Received: 25 June 2020 | Revised: 9 December 2020 | Accepted: 16 January 2021

DOI: 10.1111/desc.13090

Developmental Science WILEY

PAPER

Vocal development in a large-scale crosslinguistic corpus

Margaret Cychosz[1] | Alejandrina Cristia[2] | Elika Bergelson[3] | Marisa Casillas[4] | Gladys Baudet[3] | Anne S. Warlaumont[5] | Camila Scaff[2,6] | Lisa Yankowitz[7] | Amanda Seidl[8]

[1]Department of Hearing and Speech Sciences & Center for Comparative and Evolutionary Biology of Hearing, University of Maryland, College Park, MD, USA

[2]Laboratoire de Sciences Cognitives et de Psycholinguistique, Département d'études cognitives, ENS, EHESS, CNRS, PSL University, Paris, France

[3]Department of Psychology & Neuroscience, Center for Cognitive Neuroscience, Duke University, Durham, NC, USA

[4]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

[5]Department of Communication, University of California, Los Angeles, Los Angeles, CA, USA

[6]Human Ecology Group, Institute of Evolutionary Medicine, University of Zurich, Zurich, Switzerland

[7]Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA

[8]Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, IN, USA

Correspondence
Margaret Cychosz, University of Maryland, 0100 Samuel J.LeFrak Hall, College Park, MD 20742, USA.
Email: mcychosz@umd.edu.

Amanda Seidl, Purdue University, Lyles-Porter Hall, Room 3142, West Lafayette, IN 47907, USA.
Email: aseidl@purdue.edu

Funding information
This work was supported by two Oswalt Documenting Endangered Languages grants and the Raymond H. Stetson Scholarship in Phonetics and Speech Science to MCy; Agence Nationale de la Recherche (NR-17-CE28-0007 LangAge, ANR-16-DATA-0004 ACLEW, ANR-14-CE30-0003 MechELex, ANR-17-EURE-0017), the James S. McDonnell Foundation Understanding Human Cognition Scholar Award, a Trans-Atlantic Platform "Digging into Data" collaboration grant (ACLEW: Analyzing Child Language Experiences Around The World), with the support of Agence Nationale de la Recherche (ANR-16-DATA-0004) to AC; a Netherlands Organization for Scientific Research Veni Innovational Research Scheme Grant (275-89-033) to MCa; the National Endowment for the Humanities (HJ-253479-17) and NIH Grant DP5-OD019812 to EB; National Science Foundation grants BCS-1529127 and SMA-1539129/1827744 and a James S. McDonnell Foundation Scholar Award to ASW; and by the University of Zurich to CS. The authors do not have any conflicts of interest to report.

Abstract

This study evaluates whether early vocalizations develop in similar ways in children across diverse cultural contexts. We analyze data from daylong audio recordings of 49 children (1–36 months) from five different language/cultural backgrounds. Citizen scientists annotated these recordings to determine if child vocalizations contained canonical transitions or not (e.g., "ba" vs. "ee"). Results revealed that the proportion of clips reported to contain canonical transitions increased with age. Furthermore, this proportion exceeded 0.15 by around 7 months, replicating and extending previous findings on canonical vocalization development but using data from the natural environments of a culturally and linguistically diverse sample. This work explores how crowdsourcing can be used to annotate corpora, helping establish developmental milestones relevant to multiple languages and cultures. Lower inter-annotator reliability on the crowdsourcing platform, relative to more traditional in-lab expert annotators, means that a larger number of unique annotators and/or annotations are required, and that crowdsourcing may not be a suitable method for more fine-grained annotation decisions. Audio clips used for this project are compiled into a large-scale infant vocalization corpus that is available for other researchers to use in future work.

KEYWORDS
babbling, crosslinguistic, crowdsourcing, infants, naturalistic recording, speech, vocal development

Research Highlights

- Using naturalistic audio recordings of infants' daily environments, we measured vocal development in five culturally diverse settings.

- The ratio of clips containing canonical transitions ("ba") increased as the children aged, irrespective of cultural setting.
- Canonical transitions were found in most infants' speech by 7 months, and most infants displayed a canonical proportion at or above 0.15 by 10 months.
- The collaboration of citizen scientists permitted the annotation of over 60,000 audio clips, which are now available in a publicly shared corpus of infant vocalizations.

1 | INTRODUCTION

1.1 | The emergence of canonical babble: An important stage in vocal development

Although infants begin vocalizing from birth, their vocalizations change markedly over the first year of life. Children's early vocal production is thought to follow a universal sequence of development, with the proportion of speech-like vocalizations increasing with age (Oller, 2000). A critical milestone in this developmental sequence is the use of adult-like consonant-vowel (CV) transitions ("canonical" syllables; Oller et al., 1998). Specifically, while very young infants readily produce vowels (e.g., "ooo"), squeals (e.g., a high-pitched "eee"), some articulatorily less-demanding, isolated sonorants (e.g., "mmm"), and various other sounds, they do not begin to produce neatly-timed CV or vowel-consonant syllables until the latter of half of the first year (Oller, 1980).

Several studies report that vocal development before 9 months of age, including the emergence of canonical syllables, is language-general and consistent across languages (de Boysson-Bardies et al., 1984; Vihman et al., 2006; Whalen et al., 2007). As a child ages, these works argue that vocalizations become progressively more language-specific and attuned to the unique sounds of the ambient language. For example, at 10 months, French-learning infants may produce more nasal segments than English learners, and French infants' stop consonants have different voice onset times from English infants', both of which are attributable to the structure of French and English (Blake & de Boysson-Bardies, 1992; de Boysson-Bardies et al., 1984).

Given its adult-like CV structure, vocalizations with canonical syllables are considered to be a starting point on the path to recognizable speech. Specifically, after infants begin to produce syllable sequences featuring one unique consonant (e.g., "baba" or "dada"), they begin to produce different consonants mixed together (such as "bada"; Oller, 1980). The former is called canonical babble, and the latter variegated babble. Variegated babble is similar to combinations that occur in many words (e.g., "bunny") and commonly occurs around the same time children begin to produce words, typically close to their first birthdays (de Boysson-Bardies & Vihman, 1991). First words are often indistinguishable from sequences of canonical babble (e.g., "mama", "dada"). Thus, there appears to be a smooth developmental transition between canonical babble, variegated babbling, and lexical speech (de Boysson-Bardies & Vihman, 1991), which implies a strong relationship between early non-lexical production and later lexical production.[1]

The development of canonical babble is typically assessed in two ways. One approach is to note the age when canonical babbles first appear (canonical babbling onset; CBO). CBO can be identified by looking for reduplicated CV syllables, for example, "bababa", in infants' vocalizations (Fagan, 2009; Holmgren et al., 1986; Schauwers et al., 2004; van der Stelt & van Beinum, 1986). Alternatively, CBO can be determined by asking caregivers to provide a yes/no response (i.e., "is your child producing adult speech-like syllables?"; Eilers et al., 1993; Oller et al., 1998). When such questions are asked frequently over the course of an infant's development, they can reveal the age of CBO.

A second approach is to measure the ratio of canonical to other vocalizations, including non-canonical vocalizations such as standalone vowels. This is the canonical babbling ratio (CBR). Notably, the exact calculation for CBR varies across the literature (Eilers & Oller, 1994; Oller & Eilers, 1988). Generally speaking, CBR quantifies the relative use of CV vocalizations that are "canonical" (defined as adult-like transitions between consonants and vowels) to those that are not. A traditional approach is to count the number of canonical syllables and divide that by the total number of syllables produced by the infant (e.g., Lee et al., 2018). The metric employed in this paper, canonical proportion, (operationalized further below) is thus somewhat conceptually related to this CBR, but canonical proportion is not necessarily calculated on the basis of syllables, and vocalizations that are meaningful are not excluded.

Canonical babbling onset may be more difficult to determine than CBR because it requires repeated questionnaires or recordings, whereas a cross-sectional recording generally suffices for estimating CBR. Previous work suggests that CBO in typically developing children tends to occur around age 7 months (McGillion et al., 2017; Oller et al., 1997), while a CBR of 0.15 is typically expected by 10 months, (meaning that at 10 months 15% of the child's syllables are canonical; Oller & Eilers, 1988). For English- and Spanish-learning North American infants, CBR increases more or less linearly from 3 to 20 months of age (Oller et al., 1997; Warlaumont & Ramsdell-Hudock, 2016). While there is a rich literature on canonical babble development, the frequency with which canonical transitions are employed throughout the first years of childhood has received less attention. As children start using more diverse consonants and saying meaningful words in the second year, the focus of research shifts to these topics. We therefore have little information about how prevalent canonical transitions are in the second and third years of life, including whether the frequency of canonical transitions plateaus, or whether it continues to increase through middle childhood.

Finally, both CBO and CBR have been shown to predict language outcomes in typically-developing infants (Lang et al., 2019; McDaniel et al., 2019; McGillion et al., 2017; Oller et al., 1998, 1999). A delayed CBO or reduced CBR has been found in children who go on to develop speech/language delays and autism spectrum disorders (Fasolo et al., 2008; Lang et al., 2019; Patten et al., 2014; Stoel-Gammon, 1989) and children who have genetic disorders linked to language disorders (e.g., Fragile X syndrome; Belardi et al., 2017). In addition, Oller et al. (1999) find that children who failed to produce an age-appropriate CBR of 0.15 by 10 months of age had smaller vocabularies later in development.

1.2 | Cross-cultural comparisons

Recent work has found complex relationships between culture, social context, and infant age on vocal development, including canonical babble. For example, Lee et al. (2018) studied canonical babble development in 6- and 11-month-old English- and Mandarin-learning infants in the United States and Taiwan, respectively.[2] Each family completed a daylong recording which captured the infants' naturalistic interactions. Although some trends were similar across the two groups of infants (e.g., that CBR increased with age), others were not (the size of the increase, and its stability across situations). Those authors concluded that additional cross-cultural work on child vocal development is needed.

Further evidence of the effect of acculturation on vocal development comes from studies of infant-caregiver interactions (e.g., Albert et al., 2018; Bornstein et al., 1992; Goldstein & Schwade, 2008; Gratier & Devouche, 2011; Gros-Louis et al., 2006; Ramirez et al., 2019; Warlaumont et al., 2014). In Goldstein and Schwade (2008), caregivers of 9.5-month-olds were asked to produce speech in two conditions: contingent on their child's vocalization and non-contingent on the vocalization. The authors then measured infants' vocal responses in the two conditions. The infants in the contingent condition restructured their syllable shapes to match the caregivers' productions, for example increasing the proportion of CV syllables. However, this change was not observed for the infants in the non-contingent condition, perhaps, the authors suggest, because only the interactive nature of contingent response allowed the infants to focus on the caregiver and mimic the statistical regularities of caregiver speech (also see Laing & Bergelson, 2020; McGillion et al., 2017; Warlaumont et al., 2014). Other relevant work in this realm has found that infants' vocalizations can affect their caregivers' speech (Albert et al., 2018; Pretzer et al., 2019), though the frequency of these interactions are contingent upon culture (Bornstein et al., 1992) and recording environment (naturalistic or lab-based, Gros-Louis et al., 2006). Together these results suggest a "vocal feedback loop" where early speech-like vocalizations encourage caregiver responses, which, in turn, facilitate speech-like infant vocalizations over the first year or two of life.

If there is a critical feedback loop between infants and their caregivers, this could be expected to vary crosslinguistically and/or cross-culturally because there is great cultural diversity in the amount of speech directed to infants and young children (see especially figure 4 from Casillas et al., 2019; Cristia, 2020; Cristia et al., 2019; Klein et al., 1977; Konner, 1977; Lieven, 1994). Convergently, the datasets used in the current work include children who differ widely in the amount of child-directed speech that they hear (Bergelson, Casillas, et al., 2019; Casillas et al., 2019, in press; Cristia et al., 2019). Furthermore, verbal exchange is just one component of social feedback that could vary cross-culturally (de León, 1998). The ways that children are encouraged to engage in social interaction, and what they are led to expect as appropriate social action, may also differ. Caregiver responsivity, attentional patterns (e.g., joint attention), and tactile cues also vary across cultures (Gaskins, 2006). For example, there is ample evidence that touch is highly frequent in mother-infant exchanges (Stack & Muir, 1990) but that mothers' use of infant-directed touch varies with culture (Carra et al., 2014). This observation may be relevant because touch in mother-infant exchanges impacts social and biological development broadly (Field, 2010) and may even aid in language learning when combined with speech (Seidl et al., 2015). Thus, cultural effects on vocal development could have multiple sources (e.g., tactile practices, quantity of verbal input).

Taken together, previous work suggests a potential cultural influence on typical vocal development. And while some previous studies have not found substantial effects of culture, language, or socioeconomic status on CBO (Gros-Louis & Miller, 2018; Lee et al., 2018), that work has not studied vocal development across a wide range of cultures, but instead has focused almost entirely on highly industrialized populations.

This gap in the literature is notable given the influence of culture on other areas of infants' speech and motor development that were, historically, not apparent to researchers. For example, early work on gross motor movements, like crawling, suggested uniformity in the onset of motor milestones across cultures. But more recent work finds clear cultural differences (as summarized in Adolph et al., 2009). These differences are likely driven by different cultures' caregiving practices, some of which encourage more independent motor behaviors (e.g., through infant massage or manipulated movement) while others discourage them (e.g., through restricting early child movement). Such cultural practices drive Ugandan infants to tend to crawl at 5.5 months (Super, 1976), while Tajik infants, whose movement is generally more restricted, may not crawl until 1;0 (Karasik et al., 2018). Like early movement milestones, babbling and some types of early vocalizations have been argued to be other kinds of stereotypic motor behavior as they involve rhythmic jaw oscillations (MacNeilage & Davis, 1993). Since culture has been shown to impact gross motor milestones, it is likewise possible that it affects the development of early vocalizations, and more broadly the emergence and frequency of canonical transitions.

