Child Development, 2026, 97(4), 1233–1254
https://doi.org/10.1093/chidev/aacag046
Published: April 28, 2026
Empirical Article

Child-directed speech and infant vocal development in rural, highland Bolivia and in immigrant families in urban United States

Margaret Cychosz[1,*], Anele Villanueva[2], and Adriana Weisleder[3]

[1]Department of Linguistics, Stanford University, Stanford, CA, United States
[2]Department of Developmental Psychology, Loyola University Chicago, Chicago, IL, United States
[3]Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL, United States
*Corresponding author: Department of Linguistics, Stanford University, Margaret Jacks Hall, Building 460, Stanford, CA 94305, USA. Email: mcychosz@stanford.edu

Abstract

Decades of research have established links between speech input and children's vocabulary growth. However, it is unclear if input also facilitates phonological development. Phonology has a strong biological component—implicating fine motor development—so language socialization factors like speech input may matter less. Further complicating matters, children in many cultures are exposed to vastly different amounts of speech input, and yet these children still reach many major language development milestones on age-appropriate timelines. How does speech input relate to infant phonological development in the first years of life? We estimated infants' (1) speech input and (2) vocal maturity using daylong audio recordings taken in an Indigenous Quechua- and Spanish-speaking community in Bolivia ($n = 10$, *M* age = 12 months, 5 females, 5 males) and an immigrant Spanish- and English-speaking community in the United States ($n = 10$, *M* age = 9 months, 4 females, 6 males; all Hispanic or Latino). Although we found no differences in the *overall* amount of speech input to adults and children between communities, infants in the United States were 2.5× more likely to hear speech directed to them than the infants in Bolivia. When child-directed speech (CDS) was instead characterized to include all speech directed to any child within the infants' vicinity, there were no differences between communities. When employing these different definitions of CDS, we found positive relationships between quantity of speech input and different metrics of the Bolivian infants' vocal maturity. These results paint a nuanced picture showing that directed speech input, even when less common, *is* related to early phonological development, and that by expanding the definition of speech input to accommodate diverse cultural settings we can understand how infants' language development is resilient to differences in speech input.

Keywords language development, phonology, babbling, infancy, cross-cultural, socialization

Lay summary

Babies in rural Bolivia hear far less speech directed to them from caregivers than infants from Spanish-speaking immigrant homes in the United States, yet infants in Bolivia reach the same babbling milestones. When Bolivian caregivers do talk directly to their babies, it has a stronger effect on babbling development than in the United States. This suggests there's no single "right" amount of baby talk; infants adapt to their cultural environment, challenging Western assumptions about optimal ways to interact with infants.

Many traditional theories of child language acquisition emphasize the caregiver's role in language learning (Hoff & Naigles, 2002; Snow, 1972; Tomasello & Farrar, 1986). Support for these theories comes from findings that children who hear more speech input from caregivers develop stronger language skills in a number of domains, especially vocabulary (Ferjan Ramírez et al., 2020; Hoff, 2003; Rowe, 2008; Shneidman et al., 2013) but also language processing (Hurtado et al., 2008; Mahr & Edwards, 2018 ; Weisleder & Fernald, 2013) and morphosyntax (Hoff-Ginsberg, 1986; Huttenlocher et al., 2002, 2010). The majority of this work has focused on the impact of child-directed speech (CDS) for children's later language, a speech register known for its dynamic prosody, simplified vocabulary, and shorter utterances (Kitamura & Burnham, 2003; Soderstrom, 2007). Indeed, CDS characteristics provide important cues for word segmentation and grammatical structure in children's speech input (Bortfeld et al., 2005; de Carvalho et al., 2016; Thiessen et al., 2005). And experimental studies show that infants preferentially attend to stimuli with the

Handling Editor: Adam Winsler
Editor: Adam Winsler
Received: July 22, 2025. **Revised:** February 19, 2026. **Accepted:** March 2, 2026

prosodic characteristics of CDS over adult-directed speech (ADS) (ManyBabies, 2020), suggesting that infants may be more likely to learn from it. Together, these findings provide further support for the idea that speech input, and the CDS register in particular, may facilitate children's language learning (Fernald et al., 2013; Hart & Risley, 1995; Ma et al., 2011). In this paper, we will refer to this caregiver-centered mechanism of language acquisition as Directed Acquisition.

There are two outstanding issues with the Directed Acquisition explanation as a mechanism of child language development. First, many children in the world are exposed to very little directed speech, especially the one-on-one, dyadic interactions consisting of a single caregiver and infant that characterize typical studies measuring impacts of CDS (Cristia, 2023; Cristia et al., 2019; de León, 1998; Dixon et al., 1984; Keller, 2007; Lieven, 1994; Richman et al., 1992; Schieffelin & Ochs, 1986; Shneidman & Goldin-Meadow, 2012). Indeed, the majority of the studies linking CDS and child language outcomes were conducted on a sliver of humanity, namely Western, urbanized families living in single-generation households. In contrast, both psycholinguistic and anthropological reports find that young children in some cultural settings primarily *overhear* speech, for example, between two caregivers. Or children may alternatively hear directed speech, it might just originate from several different older siblings/child caregivers, or they could engage in/observe more multi-party interactions instead of dyadic ones with a single caregiver (Bunce et al., 2025; Casillas et al., 2021; Cristia et al., 2019; Foushee & Srinivasan, 2024; Mastin & Vogt, 2016; Shneidman & Goldin-Meadow, 2012). These are all alternatives to the dyadic verbal interactions between a single (often female) adult caregiver and infant that have been the focus of most prior research. We will broadly refer to these language learning experiences—which are not monolithic and still exhibit large amounts of variability—as Embedded Acquisition.

There is no doubt that typically developing children acquiring language in all of these environments become mature, competent speakers and listeners of their native language(s). Yet the degree to which children can learn from some speech contexts, such as child-directed, better than others remains an open question (Akhtar, 2005; Foushee et al., 2021; Shneidman & Goldin-Meadow, 2012; Shneidman & Woodward, 2016). Moreover, this may be culturally dependent: children exposed to primarily speech between adults, or speech directed to a group of children varying in age, may develop language learning strategies to accommodate those speech registers (Gaskins, 2006; Ochs & Kremer-Sadl, 2020; Scaff et al., 2024; Sperry et al., 2019b).

There are a number of reasons why the amount of CDS differs by community and socio-cultural setting, including family configuration, caregiving tradition, and economic activity (i.e., primarily subsistence agricultural vs. more market-integrated) (Cristia et al., 2023; Frye, 2022; Vogt et al., 2015; Weisner, 1987). Infants from larger families, with a greater number of older siblings are, naturally, going to hear more speech from other children (in any register) than infants with fewer siblings. But differences in the amount of exposure to different speech registers, such as CDS, and how these registers are distributed across different speakers (adults, children) also vary as a function of cultural belief within the community (i.e., if a particular community does not consider pre-verbal infants to be valid conversational partners or believe that infants can understand language before they can produce it (Ochs, 1988; Schieffelin, 1990; Weber et al., 2017), as well as norms surrounding childcare (Weisner, 1987; Weisner & Gallimore, 1977). For example, one study of French-learning toddlers in France and Sesotho-learning toddlers in South Africa showed that the French-learning children heard most of their input from mothers while the Sesotho-learning children heard more of their input from other children (Loukatou et al., 2022). The authors report that the toddlers in Lesotho "reside close to extended families, with grandparents and nieces or nephews typically present," (p. 1178) exposing the children to speech input from a greater number of distinct interlocutors (other children as well as adults) compared to the toddlers in France who were primarily exposed to speech from a single, female caregiver. Quantitative time allocation reports likewise confirm that this alloparental (i.e., parent-external care from aunts, other children, grandparents) and distributed caregiving is the norm in many communities: allomothers provide 40%–50% of caregiving among Mbendjele BaYaka Hunter-Gatherers (Chaudhary et al., 2024), and 60% of Efe 4-month-old infants' contact is with allomothers (Tronick et al., 1987) (see also Scaff et al., 2024). These are not simple dichotomies between more industrialized, Western communities and less-industrialized settings: in Western, industrialized societies such as the United States, children can also differ in their exposure to directed or overheard speech as a function of their living situation (e.g., multi-party or single-caregiver households), whether they attend daycare or are cared for at home, whether or not they have siblings, and other socioeconomic and cultural factors (Larson, Barrett, & McConnell, 2020; Shneidman et al., 2013; Soderstrom et al., 2018-12; Sperry et al., 2019b).

Besides large differences in children's exposure to directed speech input, a second issue with the Directed Acquisition approach is that most studies connecting quantity/content of speech input and child language only focused on children's vocabulary as an outcome (Hart & Risley, 1995; Hoff, 2003; Mastin & Vogt, 2016; Rowe, 2008; Shneidman et al., 2013; but see Huttenlocher et al., 2002, 2010; Luo et al., 2022). Even some studies that examined other domains, such as real time language processing, examined processing with the goal of understanding how it supports lexical development (Hurtado et al., 2008; Weisleder & Fernald, 2013).

But language acquisition is much more than the accumulation of words. There are other areas of language where the connection between caregiver speech input and children's language learning is less clear. In particular, we know little about the relationship between caregiver speech input and children's *phonological* development (Core et al., 2017; Cristia, 2020; Farabolini et al., 2022; Ferjan Ramírez et al., 2019; Parra et al., 2011; Ramírez-Esparza et al., 2014; Singh et al., 2023). On the one hand, we might expect caregiver input to facilitate phonological development as it does vocabulary because phonology and vocabulary are tightly linked throughout development (Beckman & Edwards, 2010; Edwards et al., 2011; Laing & Bergelson, 2020; Majorano et al., 2014; McCune & Vihman, 2001). There is also evidence that social interactions (vocal feedback loops between infant and caregiver) drive some early phonological outcomes like babbling maturity, independent of motor development (Goldstein & Schwade, 2008; Warlaumont et al., 2014). And there is even some experimental evidence that coaching caregivers to provide more CDS leads to increased infant babbling from 6 to 14 months (Ferjan Ramírez et al., 2020). Importantly, Goldstein and Schwade (2008) showed that it was only caregiver speech that immediately *followed*

infants' vocalizations that produced real-time changes in infant vocal production; speech at other times did not result in changes to infants' vocalizations. This finding suggests that socially guided learning in phonological development depends on the temporal contingency of input, not simply its presence or quantity. Overall, these social feedback mechanisms can produce real-time changes in infant vocal production, suggesting that phonological development could be shaped by socially guided learning from caregiver responses, including CDS.

On the other hand, phonology differs from all other levels of language studied because it has a strong biological component: progression through early phonological milestones such as babbling and consonant emergence requires fine motor mastery and coordination of the primary speech articulators (tongue, lips, and jaw). So, language socialization factors, like caregiver input and directed interaction, may matter less for phonological outcomes, especially phonological production. Overall, we know very little about how robust phonological development is to differences in input and language socialization practices such as CDS (Oller et al., 1995; Vanormelingen et al., 2020).

These two outstanding issues with the Directed Acquisition approach—the over-emphasis on children exposed to higher levels of directed speech input and the unclear connection between speech input and phonological development—lead to the current study. In this paper, we evaluate differences in infants' exposure to different types of speech input, and examine how this exposure impacts early precursors to phonological development in two communities representing different language socialization environments: an Indigenous Quechua- and Spanish-speaking community in southern Bolivia and an immigrant Spanish and English-speaking community in the United States.

Overview of child language socialization environments in the speech communities studied

Here we present qualitative observations of the socialization environments in the two communities studied, paying attention to how these socialization environments could influence the types of speech input typically examined in language acquisition research. While we describe the two communities separately, we in no way attempt to set up a dichotomy between them. Instead, we describe their practices of caregiving, as it relates to child language socialization.

