014). CDS from adults is typically limited until the infants themselves start to seek out verbal interaction (Brown 2014) and continues to be relatively infrequent through 3;0 (Casillas et al. 2020). The Tseltal data used

in the current study include 20 children (M = 10 months; median = 9; range = 5–19), including four children whose recordings were used in the Casillas et al. (2020) study.

The Yélî Dnye-speaking children live in a collection of small settlements on the north-eastern shore of Rossel Island, which is located 250 nautical miles off the south coast of mainland Papua New Guinea. Children grow up in hamlets with patrilocally organized household clusters, where there is often a shared open space between households. During the day, children are carried in their caregivers' arms, and they are frequently passed around between community members— even those from far outside the natal hamlet—who return the child to the mother for feedings (Casillas et al. 2021; Brown & Casillas in press). Yélî children mostly grow up speaking Yélî Dnye at home, although English, Tok Pisin, and other regional languages are often spoken by adults and school-aged children (Brown & Casillas in press; Casillas et al. 2021). Children begin to learn English once they start school. The linguistic environment of Yélî children can be characterized as child-centered (Brown & Casillas in press; Ochs & Schieffelin 1984); children are considered a shared responsibility, as well as a source of joy and entertainment for caregivers, and, as such, interaction with infants and young children on Rossel Island is initiated by women, men, girls, and boys alike (Casillas et al. 2021). The Yélî data we use in the current study include all 12 children in the 5–20-month age range who were reported as acquiring Yélî Dnye monolingually in the 2016 Yélî Dnye Casillas HomeBank corpus (M = 12.4 months; median = 12.5; range = 8–17), including four children whose recordings were used in the Casillas et al. (2021) study.

Language	N	Mean age in months (Median, SD, range)
Tseltal	20	10 (9, 4.0, 5–19)
Yélî Dnye	12	12.4 (12.5, 2.6, 8–17)

Table 1: Descriptive statistics of the age in months of the children in the present dataset.

2.2 The data

For both datasets, the recordings were made using an audio recorder (Olympus WS-832 or WS-853) and photo camera (Narrative Clip 1) strapped to the child's chest during most of a waking day at home. For young infants and very small children, the primary caregiver wore the photo camera (see Casillas et al. 2020 for details). The recordings document language use over the course of multiple home activity contexts. We do not use the image data in this study—only the audio recordings.

We base our analyses on a subsample of each child's spontaneous vocalizations from the day. We selected short random clips from each recording, following the random sampling procedure used for the eight children whose recordings were previously annotated for Casillas et al. (2020) and (2021). That is, we focused our analyses on spontaneous vocalizations by the target child

within nine randomly selected and non-overlapping clips for each recording; clips were 5 minutes long for the Tseltal data and 2.5 minutes long for the Yélî Dnye data. The reason for this disparity in clip duration is that the number of speakers and amount of background noise in the Yélî dataset makes it particularly time consuming to annotate, limiting what can be feasibly accomplished in transcription during researcher visits to the island (Casillas et al. 2021). We take this disparity into account in the analysis, as explained below. Considering that, for Yélî Dnye, we also have fewer recordings from children in the target age range (12 vs. 20 for Tseltal) *and* fewer recordings from young children (**Table 1**), the Yélî Dnye data should be regarded as very preliminary, especially at the younger end of the studied age range.

Each target child vocalization in each clip was classified and broadly phonetically transcribed. For the eight recordings with existing annotations, phonetic transcriptions were added to the existing speech annotations (completely independently from their previously added vocal maturity classification). For all other recordings, the child's vocalizations within each clip were first diarized, and subsequently classified and phonetically transcribed. Vocalization annotation followed the scheme shown in **Table 2**; non-canonical vocalizations were labeled as such when a vocalization was not classifiable as laughing or crying but also did not include canonical babble (e.g., vowel-only vocalizations).

Annotation	Meaning
hamuwa	example of a transcribed vocalization with canonical babble
N	non-canonical vocalizations
L	laughing
Y	crying

Table 2: Vocalization annotation scheme.

The first author—a linguist with training in phonetics who lacked previous experience with either Tseltal or Yélî Dnye—transcribed children's canonical babble using the International Phonetic Alphabet (International Phonetic Association 2020) on the basis of the phoneme inventory of each language, as it is spoken by adults. Instances in which children produced phones that are not present in the adult language were transcribed as perceived by the first author. We note that phone production by infants and very young children is typically limited to a subset of phones (Prather et al. 1975; Ingram 2008; DePaolis et al. 2016) that commonly appear across the world's languages (Moran & McCloy 2019), including simple labial, coronal, and velar oral and nasal stops with occasional fricatives, all of which occur in the languages native and highly familiar to the annotator. It is of course possible that children learning Tseltal and Yélî Dnye produce other phones outside of this set; phonological development data

for these languages is highly limited. We were unable to have our annotations completed or corrected by native speakers of each language given the impedance of travel caused by the COVID-19 pandemic. Instead, the second author, who is more familiar with the two languages and is also a linguist with phonetic training, independently annotated a random two clips (22%) of each child's recording data and found high agreement with the first annotator (see Appendix for details).

While annotating phones in the data, the transcriber kept in mind the native phonemes of the respective adult language for each dataset, which are shown in **Tables 3–6** (Tseltal phonology information derived from Polian's (2013) grammar; Yélî Dnye phonology information derived from Levinson's (2022) grammar). Occasionally, the children were observed to produce phones that are not phonemic in the adult language; these phones have been added to their respective tables and are shown in bolded and underlined text (e.g., [dʒ] is not a native phone in Tseltal but was observed in a child's production and so is added to **Table 3**).

	Bilabial	Alveodental	Palatoalveolar	Velar	Glottal
Plosives	p b p'	t **d** t'		k k'	ʔ
Affricates		ts ts'	tʃ **dʒ** tʃ'		
Fricatives		s	ʃ **ʒ**	x	h
Nasals	m	n			
Laterals		l			
Rhotics		r			
Approximants	β		j	w	

Table 3: Tseltal consonants (non-bold, no line), with the additional, non-native phones we found in Tseltal children's spontaneous vocalizations (bold, underlined).

	Front	Central	Back
Close	i		u
Close-mid	e	**ə**	o
Open		a	

Table 4: Tseltal vowels (non-bold, no line), with the additional, non-native phones we found in Tseltal children's spontaneous vocalizations (bold, underlined).

We note that the majority of Yélî Dnye phones were not attested in young Yélî Dnye-acquiring children's vocalizations (e.g., we found no doubly articulated consonants).

Table 5: Yélî Dnye consonants (non-bold and no line) with additional, non-native phones found in Yélî children's spontaneous vocalizations (bold, underlined). SBC stands for 'simultaneous bilabial closure'.