1.3 | Gender comparisons

Few studies have examined the role of gender on vocal development (cf. Oller et al., 2020). Yet there is a large literature

documenting gender differences in language outcomes and language disorders (Barbu et al., 2015; Eriksson et al., 2012; Frank et al., 2017; Hadley et al., 2011; Huttenlocher et al., 1991; Whitehouse, 2010). Males are more likely to manifest with a language disorder than females (Whitehouse, 2010). Many studies find that girls outpace same-aged boys in passing linguistic milestones such as lexical and morphosyntactic growth (Barbu et al., 2015; Eriksson et al., 2012; Frank et al., 2017; Hadley et al., 2011; Huttenlocher et al., 1991). These differences may result from early effects of sex hormones on articulatory skills (Quast et al., 2016) or sex-specific development of brain regions associated with language (Etchell et al., 2018). Another possibility is that these gender differences in language outcomes are the result of early socialization, for example if the quantity or quality of caregiver responses varied systematically by gender (Johnson et al., 2014; Sung et al., 2013; Warlaumont et al., 2014).

Given differences between boys' and girls' early lexical production (Frank et al., 2017), meaningful differences by gender may also appear in early vocal development, including in the emergence of canonical CV transitions. Nonetheless, such gender-related differences in vocal development are rarely discussed. Just two previous studies have evaluated this question for infant vocalizations, concluding that there were no notable differences between boys' and girls' early vocalizations (Sung et al., 2013) or vocal maturity (Oller et al., 2020). However, there were differences in the number of vocalizations produced, with boys vocalizing more than girls between 0 and 13 months, and between 4.5 and 6.5 months in particular (Oller et al., 2020).[3] Nevertheless, conclusions from these studies remain limited in scope given that the samples were fairly homogenous. It is thus premature for the field to conclude an absence of gender-related differences in infant vocalization development. More work is needed to explore possible gender effects on early vocalization development generally, and with respect to canonical transitions in particular. The current study helps address this gap.

2 | CURRENT STUDY

The literature suggests that canonical transitions emerge at about 7 months, and that CBRs at or above 0.15 are apparent by 10 months. Failure to achieve these milestones has been related to poorer language outcomes. However, the past literature has relied on data gathered almost exclusively from children in child-centered cultures in industrialized nations, often with limited sample sizes and short recordings made in the lab or other semi-naturalistic settings.

Furthermore, the potential relationship between gender and vocal development is under-explored. Moreover there is little work attempting to study the prevalence of canonical transitions in the second and third years of life. Taken together, these factors limit broader conclusions concerning the trajectory of vocal development. Given that vocal development is claimed to follow a universal timeline, it is important to verify these previous findings in a larger, naturally gathered, crosslinguistic, and culturally diverse sample.

2.1 | Motivation

One notable limitation of previous work on the emergence of canonical babble and transitions has been the geographic and cultural homogeneity of the research participants. Though previous work has incorporated some diversity in language (e.g., French, Swedish, Cantonese, Arabic; de Boysson-Bardies et al., 1984; Roug et al., 1989) and socio-economic status (Eilers et al., 1993; McGillion et al., 2017), the samples remain relatively small and lacking in cultural diversity. This lack of diversity could be problematic because, for example, over-sampling from infant families from university towns may result in a sample biased towards higher socio-economic classes. Furthermore, caregivers inclined to participate in scientific studies may be more prone to child-centric or pedagogical caregiving characteristics (see Rogoff, 2003: 141–146). These factors could lead to biased samples (Nielsen et al., 2017) that are not representative of much of the world. Unrepresentative populations such as these can lead to false conclusions about what is developmentally typical for human development at large.

Previous work on vocal development is also somewhat limited by the short duration and limited contexts of the recording samples. For instance, even in one of the most intensive longitudinal data collection schedules, which sampled infants weekly for a 7-month period, the data collection was limited to a 30-min parent-child interaction and 10–15 min free play session (Vihman et al., 1985). Although this longitudinal data collection schedule is laudable, recent technological advances permit longer duration recordings that capture the entirety of the infants' daily experiences. Other studies, such as Eilers et al. (1993), also relied on relatively short recordings (20–30 min), but the recordings were gathered in a soundproof room in a laboratory. During these recordings, investigators actively attempted to elicit vocalizations from the child. Current recording technologies and data storage systems allow researchers to collect longer recordings and speech samples that closely represent infants' spontaneous behavior and interaction.

Measuring infant vocalizations in language samples that are (1) culturally and socio-economically diverse and (2) representative of infants' naturalistic environments is crucial to understanding vocal development. The presence of cultural effects in other motor development areas underscores the need to analyze speech development in diverse socio-cultural settings to gain information either supporting or refuting previous studies suggesting a relative universality in vocal development. Furthermore, given that variation in early vocalizations predicts later language outcomes (McCathren et al., 1999; Oller et al., 1999; Ramírez et al., 2019; Ramírez-Esparza et al., 2014), it is essential that we understand which exogenous factors impact infants' early speech patterns.

Previous work on early vocal development in typically and non-typically developing populations has included children up to 36 months (de Boysson-Bardies & Vihman, 1991; Fagan, 2015; Jung & Houston, 2020; McDaniel et al., 2020; Patten et al., 2014). In the current work, the decision to include children as old as 36 months was made for several reasons. First, previous work on cultural effects

on infant vocalization has argued that these effects are unlikely to apply uniformly throughout the first years of life (Lee et al., 2018). Consequently, to discern the potential role of culture and/or language upon infant vocalization patterns, a wide range of ages must be considered. This is particularly true given that the languages and cultures examined here differ widely from those studied in previous work. Another important reason to include a wide age range in this study is to contribute to comparative studies of typically-developing and non-typically developing children. For example, (canonical) babble is late to emerge in children with ASD (Patten et al., 2014) and Fragile X Syndrome (Belardi et al., 2017), so it is frequently studied in non-typically developing populations and their age-matched typically developing peers well into the third year of life. The current work presents crosslinguistic data from typically developing children that can be used to compare to these populations, who may receive a diagnosis only at age two or three years. Finally, previous studies on CBR have not made it clear when CBR is expected to plateau, nor whether this would happen at similar ages for different languages and populations. For these reasons, we included children up to 36 months in the current study.

This work takes an important step in studying vocal development across highly diverse cultural and linguistic contexts, focusing on a representative sub-sample of children's spontaneous vocalizations produced in their home environments. For this work, we define a vocalization as all speech-like vocalizations, including isolated vowels, consonants, or CV transitions, well-formed or not, and excluding crying and laughing. While children's vocalizations are increasingly meaningful and lexical past the age of 12–24 months, we focus here on the speech properties of the utterances rather than their potentially meaningful content. We examined possible effects of linguistic context and infant gender on vocal development by collecting vocalizations produced during daylong (6–16 h) audio recordings that were made in children's homes in six culturally and linguistically diverse child-rearing contexts around the world (see Methods). Daylong recording technology permits naturalistic observation of these infants using much more uniform data collection protocols across variable economic and cultural contexts given that these recordings are collected at field-sites, freeing researchers to include participants outside the more typical recruitment zone close to a research lab by a university.

In the current study we ask two questions:

1. In a large culturally and linguistically diverse sample, does the proportion of canonical transition vocalizations to all vocalizations—the canonical proportion—grow as children age, as reported in CBR findings sampling a narrower range of linguistic and cultural contexts? More specifically, do children reach a 0.15 ratio of canonical to non-canonical observations by 10 months, independent of culture and language of exposure?
2. Previous work suggests that female children reach linguistic milestones earlier than males once they begin to produce lexical vocalizations. In this diverse sample, does the canonical proportion vary by child gender?

Regarding the first question, based on past and ongoing work, we anticipated that the diverse cultural settings experienced by the children in each of the six linguistic settings could affect vocal development. The goal of the current study is not to distinguish between different sources of cultural differences (e.g., caregiving practices, quantity of child-directed speech) but rather to determine if cross-cultural differences in vocal development actually exist.

The precise 0.15 threshold for canonical vocalizations was drawn from work using CBR (Belardi et al., 2017; Lee et al., 2018; Patten et al., 2014), though there are important differences between CBR and the canonical proportion employed in this paper. CBR has been used among pre-linguistic infants, thus de facto excluding meaningful speech, and is derived from syllables as a measure of the presence of canonical and non-canonical babbles in a child's repertoire. In contrast, the canonical proportion used here includes all of the children's speech-like vocalizations, which may or may not overlap with individual syllables. This characteristic of canonical proportion is an essential component of the crowdsourcing methodological design: the clips of children's vocalizations were divided into smaller clips (around 400 ms) that did not necessarily correspond to syllable shapes in order to protect participant privacy on the crowdsourcing platform. Furthermore, some of the children's vocalizations in this study may be linguistically meaningful since we thought it important to test for potential cultural and gender effects in children up to 36 months of age. In all, canonical proportion is comparable to CBR but there are notable differences between the two outcome measures.

Regarding the second question, we predicted girls might reach a canonical proportion threshold of 0.15 prior to boys based on their more advanced lexical productions in prior research. However, if gender differences in language development outcomes instead relate to other aspects of language acquisition, such as the contents of the lexicon, the canonical proportion among girls and boys might not differ.

3 | METHODS

3.1 | Corpora

The dataset used for this study consists of infant vocalizations drawn from subsets of six daylong audio recording corpora (Bergelson, 2017; Casillas et al., 2017; Cychosz, 2018; Scaff et al., 2018; Warlaumont et al., 2016), some of which are housed in HomeBank (VanDam et al., 2016) and Databrary (Databrary, 2012). See Table 1 for details. Across the corpora, 52 typically developing children, aged 0;1–3;0 (M = 1;4, SD = 0;9, 24 female, 28 male) were considered for the present study. For all of these corpora, the child participants wore lightweight recorders throughout a large portion of their day at home. Each child contributed one daylong audio recording to the dataset.

The children were exposed to a range of languages: American English, multiple varieties of Spanish, Tsimane′, Tseltal, Yélî Dnye, and Quechua. All families whose data are included here consented

TABLE 1 Summary of demographic information from each corpus

Corpus + location	Language(s)	*N*	Age (months)	Gender	Maternal education (years)	Avg. dur. (range)
Bergelson, New York, USA	English	10	7–17	4 M 6F	12–22	13.4 h (11.1–16)
Casillas, Chiapas, Mexico	Tseltal	10	2–36	5 M 5F	0–12	9.2 h (8.2–9.6)
Casillas, Milne Bay, Papua New Guinea	Yélî Dnye	10	1–36	5 M 5F	06–14	8.1 h (7.2–9.2)
Cychosz, Chuquisaca, Bolivia	Spanish + Quechua	3	22–25	3 M 0F	06–12	10.4 h (5.4–14.3)
Scaff, Beni, Bolivia	Tsimane′	16	8–32	10 M 6F	0–09	15.6 h (10.9–16)
Warlaumont, California, USA	English + Spanish	3	3	1 M 2F	10–16	12.5 h (10–16)

to data collection and semi-public sharing of the recordings as described below. The subsequent analyses were additionally approved by each author's respective institutional ethics review board. To the best of our knowledge, all children were full term with normal speech and hearing development, per parental report. The Tsimane′, Tseltal, Yélî Dnye, and Quechua speech communities are medically underserved and developmental delays may thus be under-reported.

The English-Bergelson corpus contains longitudinal observations of infants exposed primarily to American English. The data were collected in and around Rochester, New York where families were followed for a year of monthly observations beginning when infants were 6 months. This data collection included daylong audio and hour-long video recordings of the infants' daily environments, as well as in-lab experiments and parent questionnaires to evaluate lexical development (see further detail in Bergelson, Amatuni, et al., 2019).

The English-Spanish-Warlaumont corpus contains samples of primarily monolingual English-learning and bilingual English- and Spanish-learning infants from the Central Valley, California. They were recruited via word-of-mouth, flyers on the UC Merced campus and in the surrounding community, and through recruitment events, including at the local hospital. The broader corpus and study included longitudinal recordings, but for the present work, only a subset of the earliest recordings, made when the infants were approximately 3 months old, are included (Vallomparambath PanikkasserySu, 2020; Warlaumont et al., 2016).

The Tseltal Mayan corpus was made in 2015 in a rural subsistence farming community in the Chiapas highlands in southern Mexico. The vast majority of children in this community, including all of the children whose data are analyzed here, grow up speaking Tseltal monolingualllly at home. Children typically begin to learn Spanish later, in primary school (Brown, 1998), though lexical borrowings and expressions in Spanish are common in everyday Tseltal conversation. All children between ages 0;0 and 4;0 in the region around the main participating village were invited to participate via word of mouth and with the help of a Tseltal community member; participants completed a daylong recording and then, several days later, participated in a short battery of experiments evaluating their implicit language knowledge (Casillas et al., 2017, 2019).

The Yélî Dnye recordings were made in 2016 in a rural subsistence farming community, located on a remote island in Milne Bay Province, Papua New Guinea. Approximately 80% of the households with young children in the sampled region use Yélî Dnye monolingually at home, with the roughly 20% of multilingually-raised children typically also hearing English and sometimes a third, usually Papuan, language (overall: approximately 14% bilingual and 6% trilingual in this region of the island); otherwise children only begin to learn English formally when they enter primary school (Brown & Casillas, in press). That said, again, lexical borrowings and expressions in English are common in everyday Yélî Dnye conversation. The same recruitment strategy was used as in the Tseltal context described above (Casillas et al., 2017). In both communities, speech directed to young children occurs infrequently throughout the waking day (3.6 and 3.13 min/h respectively, for Tseltal and Yélî children under 3;0), though ethnographic analyses have revealed meaningful cross-site differences in early caregiver-child responsiveness patterns (Brown, 2011; Casillas et al., 2019, in press).

The Tsimane′ corpus includes audio recordings of children from two different Tsimane′ villages in the lowlands of northern Bolivia. The Tsimane′ are an indigenous group residing in the forest, riverine, and savanna areas in the Beni province (Gurven et al., 2017). While they are experiencing a fast market integration into broader Bolivian society, most Tsimane′ are monolingual in the Tsimane′ language. Speech directed to children appears to be relatively rare in Tsimane′ villages, with children receiving <1 min of speech directed to them per hour (Cristia et al., 2019). However, more recent estimates suggest that this amount is higher, between 3 to 7 min/h, depending on how input is calculated.