Prior to walking in the Bolivian community, infants spend the majority of their sleeping and waking hours on their mother's back, facing forward in a cloth swaddle (*q'epi*) that wraps around the mother's back and shoulders. Thus, the infant spends the majority of their waking hours positioned nearly at eye level with the caregiver, and with the same vantage point. The infant infrequently engages in face-to-face contact with the primary caregiver, but they are engaged in near-constant physical contact with her (de León, 1998; Keller et al., 2009). In terms of speech input, when an infant is positioned on the mother's back in the *q'epi*, speech is infrequently directed to them unless the caregiver is attempting to soothe or feed. The infant would sometimes hear directed speech if the caregiver shifts the infant forward in the *q'epi* for nursing. Infants in the *q'epi* undoubtedly can overhear speech input from a large number of distinct speakers, including children but also multiple adult females as the mother conducts housework, engages in agricultural work or weaving, and/or prepares food in a communal, primarily female, setting throughout the day. This positioning may limit opportunities for contingent vocal exchanges and turn-taking between the caregiver and the infant, as the caregiver cannot easily see the infant's facial expressions or hear subtle vocalizations that might elicit a vocal response (Keller et al., 2008). In contrast, the constant physical contact and shared vantage point has been shown to support other forms of social learning and communication in rural societies with similar modes of infant caretaking (Carra et al., 2014). Nevertheless, moments when directed speech to the infant *does* occur, for example, during nursing, soothing, or in preparation for (co-)sleeping, may be when the caregiver is most responsive to the infant's vocalizations. In other words, if Bolivian infants receive less overall directed speech, as we hypothesize, the directed speech they *do* receive may be more concentrated in exactly those contexts that appear to bolster vocal development in in-lab settings as examined in previous work in North America (Goldstein & Schwade, 2008).

Bolivian families tend to have a larger number of siblings (0–7 in the current sample) than those in the United States community studied in this work, and young children in the Bolivian sample are nearly as likely to socialize with neighboring children as their own siblings. There would thus be a larger number of distinct speakers in the infant's everyday speech input, and the infant would overhear them engaging in multi-party ADS and CDS (including CDS from adult females to groups of children within the vicinity).

Once infants walk independently, after approximately 12 months, they are still swaddled on the mother's back for extended periods, for example, during transit as the mother walks to home or market which can last several hours throughout the day; qualitative observations do not suggest that children receive more directed speech input in the *q'epi* with age. Otherwise, once the infants are semi-autonomous and walking, directed speech could also begin to originate progressively more from older children (e.g., neighboring children in the village or siblings).

The United States community consists of Spanish-speaking, first-generation immigrants from Latin America that have settled in a large, United States city. As such, infants' environments are influenced both by the socialization practices of their communities of origin, the realities of urban contexts in the United States, and the context of immigration more specifically (Baquedano-López & Mangual Figueroa, 2011; García Coll et al., 1996; Ramírez-Esparza et al., 2017; Zentella, 1998). Infants spend most of their time indoors in various mid-size, non-reverberant spaces. When the infants are outdoors with their caregivers, they are surrounded by noise from an urban, industrial lifestyle, including public transportation and cars. When in transit, the infants are variably placed facing forward or backward in a stroller, or are held in a caregiver's arms; thus, these infants have more face-to-face, dyadic experience with their primary caregiver, but typically rest further away from, and experience less physical contact with, the primary caregiver(s) than the infants in Bolivia.

Although most of the infants in the United States community have more than one caregiver, including their mother, father, older siblings, and/or grandparents, the majority are taken care of primarily by the mother. The United States infants have relatively few siblings (0–3 in the current sample) and, especially in the first year of life, very few of them attend daycare due to the high cost of private childcare and limited availability of publicly-funded

childcare in the United States (Allen & Hutton, 2023). Thus, most of them are cared for in the home and have limited opportunities to interact with, and overhear speech directed to, other children. As such, when it occurs, overheard speech would primarily originate from ADS in adult-to-adult interactions, and CDS is commonly directed to the infant and less frequently to another child (since there are fewer children around the infant). In addition, because these are immigrant families, most of them experience separation from their extended families in Latin America, which may interrupt more communal forms of caregiving and opportunities to encounter overheard speech.

A previous study characterizing the interaction contexts of Spanish-English bilingual infants in the United States found that the most common context for directed speech consisted of the mother talking to the infant and the most common context for overheard speech consisted of speech between adults (Ramírez-Esparza et al., 2017). Yet this study also found that, compared to monolingual English-speaking infants in the United States, infants from Spanish-English bilingual families spent significantly less time in one-on-one, dyadic interactions and significantly more time in group interactions. As these descriptions illustrate, infants in this immigrant United States community are likely to experience caregiving environments that are significantly more westernized than those of the Bolivian infants, even as their environments also differ from those of white, middle-class infants in the United States.

Current study

In this study, we use infant-centered, longform audio recordings to examine differences in infants' exposure to various forms of speech input, and its potential impact on early phonological development, in a cohort of infants (<24 months) residing in the rural highlands of Bolivia and the urban United States. Critically, these two communities have different language socialization environments that are hypothesized to impact amounts of speech directed to the infants. In this study, we first quantify differences by community in infants' exposure to *any* speech. Then, we evaluate differences by community in infants' exposure to various speech contexts, distinguishing between speech directed to *any* child, speech directed to the *target* child, and speech directed to an adult. Finally, we relate these differences in speech input to relatively culturally- and linguistically-neutral metrics of phonological development applicable to infants of this age range: infants' speech volubility and canonical babbling ratios (CBRs). We ask the following questions:

1 Are there differences between communities in the total amount of infants' daily speech exposure?

For this question, we take a *broad* definition of speech input, encompassing any speech within the infant's vicinity, including ADS, speech directed to any child including the target child (CDS), and speech directed *only* to the target child ("target child-directed speech," or TCDS). With this broad definition of speech, we anticipate that we will *not* find differences in overall amounts of speech input to infants between communities.

2. How does the daily amount of various types of speech input (CDS, TCDS, and ADS) differ between the two communities (Bolivia and the United States)?

For this research question, we differentiate between *broad* CDS (speech to all children including the target child) versus *narrow* CDS (only speech directed to the target child or TCDS). Given differences in caregiving environments, we predict that infants in the Bolivian community will be exposed to less TCDS than infants in the United States. However, given that infants in the Bolivian community are consistently surrounded by more children, when we instantiate CDS to include speech input to any child in the target infant's vicinity, we predict that infants in the Bolivian community will be exposed to more CDS than infants in the United States. Critically, most previous work on this topic has not taken this broad view of CDS, and generally limits the study of CDS to caregiver speech input to the target child, not input directed to other children as well (cf. Scaff et al., 2024).

3. Does speech input quantity predict infant vocal development in the United States and/or Bolivian communities?

As outlined in the introduction, it is unclear if speech input facilitates children's phonological development as it does vocabulary development (Cristia, 2020), and the extent to which different types of speech input are more or less influential. Furthermore, the mechanisms by which speech input might relate to infants' vocal development remain unclear: effects could stem from overall exposure to speech but also from opportunities for turn-taking or contingent vocal exchanges. To evaluate this question, we take the estimates of infant speech exposure from research questions 1 and 2, and relate it to two measures of infant vocal maturity: infants' speech volubility and CBR (Oller, 2000). We measure speech volubility as the number of speech-like vocalizations that the infants produced within each full hour of a daylong audio recording. The CBR refers to the ratio of canonical, well-formed syllables consisting of a consonant and vowel (e.g. CV, VC) produced in quick succession, to all of an infant's vocalizations, and is one of the most common metrics of early infant vocal development between 5 and 24 months (Cychosz & Long, 2025; Long et al., 2022).

We examine these two outcomes, speech volubility and CBR, as our metrics of vocal development for a number of reasons. First, our metrics of phonological development needed to be robust across the age range examined (6–23 months), ruling out certain babbling landmarks such as a predetermined CBR threshold or variegated babbling, which is not typically reached until 10–12 months (Oller et al., 1999) (age estimates for the emergence of these landmarks is also based on primarily Western infants acquiring a narrow sample of languages). Second, the metrics of vocal development needed to be robust to differences in the phonological structure of the ambient language since we were examining infants exposed to different languages (Quechua+Spanish in Bolivia and Spanish+English in the United States). Even seemingly "universal" aspects of phonological development, such as the onset of bilabial consonants, are malleable and subject to frequency differences in the ambient language (de Boysson-Bardies et al., 1991). Finally, while all infants vocalize from birth, *speech-like* vocal volubility increases in frequency starting from approximately 3 months (Nathani et al., 2006). And, most critically, its emergence does not follow a universal trajectory, but rather shows variability as a function of biological and environmental conditions: infants with various communicative disorders (e.g., hearing loss, Angelman's syndrome, those receiving later autism spectrum

diagnoses) produce fewer speech-like vocalizations across the first year of life (Bergelson et al., 2023; Hamrick et al., 2023; Warlaumont et al., 2014), as do children from lower socioeconomic status homes in Western contexts (Gilkerson et al., 2017; Warlaumont et al., 2014). Thus, both speech volubility and CBR are sufficiently robust measures across the age range examined, are relatively immune to ambient language structure, and could feasibly be influenced by socialization factors such as directed speech.

A note on overheard versus adult-directed speech

The category *overheard speech* is often employed in studies testing the robustness of Directed Acquisition approaches to child language development (Akhtar, 2005; Shneidman & Goldin-Meadow, 2012; Weisleder & Fernald, 2013). For example, among Yucatec Mayan infants, who are exposed to low amounts of directed speech, Shneidman and Goldin-Meadow (2012) found that CDS—but *not* overheard speech—predicted infants' later vocabulary development despite the infants hearing significantly more overheard speech in their input.

However overheard speech is not a particular speech register with identifiable acoustic or lexical qualities. The speaker and addressee are not even easily defined: from an infant's perspective, overheard speech could (and does) encompass speech between two adults, but also speech from a child to one or more adults, speech between multiple interlocutors having a barely-audible conversation in the next room, etc. These are all completely different speech "registers" and communicative contexts with different acoustic and linguistic properties, and different speaker–addressee relationships. Foushee and Srinivasan (2024) suggest that "such heterogeneity could help explain why previous studies have failed to detect correlations between the quantity of other-directed speech in children's environments and their subsequent vocabulary growth" (p. 6). Consequently, in this work, we limit our analysis to ADS, spoken between two adult interlocutors. Note that this narrow definition of ADS does differ from some other work (Bunce et al. 2025 who examined ADS spoken by other adults as well as children). However, our goal was to quantify ADS amounts to understand its potential as a language learning medium, and we anecdotally observed that speech from a young child to an adult was completely different in style (speech rate, volume, acoustic modifications such as hypoarticulation; Lindblom, 1990) and content (word choice, complexity of grammatical structure) than speech from an adult to another adult. We thus employ this narrow version of ADS (speech from one adult to one or more other adults) in this work.

Method

Participants

Infants in this work were from two different speech communities: an indigenous Quechua- and Spanish-speaking community in southern Bolivia (Cychosz, 2018), and a Spanish- and English-speaking immigrant community in the United States (Weisleder & Mendelsohn, 2019). This research was approved by the institutional review boards at the authors' institutions at the time of data collection.

Bolivian sample Families with infants and small children were recruited through the first author's personal contacts in communities surrounding a mid-size town in southern Bolivia. Approximately 100 children, age 6 months to 8 years, were enrolled in a study on children's language development and socialization. Children were eligible for inclusion in the study if they were (1) typically developing, per caregiver report (this community is medically underserved so some speech and language delays may go undiagnosed or treated) and (2) spoken to at least some of the time in Quechua at home by a primary caregiver. $N = 10$ infants (<24 months) were successfully recruited and matched to the United States infant sample as closely as possible by age, infant gender, and number of older siblings (see Table S2 in Online Supplementary Materials for detail). Caregivers of all infants in this study spoke Quechua as a first language, and some additionally spoke Spanish (see Table S2 in Online Supplementary Materials for information on maternal language dominance).

United States sample Families were participating in a larger study of a pediatric-based early childhood program (Canfield et al., 2020). Participant recruitment took place in the pediatric clinic of a public hospital in a major United States city. The clinic, which is in a neighborhood that scores high on the Area Deprivation Index and Social Vulnerability Index, provides primary and specialty care for children and adults regardless of medical insurance status or ability to pay, and serves primarily low-income families. Caregivers were approached in the pediatric clinic waiting room and asked if they were interested in participating in a program aiming to support early child development via promotion of positive parenting during routine pediatric visits (Mendelsohn et al., 2013). Eligibility for the larger study included having a child between 5 days to 3 years of age and being able to communicate in English and/or Spanish.