	Bilabial				Alveolar		Alv.+SBC		Post-Alv.		Post-Alv.+SBC		Velar			
		Pal.	Lab.	Both		Pal.		Pal.		Pal.		Pal.		**g̲**	Pal.	Lab.
Plosives	p	pʲ	pʷ	pʲʷ	t	tʃ	tp	tpʲ	t̠	t̠ʲ	tp	tpʲ	k	**g̲**	kʲ	kʷ
Prenasalized plosives	mb	mbʲ	mbʷ	mbʲʷ	nd	ndʒ	nmdb	nmdbʲ	n̠d̠	n̠d̠ʲ	n̠m̠db	n̠m̠dbʲ	ŋg			ŋgʷ
Nasally-released plosives									t̠n̠	t̠n̠ʲ	tpn̠m̠	tpn̠m̠ʲ	kŋ			kŋʷ
Nasals	m	mʲ	mʷ	mʲʷ	n	nʲ	nm	nmʲ	n̠	n̠ʲ	n̠m̠	n̠m̠ʲ	ŋ			ŋʷ
Approximants	w~β	βʲ			**r̲**				jl	lʲ	lβ					
Fricatives									ʃ				ɣ	**x̲**		

	Velar+SBC		Glottal
		Pal.	
Plosives	kp	kpʲ	**ʔ̲**
Prenasalized plosives	ŋmgb		
Nasally-released plosives	kpn̠m̠		
Nasals	ŋm		
Fricatives			**h̲**

	Front		Central		Back	
	Oral	**Nasal**	**Oral**	**Nasal**	**Oral**	**Nasal**
Close	i	ĩ	ɨ		u	ũ
Near-close	e		ə	ə̃		
Open-mid	ɛ	ɛ̃			ɔ	ɔ̃
(Near-)open	æ		ɐ	ã	ɑ	ɑ̃

Table 6: Yélî Dnye vowels.

2.3 The framework

In order to compare the onset and development of canonical babble produced by Tseltal and Yélî children to each other and also to other children discussed in previous literature (e.g., Lee et al. 2017; Cychosz et al. 2021), we calculated their canonical proportion (CP; see formula below). To make this calculation, we looked per child at what percentage of their vocalizations included canonical babble. Lee et al. (2017) used a similar measure to quantify the amount of canonical babble, but looked at canonical productions by syllables, a measurement they called the canonical babbling ratio (CBR). In contrast, we looked at canonical babble by whole vocalization, in line with the approach of Cychosz et al. (2021). To calculate CP in the present study, any vocalization containing one or more canonical syllables was counted as an instance of canonical babble, and we subsequently divided the number of canonical babble-containing vocalizations by the total number of vocalizations, excluding vocalizations comprised solely of laughing or crying (**Table 2**).

Formula for calculating CP:

$$\frac{\#voc\text{'s with canonical babble}}{\#voc\text{'s with canonical babble} + \#non\text{-}cry/non\text{-}laugh\ voc\text{'s without canonical babble}}$$

We deployed VMS as a measure of consonant development, though we made some adjustments to McCune and Vihman's (2001) original measure in order to accommodate our cross-sectional, daylong recording dataset. Our lack of longitudinal data prevents us from observing individual children's development. However, the daylong recording collection affords us more data per individual under a highly natural production setting. Laing and Bergelson (2020) define VMS as achieved at a single time point if the child produces 50+ tokens of a phone within the highest 30-minutes of vocal activity in the day. However, without automated annotation output like they had, we were unable to replicate this method. Instead, we created a new adaptation of VMS, which examines children's consonant production across the day. We define VMS as follows: If a Tseltal child produced 10+ realizations of a consonant within their total 45 randomly sampled minutes, or if a Yélî child produced 5+ realizations of a consonant within their total 22.5 randomly sampled minutes, then we consider the child to have acquired the VMS for that consonant. A visual aid of this process using hypothetical data is shown in **Figure 1**. On the

basis of vocalization rates in the data from Casillas et al. (2020) and (2021) we had anticipated that our random clip sampling would typically yield 100+ vocalizations per child, a lower-bound estimate for what has previously been required to secure stable VMS scores (Vihman et al. 1985; Vihman et al. 1994; Vihman personal communication). As noted below, there were a few children for whom we found fewer than 100 vocalizations in their random clips, namely 1 Yélî Dnye-acquiring child and 3 Tseltal-acquiring children.

While a departure from prior definitions of VMS, our present measure capitalizes on children's productions over the course of the day, which affords a new perspective on consonant production stability. We verified that our adapted VMS scores highly correlate with a second stability measure: the number of distinct consonant types each child produced across four or more of their nine clips ("cross-clip consonants"; see Appendix). Cross-clip consonant production, the secondary measure we computed, correlates with VMS counts and produces the same pattern of statistical results reported below with VMS. That said, our direct comparisons to previous VMS outcomes in what follows should be taken with a pinch of salt—the best comparison we can make is between the two populations we directly examine here, where the method of data collection and VMS measurement is the same.

Following McCune and Vihman (2001), we collapsed voiced and voiceless variants of produced consonants, as this distinction is often not yet mastered by young children (Eilers et al. 1984). Voicing is also not contrastive in either language, which further emphasizes the irrelevance of this distinction for the present study. We also only counted supraglottal consonants toward children's VMS score, following prior work: Glottals and glides already occur frequently
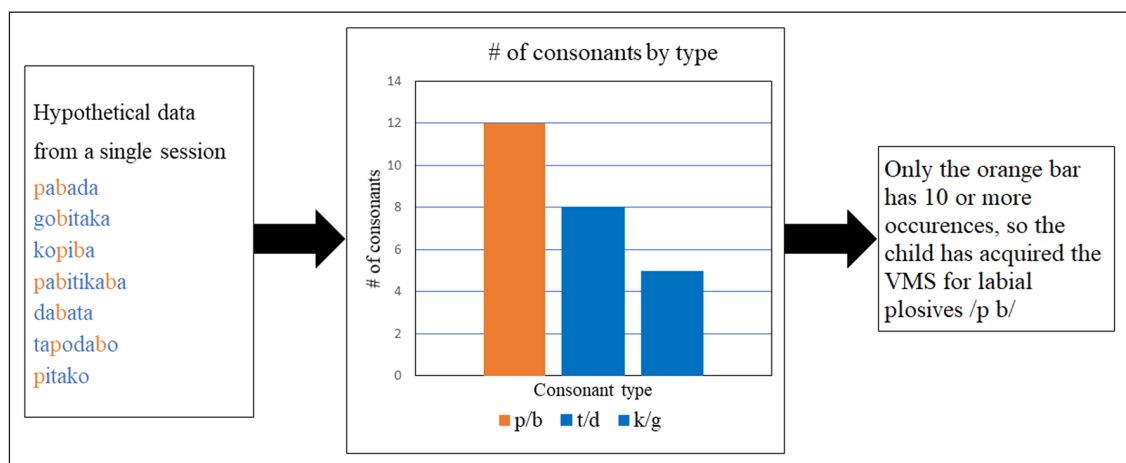