The Quechua corpus contains cross-sectional samples of bilingual children acquiring Quechua and Spanish in the south Bolivian highlands. Children in these speech communities are typically exposed to Spanish and Quechua from birth. Most will eventually speak Quechua in the home and Spanish at school and with same-aged peers; however, the languages have been in heavy contact for centuries so there is frequent mixing and lexical borrowing (Muysken, 2012). The degree of children's exposure to the two languages varies and depends on maternal education and the presence of monolingual speakers in the children's environments (Cychosz, 2020). A quantitative estimate of the quantity of child-directed speech in these communities is ongoing, but early results suggest that child-directed speech is infrequent for the first year of life, though it increases as children age.

3.2 | Publicly available vocalization corpus

To address questions about vocal development in a large, cross-cultural sample, we first created a large crosslinguistic corpus that contains annotated clips of early vocalizations.

This corpus is now publicly available for reuse and further analyses (https://osf.io/rz4tx/). The corpus can also provide training data to support methodological and computational advances to address current barriers to large-scale vocalization analysis (segmentation and annotation); this is critical because there is very little openly available tagged data on early phonological acquisition. One exception is PhonBank (https://phonbank.talkbank.org/), which has large amounts of crosslinguistic data. However, PhonBank is not ideal for assessing vocal development across diverse settings since there are few data from children under 1;0 and the data originate exclusively from industrialized cultures.

3.3 | Procedures

For four of the corpora, namely English-Bergelson, Tsimane′, Quechua, and English-Spanish-Warlaumont, the audio recordings of the children were made with the Language ENvironment Analysis (LENA) Digital Language Processor (Xu et al., 2014). LENA is a lightweight, wearable (<60 g, 5.5 × 8.5 × 1.5 cm) recording device made popular in part by its accompanying software for processing audio to extract some automated measures of children's language environments, such as the estimated number of words heard throughout the day (Xu et al., 2014). For a detailed overview of LENA's system, see Ganek and Eriks-Brophy (2018). In the Tseltal and Yélî Dnye corpora (Casillas et al., 2017, 2019, in press), recordings were instead made with a small, wearable Olympus audio recorder (WS-832, 50 g, 4 × 10 × 1.5 cm or WS-835, 80 g, 4 × 11 × 2 cm, with batteries included). Across all six corpora, children wore the recording device across their chest inside a specially-designed clothing pocket (Figure 1). Average recording lengths and ranges by corpus are listed in Table 1. An overview of these recording procedures, including data collection and pre-processing, is shown in Figure 1.

3.4 | Data pre-processing

Before annotating children's vocalizations for the prevalence of canonical transitions, we had to first (1) identify when vocalizations occurred during these multi-hour recordings and (2) extract a representative sample of the vocalizations for further annotation and analysis. Because there were two recording set-ups across our six corpora (i.e., LENA and Olympus), we identified child vocalizations in two different ways.

Recordings made with the LENA device were processed using the proprietary LENA algorithm which assigns short audio segments to one of 15 speaker categories in the child's environment (e.g.,

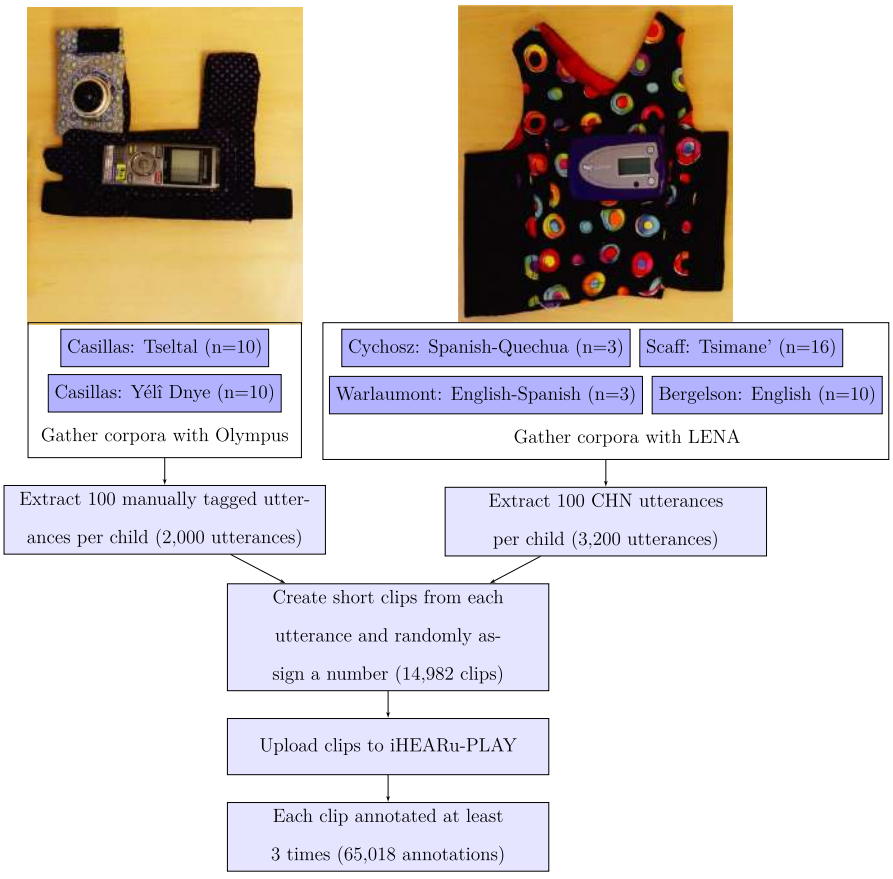

FIGURE 1 Overview of the methods showing recording devices used and stages of processing. LENA, Language ENvironment Analysis

Female-Adult-Near, Male-Adult-Near) or to the target child (the one wearing the recorder). For the rest of this paper, these audio segments of complete child vocalizations from the recordings will be referred to as "utterances." Importantly for this project, the LENA-derived output file indicates each instance throughout the day in which a child utterance was detected.

The LENA algorithm was trained on data from children learning American English. Crosslinguistic and cross-cultural validation of LENA's labels and automated counts is a focus of recent and ongoing research (e.g., Canault et al., 2016; Cristia et al., 2020; Elo, 2016; Ganek & Eriks-Brophy, 2018; Jones et al., 2019; Lehet et al., 2020; Orena et al., 2019). Those studies that examined precision of child vocalization identification in particular (Cristia et al., 2020; Elo, 2016; Jones et al., 2019) confirm that the LENA algorithm identifies child vocalizations fairly well (64% precision and 55% recall of child vocalization identification). This could be because child speech contains anatomical cues (e.g., higher fundamental frequency and ensuing irregular harmonic structure, breathiness, spectral instability from the lack of established motor routines and non-uniform vocal tract growth) that are not expected to differ greatly cross-culturally. However, the exact acoustic dimensions that the algorithm uses to identify child speech are still unknown because of the proprietary nature of the LENA system. We return to this point about the identification of child vocalizations in the Discussion where we bring to bear some recent findings seeking to validate LENA's child speaker tag in a crosslinguistic corpus containing many of the languages studied here. Furthermore, we included a "junk" annotation option so that in the event that the LENA algorithm did incorrectly tag a child, the utterance would not inadvertently be included in our description of vocalizations.

For the Tseltal and Yélî Dnye corpora, we capitalized on manual annotations that were already completed (Casillas et al., 2017, 2019, in press). At the time of data processing for the current study, the Tseltal corpus included manual annotations of 1 h of audio per child. The 1 h per child included nine 5-min sections of the recording that were randomly selected from the entire daylong recording—that is, regardless of the ongoing activity—plus 15 min of 1–5 min portions of the recording featuring the peak turn-taking and infant vocal activity for the day (see Casillas et al., 2019 for details). The Yélî Dnye utterances used in the current study were selected from an available 22.5 min of audio per child sampled over nine 2.5-min portions of the audio randomly selected from the day—again, regardless of activity (see Casillas et al., in press, for details). Overall, this processing resulted in timestamps for the onset and offset of each child utterance produced during the annotated regions of each child's recording in the Tseltal and Yélî Dnye corpora.

From this collection of utterances found for each child in each corpus, we randomly sampled 100 child utterances per child. Thus, with 100 utterances from each of 52 children, this processing resulted in 5200 child utterances from the six corpora. The child utterances drawn from the daylong recordings varied in length from 36 ms to 26,737 ms. Utterance length details by corpus are reported in Table 2.

3.5 | From utterances to clips

We next partitioned the child utterances into shorter units. For the rest of this paper, these shorter audio units, derived from the longer child utterances, will be referred to as "clips" (details are below). This was done to meet the challenge of manually tagging a large-scale dataset using a web-based, crowdsourcing citizen science platform. Specifically, publicly sharing even short utterances from recordings of natural human interaction poses a risk of privacy invasion and confidentiality breach. Participants' personal identifying information could be exposed if the recordings have not been pre-vetted by trained native speakers using clear guidelines for personal information content. In contrast, clips that are, at most, 499 ms in duration are highly unlikely to contain more than two syllables, and are thus too short to contain personal identifying information such as names or addresses. Using shorter clips (as detailed below) in this study permitted large-scale annotation beyond what could be typically completed by a single research group. At the same time, using such short clips allowed families' confidentiality and privacy to be safeguarded.

Seidl et al. (2019) provides validation of this method of tagging vocal maturity.[4] The authors evaluated two variables that could affect annotation accuracy of spontaneous child vocalizations: annotator expertise (minimally trained, semi-trained, expert) and clip length (200 ms, 400 ms, 600 ms, full utterance). Results for annotator expertise showed that both minimally-trained (naive) and semi-trained (undergraduate research assistants) annotators obtained strong correlations (reliability) with the expert annotators, suggesting that annotators did not require extensive background in child language or phonetic analysis to identify canonical transitions. Of

TABLE 2 Utterance and clip length measurements by corpus. Asterisks indicate manual utterance segmentation. All units are ms

Corpus	Child utterance length: *M* (*SD*)	Child utterance length: range	Clip length: *M* (*SD*)	Clip length: range
English-Bergelson	1035 (706)	600–9130	355 (84)	100–500
Tseltal	854 (842)*	36–11,314*	351 (95)	36–500
Yélî Dnye	927 (1701)*	53–26,737*	359 (89)	53–500
Quechua	1234 (637)	600–4760	364 (79)	100–500
Tsimane′	1124 (920)	600–18,340	359 (81)	100–500
English-Spanish-Warlaumont	1311 (668)	600–5210	366 (75)	100–500

the tested clip lengths, the 400 ms length led to an agreement on canonical transition identification that was as high as estimates made from full utterances (minimally-trained: r = 0.55 for full clips, r = 0.55 for 400 ms clips; semi-trained: r = 0.66 for full clips, r = 0.69 for 400 ms clips). Thus, Seidl et al.'s (2019) results were consistent with a growing body of language development research showing that aggregated groups of citizen scientists annotate speech production data reliably and on par with highly trained and/or expert annotators (Fernández et al., 2019; Harel et al., 2017; McAllister Byun et al., 2016), provided that the task is made small enough to benefit from categorical decisions.

To convert the longer utterances into the much shorter clips, each utterance was first cut into 400 ms clips, with the remainder (always <400 ms) included as a separate, short clip of its own (100–399 ms) except when the remainder was shorter than 100 ms. In that case, the remainder was appended onto the final 400 ms clip (1–99 ms). A clip could therefore be maximally 499 ms long. For example, a 1400 ms child utterance would be converted into 4 clips under 500 ms (400 ms + 400 ms + 400 ms + 200 ms). A 944 ms utterance would be partitioned into 3 clips (400 ms + 400 ms + 144 ms). The only exception to this procedure was for the two Casillas corpora which contained a few child utterances <100 ms in length (Yélî Dnye: N = 8, M(SD) = 78 ms (16); Tseltal: N = 22, M(SD) = 81 ms (16)). Finally, we imposed a 5 ms fade-in and -out to each clip to avoid click sounds. This process resulted in a total of 14,982 short clips from the 52 children. This crosslinguistic corpus of child vocalizations is available for use and replication (https://osf.io/rz4tx/).

3.6 | Procedures

All of these short clips were shared on a web-based citizen science platform called iHEARu-PLAY (Hantke et al., 2013), where they were annotated into one of five categories: (1) canonical (CV sequences with rapid, adult-like transitions, fully resonant vowels, and supraglottally generated consonants), (2) non-canonical (e.g., isolated vowels, isolated consonants, raspberries, squealing, CV sequences with subglottally-generated consonants, and CV sequences with slow, weak transitions and/or vowel sounds that are not fully resonant), (3) crying, (4) laughing, and (5) junk/other (vegetative sounds like coughs, all speech not from a child, speech overlap, television, and radio).

It may be relevant to clarify that our definition of non-canonical is most aligned with recent work (Belardi et al., 2017; Ha & Oller, 2019; Lee et al., 2018; Nathani et al., 2007; Oller, 2000; Patten et al., 2014) which categorized non-canonical as (1) syllables "lacking any margin (i.e., vowel-like sounds only)," (2) syllables with "vowel-like nuclei but no supraglottal articulation," (3) marginal babbles where "the formant transition between the nucleus and the margin is slow … or the vowel-like sound is not fully resonant," and (4) "syllables consisting throughout of supraglottaly-generated sound sources such as in raspberries, isolated fricatives or affricatives" (Lee et al., 2018: 9).

Prior to beginning annotation, each annotator completed a training module, linked from the iHEARu-PLAY platform and housed on Qualtrics (Qualtrics, 2019; purdue.ca1.qualtrics.com/jfe/form/SV_brsqXckmH73EpDf). The training module explained basic concepts of child vocalizations and vocalization maturation for a non-specialist audience and included multiple audio examples and definitions of the different types of canonical and non-canonical clips as well as examples of crying, laughing, and all of the categories to be classified as junk. Annotators were additionally reminded that the clips were taken from larger audio utterances and that they could be annotating clips taken from the middle of an utterance. Examples were also provided of such truncated clips. This training module is included in this project's affiliated OSF project (https://osf.io/ca6qu/).

The categorization task was shared widely throughout the language and cognitive development community via the CHILDES, Cognitive Science Society, and other psychology listservs. The task was available for anyone over the age of 18 years of age to participate in anonymously. The 136 total annotators included language, speech, psychology, and cognitive science researchers, undergraduate students, and research assistants, but also other users of the iHearuPlay platform for whom we do not have background statistics. Annotators' backgrounds and experience with language development, and behavioral research more broadly, could vary; the annotation task was designed to accommodate all levels of experience with the subject matter. There was no minimum or maximum threshold for the number of annotations to be completed by each annotator. Generally, a given clip was tagged by a unique set of annotators. However, due to a workflow characteristic in the iHEARu-Play platform,[5] some clips were annotated two times by the same coder; this occurred only for 27 clips (0.002% of all clips). No clips were annotated by the same coder more than twice.