For the current work, a subset of participants from the larger study were invited to participate in a longitudinal study of infants' language development. Inclusion criteria were families who used Spanish as the primary language in the home, full-term infants (at least 37 weeks gestational age; 2,500g) with no known neuro-developmental disorders or genetic conditions, and mothers who were the primary caregiver. Twenty-six families enrolled in the study. Of those, 10 families were selected for the current analyses to correspond to the matching criteria to the Bolivian sample (infant age, gender, and number of older siblings). All United States sample caregivers identified as Hispanic and/or Latino and were first-generation immigrants from Latin America ($N = 2$ Mexico, $N = 3$ Ecuador, $N = 3$ the Dominican Republic, $N = 1$ El Salvador, $N = 1$ Honduras). All mothers except one reported the fathers' country of birth ($N = 3$ Mexico, $N = 3$ the Dominican Republic, $N = 1$ Ecuador, $N = 1$ El Salvador, $N = 1$ Honduras). At the time of enrollment, none of the infants were attending daycare and the mother planned to be the primary caregiver. Because some of the recordings were conducted several months after this information was collected, we cannot be certain whether some infants may have had different childcare arrangements at the specific timepoint of recording. However, given the limited access to childcare in this community (see above) and the families' plans when they enrolled in the study, we expect most of the infants were cared for in the home (and all the recordings were made when the infant was in the primary caregiver's care).

Matching participants

We attempted to match the participants across the two communities by age (Bolivia M_{age} = 12.5 months, range: 5.7–23.4, United States M_{age} = 9.1 months, range: 6.4–12.6), gender (Bolivia, $N_{females}$ = 5, N_{males} = 5; United States, $N_{females}$ = 4 and $_{males}$ = 6), and, to the extent possible given differences in family size by community, number of siblings (Table S1 in Online Supplementary Materials). We did not attempt to match participants by maternal education, a common proxy for socioeconomic status in language development research (Hoff, 2003, because of differences in educational opportunities for women in these two communities. Still, most mothers from both communities had less than a high school education. We acknowledge that we were unable to successfully match the infants by age, in particular, across sites. Two of the Bolivian infants were >12 months of age, but all of the United States infants were <12 months. In the Online Supplementary Materials, we successfully replicate our results removing these two older Bolivian infants. Nevertheless, this is a limitation of the current work that we hope future work can improve upon.

Data from longform audio recordings of these infants were reported in Cychosz et al. (2021) and Cychosz et al., 2025b; work reported here builds upon that work, but the work presented here is completely novel and non-overlapping with previous reports. Cychosz et al. 2021 was a primarily methodological paper establishing that random sampling of daylong audio recordings of children's language environments resulted in stable estimates of parameters such as dual language use and speech registers. Unlike the current paper where we do not de-aggregate by infants' language of exposure, Cychosz et al. (2025b) reported how the interactions between speaker and interlocutor intersected in these infants' dual language environments. For example, using random sampling from the infants' recordings, we established what proportion of each infant's input target CDS was from the minoritized language (Quechua for Bolivia and Spanish for the United States) versus the majority language (Spanish for Bolivia and English for the United States).

Collecting daylong audio recordings

Measures of the infants' speech input and vocal production were taken using infant-centered, daylong audio recordings. For the procedure, each infant wore a small, lightweight (2″ × 3″; 2 oz.) Language ENvironment Analysis (LENA) audio recorder in shirt/vest with a special pocket sewn on the front to house the recorder (Xu et al., 2009). Caregivers in both communities were given instructions as to how to turn the recorder on and off, as well as the device's recording radius (approximately 10 feet). Families were asked to record for an entire day, aiming for 16 total hours (the device's battery life). The average duration of the recordings for the Bolivia sample was 11.88 hr (*SD*: 3.04; range: 7.77–16). Recordings from the United States sample were all 16 hr.

Recordings from the United States sample were taken at four time points: 2, 6, 9, and 12 months after study enrollment. Recordings from the Bolivia sample were collected at a single time point (study enrollment). United States sample caregivers also filled out a logbook over the course of the day to note where the infant was and with whom, and the activities they were engaging in. Caregivers in Bolivia were not asked to complete a written log, given variable levels of access to education and literacy between households.

Daylong recording annotation

The LENA system can estimate some important linguistic elements of a child's daily speech environment, such as conversational turns between an adult and the key child. However, the system cannot reliably distinguish between different speech registers, including CDS and ADS. Additionally, since the proprietary LENA algorithm was trained on child-centered *English* audio data, its performance on many of its linguistic counts, such as its Adult Word Count, are unknown for languages such as Quechua. Consequently, we devised a manual annotation procedure to derive the estimates of infants' speech exposure.

We first divided each daylong audio recording into 30-s segments. A vocal activity detector was run over each segment; low-vocal segments were candidates for nap periods. We then manually listened to each candidate nap period and removed those segments where the infant was sleeping from analysis. We then annotated the remaining clips for speaker (adult, other child, unsure), addressee (target child, adult, other child, unsure), and language (Spanish, Quechua/English, Mixed, no speech, unsure). For the purposes of annotation, an adult was any speaker who had gone through puberty (as indicated in their voice but also language and/or word use), including young teenagers. A child would thus be any speaker who had not yet entered puberty, again determined primarily by vocal characteristics. We do not report further on the bilingual language exposure measures in this work but refer readers to Cychosz et al. (2021).

Using these annotations, we then categorized the speech register(s) present within each 30-s segment: speech from an adult or other child to *any other* child, including the target infant, was classified as CDS. Speech from an adult or another child to the *target* child only was classified as TCDS. Speech spoken by an adult to another adult was classified as ADS. CDS could have some canonical characteristics of the CDS register, including higher-pitched voices and sing-song-like intonation, simplified speech, and use of diminutives (e.g., "doggy"/"perrito" instead of "dog"/"perro"). However, it was not required to have these characteristics, and could also include "typical-sounding" speech as long as there were other contextual cues indicating the speech was directed to one or more children. Vocal markers of ADS included typical-sounding voices, monotone voices, regular speech (minimal pitch modification), as well as other contextual cues that speech was directed to an adult, such as content and topic matter.

We annotated the recordings in two ways: (1) randomly sampling 30-s segments from each recording and (2) annotating every other 30-s from the entire recording. Random sampling was conducted for all 20 infants; all-day sampling was conducted for 10 infants (5 infants/community), still matched for age, gender, and sibling number. Previous work has shown that randomly sampling 30-s samples of speech from infants' daylong audio recordings provide reliable estimates of speech register exposure (Cychosz et al., 2021). We annotated an entire day of recording for five children/sample because we anecdotally observed low rates of TCDS in the Bolivia sample and wanted to ensure that we were accurately reporting on that register.

For random sampling annotation, we sampled 30-s clips from each recording to estimate the proportion of each speech register

that the infant was exposed to and stopped annotation once the estimated proportion stabilized (see Cychosz et al. [2021] for methodological detail). In all, an average of 245.2 clips/child ($SD = 140$, range = 103–701) were annotated via random sampling. For all-day annotation, we annotated every other 30-s clip from the entire recording. $N = 960$ clips were annotated from each infant in the United States and an average of 609.4 clips/child ($SD = 201.1$, range = 465–955) were annotated from the Bolivian infants.

Recordings from the Bolivia sample were annotated by the first author and those from the United States sample were annotated by the second author. The first author is fluent in Spanish, and understands Quechua, which they learned while conducting fieldwork in southern Bolivia. This author had personally collected all of the data and knew each family, which facilitated the identification of individual voices and ages (e.g., adult versus child). The second author is a fluent Spanish-English bilingual, and was thus able to use contextual cues to identify speaker and addressee. The second author also employed the logbooks (Section "Collecting daylong audio recordings") to facilitate annotation and understand when other speakers were present in the recordings. Overall, the annotators' familiarity with the languages and, in the first annotator's case, the families, allowed them to discern adult- versus (target) CDS by content (e.g., a caregiver is telling a child something), context (e.g., if a child was speaking and a caregiver responds in quick succession), and acoustics (e.g., dynamic prosody).

The team took a stringent approach to data annotation to ensure consistency between samples and annotators. First, the team constructed a strict coding protocol that outlined specific speaking situations that the annotators would face and how to approach each situation to ensure consistency between annotation of samples (e.g., unclear if a speaker should be classified as an adult or child). This written coding protocol is available in the OSF repository associated with this work. During the annotation period, the team held weekly meetings where they would discuss uncommon or difficult situations encountered during the annotation procedure to ensure consistency throughout the entire annotation period.

For the United States sample, a second bilingual Spanish-English speaker re-annotated a subset of recordings to assess inter-rater reliability and thus consistency of the annotation protocol within that sample. The second annotator first trained on two recordings, including the use of the written coding scheme and logbooks. During the training phase, inter-rater agreement for the type of speaker and speech register categories ranged from 72% to 100% depending on the category, ($M = 90.92\%$). All disagreements were discussed among coders. Upon completion of the training phase, the second annotator classified a subset of clips from a new infant recording ($N = 205$). Inter-rater reliability (% agreement) for speaker and speech register was (1) adult–child 98.03%, (2) adult–adult 96.08%, (3) child–child 93.14%, and (4) child–adult 91.18%. We also computed Cohen's κ to determine inter-rater agreement for the type of speaker and speech register: $\kappa = 0.90$ for adult–adult speech, $\kappa = 0.95$ for adult–child speech, $\kappa = 0.82$ for child–adult speech, and $\kappa = 0.82$ for child–child speech. For the Bolivian sample, intra-rater reliability (between the first author) was conducted for 20% of the original data (two infants' recordings, or $N = 202$ clips), 2–4 years following the initial annotation (the annotation scheme requires language fluency to understand context and addressee, so inter-rater reliability would have had to be conducted in-person in Bolivia but data annotation was conducted during the COVID-19 pandemic). Intra-rater reliability (% agreement) for speaker and speech register was (1) adult–child 95.22%, (2) adult–adult 94.74%, (3) child–child 99.04%, and (4) child–adult 98.09%. Cohen's kappa to determine intra-rater agreement for the type of speaker and speech register was $\kappa = 0.89$ for adult–adult speech, $\kappa = 0.87$ for adult–child speech, $\kappa = 0.82$ for child–adult speech, and $\kappa = 0.85$ for child–child speech. These results suggest that the speaker and addressee categories were well-defined and that different coders could employ them across different recordings.

It is important to note that this annotation procedure captured the presence or absence of different speech registers, from different speakers, within 30-s segments. In this work, we did not code for temporal vocal contingency between caregivers and their infants. Consequently, we do not distinguish between directed speech that immediately follows (or precedes) an infant's vocalization versus speech that is used at other times (the literature typically defines a contingent response as one happening within 2 s). Therefore, two infants could have similar amounts of TCDS in their daily environments, but different amounts of contingent speech. We leave analyses at this level to future research.

Computing speech exposure

For each infant, we computed the proportion of the day that they were exposed to a given speech register (after removing periods of sleep; see Section "Daylong recording annotation"). So, the denominator in our calculations is the total number of annotations made, including times with no speech input, and the numerator is the number of 30-s segments that contained the particular speech register. Because the denominator was the total number of annotations, if a 30-s segment contained multiple examples of the same speech register, it was still only counted once. Note that this denominator calculation differs just slightly from previous work estimating children's exposure to various speech registers where the denominator was the total number of annotations that contained any speech input (Cychosz et al., 2021). Nevertheless, we wanted to ensure that estimates of random and all-day sampling were still robust when the denominator is the total number of annotations. In online Supplementary Material I, we perform correlations between the random and all-day estimates of speech register exposure for the $N = 10$ infants whose entire recording was annotated and replicate our earlier results. This result demonstrates that this slightly revised method still produces reliable estimates of the proportion of each infant's exposure to various speech registers.

Computing infant vocal maturity

Computing speech volubility

To compute infant speech volubility, we used LENA's proprietary speaker diarization algorithm to identify vocalizations from the target infant in each daylong recording. After removing periods of sleep (which had been manually verified), we counted the number of total vocalizations from the target infant for each waking hour of their recording (removing hours <60 min at the beginning/end of the recording so as not to undercount vocalizations from incomplete hours). This was our measure of speech

volubility, and it allowed us to sample repeatedly from the same infant instead of a single sample of, for example, the average number of vocalizations per hour. We account for the nested structure of the data (repeated measures within infants) in the statistical modeling.

The confusion rates in the LENA system between target infant speech and speech from another child in the infant's vicinity are 4% (precision, or the accuracy of positively predicting the target infant) and 7% (recall, or the accuracy of negatively predicting the target infant). Consequently, we considered that we could be overestimating the infants' volubility, perhaps especially so for the infants with more siblings and/or the Bolivian infants who we anecdotally observed during fieldwork and manual data annotation to be surrounded by relatively more speech from other children. We report on this possibility, and the confusion rate between the target child and other children, by community, in Results.