Figure 1: Illustration of how VMSs are derived from the data. The illustration uses artificial strings and limits consonants to labial, alveolar, and velar stops in order to clearly show our process. Note again that, because of the difference in the amount of annotated data between corpora, we define 5 occurrences of a consonant as a VMS in the case of Yélî Dnye, instead of 10 occurrences as we do for Tseltal.

in the early part of children's first year while supraglottal consonants only begin to occur around 6–8 months (McCune & Vihman 2001). Notably, the majority of non-native phones attested in the data fall into these two categories (i.e., voiced equivalents of native voiceless phones and subglottal consonants; see the bolded, underlined phones in **Tables 3** and **5**), supporting the omission of these distinctions in the analysis and in related past work.

3. Results

We modeled our two dependent measures—canonical proportion and the number of VMSs produced by each child—using a linear regression with fixed effects of age (numeric, in months), language (Tseltal/Yélî Dnye), and their interaction (i.e., measure ~ age * language). Because we only had one datapoint per child (e.g., the number of VMSs that child displayed), we were unable to include a random effect of child. The analysis was conducted in R (R Core Team 2018) using lme4 (Bates et al. 2015), and all plots were generated with ggplot2 (Wickham 2009). Analysis scripts associated with this project can be found in its public repository at https://github.com/marisacasillas/TS_and_YD-VMS; the raw input files (in .eaf ELAN format) contain potential identifying information via utterance transcription, so can only be securely accessed in the HomeBank Casillas corpus (Casillas et al. 2017; VanDam et al. 2016). We first review findings regarding canonical proportion, then findings for VMS acquisition.

3.1 Canonical proportion

All Tseltal- and Yélî-speaking children 8 months and older used canonical babbles, consistent with previous findings (Oller 1980; Oller et al. 1998; Oller 2000; Lee et al. 2017). Furthermore, and consistent with other cross-linguistic data, both populations of children had a CP greater than 0.15 after age 0;10 (Lee et al. 2017; Cychosz et al. 2021), where CP less than 0.15 might indicate delay in development (Oller et al. 1995). Differences in CP between the youngest and oldest children in each corpus were apparent (**Figure 2**), with a gradual increase more apparent among the Tseltal children. Note, however, that the youngest children in our Yélî sample are older than the youngest children in the Tseltal sample, and as such already produce vocalizations with canonical babble quite frequently.

A linear regression of CP (see **Table 7**) revealed a significant positive effect of age ($p < .01$), no significant effect of language ($p = .15$), and no age-by-language interaction ($p = .15$). In other words, while the proportional use of canonical babble increased with age overall, there is no evidence for differing developmental rates per language group, neither overall nor specifically for age-related increase between languages.

3.2 VMS consonants

Prior work suggests that American English- and British English-acquiring children typically acquire 2 VMS consonants between 9 and 15 months, on average reaching this benchmark
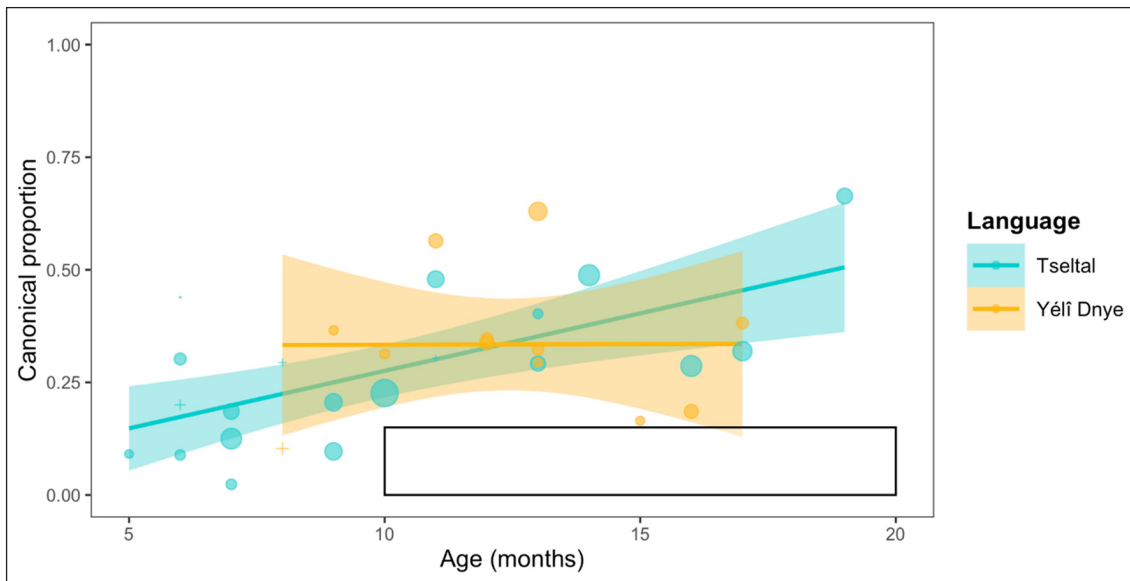
Figure 2: CP of Tseltal- and Yélî Dnye-speaking children. Scores falling in the white box would be unexpected for typically developing children based on the benchmark of 0.15 CP at age 0;10 and later (Oller et al. 1995; Lee et al. 2017; Cychosz et al. 2021). Point size indicates the number of vocalizations (range: 64–643), and those shaped '+' indicate fewer than 100 vocalizations found.