4 | RESULTS

Our primary research question concerns the time course of vocal development as measured by the prevalence of canonical transitions. Specifically, analyzing a large, culturally-diverse sample, we investigated whether canonical transitions emerge in a developmental time course similar to what has been reported in previous work. We begin the results with descriptive statistics concerning the clip annotations before turning to analyses of canonical proportion by age, corpus, and child gender.

All analyses were conducted in the RStudio computing environment (version: 1.2.5033; RStudio Team, 2020). Data visualizations were created with ggplot2 (Wickham, 2016). Modeling was conducted using a combination of the lme4 and lmerTest packages (Bates et al., 2015; Kuznetsova et al., 2017) and summaries were presented with Stargazer (Hlavac, 2018). The significance of potential model parameters was determined using a combination of log-likelihood comparisons between models, AIC estimations, and p-values procured from the lmerTest package. The alpha level for log-likelihood comparisons was corrected to 0.017 to account for the multiple comparisons (0.05/3 for three planned tests, including interactions). Continuous predictors were mean-centered to facilitate

model interpretation. All scripts to replicate these analyses are publicly available in our OSF project (https://osf.io/ca6qu/).

4.1 | Pre-processing of annotations

All 14,982 clips were posted for annotators on the iHEARu-PLAY platform. Each clip was annotated at least three times (range = 3–17 annotations, *M* = 4.34, *SD* = 2.25) for a total of 65,018 annotations. In the analyses below, we only included clips where a majority of the annotations were in agreement (i.e., 66%–100% of the annotation tags for the clip were the same). *N* = 6848 (45.71% of the original clips) had 100% agreement and *N* = 7257 (48.44%) had >66% but <100% agreement. Finally, a total of *N* = 877 clips lacked majority agreement and were removed from analyses (5.85% of original clips).[6] See Table 3 for the distribution by corpus of 100% agreement clips, majority agreement clips, and no majority agreement clips. Overall, each corpus had a similar percentage of clips across agreement categories (full agreement, majority agreement, no majority agreement). For the remainder of the analysis, we do not differentiate between clips with 100% rater agreement and those with >66% but <100% agreement, referring to both as "majority" agreement clips. Of the majority-labeled clips, *N* = 5285 (35.28%) were categorized as junk and *N* = 11 did not receive an answer due to a technical error on the platform. Those clips annotated as "junk" and "no answer" were also removed from further analyses.

Figure 2 and Table 4 display the distribution of vocalization categories across the six corpora.

Canonical clips made up between 2% to more than 20% of the clips across the six corpora. Non-canonical clips varied more in frequency across corpora, from 5% to more than 60%, which may relate to differences in age coverage across corpora. Both crying and laughing were relatively rare and will not be discussed further.

Surprisingly, the English-Spanish-Warlaumont corpus contained a higher than expected percentage of clips labeled as junk (92%). In comparison, approximately 30% of the clips in the English, Tseltal, Tsimane, and Yélî Dnye corpora contained junk clips. While difficult to determine definitively, differences in the prevalence of junk clips may be due to the younger age of the participants in the English-Spanish-Warlaumont corpus (3 months), the recording setting, a low number of speech-like clips, or a combination of these and other factors. As it was not possible to determine the cause of the junk

TABLE 3 Percentage of each corpus that contained complete agreement, majority agreement, or no agreement clips

Corpus	Complete agreement	Majority agreement	Not majority agreement
English-Bergelson	45.88	51.06	3.05
English-Spanish-Warlaumont	81.58	17.49	0.93
Quechua	65.03	33.1	1.87
Tseltal	41.39	52.03	6.58
Tsimane′	40.84	53.63	5.53
Yélî Dnye	36.3	51.01	12.69

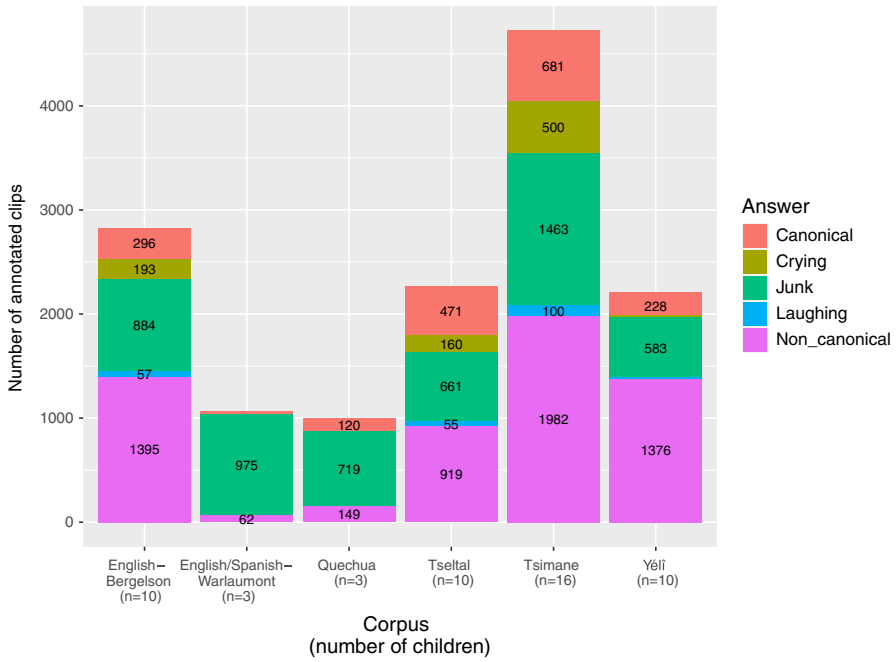

FIGURE 2 Annotations by corpus: raw counts

TABLE 4 Percentages of annotation categories by corpus

	English-Bergelson	English-Spanish-Warlaumont	Quechua	Tseltal	Tsimane′	Yélî Dnye
Canonical	10.48	2.16	12.01	20.79	14.41	10.30
Non-canonical	49.38	5.83	14.91	40.56	41.94	62.15
Laughing	2.02	0.09	0.30	2.43	2.12	0.54
Crying	6.83	0.28	0.80	7.06	10.58	0.68
Junk	31.29	91.64	71.97	29.17	30.96	26.33

clips in the Warlaumont corpus, we decided to remove the three Warlaumont corpus recordings from further analysis. This decision was justified because the Warlaumont recordings were unique in their high percentages of junk clips and low number of usable canonical + non-canonical clips (<35 clips, the lowest of all the recordings). Removing this corpus still leaves a large sample size (49 children), and 6 languages represented in the final analysis. A complete analysis that includes the three removed children is included in Supporting Information. In the Discussion, we elaborate further on possible explanations for the large amounts of junk present in those recordings.

4.2 | Results by age

As canonical proportion is predicted to increase with age, we first examined its growth over time, irrespective of corpus of origin or individual child. To calculate individual children's canonical proportions, all of the clips labeled as canonical were divided by the total number of clips labeled as canonical or non-canonical (Table 5). See the appendices for tables displaying canonical proportion by child age and an additional visual plotting proportion of canonical clips to non-canonical clips by individual child and age group (Appendices A1, A2).

As seen in Figure 3, across children, the proportion of clips labeled as canonical increased over development in this age range. To quantify this, we fit a regression model predicting canonical proportion by child age (in months): (β = 0.01, t = 5.91, p < 0.001). Results showed that for each month of development, canonical proportion increased by 0.01 (adjusted R^2 = 0.41). A canonical proportion of 0.15 was achieved at approximately 7 months.

More specifically, between the ages of 0;1 and 0;6 (inclusive; N = 6), participants' canonical proportions averaged just 0.07 (SD = 0.04). The average canonical proportion increased to 0.15 (SD = 0.11) for infants aged 0;7–1;0 (n = 11). Figure 4 plots those children who have reached the 0.15 threshold, against those who have not, by age. As anticipated, most children under 7 months have a canonical proportion <0.15, but this becomes rarer as children age: only two children over 1;5 (aged 30 and 31 months) did not show a canonical proportion at or above this 0.15 threshold.

Cross-corpus differences in canonical proportion growth were relatively small (Figure 5). Canonical proportion increased with age in each cross-sectional corpus with the following Pearson correlations: Tsimane′ (R = 0.11, [CI = −0.4, 0.58], p = 0.68, spanning 8–32 months), Tseltal (R = 0.90, [CI = 0.64, 0.98], p < 0.001, 2–36 months), Yélî Dnye (R = 0.89, [CI = 0.58, 0.97], p < 0.001, 1–36 months), and Bergelson (R = 0.39, [CI = −0.31, 0.82], p = 0.26, 7–17 months).[7] Two Tsimane′ children, one aged 30 months and another 31 months, were notable exceptions within the entire dataset with canonical proportion of 0.12 and 0.09, respectively. We explore possible explanations for this pattern in the Discussion. Additionally, one child from the Tseltal corpus, aged 0;11, had a higher-than-anticipated canonical proportion, with respect to the entire dataset, of 0.40.

TABLE 5 Counts of canonical to non-canonical clips and canonical proportion by child age (months): all corpora. Note that each age bracket can contain children from multiple corpora

Age in months	Canonical	Non-canonical	Total	Canonical proportion
1	6	145	151	0.04
2	5	120	125	0.04
4	25	281	306	0.08
6	10	103	113	0.09
7	37	384	421	0.09
8	52	420	472	0.11
9	57	320	377	0.15
10	16	89	105	0.15
11	63	93	156	0.40
12	35	131	166	0.21
13	53	306	359	0.15
14	49	135	184	0.27
15	97	516	613	0.16
16	138	308	446	0.31
17	62	322	384	0.16
18	99	223	322	0.31
19	23	102	125	0.18
20	61	328	389	0.16
22	68	133	201	0.34
23	147	185	332	0.44
24	154	184	338	0.46
25	14	25	39	0.36
26	77	122	199	0.39
27	81	75	156	0.52
30	21	158	179	0.12
31	16	164	180	0.09
32	120	212	332	0.36
36	210	237	447	0.47

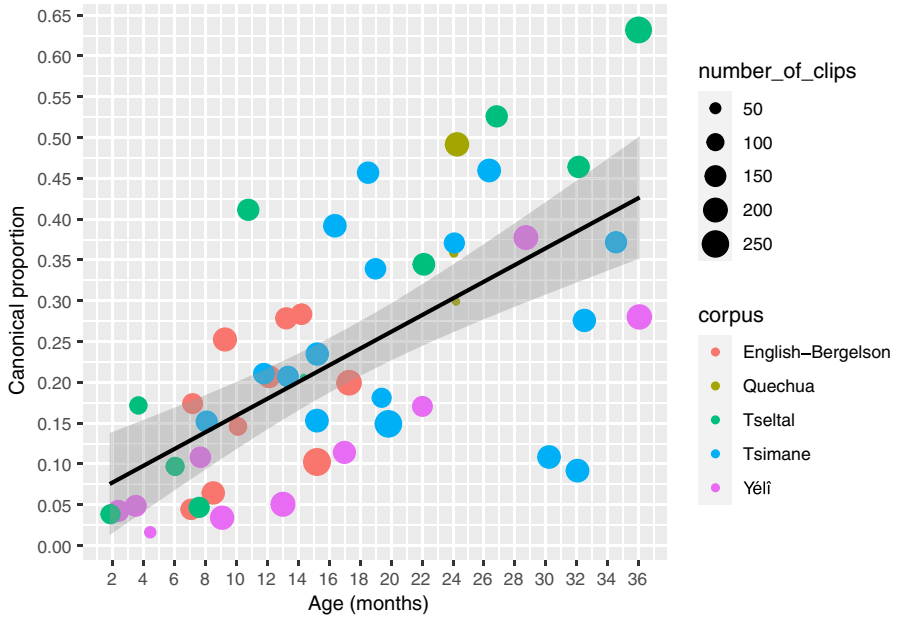

FIGURE 3 Canonical proportion by child age and corpus. Shaded band surrounding regression line represents 95% confidence intervals. Each point represents one child and point size refers to the number of clips used to calculate canonical proportion

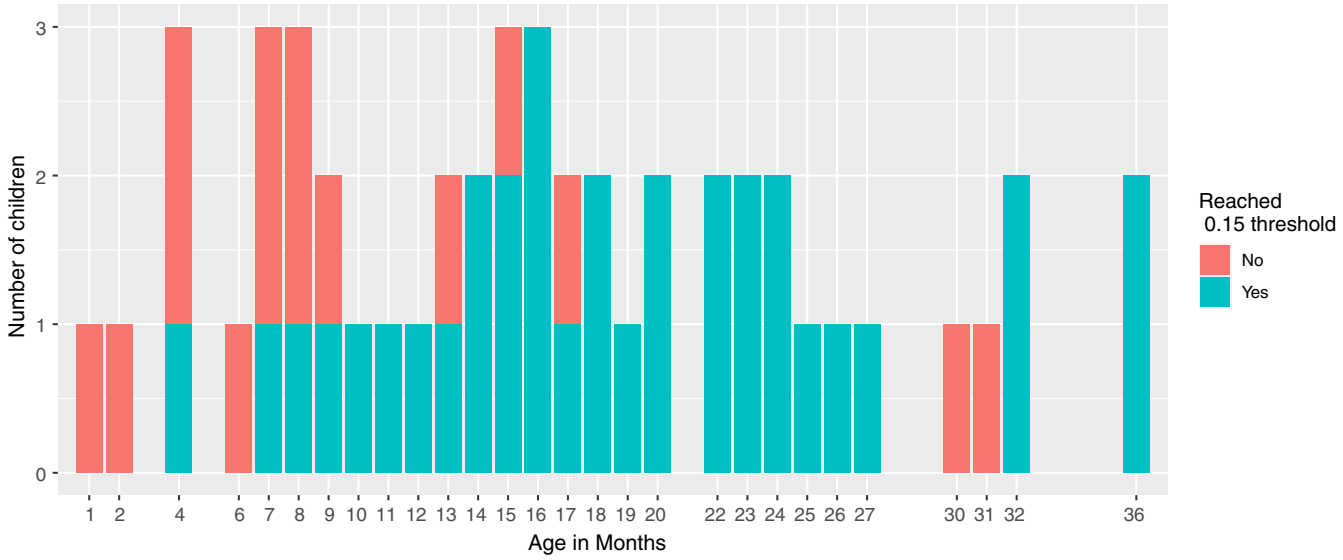

FIGURE 4 Children in the current study whose canonical proportion is above 0.15, plotted by age (in months)

The weakest relationship between canonical proportion and age was evident in the Tsimane′ corpus, which showed overall relatively high canonical proportions (estimate before 15 months at about 0.25) and high variability between participants which seemed unrelated to age. Indeed, almost all of the Tsimane′ children had a canonical proportion at or above 0.15: even the youngest child in the Tsimane′ corpus, aged 0;8, had a canonical proportion of 0.16. Consequently, the lack of age-related change could be due to these children reaching the 0.15 threshold at a slightly younger age than previously reported in North American and other Western samples, though future crosslinguistic work will be needed to verify this.