Computing canonical babbling ratio

To compute each child's CBR, we again employed the LENA speaker diarization algorithm as a first-pass to identify vocalizations from the key child. We did not draw from either high-volubility or high-interaction periods of the recording because we wanted to sample from each infant's total vocal repertoire without introducing bias into the sampling method. We then removed all vocalizations algorithmically-identified as containing cry because, although the line between speech-like and cry can be continuous, we were interested in vocalizations that were primarily speech-like and our method of estimating CBR only required obtaining a sub-sample of the infants' speech-like vocalizations.

Next, employing a graphical user interface developed for this coding scheme, we randomly sampled the audio clips identified by LENA as having key child vocalizations and manually annotated them for Speaker (key child, other child, adult female, adult male, unsure, overlap, or no speech). In practice, the "unsure" category was used in the event that the speech produced in the clip was too short to reliably identify the speaker and the "overlap" category was used if two speakers were overlapping such that the coder could not make out their identities; if the key child was overlapping with another speaker, and the coder could identify the key child, then the child's vocalizations progressed for additional annotation. If not, the clip was marked as "overlap" and not employed for further analysis.

If the key child was identified in the clip, we next annotated the clip for Vocalization Type (speech-like, cry, laugh, raspberry/gurgling, creak, squealing, or vegetative [i.e., grunting or eating]). We closely followed previous work in the measurement of CBR and considered speech-like to contain a fully resonant, supraglottal vowel/vocalization and/or consonant-vowel transition (Lee et al., 2018; Morgan & Wren, 2018). A given utterance could contain multiple types of vocalizations (e.g., speech-like and squeal); we follow Nathani et al. (2006) in classifying the utterances according to the highest vocal level achieved in the vocalization. So, the squeal+speech-like vocalization would be classified as "speech-like." Finally, for all vocalizations coded as "speech-like," we classified whether or not the vocalization contained any canonical syllables (a C-V or V-C sequence), and we further noted the number of CV/VC transitions in the vocalization. Following most work in this domain, glide-vowel sequences were not considered canonical syllables but laterals, rhotics, fricatives (though rare), plosives, and nasals were (Cychosz & Long, 2025; Morgan & Wren, 2018). To compute the CBR, we took the number of canonical vocalizations (where each canonical vocalization could only be counted once, regardless of how many CV/VC transitions it contained), and divided it by the total number of speech-like vocalizations that the infant produced (excluding cry, vegetative sounds, etc.).

Following Molemans et al. (2012), who systematically analyzed the number of vocalizations required to reliably estimate CBR in infancy, we annotated until we had sampled $N = 100$ speech-like vocalizations from each infant, and then calculated the CBR. The only exception was for 1 Bolivian infant for whom we annotated all available speech-like vocalizations in the recording ($N = 45$). In total, we hand annotated $N = 3{,}734$ vocalizations that the algorithm had identified as the key child ($M = 186.7$ vocalizations/infant, $SD = 45.54$, range = 120–286).

Then, to obtain robust estimates of each infant's CBR, and account for potential variability in vocalization sampling, we employed bootstrap sampling. For each infant, we generated 100 CBR estimates by repeatedly sampling 90% of their speech-like vocalizations with replacement. This bootstrapping approach provided distribution information about the stability of the CBR measure while maintaining the nested structure of the data in our statistical models.

Finally, because our dataset comprised infants of different ages, we normalized (z-score) each infant's CBR using age-normed (population-level) values reported in Cychosz and Long (2025) so that we could evaluate the relationship between speech input and CBR in our sample that spanned an age range across the first two years of life.

The first author, trained in phonetic and phonological analysis of infant and child speech, hand-annotated all of the infants' data. A second annotator, trained in phonetics and child speech development but who was unfamiliar with the research project and questions, hand-annotated 20% of the original data (all of the vocalizations annotated by the first author from $N = 2$ infants from Bolivia and $N = 2$ from the United States). The stages of vocal development are continuous, with transitions between milestones such as marginal to full babbling happening gradually. Consequently, both annotators were blind to infant age during annotation (beyond being aware that the sample spanned 6–23 months) so that they would not be biased to classify the speech of older infants as more speech-like. To ensure fidelity to the coding protocol, we conducted inter-rater reliability using Cohen's kappa: for speaker it was $\kappa = .84$, for vocalization type (canonical versus non-canonical) it was $\kappa = .82$.

Results

Research Questions 1 and 2: Are there differences between communities in the total amount of infants' speech exposure? How does the daily amount of various types of speech input (CDS, TCDS, and ADS) differ between the two communities (Bolivia and the United States)?

Descriptive statistics

For these analyses, we used the data from the $N = 10$ infants with all-day annotation (5 infants/community, every other 30-s

Table 1 Proportion of daily speech input, by type of speech input and community.

	Bolivia *M (SD)* range	U.S. *M (SD)* range
Any speech	0.74 (0.09) 0.64–0.85	0.68 (0.14) 0.50–0.83
Speech directed to any child (CDS)	0.40 (0.08) 0.35–0.55	0.48 (0.15) 0.32–0.65
Speech directed to the target child (TCDS)	0.12 (0.06) 0.05–0.2	0.39 (0.08) 0.30–0.53
Adult-directed speech	0.30 (0.14) 0.12–0.46	0.16 (0.12) 0–0.3

Note. Here the numerator for each computed proportion reflects the number of 30-second segments containing the speech register, and the denominator is the total number of annotations made (which includes segments without speech input but excludes when the infant was sleeping). Proportions are derived from the 10 infants with all-day annotation.

segment annotated) because we wanted to ensure we had sufficient data to reliably estimate TCDS in both samples even if it was infrequent, requiring us to look over the course of the entire day.

We first present descriptive statistics and visualizations that reflect infants' daily exposure to various types of speech input. Table 1 presents the proportion of the types of speech present in the infants' daily environments, by community. Infants in the United States were exposed to slightly less speech overall—an average proportion of 0.68 ($SD = 0.14$) of the entire day (here the proportion number refers to the proportion of 30-s segments that contained speech) excluding sleep compared to 0.74 (0.09) for the Bolivian infants. However, infants in the United States were exposed to more TCDS. In fact, the range of TCDS levels in the United States (avg. = 0.39 (0.08) 0.3–0.53 proportion of 30-s clips over the entire day) did not even overlap with the values for the Bolivian infants (avg. = 0.12 (0.06) 0.05–0.2 proportion of 30-s clips over the entire day).

Figures 1 and 2 present the data further de-aggregated by infant, showing distributions of speech input over the course of the entire day: Figure 1 presents the *total* speech input that each child was exposed to while Figure 2 presents the data split by types of speech input: CDS, TCDS, and ADS. In these figures, denser concentrations signify that the particular 5-min epoch was full of the given speech register.

Modeling exposure to types of speech input by community

To model the impact of community upon the presence of overall speech and the various types of speech input, we fit a series of logistic mixed-effects regression models. Each model had a two-level nested structure with random intercepts for each participant and a fixed effect of community (Bolivia vs. United States). The dependent variable was each 30-s segment (after excluding sleep) annotated for presence or absence of any speech (model 1), or the particular type of speech input: speech to any child (CDS; model 2), speech to the target child only (TCDS; model 3), and speech to adults (ADS; model 4) in every annotation made. The model outcome was then the log-odds of the presence of any speech/that particular type of speech input within each 30-s segment. Treatment contrasts were applied; the main effect of community is binary, so the reference level can be derived from the model summary in Table 2. The intercept represents the log-odds of speech presence for the reference community (Bolivia). For example, in Table 2, an intercept of 1.10 indicates that the log-odds of speech presence for Bolivian infants is 1.10. The community parameter represents the difference in log-odds between the United States and Bolivia communities; when added to the intercept, it gives the log-odds for United States infants. For Table 2, the community estimate of −0.30 means that the United States infants have a log-odds of 0.80 for any speech presence.

For the model predicting the presence of *any* speech, the effect of Community did not improve upon a baseline, random-effect only model suggesting that there are no differences between communities in infants' overall speech exposure (log-likelihood test comparison: $\chi^2 = 0.8$, $df = 1$, $p = .37$; see Figure 1 for illustration of individual infants). We likewise did not find a reliable effect (no improvement from the baseline model) of Community for the likelihood of speech directed to *any* child in the infants' environments (CDS), again suggesting that infants in both communities

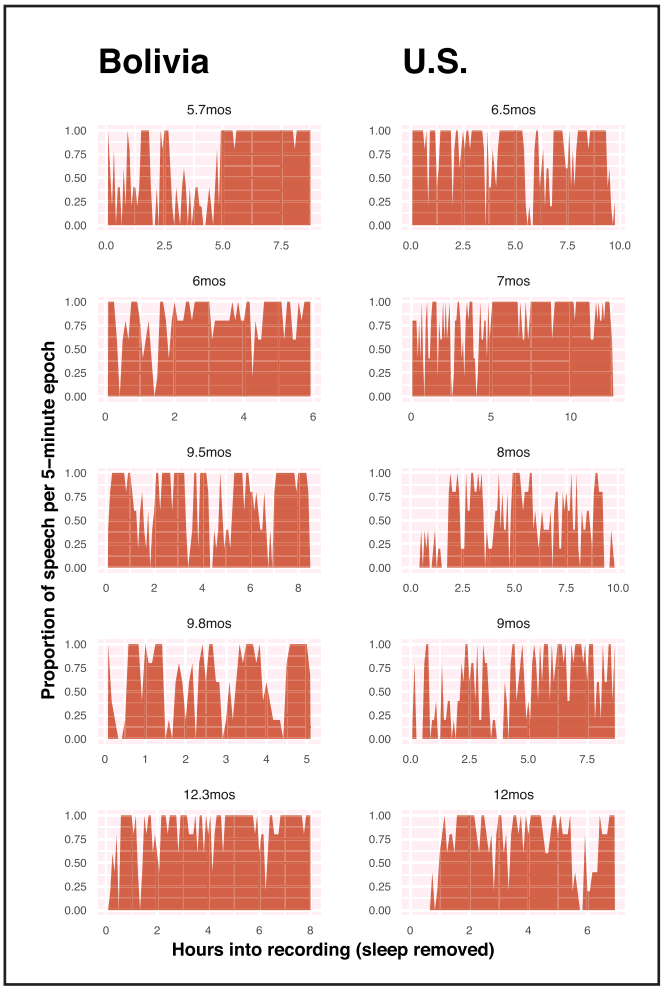

Figure 1 Distribution of total speech input for individual infants in Bolivia (left) and the United States (right) with all-day annotation. To plot the daily trends for each infant, data were binned into 5-min epochs and the proportion of speech within each epoch was computed. Both visuals and statistical modeling (see Table 2 for model summary) disregard periods of sleep.