Coefficients:				
Term	**Estimate**	**Std. Error**	**t-value**	**p-value**
(Intercept)	0.02	0.082	0.243	0.81
Age in months	0.026	0.026	3.339	<.01
Language	0.311	0.211	1.474	0.152
Age in months:Language	–0.025	0.017	–1.478	0.151

Table 7: Output of the CP regression analysis.

between 10 and 12 months (McCune & Vihman 2001; DePaolis et al. 2016). We see that most Tseltal- and Yélî Dnye-acquiring children showed VMS scores aligning with prior findings based on English, though 1 Tseltal-learning child and 3 Yélî Dnye-learning children produced fewer VMS consonants than expected (**Figure 3**). These findings are not necessarily anomalous: in McCune & Vihman's (2001) study, 6 out 20, or 30%, of the English children did not reach 2 VMS before 15 months, which is comparable to the 27.3% of Tsetal and 45.4% of Yélî children found here. In contrast, in the study by DePaolis et al. (2016) all of the children did achieve 2 VMSs before 15 months. An overview of all VMS consonants found in the Tseltal and Yélî Dnye data is shown in **Table 8** (the data tables used to generate **Figure 3** and **Table 8** are shown in Appendix Table 1 and Appendix Table 2).
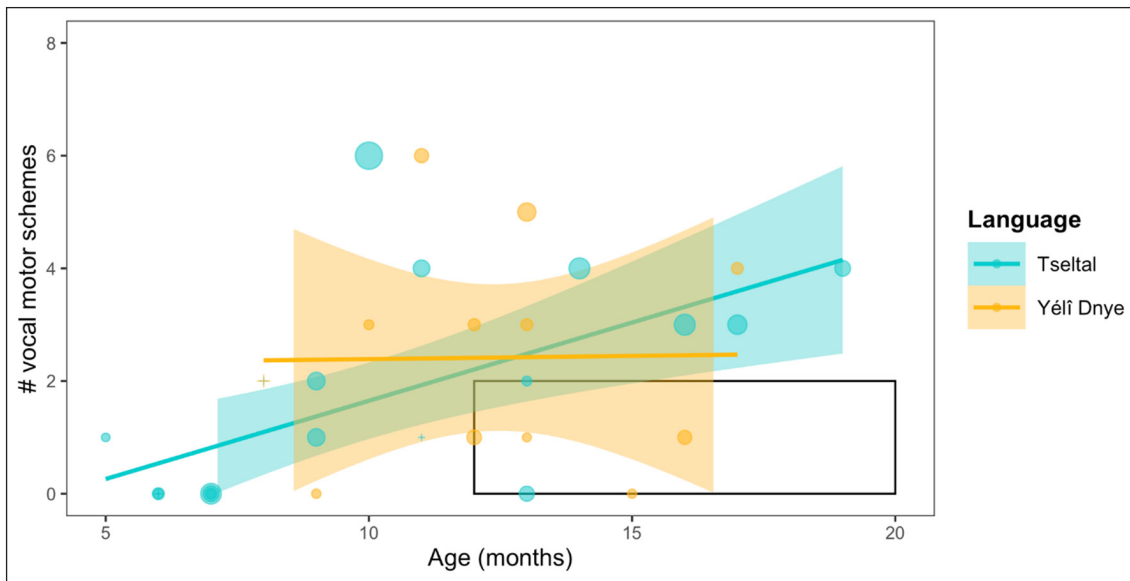
Figure 3: VMS count of Tseltal- and Yélî Dnye-speaking children; the white box indicates <2 VMSs beyond 12 months. Point size indicates number of vocalizations (range: 64–643), and those shaped '+' indicate fewer than 100 vocalizations found.

In both Tseltal- and Yélî Dnye-acquiring children, all phones reaching VMS were part of the native phonological inventory, voicing contrasts aside. The most common VMS consonants produced by children in both samples were [t/d] and [m], followed by [p/b] and [k/g] (see **Table 8**). The [m] VMS is more prevalent here than in prior studies, while [p/b], [t/d], and [k/g] are well-attested early VMS consonants in English, Italian, and Welsh (DePaolis et al. 2011; Majorano et al. 2014; McCune & Vihman 2001; DePaolis et al. 2013). The difference in distribution of the plosives across these two languages shows an interesting pattern: while in both cases [t/d] is the most prevalent type of plosive (in line with the hypothesis that the alveolar place of articulation is universally less marked, (Shaw 1991; Tsuji et al. 2015), in Tseltal the second most prevalent is [p/b] with few instances of [k/g], while in Yélî Dnye [k/g] is the second most prevalent plosive type and there are few instances of [p/b]. The graphs for the total number of phones used across tokens analyzed in each language are shown in Appendix Figure 4.

	p/b	t/d	k/g	m	n	l
Tseltal	7	8	2	8	5	3
Yélî Dnye	3	7	6	7	2	4

Table 8: Distribution of all VMS consonants acquired by language sample.

A linear regression of VMS counts (see **Table 9**) revealed a significant positive effect of age ($p < .01$), no significant effect of language ($p = .20$) and no age-by-language interaction ($p = .21$). In other words, while the number of VMSs acquired increased with age overall, there is no evidence for different patterning by language group, neither overall nor specifically for age-related increase between languages.

Coefficients:				
Term	**Estimate**	**Std. Error**	**t-value**	**p-value**
(Intercept)	−1.128	1.002	−1.125	0.27
Age in months	0.278	0.093	2.987	<.01
Language	3.404	2.565	1.327	0.195
Age in months:Language	−0.266	0.208	−1.281	0.211

Table 9: Output of the VMS regression analysis.

4. Discussion

Children living in the Tseltal and Yélî communities are less frequently directly spoken to by adults, yet prior work shows no apparent delay in their early linguistic development (Casillas et al. 2020; 2021). This finding may seem counterintuitive, as adult caregiver CDS has previously been strongly associated with faster-growing receptive and productive vocabularies (Hart & Risley 1995; Hoff 2003; Shneidman & Goldin-Meadow 2012; Weisleder & Fernald 2013). The apparent discrepancy may lie in the fact that prior work on Tseltal and Yélî Dnye used vocal maturity measures that are, in fact, fairly robust to environmental variation. In the current study, we investigated whether there was evidence of delay on a measure of early phonological development that has been shown to relate to early lexical development—stable early consonant production—by adapting a measure of vocal motor scheme acquisition. We also tested whether the prior finding of non-delayed canonical babble onset would hold up with a larger sample of children than was studied previously.

We predicted that, replicating prior work, canonical babble development would show no delays in either corpus (Casillas et al. 2020; 2021; Cychosz et al. 2021). Following prior results showing no delays in the emergence of single- and multi-word utterances (Casillas et al. 2020; 2021), we further hypothesized that VMS acquisition by Yélî Dnye- and Tseltal-learning children would be on par with previous results from English-learning children, who typically acquire at least 2 VMS consonants around age 10 to 12 months, almost certainly doing so by 15 months. Given the large difference in the size of these two languages' phoneme inventories and the relatively new recording type (i.e., daylong audio), we also opportunistically explored the possibility that Yélî children would acquire their VMS consonants slightly later than Tseltal children.