In the English-Bergelson corpus, the canonical proportion increased from an intercept of 0.14 to 0.22 between 7 and 17 months. Thus, the weaker relationship between age and canonical proportion in the English-Bergelson corpus than the Tseltal and Yélî Dnye corpora could be due to the smaller range of ages sampled (7–17 months in English-Bergelson vs. 2–36 and 1–36 months in the other two). The Tseltal and Yélî Dnye canonical proportion results also differed numerically, with lower initial and final values for the latter than the former. Future work exploring whether such differences are related to syllable structure and/or phonotactic differences between the two input languages given the highly distinct phonological systems of Tseltal and Yélî Dnye would be a welcome addition to the literature.

Overall, these analyses by corpus show that children reached a 0.15 canonical proportion threshold before 10 months of age. This held for a diverse set of cultural groups, including ones

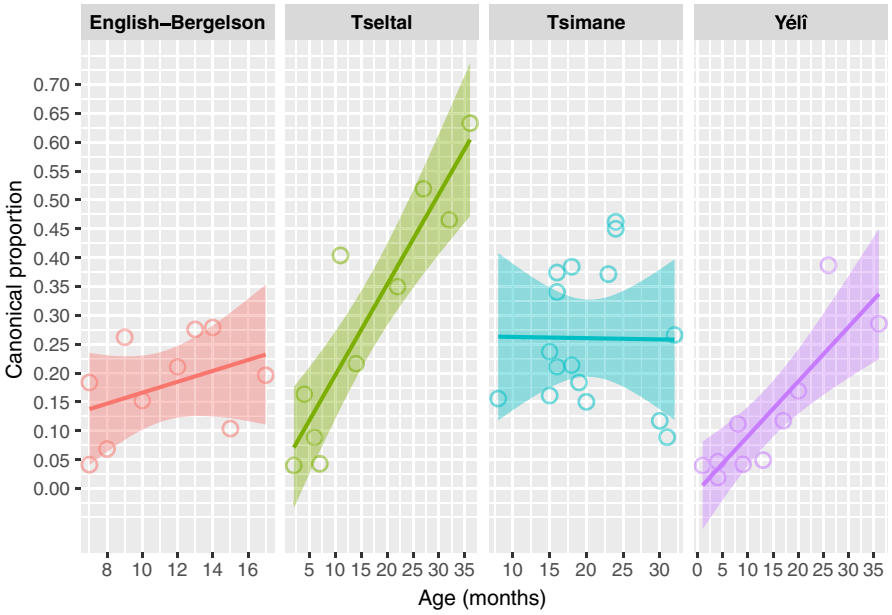

FIGURE 5 Canonical proportion by child age (months) across the four corpora that contained cross-sectional age samples. Note that *x*-axis scales differ by corpus

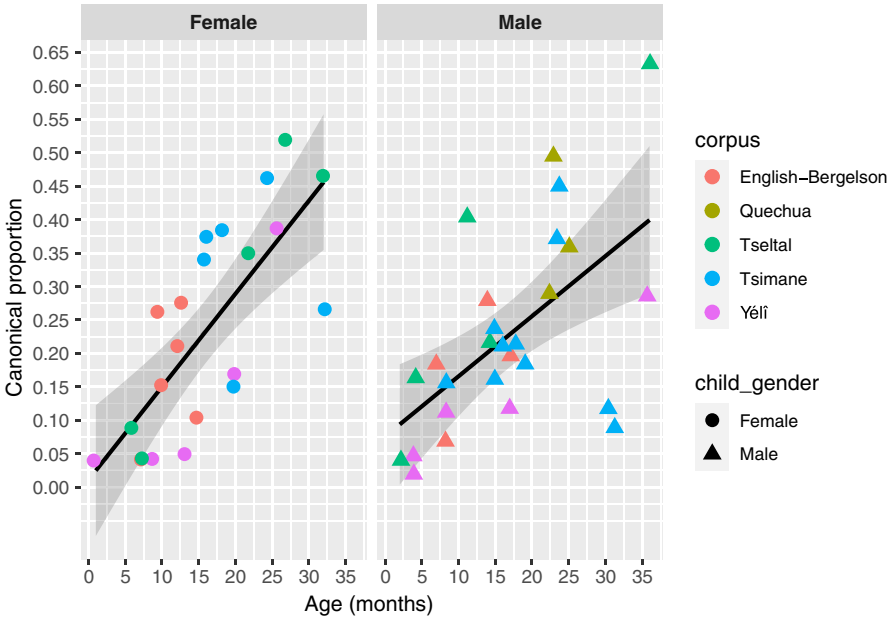

FIGURE 6 Canonical proportion by child age (months) and gender

previously reported to have very low quantities of child-directed speech.

4.3 | Results by gender

Finally, we analyzed how canonical proportion varied with respect to child gender. Figure 6 plots canonical proportion for all corpora, split by gender for the n = 27 boys and n = 22 girls. Canonical proportion was positively correlated with child age for girls (R = 0.75, [CI = 0.49, 0.89], p < 0.001) and boys (R = 0.58, [CI = 0.26, 0.79], p = 0.001). Though the correlation appears slightly stronger for the female children, the confidence intervals of these correlation statistics overlap widely.

On the basis of the linear relationship between canonical proportion and child age in both the female and male groups, a linear mixed effects model was fit to predict canonical proportion. After controlling for corpus in the random effects structure and including child age (in months) as a fixed effect, a log-likelihood test demonstrated that the addition of a covariate for child gender did not improve model fit (df = (1), χ^2 = (0.31), p = 0.58) (Table 6). Note that at one data point per child, these analyses do not permit random slopes of child nested within corpus. The interaction between child age (in months) and child gender did not improve a model with child age either (df = (1), χ^2 = (1.75), p = 0.19). We thus conclude that in our sample there is no evidence for differences in canonical proportion by gender.

TABLE 6 Canonical proportion growth by child age (months) and assigned gender

	Models		
	(1)	(2)	(3)
Intercept	0.24 (0.18, 0.29)***	0.25 (0.18, 0.31)***	0.25 (0.18, 0.31)***
Child age (months)	0.01 (0.01, 0.01)***	0.01 (0.01, 0.01)***	0.01 (0.01, 0.02)***
Child gender:male		−0.02 (−0.08, 0.04)	−0.02 (−0.08, 0.04)
Child age × child gender:male			−0.005 (−0.01, 0.002)
Observations	49	49	49
Log likelihood	29.59	27.24	23.36
Akaike Inf. Crit.	−51.18	−44.47	−34.71
Bayesian Inf. Crit.	−43.61	−35.01	−23.36

*** $p < 0.01$.

5 | DISCUSSION

In this study, we found a high degree of consistency within our culturally and linguistically diverse dataset: we found that a canonical proportion of 0.15 was reached on average at about 7 months, and for most children before 10 months, across the corpora. Since the large-scale annotation required for this project took place on a public crowdsourcing website, the canonical proportion metric was necessarily based on short (around 400 ms) clips rather than syllables. The canonical proportion metric also included all speech-like vocalizations, and yet the threshold found in previous reports of CBR in more culturally- and linguistically homogeneous datasets (Oller et al., 1997, 1998, 1999) remained meaningful. This finding not only increases confidence about the universality of vocal development, and the prevalence of canonical transitions in particular, but also helps validate automatic extraction and explore crowdsourced labeling as viable methods for data processing and annotation of naturalistic daylong audio recordings of children's language environments.

It is worth underscoring that the crowdsourcing method used in this project appears to be a promising approach for other questions of interest in cognitive development. Our collaboration with citizen scientists allowed us to acquire more than 60,000 annotations from scores of annotators who were intrinsically interested in contributing to this effort. Furthermore, the crowdsourcing platform we employed allowed annotators to be quickly trained, permitting more unique users to join the effort. This approach made the production of a relatively large, well-tagged dataset of infant vocalizations from around the world feasible, and may provide training data for future speech parsing algorithms.

5.1 | Cross-corpus comparisons

The dataset employed in this study compiles spontaneous child vocalizations from linguistically- and culturally-diverse corpora. Results demonstrate that in our crosslinguistic sample, children appear to reach a 0.15 canonical proportion before the age of 10 months. One reason why we anticipated possible cross-cultural differences in canonical proportion trajectories was because previous research has found a role of culture, specifically caregiving practices, on other motor behavior in early childhood (Adolph et al., 2009; Karasik et al., 2018; Super, 1976). Furthermore, there has been some limited discussion of possible cultural reasons behind differences in vocal development in infants from Taiwan and the United States (Lee et al., 2018). However, unlike previous reports of cultural differences in gross motor milestones such as crawling, our results do not support an interpretation of cultural differences in vocal milestones—at least for canonical transitions. As with all null effects in developmental research, this conclusion will require further exploration, via different data collection methods or in additional cultural contexts. However, the current sample suggests that canonical transitions increase in prevalence along a similar timeline cross-culturally.

The similarities in vocal development across multiple cultural contexts in this study mirror previous work on the robustness, or *canalization* (Oller, 2000), of vocal development in a variety of language learning environments. Previous work has not demonstrated significant effects of bilingual status, infant prematurity, or family socioeconomic status upon the development of canonical transitions or babbling (Eilers et al., 1993; Oller et al., 1994, 1997). This study augments the conclusions from this previous research by showing similarities across a very diverse set of cultures, with distinct caregiving practices (e.g., quantity of child-directed speech).

There were, however, some differences of note between corpora. One difference concerned the relative quantity of usable data within each corpus. Specifically, the English-Spanish-Warlaumont and Quechua corpora had higher percentages of 'junk' clips than the other corpora, even relative to the other automatically speaker tagged LENA corpora (English-Bergelson and Tsimane'). It is reasonable to think that age differences between corpora could explain the differences in quantity of'junk' clips. However, the English-Spanish-Warlaumont and Quechua corpora captured quite different age ranges. The English-Spanish-Warlaumont contained children on the younger end of our sample (3 months) and the Quechua corpus contained children on the older end (22–25 months). Therefore, it is unlikely that the high prevalence of 'junk' in these corpora is related solely to age.

One may wonder whether the Quechua and English-Spanish-Warlaumont corpora were gathered in noisier environments, or with more speaker overlap, than the other corpora. However, this explanation also does not fit the data. The English-Spanish-Warlaumont corpus, which was collected in North America, likely captures the child at home (similar to the English-Bergelson corpus), whereas the Quechua corpus was collected in a community in Bolivia where children typically spend a large portion of time outside and around high volumes of multi-talker conversation during the day (similar to the Tsimane′, Tseltal, and Yélî Dnye corpora). Yet we see larger amounts of junk in the Quechua and English-Spanish-Warlaumont than the other automatically speaker-tagged corpora: English-Bergelson and Tsimane′. Thus age and environment do not clearly explain the different quantities of "junk" in some corpora in the dataset.

Another key difference across sub-corpora is the relationship between age and the canonical proportion outcome. Of the four corpora with more than three participants, the English-Bergelson and Tsimane′ corpora showed a somewhat weaker relationship between canonical proportion and age than the Yélî Dnye and Tseltal corpora, although all four corpora had overlapping age ranges.

Note that here too the corpora that patterned together, English-Bergelson and Tsimane′, did not come from similar cultural contexts or environmental settings. The English-Bergelson corpus contains children from the suburban United States, generally within small family units with one or more adult caregivers. In contrast, the Tsimane′ families lived in open households in a small village where as soon as children can walk, they spend substantial portions of the day with other children (including siblings). In this sense, the Tsimane′ setting is more similar to that of the Tseltal and Yélî settings.

One might also ask whether cross-corpus differences in the relationship between age and canonical proportion, or the prevalence of "junk" in some corpora, is attributable to how the data were pre-processed. Specifically, in the Yélî Dnye and Tseltal corpora, the key child utterances were hand-identified while in the remaining corpora the LENA algorithm automatically identified the child utterances. However, there are several reasons why it is unlikely that the observed differences are attributable to data pre-processing. First, all of the child utterances were chopped into smaller clips, and subsequently annotated, in the exact same manner. All of the processed clips were also annotated together, intermingled on the same online platform. Given the similarity in annotation methods, and the short duration of the audio clips (clips were around 400 ms in length), it seems unlikely that cross-corpus differences could have arisen in the pre-processing step.

Another reason why it seems unlikely that pre-processing could explain these results is because LENA's annotation of child utterances has been validated on several of the corpora studied here (Cristia et al., 2020). The LENA annotation algorithm is trained on English data, and while the specifics of the underlying annotation technique remain a blackbox to developmental researchers, the annotation of child vocalizations in particular appears crosslinguistically robust. This is because unlike other LENA-derived annotations, such as Adult Word Count, which could rely on a specific language's phonotactic structure or stress placement, the child vocalization tag likely instead relies on anatomically-based acoustic cues, such as the heightened fundamental frequency and irregular harmonic structure of children's voices, which are not expected to vary much across our participant populations.

We did nevertheless entertain the possibility that there were some false alarms in LENA's annotation technique that could have resulted in a high proportion of "junk". Additional "junk" labeling might have occurred, for example, if the citizen scientists noticed that the mis-attributed clips contained a male or female adult, or a non-human noise. However, crucially, the confusion between an adult and a child could have been harder for citizen scientists to hear if the adult was using infant-directed speech so extreme that it sounded like a child. This would be more likely in populations with a very marked infant-directed speech register, such as that found in middle-to-upper class North American contexts. In this case, one could end up having a flat regression line against age because even young infants would inappropriately get canonical babble tags that in reality reflected female adults or older children.