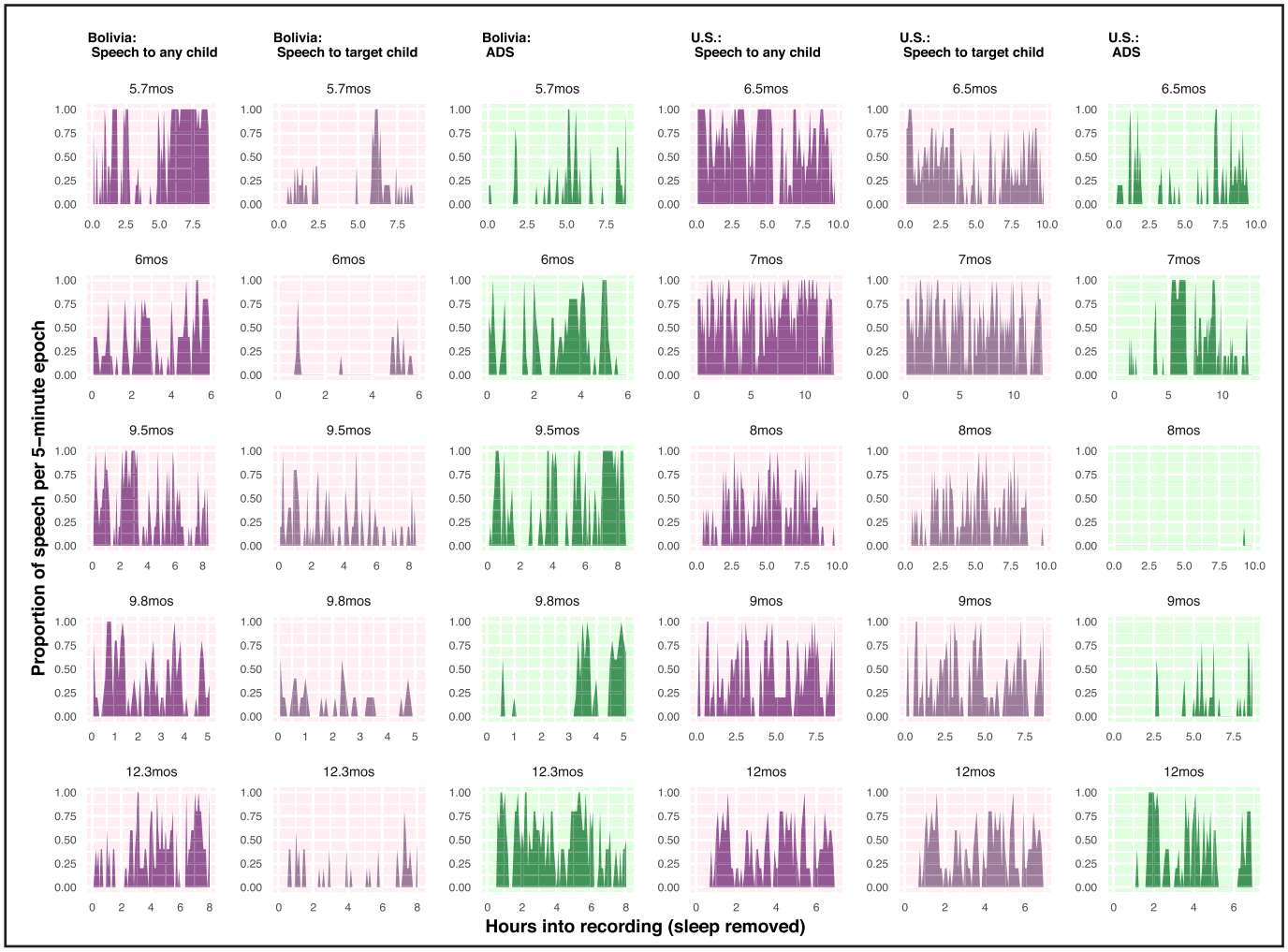

Figure 2 Proportion of speech registers for individual infants in Bolivia (left) and United States (right) with all-day annotation. To plot the daily trends for each infant, data were binned into 5-min epochs and the proportion of speech within each epoch was computed. Both visuals and statistical modeling (see Table 2 for model summary) disregard periods of sleep.

Table 2 Model summaries predicting log odds of different types of speech input over the course of the day as a function of community.

Model no.	Outcome predicted	Parameter	Estimate	*SE*	*z value*	*p* value
1	Any speech	Intercept	1.10	0.23	4.68	<.001
	Any speech	Community: U.S.	−0.30	0.33	−0.92	.36
2	Speech directed to any child	Intercept	−0.40	0.20	−1.95	.05
	Speech directed to any child	Community: U.S.	0.29	0.29	1.03	.30
3	Target-child-directed speech	Intercept	−2.09	0.19	−11.33	<.001
	Target-child-directed speech	Community: U.S.	1.63	0.25	6.44	<.001
4	Adult-directed speech	Intercept	−0.93	0.57	−1.63	.10
	Adult-directed speech	Community: U.S.	−1.38	0.81	−1.70	.09

Note. Estimates are derived from the 10 infants with all-day annotation.

are exposed to similar amounts of CDS when using a broader definition ($\chi^2 = 1.00$, $df = 1$, $p = .32$). On the other hand, we did find that infants in the United States have a higher likelihood of exposure to TCDS than infants in Bolivia (reference level = Bolivia: $\beta = 1.63$, $p < .001$). Thus, confirming our hypothesis, infants in the United States community have greater exposure to speech directed to them than infants in Bolivia, but there are no differences in the likelihood of CDS exposure over the course of the day between communities when using a broader definition of CDS that encompasses speech to *any* child.

Compared to the speech exposure estimates for the 10 infants (5/community) in Table 1, there was slightly larger variability in the

Table 3 Infant vocal maturity measures and proportion of the waking day with exposure to different types of speech input, by community.

	Bolivia *M (SD)* range	U.S. *M (SD)* range
Any speech	0.69 (0.09) 0.54–0.83	0.66 (0.17) 0.35–0.85
Speech directed to an adult (ADS)	0.27 (0.11) 0.10–0.43	0.17 (0.16) 0–0.46
Speech directed to any child (CDS)	0.39 (0.06) 0.3–0.5	0.49 (0.17) 0.31–0.76
Speech directed to the target child (TCDS)	0.18 (0.08) 0.08–0.32	0.42 (0.16) 0.22–0.76
Hourly vocalizations from target infant[a]	108.88 (87.92) 0–417	97.84 (77.14) 0–345
Canonical babbling ratio (CBR)	0.34 (0.13) 0.06–0.67	0.19 (0.1) 0–0.50
Age-normed canonical babbling ratio[b]	0.09 (0.97) −2.80–2.96	−0.19 (1.12) −1.87–2.75

Note. Here the numerator for each computed proportion of speech input reflects the number of 30-second segments containing the type of speech input, and the denominator is the total number of annotations made (which includes segments without speech input but excludes when the infant was sleeping). Proportions are derived from all 20 infants via random sampling with replacement from each infant's recording. Hourly vocalizations were computed algorithmically from the entire recording and canonical babbling ratio was derived from manual annotations of vocalizations sampled from the same recording.
M = Mean; *SD* = Standard Deviation.
[a]Repeated measures within infants (for each hour of the day), so here 0 vocalizations does not mean that the infant did not vocalize at all, but instead refers to an hour of the day where the infant did not vocalize.
[b]Age normalization conducted via z-score normalization so the age-specific, population level average is 0, and each SD = 1.

daily proportion of CDS and TCDS exposure for the 20 infants with random sampling than the subset with all-day sampling (e.g., 0.3–0.53 TCDS exposure in the all-day sampled United States cohort versus 0.22–0.76 in the randomly-sampled United States cohort; see Table 1). To ensure that our findings by community in daily speech exposure were consistent across all 20 infants, we fit the same logistic mixed-effects models as for research question 2 to predict the likelihood of CDS and TCDS exposure for each 30-s segment annotated (after removing sleep; not tabled). For CDS, a model with the fixed effect of Community did not improve upon either baseline model that we tested (Model 1: just random intercept of Child $\chi^2 = 3.07$, $df = 1$, $p = .08$; Baseline model 2: random intercept of child and fixed effect of child age to account for the larger age variability in the 20-infant sample $\chi^2 = 3.32$, $df = 1$, $p = .07$). For TCDS, however, Community did improve upon the baseline models, both for the model with ($\chi^2 = 18.63$, $df = 1$, $p < .001$) and without the effect of child age ($\chi^2 = 15.53$, $df = 1$, $p < .001$). In other words, the results reaffirm a lack of evidence for differences in CDS exposure between the United States and Bolivian communities, but reaffirm that infants in the United States are exposed to more TCDS.

Research Question 3: Does speech input to infants predict infant vocal maturity in the United States and/or Bolivia?

Descriptive statistics

We next ask if there is a relationship between infants' daily exposure to different types of speech input and their vocal maturity. For this analysis, we looked at all 20 infants (10/community). The speech input estimates were made via random sampling from each infant's recording and the vocal maturity metrics—the number of vocalizations from the target infant for each hour of the day (speech volubility) and CBR—were derived from vocalizations over the entire recording (see Methods for detail). Table 3 presents the proportion of CDS and TCDS present within each infant's recording (for all 20 infants) as well as the proportion of each day that the infants were exposed to *any* speech (encompassing TCDS, CDS, and ADS).

Modeling the relationship between speech input and infant vocal maturity

Finally, we address the third research question: is there a relationship between speech input and infants' vocal maturity? We assess vocal maturity as the (1) average number of vocalizations from the target infant for each hour of the recording and (2) CBR. These two measures of vocal maturity were unrelated to one another in our sample ($r = -0.06$, $p = .78$).

We fit a series of linear mixed effects models with the following hierarchical structure: to account for child- and (when relevant) hour-of-day level variability in the relationships between speech input and vocal maturity, we fit models with the maximal random effects structures that the design permitted (Barr et al., 2013). These models included random intercepts for child and random slopes of speech input proportion (CDS or TCDS, depending on the model) by child, allowing both baseline vocal maturity and the strength of the relationship between speech exposure and vocal maturity to flexibly vary by child. For volubility models, we additionally included random intercepts for hour-of-day to account for differences in amounts of infant vocalization throughout the day. Adding random slopes of hour by child would require estimating child-specific deviations for each of the 8–16 hr per child, which would be overparameterized given the sample ($N = 20$). The final random effects structures were: Volubility: (1 + prop_speech | child_id) + (1 | hour), CBR: (1 + prop_speech | child_id). We initially encountered convergence issues so we employed the "bobyqa" optimizer during model fitting to facilitate convergence. With bobyqa, all models converged successfully without warnings.

For the fixed effects, treatment contrasts were again applied, so the reference level can be derived from the model summary in Table 4. All continuous variables were centered and scaled to facilitate comparison between different effects and fit with REML to further ensure this comparison between models. We employed the z-score normalized CBR values (by age) (see Methods). We retained child age as a covariate as a conservative measure to ensure that any remaining age-related variance did not confound our interpretation of speech input effects. This covariate also accounts for the different age distributions between our Bolivia and United States

Table 4 Model summary predicting relationship between proportion of daily speech input and target infant speech volubility (left) and age-normed canonical babbling ratio (CBR) (right).

	Speech volubility				Canonical babbling ratio			
Parameter	**Estimate**	***SE***	***t* value**	***p* value**	**Estimate**	***SE***	***t* value**	***p* value**
Intercept	143.93	18.13	7.94	<.001	1.48	0.50	2.94	.005
Prop. of Speech Input	514.60	181.24	2.84	.013	9.52	4.09	2.33	.024
Corpus:U.S.	−50.10	22.92	−2.19	.043	−1.68	0.60	−2.80	.008
Age (mos)	0.86	2.01	0.43	.675	−0.22	0.04	−5.25	<.001
Num. Siblings	−3.33	4.73	−0.70	.497	0.21	0.12	1.78	.082
Prop. of Speech Input*Corpus: U.S.	−460.55	195.68	−2.35	.033	−11.31	4.15	−2.73	.009

Note. Speech directed to any child (CDS) was related to speech volubility, while speech directed to the target child (TCDS) was more strongly related to CBR. Estimates of daily speech input are derived from all 20 infants via random sampling, estimates of hourly infant speech volubility are derived algorithmically from each infant's entire recording, and CBR measures are derived from manual annotation of target infant vocalizations.

samples. Finally, to be maximally conservative, our models predicting infant vocal maturity always include a term for Number of Siblings (even when this term did not improve upon model fit), to control for potential effects of sibling number across children and communities (see following section for explanation). As the final model summaries will illustrate, number of Siblings was not significantly related to either measure of infant vocal maturity.

After baseline model construction, model fitting of the key fixed effects—type of speech input—was conducted in a step-wise, additive fashion. For both vocal maturity outcomes, we evaluated models with and without TCDS and CDS both to see if there was an effect of speech input and if so, which speech register was the stronger predictor. Table 4 presents the final model summaries and Figures 3 and 4 plot the relationship between speech input and infant vocal maturity for the two communities. The best speech volubility model included main effects of proportion of CDS exposure ($\beta = 514.60$, $p = .013$), as well as the interaction of proportion of CDS and Community, suggesting that the relationship between CDS and infant volubility varied by community (reference level Bolivia: $\beta = -460.55$, $p = .33$). TCDS was not reliably related to infant speech volubility in either community. The positive estimate for the effect of proportion CDS, and the negative estimate for the interaction of proportion CDS and Community, indicates that while infants in both communities who are exposed to relatively more CDS in their daily lives vocalize more, this relationship is stronger for the Bolivian infants than the United States infants (Figure 3).