4.1 Canonical babbling

Prior work has shown that most children start using canonical babble after 8 months, with canonical babble making up at least 0.15 of their total syllables by 0;10 (Oller 1980; Oller et al. 1998; Oller 2000; Lee et al. 2017; Cychosz et al. 2021). Using a related measure, canonical proportion (CP), we found that both Tseltal- and Yélî Dnye-acquiring children have indeed already surpassed this benchmark of 0.15 at 7 months, in line with the cross-linguistic findings of Cychosz et al. (2021). This is significantly earlier than the 10 months reported by Lee et al. (2017), which might be because CP is a less fine-grained measure than the canonical babbling ratio they use (CBR). These results are consistent with prior work on these populations (Casillas et al. 2020; 2021) and from other language communities, despite the lower rate of directed speech from adult speakers in the present communities of study.

4.2 Vocal motor schemes

Prior work on typically developing children learning English suggests that children typically use two stable consonants in their pre-lexical babble (i.e., vocal motor schemes) beginning between 10 and 12 months of age, almost certainly reaching this milestone by 15 months (McCune & Vihman 2001; DePaolis et al. 2016). Most of the Tseltal and Yélî Dnye-learning children followed this pattern, and the consonants qualifying as having reached VMS status are consistent with past work (i.e., [t/d] > [p/b], [k/g] > [m], [n], [l]). Importantly, 1 Tseltal-learning child (1;1) and 3 Yélî Dnye-learning children (1;1, 1;3, and 1;4) produced fewer than two VMS consonants after 12 months (27.3% and 45.4% of those samples, respectively). Our review of individual variation among typically developing children in the prior studies on English suggests that this pattern is comparable to between-individual variation reported by McCune and Vihman (2001; 30% of the sample) but represents a larger-than-expected share of the sample with low VMS counts compared to DePaolis et al. (2016; 0% of the sample). Because our adapted VMS measure and our cross-sectional data differ from these past studies on English, direct comparison of this milestone is noisy at best. However, we take the overall similarity in patterning across age as providing no clear support for the idea that these children, who hear less adult CDS, are delayed in their early linguistic development—in this case with a pre-lexical measure that has been shown to connect to early lexical development.

Our best comparisons can be made between the two language communities we directly study here, because the methods of data collection, sampling, and VMS estimation were nearly identical. We predicted that, if there *were* a difference between the communities, Yélî Dnye-acquiring children would have a slightly slower development of VMS consonants given the larger and more complex Yélî Dnye phonological inventory. While relatively more Yélî children than Tseltal children did not reach the 2-VMS milestone by 12 months, we found no statistical evidence for an effect of language on VMS consonant counts, nor any interaction effect of language and age on

VMS consonant counts. This finding is consistent with past work showing similar types of early consonants across languages (Oller et al. 1976; Stoel-Gammon & Cooper 1984; Vihman et al. 1985). While the inventory of early VMS consonant types was similar between the two language groups, we saw preliminary evidence for differences in the distribution of those consonant types (i.e., the prevalence of labials for Tseltal speakers and velars for Yélî Dnye speakers). This type of cross-linguistic difference in VMS consonant prevalence has been documented in prior work (DePaolis et al. 2011; Majorano et al. 2014; McCune & Vihman 2001), and future work should capitalize on cross-linguistic datasets to more closely examine what drives these variations, if not frequency of those phones in the child's input (DePaolis et al. 2011; Majorano et al. 2014).

4.3 Infrequent CDS and phonological development: canonical babble vs vocal motor schemes

Keeping in mind that (a) our results are preliminary, (b) some of our VMS scores are based on fewer vocalizations than hoped for, and (c) 2 VMS consonants by 10–12 months is a less robustly attested benchmark than the one we use for canonical babble (particularly with respect to daylong recording data) and is not in any way related to clinical language delay, we now tentatively discuss the possibility that some readers may still be considering: that VMS is indeed sensitive to environmental variation and that, therefore, some children are likely to show slower VMS development in these two communities where adult CDS is relatively less frequent. We discuss this possibility because the current findings are limited. Then, an extra contribution this paper can make is to lay out a set of ideas that can be explored in future work; work with results that may differ from what we report here.

Prior research suggests that canonical babble onset is more a meaningful benchmark for motor development than it is for language development (Vihman et al. 2009). Under this view, canonical babble is a hallmark of the development of rhythmic motor tools that develop in the first year of life (Iverson, Hall, Nickel & Wozniak 2007), and develops as a result of the interaction of both proprioceptive and auditory experience with sound production (Westermann & Miranda 2004; Guenther & Vladusic 2012) but doesn't itself require stable articulatory or phonological representations. This perspective—that canonical babble demonstrates a motor skill that helps prepare the child for language, rather than an early step in language development—aligns with findings showing its cross-cultural and cross-linguistic developmental robustness (Oller 1980; Oller et al. 1998; Oller 2000; Lee et al. 2017; Cychosz et al. 2021).

In contrast, VMS gives us insight into early, stable consonant productions that prepare the child to approximate speech in the ambient language (Vihman 1993), and so may be expected to link more tightly to early linguistic representations and, thereby, early productive vocabulary. Over the first year of life, we know that children's ambient language environment comes to shape how they perceive speech (Werker & Tees 1984; Monnin & Lœvenbruck 2010). In fact,

adult CDS (and not just ambient speech in general) has been proposed to facilitate early phone discrimination (Kuhl et al. 2003; Kuhl 2007) and may thereby help children acquire phonological categories sooner. In production, we can also imagine that the elicitation of child-produced speech by caregivers engaging in interactive CDS may result in children getting more frequent practice in attempting adult-like phonological forms (Kuhl & Meltzoff 1996; Kuhl 2007). If we understand VMS as reflecting something about the initial stabilization of phonological categories rather than a simple practiced motor skill—e.g., via the coupling of articulatory parameters and auditory perception (Westermann & Miranda 2004)—we can predict that it is sensitive to the child's exposure to adult CDS and, thereby, linked to early lexical development (Vihman 1993; McCune & Vihman 2001; McGillion et al. 2017), which is also sensitive to adult CDS. We do not find evidence for this idea in the present dataset, suggesting that VMS, like CP, may be fairly robust to environmental variation despite its relationship with early lexical productions. The findings raise multiple questions: at what point does environmentally driven variation in early production milestones begin to emerge?; is it limited to lexical phenomena (see also Cristia 2020)?; and is the scope of variation primarily within and not across populations? Investigations addressing these questions shed new light on the unresolved question of how it is that children, in such variable developmental contexts, come to acquire the linguistic representations and language practices appropriate to their local communities.