To that end, Cristia et al., (2020) present some results attempting to validate LENA's child speaker tags that are relevant to the current study. The authors sampled child vocalization tags from the Tsimane′, English-Bergelson and English-Spanish-Warlaumont corpora (different samples from those examined in the current paper). Confusion matrices for precision rates, outlined with accompanying prose in Supporting Information, support the notion that child vocalization tags are crosslinguistically robust and thus are quite unlikely to account for the cross-corpus differences or prevalence of "junk" in some corpora. The confusion patterns reveal that maximally 6%–7% of the data in the Tsimane′, English-Spanish-Warlaumont, and Bergelson corpora could come from confusable speakers (not the target child; see Supporting Information for details).

As an additional precautionary measure to safeguard against differences in the method used to identify child utterances, we reran all analyses on a subset of the Tseltal and Yélî Dnye data. These results are included in Supporting Information in the affiliated OSF project (https://osf.io/ca6qu/). Specifically, for this sub-analysis, we removed all clips from the Tseltal and Yélî Dnye data that derived from utterances <600 ms. Our reasoning for this was that some of the child utterances in these two corpora were extremely short (see Table 2 for ranges), while the minimum utterance length in the remaining corpora was 600 ms. In total, this resulted in the removal of 1206 clips, or 8.05% of all clips in the entire dataset. This sub-analysis showed broadly the same patterns as the main analyses above: canonical proportion increased linearly with age, and the effect by age was slightly more notable in the Yélî Dnye and Tseltal corpora than the Tsimane′ and English-Bergelson corpora. Again, most infants reached the key 0.15 threshold by 10 months, if not earlier (as in the Tsimane′ corpus). On the basis of this analysis, we feel more confident in our initial conclusion that there were few notable crosslinguistic differences in canonical proportions.

As mentioned above, there were two notable exceptions to the canonical proportion trend in this dataset. Two Tsimane′ children,

aged 30 and 31 months, had fairly low canonical proportions of 0.12 and 0.09, respectively. Given the large number of Tsimane′ children included in this study (n = 16), most of whom followed a linear trajectory of an increased canonical proportion, we do not believe that these exceptions reflect cultural differences in canonical proportion development. In fact, these two children were the only ones in the Tsimane′ corpus with a canonical proportion below 0.15. One possible interpretation of these outlying canonical proportions is that the two children were exhibiting signs of language delay. Compared to, for example, the North American samples analyzed here, the Tsimane′ community is medically-underserved. As a result, there was no independent or locally-normed assessment to determine if the children were experiencing delays in their language development. However, longitudinal follow-ups of these children showed no evidence of atypical development a year after the recordings analyzed in this paper were collected. This leads us to conclude that there may instead have been ambient effects in these two children's recordings, like increased background noise, that affected the resulting canonical proportion estimates.

In sum, it seems these results demonstrate that crosslinguistically, children might be expected to reach a 0.15 canonical proportion before the age of 10 months. The conclusion drawn here reinforces the importance of reporting cross-cultural *similarities* in development, in addition to differences (Tamis-LeMonda & Song, 2012). Still, the number of children represented in each corpus is relatively small. This limits the interpretation of the cross-corpus differences that we preliminarily discuss. Furthermore, we implemented a novel vocal metric, canonical proportion, which is distinct from the more traditional CBR: canonical proportion is not necessarily estimated from syllables since the public-facing crowdsourcing platform required the use of very short audio clips that may or may not have encapsulated syllables. Finally, there were large amounts of "junk" classifications in some corpora that were not readily explained by the corpus language, sociocultural setting, child age, or data pre-processing steps. Researchers looking to implement citizen science annotation into their workflow should be aware that some classification decisions can result in significant quantities of unusable annotated data. It will thus be necessary for others to supplement our work with more studies and datasets in order to draw stronger conclusions about meaningful differences, or lack thereof, in speech development between these and other linguistically- and culturally-diverse corpora. In particular, there is a need for both rigorous, manual segmentation of crosslinguistic samples, as well as methodological advances in automatic vocalization segmentation to facilitate crosslinguistic research at a larger scale. We hope that the corpus generated from this project proves a useful tool for these endeavors.

5.2 | Gender

This study also sought to determine if there were gender differences in children's canonical proportion. We found no significant differences by gender in our dataset. There are a few ways to interpret this result. First, it is possible that our large cross-sectional cohort might lack the power to detect a subtle gender difference. Alternatively, it is possible that the onset and frequency of canonical transitions do not vary by gender and that other mechanisms are involved in language differences by gender later in development. Finally, if gender differences in canonical proportion were very minor, language differentiation by gender in development may be non-linear and dependent on the aspect of vocalization analyzed. For examples, in a study of American infants, girls and boys showed early differences (perhaps due to infant sex hormone surges; Quast et al., 2016) in volubility (Oller et al., 2020), which disappeared at the critical age when canonical babbling develops (at around 7–12 months). Language differences in lexical and morphosyntactic outcomes seem to reappear later in development (Barbu et al., 2015; Eriksson et al., 2012; Frank et al., 2017; Hadley et al., 2011; Huttenlocher et al., 1991). Future research could expand on the current project by analyzing more features, including volubility and more detailed phonological, lexical, and grammatical codes, to study patterns of similarity and difference between genders cross-culturally. Given that the current dataset focused on canonical transitions, our results suggest that the lack of gender differences within this aspect of language development are a cross-cultural phenomenon.

6 | CONCLUSION

This study presented the first analyses of child vocalization development across a highly linguistically and culturally diverse sample. We found that the timeline of canonical transition development does not appear to vary dramatically by cultural context or child gender. The expected age to reach a canonical proportion of 0.15 was approximately 7 months, and, overall, canonical proportion increased positively with age. However, the relationship between age and canonical proportion was stronger in some corpora (Tseltal, Yélî Dnye) than others (Tsimane′, English-Bergelson). These differences were not readily explained by differences in cultural context.

These findings replicate previous work with less diverse samples and settings, and invite further work with typical and atypical children within these populations in order to derive developmental benchmarks from child vocalizations that are independent of language and cultural exposure. In addition, the child vocalization corpus created for this project is now publicly available for other developmental and computational researchers to analyze and build on in future work.

This work also explored how crowdsourcing can be used to elicit large quantities of annotations on already existing data from citizen scientists. This workflow allowed us to efficiently and economically annotate existing data while engaging the public in science. Future practitioners should note that lower inter-annotator reliability on crowdsourcing platforms means that a larger number of annotations/annotators may be required. Lower inter-annotator reliability may also indicate that crowdsourcing may not be suitable for more fine-grained data annotation tasks. Still, incorporating large-scale annotation efforts such as these into social science research is a

crucial step towards increasing data reliability and replicability as it permits multiple, large-scale annotations on shareable datasets, across multiple labs and research sites.

ACKNOWLEDGEMENTS

The authors thank the 52 families who contributed recordings for this project, the research assistants and collaborators who participated in the creation of the corpora, and the citizen science annotators.

AUTHOR CONTRIBUTIONS

MCy and AS directed the research collaboration. MCy, AC, EB, MCa, ASW, and CS contributed data. AC, MCa, AS, GB, and MCy pre-processed the data. MCy analyzed the data. MCy, GB, and EB organized the OSF and Github documentation. All authors contributed to the design of the study and wrote the paper.

DATA AVAILABILITY STATEMENT

The audio data annotated and analyzed for this paper are publicly available for re-use (https://osf.io/rz4tx/).

ORCID

Margaret Cychosz https://orcid.org/0000-0003-3021-4707
Alejandrina Cristia https://orcid.org/0000-0003-2979-4556
Elika Bergelson https://orcid.org/0000-0003-2742-4797
Marisa Casillas https://orcid.org/0000-0001-5417-0505
Anne S. Warlaumont https://orcid.org/0000-0001-9450-1372
Camila Scaff https://orcid.org/0000-0002-7546-9538
Lisa Yankowitz https://orcid.org/0000-0003-2604-5840

ENDNOTES

[1] Lee, Jhang, Chen, Relyea, and Oller (2017) point out some methodological concerns of de Boysson-Bardies et al. (1984). These include a lack of annotator blinding to hypotheses, the presence of cues from ambient language, and differences in recording equipment across sites, all of which may have led to erroneous or biased results.

[2] One family studied also spoke Southern Min in the home.

[3] If there are small effects of gender on early vocalizations, larger samples may be required to discern them, and thus authors of previous work may have not reported them because they were not significant. We have a larger sample (when combining across cultures) than much previous work, which means that we have both more power to detect a difference if one exists, and more precision in our measure of the actual size of an effect. Thus, an additional reason to report on gender is to aid future meta-analyses seeking to quantify the true effect size of gender on vocal development.

[4] Semenzin, Hamrick, Seidl, Kelleher, and Cristia (2020) likewise validated a crowdsourcing approach to vocal maturity annotation. A group of in-lab expert and citizen science annotators classified children's vocalizations into crying, laughing, canonical, non-canonical, and junk. Results showed a high weighted accuracy correspondence (73%) between annotations performed by the two groups and estimates of canonical proportion were highly correlated between in-lab and citizen science annotators ($r = 0.92$, $p < 0.001$).

[5] We posted clips to iHEARu-Play in several data batches, based on when the pre-processed data became available. Sometimes a clip was posted in more than one batch (e.g., because the clip had not previously received at least three annotations). iHEARu-Play does not have a way to stop coders from annotating the same clip between batches, so some coders received the same clip twice.

[6] For example, a clip with three annotations, all of which were different (e.g., cry, junk, laugh) would be removed. A clip with four annotations, two of which were different (e.g., cry, cry, laugh, laugh) would be removed. A clip with five annotations, three of which were different (e.g., cry, cry, laugh, laugh, junk), would be removed. However, a clip with three annotations, two of which were the same (e.g., cry, cry, laugh) was retained. Finally, a clip with four annotations, three of which were the same (e.g., cry, cry, cry, laugh) were also retained.

[7] Here only the developmental trends for those corpora that contained cross-sectional age samples are presented (Tsimane', Tseltal, Yélî Dnye, and English-Bergelson). The Quechua corpus is not visualized as it only contained three children in our current sample, which was not sufficient to track developmental changes.

REFERENCES

Adolph, K. E., Karasik, L. B., & Tamis-LeMonda, C. S. (2009). Motor skills. In M. Bornstein (Ed.), *Handbook of cultural developmental science* (pp. 61–88). Taylor & Francis.

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.

Barbu, S., Nardy, A., Chevrot, J.-P., Guellao, B., Glas, L., Juhel, J., & Lemasson, A. (2015). Sex differences in language across early childhood: Family socioeconomic status does not impact boys and girls equally. *Frontiers in Psychology*, *6*, 1874. https://doi.org/10.3389/fpsyg.2015.01874

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Belardi, K., Watson, L. R., Faldowski, R. A., Hazlett, H., Crais, E., Baranek, G. T., Oller, D. K. (2017). A retrospective video analysis of canonical babbling and volubility in infants with Fragile X syndrome at 9–12 months of age. *Journal of Autism and Developmental Disorders*, *47*(4), 1193–1206. https://doi.org/10.1007/s10803-017-3033-4

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

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., Casillas, M., Soderstrom, M., Seidl, A., Warlaumont, A. S., & Amatuni, A. (2019). What do North American babies hear? A large-scale cross-corpus analysis. *Developmental Science*, *22*(1), e12724. https://doi.org/10.1111/desc.12724

Blake, J., & de Boysson-Bardies, B. (1992). Patterns in babbling: A cross-linguistic study. *Journal of Child Language*, *19*(1), 51–74.

Bornstein, M. H., Tamis-LeMonda, C. S., Tal, J., Ludemann, P., Toda, S., Rahn, C. W., Pecheux, M.-G., Azuma, H., & Vardi, D. (1992). Maternal responsiveness to infants in three societies: The United States, France, and Japan. *Child Development*, *63*(4), 808–821.

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

Brown, P. (2011). The cultural organization of attention. In A. Duranti, E. Ochs & B. B. Schieffelin (Eds.), *Handbook of language socialization* (pp. 29–55). Wiley-Blackwell.

Brown, P., & Casillas, M. (in press). Childrearing through social interaction on Rossel Island, PNG. In A. J. Fentiman & M. Goody (Eds.), *Esther Goody revisited: Exploring the legacy of an original inter-disciplinarian*. Berghahn. https://psyarxiv.co/rvky/GoogleScholar

Canault, M., Le Normand, M.-T., Foudil, S., Loundon, N., & Thai-Van, H. (2016). Reliability of the language environment analysis system (lenaTM) in european french. *Behavior Research Methods*, *48*(3), 1109–1124. https://doi.org/10.3758/s13428-015-0634-8

Carra, C., Lavelli, M., & Keller, H. (2014). Differences in practices of body stimulation during the first 3 months: Ethnotheories and behaviors of Italian mothers and West African immigrant mothers. *Infant Behavior and Development*, *37*(1), 5–15. https://doi.org/10.1016/j.infbeh.2013.10.004

Casillas, M., Brown, P., & Levinson, S. C. (in press). Early language experience in a Papuan community. *Journal of Child Language*. https://doi.org/10.1017/S0305000920000549

Casillas, M., Brown, P., & Levinson, S. C. (2017). *Casillas HomeBank corpus*. https://doi.org/10.21415/T51X12

Casillas, M., Brown, P., & Levinson, S. C. (2019). Early language experience in a Tseltal Mayan village. *Child Development*, *91*(5), 1819–1835. https://doi.org/10.1111/cdev.13349

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

Cristia, A., Dupoux, E., Gurven, M., & Stieglitz, J. (2019). Child-directed speech is infrequent in a forager-farmer population: A time allocation study. *Child Development*, *90*(3), 759–773. https://doi.org/10.1111/cdev.12974

Cristia, A., Lavechin, M., Scaff, C., Soderstrom, M., Rowland, C., Räsänen, O., Bunce, J., & Bergelson, E. (2020). A thorough evaluation of the Language Environment Analysis (LENA) system. *Behavior Research Methods*. https://doi.org/10.3758/s13428-020-01393-5

Cychosz, M. (2018). *Cychosz HomeBank corpus*. https://doi.org/10.21415/YFYW-HE74

Cychosz, M. (2020). *Phonetic development in an agglutinating language*. Unpublished doctoral dissertation, University of California, Berkeley, CA.