For CBR, while proportion of CDS exposure did improve upon model fit, proportion of TCDS exposure, and its interaction with Community, provided the best fit to the data (comparing model with and without interaction: ($\chi^2 = 6.28$, $df = 1$, $p = .012$). There was a strong, positive main effect of TCDS on CBR, representing the relationship for the Bolivian infants (the reference level) ($\beta = 9.52$, $p = .024$). The significant interaction between TCDS and Community again suggests that the relationship varied by community. To interpret the effect for the United States infants, we combine the main effect with the interaction term ($\beta = -11.31$), resulting in a slight *negative* relationship ($9.52 - 11.31 = -1.79$). In sum, while Bolivian infants who received more TCDS showed substantially higher CBRs for their age, United States infants showed a marginally negative relationship.

We wanted to be sure that the observed relationships were unique to TCDS/CDS, and not attributable to overall speech in the

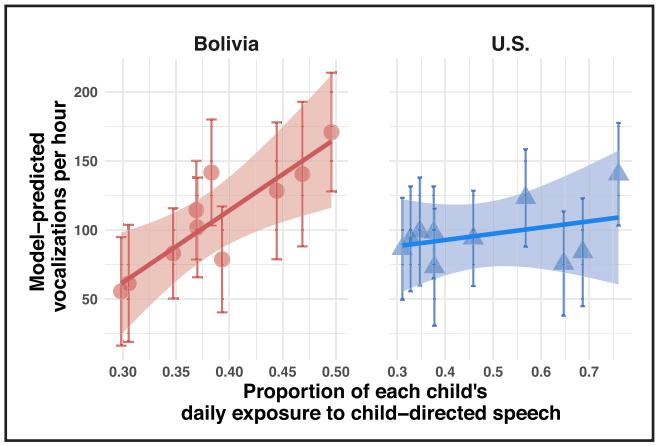

Figure 3 Model-predicted hourly vocalizations from infants as a function of the proportion of daily exposure to child-directed speech in Bolivia (left, red) and the United States (right, blue). Points represent individual infant predictions with 95% confidence intervals shown as error bars. Thick, dark lines represent population-level predictions and ribbons refer to 95% confidence intervals around the predictions. There is a strong positive relationship between CDS and vocalizations in Bolivia, while the relationship is substantially weaker in the United States. Note the different x-axis scales.

infant's environment. So we similarly evaluated potential relationships between the total proportion of *any* speech exposure (speech directed to the target child, other children, or adults) and infant volubility/CBR by adding a fixed effect of Any Speech proportion to the baseline models; we saw no improvement over the baseline model for volubility with or without the interaction with Community (with: $p = .34$; without: $p = .62$) or CBR with or without the interaction with Community (with: $p = .19$; without: $p = .11$). This analysis suggests that the observed relationship between speech input and vocal maturity is unique to the specific registers we examined and is not an effect of overall speech exposure.

Evaluating confound with number of siblings and adjacent children We considered that there could be potential confounds in our analysis related to the number of siblings in each infant's household. The Bolivian infants had on average a higher number of siblings than the United States infants ($M = 2.5$, $SD = 2.17$, range = 0–7 for Bolivia compared to $M = 1.3$, $SD = .95$, range = 0–3

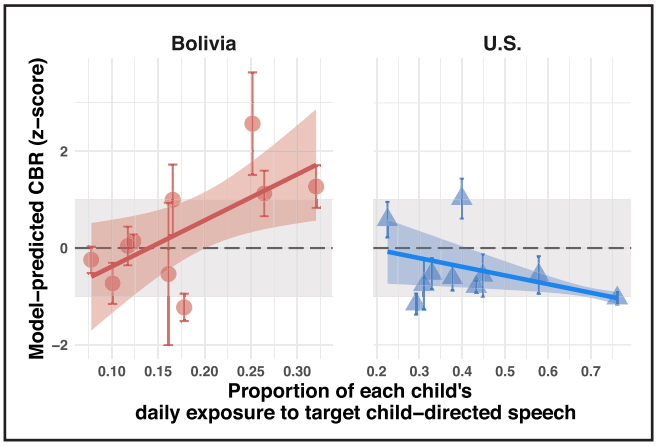

Figure 4. Model-predicted canonical babbling ratio (CBR) z-scores as a function of the proportion of daily exposure to target child-directed speech in Bolivia (left, red) and the United States. (right, blue). Points represent individual infant predictions with 95% confidence intervals shown as error bars. Thick, dark lines represent population-level predictions and ribbons refer to 95% confidence intervals around the predictions. The dashed horizontal line at CBR z-score = 0 indicates age-expected babbling development and the light, shaded gray indicates +/1 SD from population-level mean reported in Cychosz and Long (2025). There is a positive relationship between target child-directed speech and CBR in Bolivia, while the relationship is slightly negative in the United States. Note the different x-axis scales.

for the United States; see Table S1 in Online Supplementary Materials). This fact likely explains the larger amounts of CDS (to any child) that the Bolivian infants are exposed to, but it also potentially means a higher likelihood that the speaker diarization algorithm mis-identified speech from other children adjacent to the target infant, so an infant from a larger family may inadvertently be reported as more vocal. We evaluated this concern in two ways (see online Supplementary Material II for full analysis): first, as stated previously, our models always included a term for number of siblings. Second, we evaluated the potential confound of sibling number in one additional way: we fit another mixed-effects model predicting infant speech volubility with the same random effects structure listed above. To this random-effect only model, we added the Number of Siblings term and found (1) no improvement upon model fit and (2) no significant effect of Number of Siblings in the model summary (model summary included in online Supplementary Material II). Thus, we conclude that the stronger relationship between CDS and infant volubility is unlikely to be attributable to the higher number of siblings among the infants in Bolivia.

However, controlling for Number of Siblings does not entirely mitigate our concern because some vocalizations may come from neighborhood children unrelated to the target infant, as we have qualitatively observed that the Bolivian infants are exposed to more children outside of their family than the U.S. infants. So, we conducted one final analysis to ensure that our results concerning the relationship between CDS and target infant speech volubility were robust. This analysis takes advantage of the fact that, during manual annotation of the algorithmically-identified target infant vocalizations, we classified all non-target infant speech into other speaker categories (e.g., other child, adult female; see Methods). From these annotations, we computed the confusion rates of target infant and other children for the Bolivian infants and United States infants. The confusion rates were low in both samples and, although somewhat higher in the Bolivian ($M = 4.75\%$, $SD = 3.12\%$) than the United States ($M = 2.36\%$, $SD = 2.77\%$) sample, this difference was not statistically significant, $t(df) = 1.81(17.75)$, $p = .09$ (see online Supplementary Materials for detail). On the basis of these analyses, we conclude that our finding of a stronger relationship between CDS and infant volubility in Bolivia is unlikely to be entirely explained by higher rates of confusion between the target infant and other children in the Bolivian sample.

Discussion

This work examined differences in the quantity of infants' (<24 months) speech exposure—and how those differences predicted the infants' vocal maturity—in two language communities: a Quechua-Spanish community in rural Bolivia and a Spanish-English immigrant community in the urban United States. Cross-cultural work suggests there is significant variation in the frequency of infant-directed speech across populations, and particularly between rural, less-industrialized populations and urban, industrialized ones (Bunce et al., 2025; Cristia, 2023; Cristia et al., 2019; de León, 1998; Keller, 2007; Lieven, 1994; Ochs & Kremer-Sadlik, 2015; Schieffelin & Ochs, 1986; Shneidman & Goldin-Meadow, 2012). Here, when we quantified the infants' exposure to several different types of speech input, we found that (1) quantity of directed speech to young infants differed between communities and was relatively rare in the Bolivia sample, but (2) when instead a broad definition of CDS was employed, encompassing speech directed to *any* child in the infant's vicinity, there were no longer differences in CDS quantity by community, and (3) the overall amount of *total* speech input available to the infants also did not differ by community. In the second part of this work, we examined if individual differences in exposure to CDS predicted the infants' vocal maturity. We found that despite overall lower rates of speech directed to the target infants in the Bolivia sample, rates of CDS—to any child in the infant's vicinity—positively predicted the target infant's speech volubility and that TCDS predicted CBR. In the following sections, we discuss each of these research findings in detail and how they relate to current and proposed models of language development.

The overall amount of total speech input available to the infants did not differ by community

The paradoxical findings that higher quantities of directed speech appear to predict children's language development, but that large numbers of children worldwide do not receive much directed input and still reach major language development milestones, have led to a theoretical interest in understanding variation in the different types of speech input available to infants, including child-directed and overheard speech, and the extent to which these support infants' language learning (Casillas et al., 2020; Cristia, 2023; Foushee & Srinivasan, 2024; Schick et al., 2025). We first quantified the total amount of speech input *available* to the infants in this study, encompassing speech that they could possibly overhear but that was not directed to them. Our metric factored in not just child-directed or even child-adjacent speech, but

background speech, overlapping speech, as well as speech from other children in addition to speech from adult caregivers. Despite eventually finding notable differences in quantities of TCDS by community, we found no differences by community in total speech input when using this broad definition of total available input, or when using a definition of CDS that includes speech to children in the vicinity of the infant.

Scaff et al. (2024) argue that broader definitions of speech input to children, outside of the child-centered, dyadic interactions commonly measured in developmental psychology, may be necessary to properly model effects of speech input quantity and content upon language acquisition. Indeed, it may be that children's language learning mechanisms are fine-tuned to *suit the speech input available to them* and that, in all likelihood, there is no "optimal" speech register for language learning that is broadly applicable to all developmental settings. Another, not mutually exclusive, possibility is that specific types of language input are better or worse suited to support different aspects of language development. For example, caregiver speech that is specifically directed to the target child and contingent on that child's communicative attempts, objects within view, and focus of attention might be particularly beneficial for learning object names (Golinkoff et al., 2015; Tamis-LeMonda et al., 2014; Yurovsky, 2018), a skill that has been the focus of much research in developmental psychology. In contrast, overhearing speech to others, including other children, may expose infants to a greater variety of words and sentence structures, and to more exemplars of these, from which to learn the sounds and statistical patterns of their language.

Consider the speech environments experienced by the infants in the rural Bolivian and urban United States immigrant communities in this study: the Bolivian infants are, on average, exposed to more ADS than the United States infants (ADS for the Bolivian infants, as in many cultural settings, is spoken faster than CDS, with more acoustic reduction, more embeddings, and uses more complex, infrequent vocabulary items). At any given moment, the Bolivian infants are also exposed to a greater number of distinct interlocutors, with larger amounts of speaker overlap, that they must differentiate between. In the Bolivian community, mothers often engage in domestic work, agriculture, and food preparation communally, while fathers may leave for extended periods in search of work in larger urban areas. Consequently, the overlapping adult speech that the infants hear often stems primarily from groups of women, making it more difficult to distinguish between acoustically (i.e., there are larger pitch range differences between male and female adults than between female adults). Finally, the infants in Bolivia spend the majority of their day outdoors, which has several acoustic consequences for the speech signal including noise interference (from wind, traffic, and animals—the latter especially relevant in this more rural setting), omnidirectional dispersion (fewer obstructions from walls, furniture, etc. cause faster reductions in speech signal intensity and thus volume), and lack of reflection (mild reverberation in enclosed spaces can lend the perception of sustained intensity in the speech signal over longer periods of time) (Attenborough & Van Renterghem, 2021). All of these factors could contrive to make the speech signal more challenging to parse. Such a signal may be objectively more difficult to break into or learn from in the short-term, but could actually make these children more resilient language learners and perceivers in the long-term. Again, children may adapt to the particular speech input available to them.

Compare these characteristics of speech input to the Bolivian infants versus those of the infants from immigrant families in urban United States: the United States infants are exposed to large amounts of directed speech, mostly by a single caregiver in an indoor setting with moderate levels of reverberation. There is minimal speaker overlap, because there are fewer overall speakers, and when speaker overlap occurs it is often between two acoustically distinct voices (e.g., adult female and another child or adult female and male). On the other hand, infants in the immigrant United States community have fewer opportunities to interact with siblings and peers, both because family sizes are smaller and because they are separated from extended family in their countries of origin; thus, they have less access to speech from other children, which has been shown to be a particularly useful source of input in some contexts and for some aspects of language development (Schick et al., 2025). And, because there are fewer speakers overall, they are exposed to less variable input, which has been argued to support the development of phonetic categories (Place & Hoff, 2016). Thus, infants in the United States community may also have to adapt to the particular speech input available to them, relying more on directed speech from few adult caregivers and less on interactions with peers or parsing more speaker overlap.