Returning to the issue of input quantity, we note that total linguistic input in these communities is not at all sparse. On the basis of past work by Casillas and colleagues (2020; 2021), we can say that there is a great deal of other ambient speech present in these two communities for children to learn from beyond adult CDS, much of it directed to other children within earshot of the target children or to the target children from other children. While direct comparisons of VMS counts between the present work and past work are limited by different measurement and data collection approaches, the patterns of VMS count were also similar between the two communities studied here, which have *parallel* data collection and measurement approaches. Because children in these two communities hear differing quantities of CDS from adults vs. children (Casillas et al. 2020; 2021; Bunce et al. in revision), we take the comparative findings between these two communities as evidence that VMS acquisition is fairly robust to variation in input type (adult CDS, child CDS, and overhearable other speech).

While we can only draw very limited conclusions with the present data, we hope that future work will investigate stable consonant production in more language communities with diverse phonologies and caregiving contexts; together, e.g., in a meta-analysis, these studies could do much to illuminate the role of linguistic input in shaping the development of stable pre-lexical consonant production and later lexical development. For example, if future work consistently finds earlier VMS acquisition in linguistic communities where children experience not just a lot of linguistic input, but a lot of directed linguistic input from adults, it would stand as evidence

that mature speech directed to the target child, and not exposure to speech in general (nor CDS in general), critically shapes phonological development in this transitional period between babble and first words. If no such pattern emerges, it would suggest early robustness to environmental variation in this transitional period, pinpointing lexical development as deeply different from phonological development in its sensitivity to linguistic input, and raising questions about what mechanisms would drive such a difference. Of course, until we gain greater clarity on the relationship between VMS and language environment from further work along the lines suggested above, the current findings should only be taken as preliminary.

4.4 Further research

The current study has several important limitations. While our sample size is an improvement over previous work done in these communities (Casillas et al. 2020; 2021), it is still quite limited considering our cross-sectional design and broad age span (0;5–1;8). Future work should aim to gather larger and more symmetrical samples across languages than we managed here to more comparably explore within-population variability and age-based differences. We were also unable to achieve at least 100 spontaneous vocalizations for every child with our random sampling technique, which means some individual estimates may not be as stable as hoped. The main danger of having too little data per child is underestimating their VMS measures, but we note our present outcomes appeared comparable to past work. Relatedly, our lack of longitudinal data or multiple recordings per child may have yielded less stable individual estimates of canonical babbling and VMS than those presented in prior work (though see our second measure of stable consonant production in the Appendix). Because we had no measure of CDS input rate for each child in the study, we were also only able to compare individual production patterns to community-wide CDS input patterns. Individual input rate estimates would allow us to investigate whether VMS production relates to directed input *within* each community, regardless of the benchmark for English and other languages.

Finally, while phonetically trained, our transcriber did not speak the target languages and so did not benefit from the contextual and lexical information that could have rendered the transcriptions closer to what would be heard by a native annotator. Future work linking VMS to individualized input rate estimates and vocabulary size in these and comparably designed corpora from other languages and cultural communities are needed to clarify the results presented in the current study.

5. Conclusion

We find that Tseltal and Yélî Dnye-learning children show a similar pattern of CP use and VMS acquisition, compared to each other and compared to past work on English-learning children. Some children did not reach the typical milestone. These "pre-milestone" children make up a

similar proportion of the sample as was found in one prior study of English (McCune & Vihman 2011), and a larger portion than was found in another study (DePaolis et al. 2016). In addition to replicating prior work on CP development with a larger and independently annotated sample, we take the current findings as providing no support for the idea that these children, who hear less adult CDS than English-learning children, are delayed in their early linguistic development. Future work should continue to examine stable consonant production in diverse developmental environments and should ideally also more closely examine individual differences within each community, including any links to the lexical development of the same children. Such data will be key to illuminating the ways in which children's linguistic environments influence their transition from pre-lexical to lexical productions cross-linguistically.

Additional file

The additional file for this article can be found as follows:

- **Appendix.** Recording-level VMS and cross-clip consonant count data, reliability analyses, and further exploration of cross-clip consonant counts. DOI: https://doi.org/10.16995/glossa.5813.s1

Competing interests

The authors have no competing interests to declare.

References

Bates, Douglas & Mächler, Martin & Bolker, Ben & Walker, Steve. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1). 1–48. DOI: https://doi.org/10.18637/jss.v067.i01

Brown, Penelope. 1998. Conversational structure and language acquisition: The role of repetition in Tzeltal adult and child speech. *Journal of Linguistic Anthropology* 8(2). 197–221. DOI: https://doi.org/10.1525/jlin.1998.8.2.197

Brown, Penelope. 2011. The cultural organization of attention. In Alessandro Duranti, Elinor Ochs & Bambi B. Schieffelin (eds.), *Handbook of language socialization*, 29–55. Malden, MA: Wiley-Blackwell.

Brown, Penelope. 2014. The interactional context of language learning in Tzeltal. In Arnon, Inbal & Casillas, Marisa & Kurumada, Chigusa & Estigarribia, Bruno (eds.), *Language in interaction: Studies in honor of Eve V. Clark*, 51–82. Amsterdam, NL: John Benjamins. DOI: https://doi.org/10.1075/tilar.12.07bro

Brown, Penelope & Casillas, Marisa. In press. Childrearing through social interaction on Rossel Island, PNG. In Fentiman, Alicia J. & Goody, Mary (eds.), *Esther Goody revisited: Exploring the legacy of an original inter-disciplinarian*. New York, NY: Berghahn.

Bunce, John & Soderstrom, Melanie & Bergelson, Elika & Rosemberg, Celia & Stein, Alejandra & Alam, Florencia & Migdalek, Maia & Casillas, Marisa. In revision. A cross-cultural examination of young children's everyday language experiences. https://psyarxiv.com/723pr/.

Casillas, Marisa & Brown, Penelope & Levinson, Stephen. 2017. *Casillas HomeBank corpus*.

Casillas, Marisa & Brown, Penelope & Levinson, Stephen C. 2020. Early language experience in a Tseltal Mayan village. *Child Development* 91(5). 1819–1835. DOI: https://doi.org/10.1111/cdev.13349

Casillas, Marisa & Brown, Penelope & Levinson, Stephen C. 2021. Early language experience in a Papuan village. *Journal of Child Language* 8(4). 792–814. DOI: https://doi.org/10.1017/S0305000920000549

Chavajay, Pablo & Rogoff, Barbara. 1999. Cultural variation in management of attention by children and their caregivers. *Developmental Psychology* 35(4). 1079–1090. DOI: https://doi.org/10.1037/0012-1649.35.4.1079

Cristia, Alejandrina. 2020. Language input and outcome variation as a test of theory plausibility: The case of early phonological acquisition. *Developmental Review* 57. 100914. DOI: https://doi.org/10.1016/j.dr.2020.100914

Cristia, Alejandrina & Casillas, Marisa. 2022. Non-word repetition in children learning Yélî Dnye. *Language Development Research* 2(1). 69–104.