Databrary. (2012). *The databrary project: A video data library for developmental science* [Computer software manual]. http://databrary.org

de Boysson-Bardies, B., Sagart, L., & Durand, C. (1984). Discernible differences in the babbling of infants according to target language. *Journal of Child Language*, *11*(1), 1–15.

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

de León, L. (1998). The emergent participant: Interactive patterns in the socialization of Tzotzil (Mayan) infants. *Journal of Linguistic Anthropology*, *8*(2), 131–161. https://doi.org/10.1525/jlin.1998.8.2.131

Eilers, R. E., & Oller, D. (1994). Infant vocalizations and the early diagnosis of severe hearing impairment. *The Journal of Pediatrics*, *124*(2), 199–203. https://doi.org/10.1016/S0022-3476(94)70303-5

Eilers, R. E., Oller, D. K., Levine, S., Basinger, D., Lynch, M., & Urbano, R. (1993). The role of prematurity and socioeconomic status in the onset of canonical babbling in infants. *Infant Behavior and Development*, *16*(3), 297–315. https://doi.org/10.1016/0163-6383(93)80037-9

Elo, H. (2016). *Acquiring language as a twin: Twin children's early health, social environment and emerging language skills* (Unpublished doctoral dissertation). Tampere University Press.

Eriksson, M., Marschik, P. B., Tulviste, T., Almgren, M., Pérez Pereira, M., Wehberg, S., Marjanovič-Umek, L., Gayraud, F., Kovacevic, M., & Gallego, C. (2012). Differences between girls and boys in emerging language skills: Evidence from 10 language communities. *British Journal of Developmental Psychology*, *30*(2), 326–343.

Etchell, A., Adhikari, A., Weinberg, L. S., Choo, A. L., Garnett, E. O., Chow, H. M., & Chang, S.-E. (2018). A systematic literature review of sex differences in childhood language and brain development. *Neuropsychologia*, *114*, 19–31.

Fagan, M. (2009). Mean length of utterance before words and grammar: Longitudinal trends and developmental implications of infant vocalizations. *Journal of Child Language*, *36*(3), 495–527. https://doi.org/10.1017/S0305000908009070

Fagan, M. (2015). Why repetition? Repetitive babbling, auditory feedback, and cochlear implantation. *Journal of Experimental Child Psychology*, *137*, 125–136. https://doi.org/10.1016/j.jecp.2015.04.005

Fasolo, M., Majorano, M., & D'Odorico, L. (2008). Babbling and first words in children with slow expressive development. *Clinical Linguistics & Phonetics*, *22*(2), 83–94. https://doi.org/10.1080/02699200701600015

Fernández, D., Harel, D., Ipeirotis, P., & McAllister, T. (2019). June). Statistical considerations for crowdsourced perceptual ratings of human speech productions. *Journal of Applied Statistics*, *46*(8), 1364–1384. https://doi.org/10.1080/02664763.2018.1547692

Field, T. (2010). Touch for socioemotional and physical well-being: A review. *Developmental Review*, *30*(4), 367–383. https://doi.org/10.1016/j.dr.2011.01.001

Frank, M., Braginsky, M., Marchman, V., & Yurovsky, D. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, *44*(3), 677–694. https://doi.org/10.1017/S0305000916000209

Ganek, H., & Eriks-Brophy, A. (2018). Language environment analysis (LENA) system investigation of day long recordings in children: A literature review. *Journal of Communication Disorders*, *72*, 77–85. https://doi.org/10.1016/j.jcomdis.2017.12.005

Gaskins, S. (2006). Cultural perspectives on infant-caregiver interaction. In N. J. Enfield & S. Levinson (Eds.), *Roots of human sociality: Culture, cognition, and interaction* (pp. 279–298). Berg.

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, *19*(5), 515–523.

Gratier, M., & Devouche, E. (2011). Imitation and repetition of prosodic contour in vocal interaction at 3 months. *Developmental Psychology*, *47*(1), 67–76. https://doi.org/10.1037/a0020722

Gros-Louis, J., & Miller, J. L. (2018). From 'ah' to 'bah': social feedback loops for speech sounds at key points of developmental transition. *Journal of Child Language*, *45*(3), 807–825.

Gros-Louis, J., West, M. J., Goldstein, M. H., & King, A. P. (2006). November). Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of Behavioral Development*, *30*(6), 509–516. https://doi.org/10.1177/0165025406071914

Gurven, M., Stieglitz, J., Trumble, B., Blackwell, A. D., Beheim, B., Davis, H., Hooper, P., & Kaplan, H. (2017). The Tsimane' health and life history project: integrating anthropology and biomedicine. *Evolutionary Anthropology: Issues, News, and Reviews*, *26*(2), 54–73.

Ha, S., & Oller, D. K. (2019). Canonical babbling in korean-acquiring infants at 4–9 months of age. *Communication Sciences & Disorders*, *24*(1), 1–8. https://doi.org/10.12963/csd.19577

Hadley, P. A., Rispoli, M., Fitzgerald, C., & Bahnsen, A. (2011). Predictors of morphosyntactic growth in typically developing toddlers: Contributions of parent input and child sex. *Journal of Speech, Language, and Hearing Research*, *54*(2), 549–566. https://doi.org/10.1044/1092-4388(2010/09-0216)

Hantke, S., Eyben, F., Appel, T., & Schuller, B. (2013). *ihearu-play: Introducing a game for crowdsourced data collection for affective computing*. In 2015 International Conference on Affective Computing and Intelligent Interaction (ACII) (pp. 891-897), IEEE.

Harel, D., Hitchcock, E. R., Szeredi, D., Ortiz, J., & McAllister Byun, T. (2017). Finding the experts in the crowd: Validity and reliability of crowd sourced measures of children's gradient speech contrasts. *Clinical Linguistics & Phonetics*, *31*(1), 104–117. https://doi.org/10.3109/02699206.2016.1174306

Hlavac, M. (2018). *Stargazer: Well-formatted regression and summary statistics tables*. Central European Labour Studies Institute (CELSI). https://cran.r-project.org/web/packages/stargazer/stargazer.pdf

Holmgren, K., Lindblom, B., Aurelius, G., Jailing, B., & Zetterström, R. (1986). On the phonetics of infant vocalization. In B. Lindblom & R. Zetterstrom (Eds.), *Precursors of early speech* (pp. 51–63). Palgrave Macmillan.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, *27*(2), 236.

Johnson, K., Caskey, M., Rand, K., Tucker, R., & Vohr, B. (2014). Gender differences in adult-infant communication in the first months of life. *Pediatrics*, *134*(6), e1603–e1610.

Jones, R. M., Plesa Skwerer, D., Pawar, R., Hamo, A., Carberry, C., Ajodan, E. L., Caulley, D., Silverman, M. R., McAdoo, S., Meyer, S., Yoder, A., Clements, M., Lord, C., & Tager-Flusberg, H. (2019). How effective is LENA in detecting speech vocalizations and language produced by children and adolescents with ASD in different contexts? *Autism Research*, *12*(4), 628–635. https://doi.org/10.1002/aur.2071

Jung, J., & Houston, D. (2020). The relationship between the onset of canonical syllables and speech perception skills in children with cochlear implants. *Journal of Speech, Language, and Hearing Research*, *63*(2), 393–404. https://doi.org/10.1044/2019_JSLHR-19-00158

Karasik, L. B., Tamis-LeMonda, C. S., Ossmy, O., & Adolph, K. E. (2018). The ties that bind: Cradling in Tajikistan. *PLoS One*, *13*(10), e0204428.

Klein, R., Lasky, R. E., Yarbrough, C., Habicht, J., & Sellers, M. J. (1977). Relationship of infant/caretaker interaction, social class and nutritional status to developmental test performance among Guatemalan infants. In P. Leiderman (Ed.), *Culture and infancy: Variations in the human experience* (pp. 385–403). Academic Press.

Konner, M. (1977). Infancy among the Kalahari Desert San. In P. Leiderman (Ed.), *Culture and infancy: Variations in the human experience* (pp. 287–328). Academic Press.

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmerTest package: Tests in linear mixed-effects models. *Journal of Statistical Software*, *82*(13), 1–26.

Laing, C., & Bergelson, E. (2020). From babble to words: Infants' early productions match words and objects in their environment. *Cognitive Psychology*, *122*, 101308. https://doi.org/10.1016/j.cogpsych.2020.101308

Lang, S., Bartl-Pokorny, K. D., Pokorny, F. B., Garrido, D., Mani, N., Fox-Boyer, A. V., Zhang, D., & Marschik, P. B. (2019). Canonical babbling: A marker for earlier identification of late detected developmental disorders? *Current Developmental Disorders Reports*, *6*(3), 111–118. https://doi.org/10.1007/s40474-019-00166-w

Lee, C., Jhang, Y., Chen, L., Relyea, G., & Oller, D. K. (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*, 100–126. https://doi.org/10.1080/15475441.2016.1180983

Lee, C., Jhang, Y., Relyea, G., Chen, L., & Oller, D. K. (2018). Babbling development as seen in canonical babbling ratios: A naturalistic evaluation of all-day recordings. *Infant Behavior and Development*, *50*, 140–153. https://doi.org/10.1016/j.infbeh.2017.12.002

Lehet, M., Arjmandi, M. K., Houston, D., & Dilley, L. (2020). Circumspection in using automated measures: Talker gender and addressee affect error rates for adult speech detection in the Language Environment Analysis (LENA) system. *Behavior research methods*. https://doi.org/10.3758/s13428-020-01419-y

Lieven, E. V. M. (1994). Crosslinguistic and crosscultural aspects of language addressed to children. In C. Gallaway & B. J. Richards (Eds.), *Input and interation in language acquisition* (pp. 56–73). Cambridge University Press.

MacNeilage, P. F., & Davis, B. L. (1993). Motor explanations of babbling and early speech patterns. In B. de Boysson-Bardies, S. de Schoenen P. W. Jusczyk, P. F. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 341–352). Springer.

McAllister Byun, T., Harel, D., Halpin, P. F., & Szeredi, D. (2016). Deriving gradient measures of child speech from crowdsourced ratings. *Journal of Communication Disorders*, *64*, 91–102. https://doi.org/10.1016/j.jcomdis.2016.07.001

McCathren, R. B., Yoder, P. J., & Warren, S. F. (1999). The relationship between prelinguistic vocalization and later expressive vocabulary in young children with developmental delay. *Journal of Speech, Language, and Hearing Research*, *42*(4), 915–924. https://doi.org/10.1044/jslhr.4204.915

McDaniel, J., Woynaroski, T., Keceli-Kaysili, B., Watson, L. R., & Yoder, P. (2019). Vocal communication with canonical syllables predicts later expressive language skills in preschool-aged children with autism spectrum disorder. *Journal of Speech, Language, and Hearing Research*, *62*(10), 3826–3833. https://doi.org/10.1044/2019_JSLHR-L-19-0162

McDaniel, J., Yoder, P., Estes, A., & Rogers, S. J. (2020). Predicting expressive language from early vocalizations in young children with autism spectrum disorder: Which vocal measure is best? *Journal of Speech, Language, and Hearing Research*, *63*(5), 1509–1520. https://doi.org/10.1044/2020JSLHR–19–00281.

McGillion, M., Herbert, J., 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

Muysken, P. (2012). Contacts between indigenous languages in South America. In L. Campbell & V. Grondona (Eds.), *The indigenous languages of South America: A comprehensive guide* (pp. 235–258). Walter de Gruyter.

Nathani, S., Oller, D. K., & Neal, A. R. (2007). On the robustness of vocal development: An examination of infants with moderate-to-severe hearing loss and additional risk factors. *Journal of Speech Language and Hearing Research*, *50*, 1425–1444. https://doi.org/10.1044/1092-4388(2007/099)

Nielsen, M., Haun, D., Kartner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology*, *162*, 31–38. https://doi.org/10.1016/j.jecp.2017.04.017

Oller, D. K. (1980). The emergence of the sounds of speech in infancy. In G. Y. Komshian, J. Kavanagh, & C. Ferguson (Eds.), *Child phonology* (Vol. *1*, pp. 93–112). Academic Press.

Oller, D. K. (2000). *The emergence of the speech capacity*. Lawrence Erlbaum Associates.

Oller, D. K., & Eilers, R. E. (1988). The role of audition in infant babbling. *Child Development*, *59*(2), 441–449.

Oller, D. K., Eilers, R. E., Neal, A., & Cobo-Lewis, A. B. (1998). Late onset canonical babbling: A possible early marker of abnormal development. *American Journal on Mental Retardation*, *103*(3), 249–263. https://doi.org/10.1352/0895-8017(1998)103<0249:LOCBAP>2.0.CO;2

Oller, D. K., Eilers, R. E., Neal, A. R., & Schwartz, H. K. (1999). Precursors to speech in infancy: The prediction of speech and language disorders. *Journal of Communication Disorders*, *32*, 223–245.

Oller, D. K., Eilers, R. E., Steffens, M. L., Lynch, M. P., & Urbano, R. (1994). Speech-like vocalizations in infancy: An evaluation of potential risk factors. *Journal of Child Language*, *21*(1), 33–58. https://doi.org/10.1017/S0305000900008667

Oller, D. K., Eilers, R. E., Urbano, R., & Cobo-Lewis, A. B. (1997). Development of precursors to speech in infants exposed to two languages. *Journal of Child Language*, *24*(2), 407–425.