Directed speech to young infants is rare in the Bolivia sample

The finding that the Bolivian infants were exposed to significantly less TCDS than same-aged infants in the United States is consistent with comparative quantitative estimates of TCDS in other rural/less-industrialized versus more urbanized settings (Cristia, 2023; Vogt et al., 2015) (though we again emphasize that it is impossible to discretize communities into stringent categories like these or generalize over all more- or less-industrialized settings). Why is directed speech to infants more common in certain cultural contexts than others, and why is directed speech to individual children so commonplace among industrialized groups (Ochs & Kremer-Sadlik, 2015)? A number of not mutually exclusive proposals have been put forth: verbal interaction may be less common because it does not promote the developmental outcomes (e.g., motor) most valued in a given community (Frye, 2022; Vogt et al., 2015); caregivers may consider pre-verbal infants to be less valid conversational partners or do not believe that pre-verbal infants can understand any language (Gordon et al., 2020; Ochs, 1988; Schieffelin, 1990; Weber et al., 2017); cultural beliefs may emphasize that children best acquire the components of language that are of most importance to the community via observation, not directed instruction or interaction (Gaskins, 2006); or, in bilingual settings, caregivers may believe that one language is easier to learn than another and thus they should direct input to the infant in the more complex of the two languages (Padilla-Iglesias et al., 2021).

Other reasons for cross-cultural differences in quantities of directed speech to children rely less on cultural beliefs and more on ecological and socio-economic factors that influence the dynamics of caregiving more generally (Cristia, 2023). One such factor is family size, but it is not simply that larger families mean there is less attention to go around (and thus less speech directed to a single infant). Less-industrialized communities do tend to

have larger family sizes, with more siblings. More communal child-rearing, including caregiving conducted by older children outside of the child's nuclear family, is common. Directed input may have at one time stemmed primarily from older children in such settings, particularly once an infant started walking and was no longer primarily transported by their mother. However, with the introduction of formal schooling, and economic opportunities further from the home, a move to market-based economies and industrialization has resulted in older children spending less time engaging in caregiving work for younger siblings (Weisner, 1987). In a natural experiment, Padilla-Iglesias et al. (2021) present a clear illustration of this market-driven phenomenon: the authors quantified directed versus overheard speech input to two cohorts of Yucatec Mayan infants, born 6 years apart (2007–2013), who were exposed to bilingual Yucatec Maya and Spanish input. During this time, the speech community underwent significant market integration (e.g., fathers of infants in the 2007 cohort conducted significantly more agricultural labor, but less wage labor, than the 2013 cohort). Infants in the 2013 cohort received significantly less directed input *from other children* than infants in the 2007 cohort. And even when the reduced directed input was compensated for by increased input from adult caregivers, the 2013 cohort still heard significantly less directed—but not overheard—input overall than the first cohort (see also Cychosz et al., 2025b for differences in rates of directed speech input by minoritized vs. majority languages in urban versus rural settings).

Quantities of CDS are similar across the Bolivia and United States samples, and predict individual differences in infants' vocal maturity

Finding that infants in Bolivia and the United States are exposed to similar rates of CDS when using a broader definition of CDS emphasizes the importance of including multiple definitions of speech input in research in developmental psycholinguistics. It is commonly observed that some children (e.g., from lower socioeconomic status households in the United States or less industrialized communities internationally) receive less directed input—a finding that we replicate in our samples (Cristia, 2023; Dailey & Bergelson, 2022). But such estimates, including ones we have made in some of our previous work, are often made using narrow, culturally-defined definitions of CDS that did not always consider other potential sources of input outside of directed speech (see discussions in Scaff et al., 2024 ; Sperry, Sperry, and Miller, 2019a, 2019b). As Figure 2 demonstrates in an additive manner, when we take a broader definition of CDS we find that the Bolivian infants *do* hear significant amounts of CDS, while at the same time hearing lower rates of TCDS.

The next question, then, is whether this broader definition of CDS predicts children's language development. Here, we find that broad CDS is positively associated with infants' speech volubility and, for the Bolivian infants, their CBR. On the other hand, we find that TCDS is a stronger predictor of CBR than CDS, while *all speech* is not a good predictor of CBR or volubility. We draw several conclusions from these findings.

First, the overall findings support the hypothesis that infants' phonological development is sensitive to environmental influences, and particularly CDS. Most previous evidence supporting an impact of CDS upon language learning outcomes has focused on children's vocabulary development (Hoff, 2003; Rowe, 2008; cf. Huttenlocher et al., 2002, 2010) and it has been unclear if, and how, CDS impacts other domains of language development, especially phonology (Cristia, 2020). There are clear reasons why CDS might impact phonology, especially speech production, in a completely different manner than vocabulary, or not play a significant role at all: phonological production has a strong, biologically-based component that regulates children's early motoric abilities in an environmentally neutral manner. Infants initially master basic motor routines that dictate their progression through early stages of consonant emergence. For example, early mastery of simple vertical lip and jaw oscillations permits the emergence of /b/ and /m/ and eventual tongue tip flexibility leads to the often early-acquired /d/. Later-acquired segments, however, such as /l/ or /ɹ/, require multiple simultaneous articulatory configurations, and thus these segments are almost universally later-acquired. Other domains of language, such as vocabulary, are both less dependent on these biological constraints and, because they require cultural learning, more susceptible to environmental influence. In sum, because phonological production in particular passes through a vocal tract "filter," environmental effects upon early phonological production may be mitigated by early motoric limitations of the children's developing speech apparatuses (Vorperian et al., 2005).

On the other hand, there is evidence that social interactions influence early phonological development, particularly the frequency and maturity of infants' vocal productions (Ferjan Ramírez et al., 2020; Goldstein & Schwade, 2008; Ramírez-Esparza et al., 2014, 2017; Warlaumont et al., 2014). The current findings expand on this earlier evidence by showing that the amount of CDS infants hear on a typical day relates to the quantity and maturity of their vocalizations.

Second, we find different effects of broad CDS and TCDS on the two measures of vocal development. While broad CDS predicted infants' speech volubility, TCDS was a stronger predictor of CBR. This suggests there are potentially meaningful differences in the benefits of directed and non-directed speech even within CDS. That is, in addition to the acoustic and linguistic features that differentiate the CDS register from ADS, TCDS is likely to involve interactive features that are also beneficial to language learning, such as being contingent on the child's communicative attempts and more finely tailored to the child's general developmental level and focus of attention (Golinkoff et al., 2015; Rowe & Snow, 2020). Consistent with this, some studies have found that speech that uses the prosodic and linguistic features of "parentese" *and* occurred in one-on-one contexts between a caregiver and infant is associated with higher quantities of infants' speech productions, while standard speech and speech in group contexts is not (Ramírez-Esparza et al., 2014, 2017).

Another, not mutually exclusive possibility, is that our measure of TCDS more closely captured the types of vocal exchanges between caregivers and infants that have been found to facilitate the development of more advanced vocal patterns in studies of United States infants (Goldstein & Schwade, 2008; Ramírez-Esparza et al., 2014, 2017; Warlaumont et al., 2014). In particular, previous studies found that contingent vocal feedback produced changes in the likelihood and phonological form of infants' vocal productions, while caregiver speech that was not contingent on

infants' vocalizations did not (Goldstein & Schwade, 2008; Warlaumont et al., 2014). Further, analyses of caregiver–infant interactions across diverse linguistic and sociocultural contexts have found that caregivers simplify their speech in response to infants' immature vocalizations, providing input that could be particularly suitable for infants' learning (Elmlinger et al., 2023; Elmlinger et al., 2025). Although we did not examine the temporal or linguistic features of TCDS in the current study, the finding that speech *to* the target infant (which is more likely to exhibit these properties), but not CDS surrounding the infant, predicted infants' canonical babbling, is consistent with this prior literature and suggests that vocal exchanges *with* infants may play a particularly important role in the development of more mature vocalizations, *even in contexts in which these vocal exchanges are relatively infrequent*. On the other hand, our findings also suggest that exposure to speech directed to other children—which may be indicative of opportunities to observe and participate in verbal interactions with a range of interlocutors—may prompt infants to produce a greater number of vocalizations.

Finally, it is worth noting that across both communities, almost all of the infants were meeting age-appropriate milestones for vocal development, including canonical babbling (see Figure 4). Thus, despite large between-community differences in TCDS and substantial *individual* variation in exposure to CDS and TCDS, the emergence of canonical babble for infants in both of these communities occurs at ages comparable to what would be expected from normative data. This aligns with what has been reported for children in a Tseltal Mayan community and a Papuan community who have also been reported to have infrequent exposure to TCDS (Casillas et al., 2020, 2021a). Whether one thinks that variation in vocal behaviors within the normative range is meaningful or not is to some extent a matter of opinion. An outstanding question for future research is whether the observed variation in vocalization rates and canonical babbling ratios is related to children's communicative development in ways that have relevance for their health and well-being within their own communities.

The relationship between target CDS and canonical babbling ratio varies by community

An unexpected finding in this work is the difference between communities in how TCDS relates to infant babbling. For Bolivian infants, we found a strong positive relationship between the proportion of TCDS they received and their age-normed CBRs ($\beta = 67.23$, $p = .007$). The United States infants, however, showed a slightly *negative* relationship between TCDS and CBR (combined effect of approximately $\beta = -5.22$).

This finding in the United States community was unexpected and warrants cautious interpretation, particularly given the small sample. One possibility is that the United States caregivers adhered more strongly to an instructional model more closely aligned with Directed Acquisition, where they view their speech input as necessary for teaching language. Although this explanation is more consistent with what is known about middle class United States parents than the first-generation immigrant families in this study, most of the immigrant parents had migrated from urban settings in Latin America, where caregiving practices have been influenced by industrialization and schooling in ways that emphasize verbal communication between parents and infants (Keller et al., 2009; Ochs & Kremer-Sadlik, 2015; Richman et al., 1992; Van Kleeck, 1994). Further, it may be that these parents had begun to adopt the language socialization practices of the dominant United States culture (García Coll et al., 1996). In contrast, Bolivian caregivers may view language development as more self-directed by the child, responding with increased input primarily when the child demonstrates readiness through more speech-like vocalizations. These different beliefs about the role of caregiver input in language development could have prompted the United States caregivers to produce larger quantities of TCDS, or TCDS that is not responsive to the infants' communicative attempts and developmental needs, and which has little (or even a negative) impact on their vocal development. This would be consistent with a threshold model for the relation between language exposure and acquisition, in which speech input is useful for language learning only up to a certain point, after which infants can no longer make much use of it (Cristia, 2020). This is also consistent with the proposal of a curvilinear relationship between caregiver responsiveness and infant development, with an optimal level of verbal responsiveness after which additional amounts of interaction are excessive or even intrusive (Bornstein and Manian, 2013. If caregivers respond indiscriminately to all infant vocalizations, or if they frequently talk to the infant when they are not attending, this may decrease the signal value of caregivers' verbal responses for infants' vocal learning.

Another potential explanation is that the United States caregivers could be engaging in compensatory behavior—they may increase their directed speech to infants who show less mature vocalization patterns. In particular, United States caregivers may be more attentive to or concerned about language development milestones, perhaps due to greater exposure to child development information through healthcare systems, parenting resources, or cultural emphasis on early achievement. It is possible that when the United States caregivers perceive their infant as communicating less maturely, they may increase their directed input to stimulate development. It is also possible that infants with lower CBRs might elicit more caregiver speech through their communicative behaviors. Less vocally mature infants may make greater use of other communication strategies (gestures, facial expressions, or less mature vocalizations) that prompt caregivers to respond verbally, effectively increasing the TCDS they receive.

Another possibility is that the estimates of TCDS used in this study, which are based on annotations of 30-s segments rather than counts of words or conversational turns, do not provide a sufficiently close approximation to the contingent vocal exchanges between infants and caregivers that are thought to drive early phonological development (Elmlinger et al., 2023; Goldstein & Schwade, 2008; Warlaumont et al., 2014). Further, it may be that these estimates captured somewhat different behaviors in the United States and Bolivian communities. For example, as shown in Figure 2, exposure to TCDS among United States infants tended to be distributed throughout the day, while in the Bolivian community it was more concentrated during specific periods. Thus, it may be that TCDS in the United States community occurred in a larger number of 30-second segments but that each of these segments contained only a few words or conversational turns, while TCDS in the Bolivian community occurred in a smaller number of segments representing denser clusters of language interaction (Casillas et al., 2020). Our current measures

cannot distinguish between these scenarios. Future work will seek to examine links between CDS and TCDS and finer-grained measures of caregiver-infant vocal exchanges to more precisely test the mechanisms by which speech input relates to vocal development in these communities.