Cychosz, Margaret & Cristia, Alejandrina & Bergelson, Elika & Casillas, Marisa & Baudet, Gladys & Warlaumont, Anne S. & Scaff, Camila & Yankowitz, Lisa & Seidl, Amanda. 2021. Vocal development in a large-scale cross-linguistic corpus. *Developmental Science* 24(5). 1–22. DOI: https://doi.org/10.1111/desc.13090

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

DePaolis, Rory A. & Keren-Portnoy, Tamar & Vihman, Marilyn M. 2016. Making sense of infant familiarity and novelty responses to words at lexical onset. *Frontiers in Psychology* 7(715). 1–12. DOI: https://doi.org/10.3389/fpsyg.2016.00715

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

DePaolis, Rory A. & Vihman, Marilyn M. & Nakai, S. 2013. The influence of babbling patterns on the processing of speech. *Infant Behavior and Development* 36(4). 642–649. DOI: https://doi.org/10.1016/j.infbeh.2013.06.007

Eberhard, David M. & Simons, Gary F. & Fennig, Charles D. (eds.). 2020. *Ethnologue: Languages of the World*. Twenty-third edition. Dallas, Texas: SIL International. Online version: www.ethnologue.com.ru.idm.oclc.org.

Eilers, Rebecca E. & Oller, D. Kimbrough & Benito-Garcia, Carmen R. 1984. The acquisition of voicing contrasts in Spanish and English learning infants and children: A longitudinal study. *Journal of Child Language* 11(2). 313–336. DOI: https://doi.org/10.1017/S0305000900005791

Fikkert, Paula & Levelt, Claartje. 2008. How does Place fall into Place?. In Avery, Peter, Elan Dresher, B., & Rice, Keren (eds.), *Contrast in phonology: Theory, perception, acquisition*, 231–268. De Gruyter Mouton.

Gaskins, Susan & Paradise, Ruth. 2010. Learning through observation in daily life. In Lancey, David F. & Bock, Susan & Gaskins, Susan (eds.), *The anthropology of learning in childhood*, 85–117. Walnut Creek, CA: Rowman AltaMira Press.

Guenther, Frank H. & Vladusich, Tony. 2012. A neural theory of speech acquisition and production. *Journal of Neurolinguistics* 25(5). 408–422. DOI: https://doi.org/10.1016/j.jneuroling.2009.08.006

Hart, Betty & Risley, Todd R. 1995. *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H. Brookes.

Hoff, Erika. 2003. The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development* 74(5). 1368–1378. DOI: https://doi.org/10.1111/1467-8624.00612

International Phonetic Association. 2020. www.internationalphoneticassociation.org/content/full-ipa-chart.

Ingram, David. 2008. Cross-Linguistic Phonological Acquisition. *The handbook of clinical linguistics*, 626–640. DOI: https://doi.org/10.1002/9781444301007.ch38

Iverson, Jana M. & Hall, Amanda J. & Nickel, Lindsay & Wozniak, Robert H. 2007. The relationship between reduplicated babble onset and laterality biases in infant rhythmic arm movements. *Brain and Language* 101(3). 198–207. DOI: https://doi.org/10.1016/j.bandl.2006.11.004

Jakobson, Roman. 1941/1968. Child language, aphasia and phonological universals. The Hague, The Netherlands: Mouton. DOI: https://doi.org/10.1515/9783111353562

Kidd, Evan & Garcia, Rowena. 2022. How diverse is child language acquisition research? *First Language*, OnlineFirst. DOI: https://doi.org/10.1177/01427237211066405

Kuhl, Patricia K. & Meltzoff, Andrew N. 1996. Infant vocalizations in response to speech: Vocal imitation and developmental change. *The Journal of the Acoustical Society of America* 100(4). 2425–2438. DOI: https://doi.org/10.1121/1.417951

Kuhl, Patricia K. & Tsao, Feng-Ming & Liu, Huei-Mei. 2003. Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences* 100(15). 9096–9101. DOI: https://doi.org/10.1073/pnas.1532872100

Kunnari, Sari. 2003. Consonant inventories: a longitudinal study of Finnish-speaking children. *Journal of Multilingual Communication Disorders* 1(2) 124–131. DOI: https://doi.org/10.1080/14769670310000909944

Laing, Catherine & Bergelson, Elika. 2020. From babble to words: Infants' early productions match words and objects in their environment. *Cognitive Psychology* 122. 1–33. DOI: https://doi.org/10.1016/j.cogpsych.2020.101308

Lee, Chia-Cheng & Jhang, Yuna & Relyea, George & Chen, Li-mei & Oller, D. Kimbrough. 2017. Babbling development as seen in canonical babbling ratios: A naturalistic evaluation of all-day recordings. *Infant Behavior and Development* 50. 140–153. DOI: https://doi.org/10.1016/j.infbeh.2017.12.002

Lee, Chia-Cheng & Jhang, Yuna & Relyea, George & Chen, Li-mei & Oller, D. Kimbrough. 2017. Subtlety of ambient-language effects in babbling: a study of English-and Chinese-learning infants at 8, 10, and 12 months. *Language Learning and Development* 13(1). 100–126. DOI: https://doi.org /10.1080/15475441.2016.1180983

Lee, Sue Ann S. & Davis, Barbara & Peter, MacNeilage. 2010. Universal production patterns and ambient language influences in babbling: A cross-linguistic study of Korean-and English-learning infants. *Journal of Child Language* 37(2). 293–318. DOI: https://doi.org/10.1017/ S0305000909009532

Levinson, Stephen C. 2022. A grammar of Yélî Dnye: The Papuan language of Rossel Island. Berlin, Boston: De Gruyter Mouton. DOI: https://doi.org/10.1515/9783110733853

Majorano, Marinella & Vihman, Marilyn M. & DePaolis, Rory A. 2014. The relationship between infants' production experience and their processing of speech. *Language Learning and Development* 10(2). 179–204. DOI: https://doi.org/10.1080/15475441.2013.829740

McCune, Lorriane & Vihman, Marilyn May. 2001. Early phonetic and lexical development. *Journal of Speech, Language, and Hearing Research* 44. 670–684. DOI: https://doi.org/10.1044/1092-4388(2001/054)

McGillion, Michelle & Herber, Jane S. & Pine, Julian & Vihman, Marilyn M. & DePaolis, Rory A. & Keren-Portnoy, Tamar & Matthews, Danielle. 2017. What paves the way to conventional language? The predictive value of babble, pointing, and socioeconomic status. *Child Development* 88(1) 156–166. DOI: https://doi.org/10.1111/cdev.12671

Monnin, Julia & Lœvenbruck, Hélène. 2010. Language-specific influence on phoneme development: French and Drehu data. *Proceedings of Interspeech 2010*, 1882–1885. Makuhari, Japan, September 26–30, 2010. DOI: https://doi.org/10.21437/Interspeech.2010-543

Moran, Steven & McCloy, Daniel. (eds.) 2019. PHOIBLE 2.0. Jena: Max Planck Institute for the Science of Human History. (Available online at http://phoible.org, Accessed on 2021-08-13.)