Oller, D. K., Griebel, U., Bowman, D. D., Bene, E., Long, H. L., Yoo, H., & Ramsay, G. (2020). Infant boys are more vocal than infant girls. *Current Biology*, *30*(10), R426–R427. https://doi.org/10.1016/j.cub.2020.03.049

Orena, A. J., Byers-Heinlein, K., & Polka, L. (2019). Reliability of the language environment analysis recording system in analyzing French–English bilingual speech. *Journal of Speech, Language, and Hearing*

Research, *62*(7), 2491–2500. https://doi.org/10.1044/2019_JSLHR-L-18-0342

Patten, E., Belardi, K., Baranek, G. T., Watson, L. R., Labban, J. D., & Oller, D. K. (2014). Vocal patterns in infants with autism spectrum disorder: Canonical babbling status and vocalization frequency. *Journal of Autism and Developmental Disorders*, *44*(10), 2413–2428. https://doi.org/10.1007/s10803-014-2047-4

Pretzer, G. M., Lopez, L. D., Walle, E. A., & Warlaumont, A. S. (2019). Infant-adult vocal interaction dynamics depend on infant vocal type, child-directedness of adult speech, and timeframe. *Infant Behavior and Development*, *57*, 101325. https://doi.org/10.1016/j.infbeh.2019.04.007

Qualtrics. (2019). *Qualtrics Online Survey Software*. https://www.qualtrics.com

Quast, A., Hesse, V., Hain, J., Wermke, P., & Wermke, K. (2016). Baby babbling at five months linked to sex hormone levels in early infancy. *Infant Behavior and Development*, *44*, 1–10. https://doi.org/10.1016/j.infbeh.2016.04.002

Ramirez, N. F., Lytle, S., Fish, M., & Kuhl, P. (2019). Parent coaching at 6 and 10 months improves language outcomes at 14 months: A randomized controlled trial. *Developmental Science*, *22*(3), e12762. https://doi.org/10.1111/desc.12762

Ramírez, N. F., Lytle, S. R., Fish, M., & Kuhl, P. K. (2019). Parent coaching at 6 and 10 months improves language outcomes at 14 months: A randomized controlled trial. *Developmental Science*, *22*(3), e12762. https://doi.org/10.1111/desc.12762

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2014). November). Look who's talking: Speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science*, *17*(6), 880–891. https://doi.org/10.1111/desc.12172

Rogoff, B. (2003). *The cultural nature of human development*. Oxford University Press.

Roug, L., Landberg, I., & Lundberg, L.-J. (1989). Phonetic development in early infancy: a study of four Swedish children during the first eighteen months of life. *Journal of Child Language*, *16*(1), 19–40. https://doi.org/10.1017/S0305000900013416

RStudio Team, & . (2020). *RStudio: Integrated Development for R* (Version 1.2.5033). RStudio, PBC. https://rstudio.com/

Scaff, C., Stieglitz, J., & Cristia, A. (2018). *Daylong recordings from young children learning Tsimane' in Bolivia*. https://nyu.databrary.org/volume/445

Schauwers, K., Gillis, S., Daemers, K., Beukelaer, C. D., & Govaerts, P. (2004). The onset of babbling and the audiological outcome in cochlear implantation between 5 and 20 months of age. *Otology and Neurotology*, *25*, 263–270.

Seidl, A., Tincoff, R., Baker, C., & Cristia, A. (2015). Why the body comes first: Effects of experimenter touch on infants' word finding. *Developmental Science*, *18*(1), 155–164. https://doi.org/10.1111/desc.12182

Seidl, A., Warlaumont, A., & Cristia, A. (2019). Towards detection of canonical babbling by citizen scientists: Performance as a function of clip length. In *Proceedings of interspeech* (pp. 3579–3583).

Semenzin, C., Hamrick, L., Seidl, A., Kelleher, B., & Cristia, A. (2020). Towards large-scale data annotation of audio from wearables: validating zooniverse annotations of infant vocalization types. In *Proceedings of the IEEE spoken language technology workshop*.

Stack, D. M., & Muir, D. W. (1990). Tactile stimulation as a component of social interchange: New interpretations for the still-face effect. *British Journal of Developmental Psychology*, *8*(2), 131–145. https://doi.org/10.1111/j.2044-835X.1990.tb00828.x

Stoel-Gammon, C. (1989). Prespeech and early speech development of two late talkers. *First Language*, *9*(6), 207–223. https://doi.org/10.1177/014272378900900607

Sung, J., Fausto-Sterling, A., Coll, C. G., & Seifer, R. (2013). The dynamics of age and sex in the development of mother–infant vocal communication between 3 and 11 months. *Infancy*, *18*, 1135–1158. https://doi.org/10.1111/infa.12019.

Super, C. (1976). Environmental effects on motor development: The case of 'African infant precocity'. *Developmental Medicine & Child Neurology*, *18*, 561–567.

Tamis-LeMonda, C. S., & Song, L. (2012). Parent–infant communicative interactions in cultural context. In *Handbook of psychology: Developmental psychology* (Vol. *6*, 2nd edn., pp. 143–170). Wiley. https://doi.org/10.1002/9781118133880

VallomparambathPanikkasserySu, R., Pretzer, G. M., Mendoza, S., Shedd, C., Kello, C., Gopinathan, A. T., & Warlaumont, A. S.(2020). A foraging approach to analysing infant and caregiver vocal behaviour. *Scientific Reports*, *10*, 1–14.

van der Stelt, J., & van Beinum, F. K. (1986). The onset of babbling related to gross motor development. In B. Lindblom, & R. Zetterstrom (Eds.), *Precursors of early speech* (pp. 163–173). Palgrave Macmillan.

VanDam, M., Warlaumont, A. S., Bergelson, E., Cristia, A., De Palma, P., & MacWhinney, B. (2016). *Homebank: An online repository of daylong child-centered audio recordings* [Computer software manual]. https://homebank.talkbank.org/

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. https://doi.org/10.2307/414151

Vihman, M. M., Nakai, S., & DePaolis, R. (2006). Getting the rhythm right: A cross-linguistic study of segmental duration in babbling and first words. In L. Goldstein, C. T. Best, & D. H. Whalen (Eds.), *Papers in laboratory phonology viii: Varieties of phonological competence* (pp. 341–366). Cambridge University Press.

Warlaumont, A. S., Pretzer, G. M., Mendoza, S., & Walle, E. A. (2016). *Warlaumont HomeBank corpus*. https://doi.org/10.21415/T54S3C

Warlaumont, A. S., & Ramsdell-Hudock, H. L. (2016). Detection of total syllables and canonical syllables in infant vocalizations. In *Interspeech* (pp. 2676–2680). https://doi.org/10.21437/Interspeech.2016-1518

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.

Whalen, D., Levitt, A. G., & Goldstein, L. M. (2007). Vot in the babbling of french- and english-learning infants. *Journal of Phonetics*, *35*(3), 341–352. https://doi.org/10.1016/j.wocn.2006.10.001

Whitehouse, A. J. O. (2010). Is there a sex ratio difference in the familial aggregation of specific language impairment? A meta-analysis. *Journal of Speech, Language, and Hearing Research*, *53*(4), 1015–1025. https://doi.org/10.1044/1092-4388(2009/09-0078)

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Xu, D., Richards, J. A., & Gilkerson, J. (2014). Automated analysis of child phonetic production using naturalistic recordings. *Journal of Speech, Language, and Hearing Research*, *57*(5), 1638–1650. https://doi.org/10.1044/2014_JSLHR-S-13-0037

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Cychosz M, Cristia A, Bergelson E, et al. Vocal development in a large-scale crosslinguistic corpus. *Dev Sci*. 2021;00:e13090. https://doi.org/10.1111/desc.13090

APPENDIX

TABLE A1 Counts of canonical clips, non-canonical clips, and canonical babbling ratio by individual child: all corpora

Age in months;age in days (corpus child ID)	Canonical	Non-canonical	Total	Proportion
1;75 (Casillas-Yeli F07)	6	145	151	0.04
2;55 (Tseltal 643)	5	120	125	0.04
3;110 (Warlaumont 857)	7	14	21	0.33
3;94 (Warlaumont 340)	6	27	33	0.18
3;95 (Warlaumont 274)	10	21	31	0.32
4;109 (Casillas-Yeli F32)	7	143	150	0.05
4;109 (Tseltal 7176)	17	87	104	0.16
4;133 (Casillas-Yeli F28)	1	51	52	0.02
6;182 (Tseltal 8179)	10	103	113	0.09
7;214 (Seedlings 36)	6	139	145	0.04
7;221 (Seedlings 26)	25	111	136	0.18
7;228 (Tseltal 2109)	6	134	140	0.04
8;231 (Casillas-Yeli F42)	16	127	143	0.11
8;246 (Tsimane′ 14)	24	130	154	0.16
8;256 (Seedlings 4)	12	163	175	0.07
9;277 (Casillas-Yeli F34)	8	182	190	0.04
9;279 (Seedlings 44)	49	138	187	0.26
10;310 (Seedlings 28)	16	89	105	0.15
11;326 (Tseltal 8787)	63	93	156	0.40
12;371 (Seedlings 8)	35	131	166	0.21
13;394 (Casillas-Yeli F23)	10	193	203	0.05
13;402 (Seedlings 14)	43	113	156	0.28
14;433 (Seedlings 11)	41	106	147	0.28
14;435 (Tseltal 7326)	8	29	37	0.22
15;461 (Seedlings 43)	27	233	260	0.10
15;463 (Tsimane′ 41)	41	132	173	0.24
15;464 (Tsimane′ 6)	29	151	180	0.16
16;1050 (Tsimane′ 11b)	58	97	155	0.37
16;407 (Tsimane′ 36)	31	116	147	0.21
16;579 (Tsimane′ 34)	49	95	144	0.34
17;519 (Casillas-Yeli F10)	20	150	170	0.12
17;524 (Seedlings 9)	42	172	214	0.20
18;356 (Tsimane′ 7)	31	114	145	0.21
18;500 (Tsimane′ 33)	68	109	177	0.38
19;591 (Tsimane′ 11)	23	102	125	0.18
20;601 (Tsimane′ 39)	38	215	253	0.15
20;670 (Casillas-Yeli F11)	23	113	136	0.17
22;673 (Tseltal 7220)	57	106	163	0.35
22;733 (Quechua 114)	11	27	38	0.29
23;731 (Tsimane′ 9)	52	88	140	0.37
23;740 (Quechua 105)	95	97	192	0.49
24;566 (Tsimane′ 37)	73	85	158	0.46
24;799 (Tsimane′ 35)	81	99	180	0.45
25;730 (Quechua 117)	14	25	39	0.36

(Continues)

TABLE A1 (Continued)

Age in months;age in days (corpus child ID)	Canonical	Non-canonical	Total	Proportion
26;871 (Casillas-Yeli F31)	77	122	199	0.39
27;815 (Tseltal 6216)	81	75	156	0.52
30;917 (Tsimane′ 10)	21	158	179	0.12
31;975 (Tsimane′ 3)	16	164	180	0.09
32;980 (Tseltal 2625)	74	85	159	0.47
32;991 (Tsimane′ 2)	46	127	173	0.27
36;1094 (Casillas-Yeli F13)	60	150	210	0.29
36;1097 (Tseltal 3026)	150	87	237	0.63

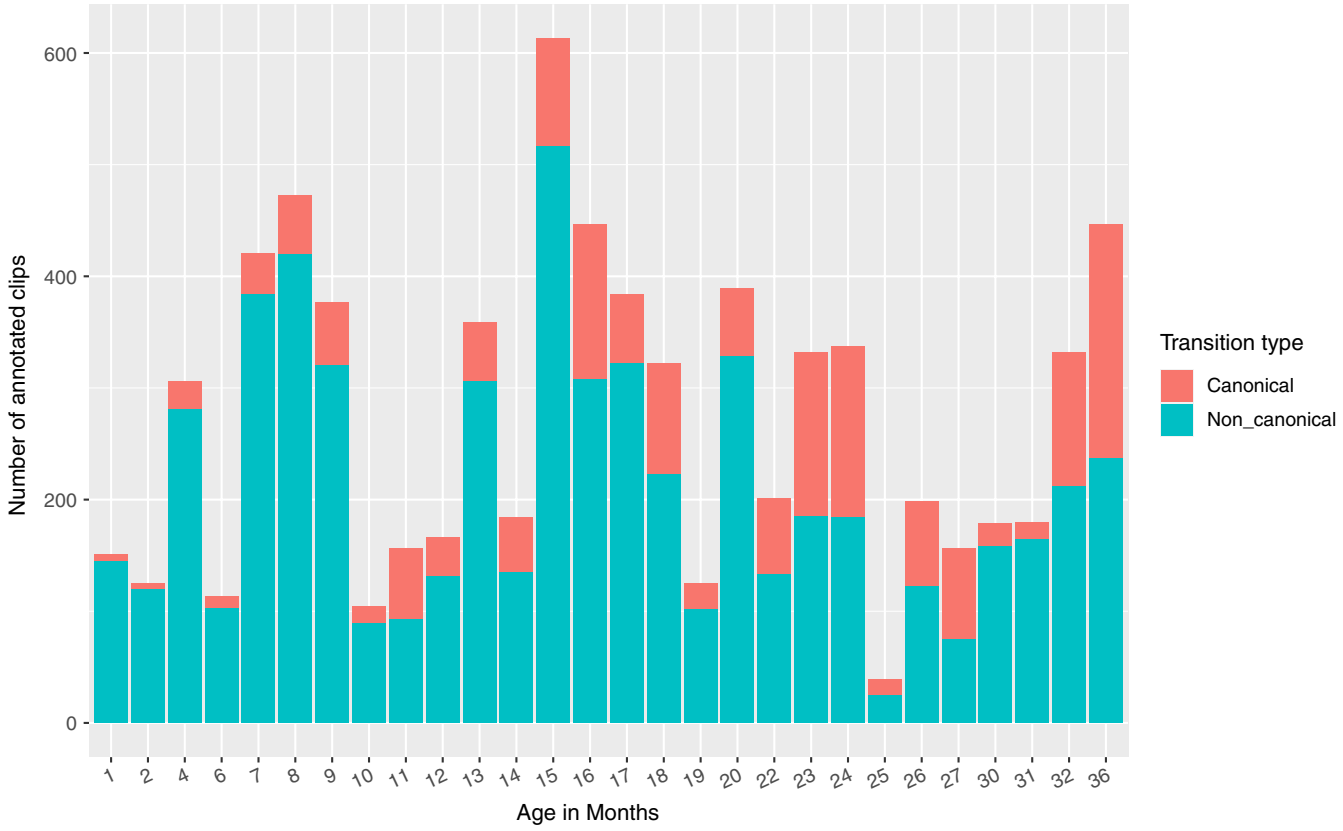

Figure A2 Canonical and non-canonical clips by age (in months)