Ochs, Elinor & Schieffelin, Bambi B. 1984. Language acquisition and socialization: Three developmental stories and their implications. *Culture theory: Essays on mind, self, and emotion*, 276–322. Cambridge: Cambridge University Press.

Oller, D. Kimbrough. 1980. The emergence of the sounds of speech in infancy. In Yeni-Komshian, Grace & Kavanagh, James F. & Ferguson, Charles A. (eds.), *Child Phonology* 1. 93–112. DOI: https://doi.org/10.1016/B978-0-12-770601-6.50011-5

Oller, D. Kimbrough. 2000. *The emergence of the speech capacity*. Lawrence Erlbaum Associates. DOI: https://doi.org/10.4324/9781410602565

Oller, D. Kimbrough & Eilers, Rebecca E. & Basinger, Devorah & Steffens, Michelle L. & Urbano, Richard. 1995. Extreme poverty and the development of precursors to the speech capacity. *First Language* 15(44). 167–187. DOI: https://doi.org/10.1177/014272379501504403

Oller, D. Kimbrough & Levine, Sharyse & Cobo-Lewis, Alan B. & Eilers, Rebecca E. & Pearson, Barbara Z. 1998. Vocal precursors to linguistic communication: How babbling is connected to meaningful speech. In Rhea Paul (ed.), *Exploring the speech/language connection*, Vol 8, 1–25. Baltimore, MD: Paul H. Brookes Publishing.

Oller, D. Kimbrough & Wieman, Leslie A. & Doyle, William J. & Ross, Carol. 1976. Infant babbling and speech. *Journal of Child Language* 3(1). 1–11. DOI: https://doi.org/10.1017/S0305000900001276

Polian, Gilles. 2013. *Gramática del tseltal de Oxchuc*. Centro de Investigaciones y Estudios Superiores en Antropología Social, México.

Prather, Elizabeth M. & Hedrick, Dona Lee & Kern, Carolyn A. 1975. Articulation development in children aged two to four years. *Journal of Speech and Hearing Disorders* 40(2). 179–191. DOI: https://doi.org/10.1044/jshd.4002.179

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/.

Rogoff, Barbara & Mistry, Jayanthi & Göncü, Artin & Mosier, Christine. 1993. Guided participation in cultural activity by toddlers and caregivers. *Monographs of the Society for Research in Child Development* 8(8). i–179. DOI: https://doi.org/10.2307/1166109

Shaw, Patricia A. 1991. Consonant Harmony Systems: The Special Status of Coronal Harmony. In Paradis, Carole, & Jean-François Prunet (eds.), *Phonetics and Phonology* 2. 125–157. San Diego: Academic Press. DOI: https://doi.org/10.1016/B978-0-12-544966-3.50013-0

Shneidman, Laura A. & Goldin-Meadow, Susan. 2012. Language input and acquisition in a Mayan village: How important is directed speech? *Developmental Science* 15(5). 659–673. DOI: https://doi.org/10.1111/j.1467-7687.2012.01168.x

Stoel-Gammon, Carol & Cooper, Judith A. 1984. Patterns of early lexical and phonological development. *Journal of Child Language* 11(2). 247–271. DOI: https://doi.org/10.1017/S0305000900005766

Tsuji, Sho & Mazuka, Reiko & Cristia, Alejandrina & Fikkert, Paula. 2015. Even at 4 months, a labial is a good enough coronal, but not vice versa. *Cognition* 134. 252–256. DOI: https://doi.org/10.1016/j.cognition.2014.10.009

VanDam, Mark & Warlaumont, Anne S. & Bergelson, Elika & Cristia, Alejandrina & Soderstrom, Melanie & De Palma, Paul & MacWhinney, Brian. 2016. HomeBank: An online repository of daylong child-centered audio recordings. *Seminars in Speech and Language* 37(2). 128–142. Thieme Medical Publishers. DOI: https://doi.org/10.1055/s-0036-1580745

Vihman, Marilyn M. 1993. Vocal motor schemes, variation and the production-perception link. *Journal of Phonetics* 21. 163–169. DOI: https://doi.org/10.1016/S0095-4470(19)31315-4

Vihman, Marilyn M. & de Boysson-Bardies, Bénédicte. 1994. The nature and origins of ambient language influence on infant vocal production and early words. *Phonetica* 51(1–3). 159–169. DOI: https://doi.org/10.1159/000261967

Vihman, Marilyn M. & DePaolis, Rory A. & Keren-Portnoy, Tamar. 2009. A dynamic systems approach to babbling and words. *British Journal of Psychology* 108(1). 1–27.

Vihman, Marilyn M. & Kay, Edwin & de Boysson-Bardies, Bénédicte & Durand, Catherine & Sundberg, Ulla. 1994. External sources of individual differences? A cross-linguistic analysis of the phonetics of mothers' speech to 1-yr-old children. *Developmental Psychology* 30(5) 651–662. DOI: https://doi.org/10.1037/0012-1649.30.5.651

Vihman, Marilyn M. & Macken, Marlys A. & Miller, Ruth & Simmons, Hazel & Miller, Jim. 1985. From babbling to speech: A re-assessment of the continuity issue. *Language* 61(2). 397–445. DOI: https://doi.org/10.2307/414151

Weisleder, Adriana & Fernald, Anne. 2013. Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science* 24(11). 2143–2152. DOI: https://doi.org/10.1177/0956797613488145

Werker, Janet F. & Tees, Richard C. 1984. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7(11). 49–63. DOI: https://doi.org/10.1016/S0163-6383(84)80022-3

Westermann, Gert & Miranda, Eduardo R. 2004. A new model of sensorimotor coupling in the development of speech. *Brain and Language* 89(2). 393–400. DOI: https://doi.org/10.1016/S0093-934X(03)00345-6

Wickham, Hadley. 2009. *Ggplot2: Elegant graphics for data analysis*. New York, NY: Springer-Verlag. Retrieved from http://ggplot2.org. DOI: https://doi.org/10.1007/978-0-387-98141-3