One final possibility is that the single daylong recordings collected in this study did not capture a sufficiently representative sample of infants' speech exposure, particularly for the United States infants. For example, it is possible that some United States infants attended childcare settings outside the home that we did not capture in the recordings, thus under- or over-estimating their exposure to CDS and TCDS and distorting this relationship.

Overall, this finding highlights that the relationship between caregiver speech input and infant babbling is not universal but may instead be culturally situated. The positive relationship in Bolivia suggests that in contexts where directed speech is generally rarer, infants may be particularly responsive to the directed speech they do receive. Conversely, the flat or slightly negative relationship in the United States sample demonstrates how different socialization patterns and exposure to models of child development may lead to distinct caregiver–infant dynamics.

Limitations and future directions

One limitation of this work is the relatively small sample size: the Bolivian sample was limited by the age of infants present in the community at the time of data collection. The samples were also not terribly well matched across communities, again due to limitations faced during data collection in Bolivia. It was not possible to match the infants for household size (families are larger in Bolivia). And while we were nearly successful at matching for gender across communities, the Bolivian infants were, on average, older than the United States infants, including two infants >12 months of age. Future data collection in the community, collected using similar infant-centered recording technology, will be able to add to the corpus and thus make more definitive conclusions about cross-site differences as well as examine additional vocal maturity metrics. This contribution will be important because another limitation of this work is that our measures of phonological development focused exclusively on phonological production, and not measures of perception. Measuring the relationship between speech input and speech perception could mitigate some of the biological constraints acting upon the relationship between speech input and phonological production measures. As discussed above, any impact of environmental factors, such as quantities of directed input, upon phonological production must pass through a vocal tract filter (i.e., cognitive perceptual speech categories are mitigated by universal motoric constraints upon infant vocal production). Indeed, given this fact about early motoric limitations in the developing infant speech apparatus, we could reasonably assume that effects of caregiver speech input would be *enhanced* for perceptual outcomes (Singh et al., 2023), so we think this is an area ripe for future research.

An additional area of research would be to examine more dimensions of the temporal structure of infants' language input. For example, recent work has demonstrated that the *distribution* of language input, above and beyond quantity, can explain individual variability in children's language learning outcomes (Cychosz et al., 2025a). That analysis was conducted over North American preschoolers acquiring English, but such differences are likely to emerge cross-culturally as well. For example, among the Bolivian infants in this sample, overall language input is saturated, without clear bursts of language activity but a high quantity of speech surrounding the infant. The United States infants, on the other hand, often exposed to language input from one or just a handful of interlocutors, might be more likely to experience language input in bursts of activity where fewer speakers produce utterances surrounded by moments of less talk or silence. This, we believe, would be an interesting avenue for future research as well. Finally, as discussed above, characterizing contingent versus non-contingent vocal exchanges in these recordings would help further advance our understanding of the mechanisms by which speech input relates to vocal development in these communities.

Conclusion

This study quantified the speech input available to infants (6–23 months) in an Indigenous Quechua- and Spanish-speaking community in Bolivia and an immigrant Spanish- and English-speaking community in the United States, and how this input related to the infants' early vocal maturity. We first employed a definition of CDS that approximates the one commonly employed in developmental psychology work conducted with Western children—speech directed to the infant by adult (and, in our case, also child) caregivers—and found that the United States infants were exposed to approximately 2.5× more directed speech than the Bolivian infants. When we expanded the definition to include speech to *any* child in the infant's vicinity, there were no differences in amounts of directed speech by community. We additionally found no differences by community in the overall amount of input available to the infants, including input with overlapping speech and ADS. Finally, despite lower rates of directed speech to the target infant, the amount of speech input strongly and positively predicted different metrics of the Bolivian infants' vocal maturity. This finding suggests that in contexts where directed speech is less common, infants may be particularly responsive to the directed speech they do receive. This work is one of the first to show how quantities of directed speech are associated with precursors to language development in infancy across diverse cultural contexts. Overall, these results reinforce the need to employ expanded, culturally-informed definitions of the types of speech input available to infants to better understand relationships between speech input and children's language development.

Supplementary material

Supplementary material is available at *Child Development* online.

Data availability

All code and anonymized data to replicate this work can be found at the Open Science Framework repository at this link: https://osf.io/9zxwt/overview?view_only=58236b23b9f149e68a72cfdc1a70be3f.

Author contributions

Margaret Cychosz (Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing—review & editing

[equal], Formal analysis, Writing—original draft [lead]), Anele Villnueva (Data curation, Methodology, Validation, Writing—review & editing [equal]), and Adriana Weisleder (Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing—review & editing [equal], Visualization, Writing—original draft [supporting])

Funding

This work was funded in part by an Oswalt Documenting Endangered Languages Grant (M.C.), National Institute on Deafness and Other Communication Disorders Grant R21DC018357 (A.W.) and Academic Pediatric Association Young Investigator Award (A.W.).

Conflicts of interest

The authors have no conflicts of interest to report.

Acknowledgments

The authors gratefully acknowledge funding from the following sources: Oswalt Documenting Endangered Languages Grant (M.C.), National Institute on Deafness and Other Communication Disorders Grant R21DC018357 (A.W.), and Academic Pediatric Association Young Investigator Award (A.W.). Additional thanks to Roxy Borg and Anika Velasco for assistance in data coding, to Jan Edwards for use of their recording equipment, and to Alan Mendelsohn and the PlayReadVIP team for support in creating the United States corpus.

References

Akhtar, N. (2005). The robustness of learning through overhearing. *Developmental Science*, *8*, 199–209. https://doi.org/10.1111/j.1467-7687.2005.00406.x

Allen, L., & Hutton, R. (2023). Opportunity gaps in early care and education experienced by children from birth to Pre-K. In L. Allen, & R. Hutton (Eds.), *Closing the opportunity gap for young children*. National Academies Press (US). https://www.ncbi.nlm.nih.gov/books/NBK596379/

Attenborough, K., & Van Renterghem, T. (2021). *Predicting outdoor sound*. CRC Press. https://doi.org/10.1201/9780429470806

Baquedano-López, P., & Mangual Figueroa, A. (2011). Language socialization and immigration. In A. Duranti, E. Ochs, & B. B. Schieffelin (Eds.), *The handbook of language socialization* (pp. 536–563). Wiley-Blackwell. https://doi.org/10.1002/9781444342901

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Beckman, M. E., & Edwards, J. (2010). Generalizing over lexicons to predict consonant mastery. *Laboratory Phonology*, *1*, 319–343. https://doi.org/10.1515/labphon.2010.017

Bergelson, E., Soderstrom, M., Schwarz, I. C., Rowland, C. F., Ramírez-Esparza, N., Hamrick, L. R., Marklund, E., Kalashnikova, M., Guez, A., Casillas, M., Benetti, L., van Alphen, P., & Cristia, A. (2023). Everyday language input and production in 1,001 children from six continents. *Proceedings of the National Academy of Sciences*, *120*, e2300671120. https://doi.org/10.1073/pnas.2300671120

Bornstein, M. H., & Manian, N. (2013). Maternal responsiveness and sensitivity re-considered: Some is more. *Development and Psychopathology*, *25*, 957–971. https://doi.org/10.1017/S0954579413000308

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychological Science*, *16*, 298–304. https://doi.org/10.1111/j.0956-7976.2005.01531.x

Bunce, J., Soderstrom, M., Bergelson, E., Rosemberg, C., Stein, A., Alam F., Migdalek, M. J., & Casillas, M. (2025). A cross-cultural examination of young children's everyday language experiences. *Journal of Child Language*, *52*, 786–814. https://doi.org/10.1017/S030500092400028X

Canfield, C. F., Seery, A., Weisleder, A., Workman, C., Brockmeyer Cates, C., Roby, E., Payne R., Levine S., Mogilner L., Dreyer B., & Mendelsohn, A. (2020). Encouraging parent–child book sharing: Potential additive benefits of literacy promotion in health care and the community. *Early Childhood Research Quarterly*, *50*, 221–229. https://doi.org/10.1016/j.ecresq.2018.11.002

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 & Development*, *37*, 5–15. https://doi.org/10.1016/j.infbeh.2013.10.004

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

Casillas, M., Brown, P., & Levinson, S. C. (2021). Early language experience in a Papuan community. *Journal of Child Language*, *48*, 792–814. https://doi.org/10.1017/S0305000920000549

Chaudhary, N., Salali, G. D., & Swanepoel, A. (2024). Sensitive responsiveness and multiple caregiving networks among mbendjele BaYaka hunter-gatherers: Potential implications for psychological development and well-being. *Developmental Psychology*, *60*, 422–440. https://doi.org/10.1037/dev0001601

Core, C., Chaturvedi, S., & Martinez-Nadramia, D. (2017). *The role of language experience in nonword repetition tasks in young bilingual Spanish-English speaking children*. Proceedings of the 41st Annual Boston University Conference on Language Development (pp. 179–185). Cascadilla Press. http://www.lingref.com/bucld/41/BUCLD41-14.pdf

Cristia, A. (2020). Language input and outcome variation as a test of theory plausibility: The case of early phonological acquisition. *Developmental Review: DR*, *57*, e100914. https://doi.org/10.1016/j.dr.2020.100914

Cristia, A. (2023). A systematic review suggests marked differences in the prevalence of infant-directed vocalization across groups of populations. *Developmental Science*, *26*, e13265. https://doi.org/10.1111/desc.13265

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*, 759–773. https://doi.org/10.1111/cdev.12974

Cristia, A., Foushee, R., Aravena-Bravo, P., Cychosz, M., Scaff, C., & Casillas, M. (2023). Combining observational and experimental approaches to the development of language and communication in rural samples: Opportunities and challenges. *Journal of*

Child Language, *50*, 495–517. https://doi.org/10.1017/S0305000922000617

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

Cychosz, M., & Long, H. L. (2026). Canonical babbling ratio development in infancy: A systematic review and meta-analysis of methodological and ambient language influences. *Developmental Science*, *29*, e70139 https://doi.org/10.1111/desc.70139.

Cychosz, M., Romeo, R. R., Edwards, J. R., & Newman, R. S. (2025a). Bursty, irregular speech input to children predicts vocabulary size. *Developmental Science*, *28*, e13590. https://doi.org/10.1111/desc.13590

Cychosz, M., Villanueva, A., & Weisleder, A. (2021). Efficient estimation of children's language exposure in two bilingual communities. *Journal of Speech, Language, and Hearing Research: JSLHR*, *64*, 3843–3866. https://doi.org/10.1044/2021_JSLHR-20-00755

Cychosz, M., Villanueva, A., & Weisleder, A. (2025b). Bilingual language input to infants in Bolivia and the United States. *Infancy: The Official Journal of the International Society on Infant Studies*, *30*, e70009. https://doi.org/10.1111/infa.70009

Dailey, S., & Bergelson, E. (2022). Language input to infants of different socioeconomic statuses: A quantitative meta-analysis. *Developmental Science*, *25*, e13192. https://doi.org/10.1111/desc.13192

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

de Carvalho, A., Dautriche, I., & Christophe, A. (2016). Preschoolers use phrasal prosody online to constrain syntactic analysis. *Developmental Science*, *19*, 235–250. https://doi.org/10.1111/desc.12300

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

Dixon, S. D., LeVine, R. A., Richman, A., & Brazelton, T. B. (1984). Mother-child interaction around a teaching task: An African-American comparison. *Child Development*, *55*, 1252–1264. https://doi.org/10.2307/1129995

Edwards, J., Munson, B., & Beckman, M. E. (2011). Lexicon–phonology relationships and dynamics of early language development – a commentary on stoel-Gammon's 'relationships between lexical and phonological development in young children'. *Journal of Child Language*, *38*, 35–40. https://doi.org/10.1017/S0305000910000450

Elmlinger, S. L., Goldstein, M. H., & Casillas, M. (2023). Immature vocalizations simplify the speech of Tseltal Mayan and U. S. Caregivers. *Topics in Cognitive Science*, *15*, 315–328. https://doi.org/10.1111/tops.12632

Elmlinger, S. L., Levy, J. A., & Goldstein, M. H. (2025). -02). immature vocalizations elicit simplified adult speech across multiple languages. *Current Biology*, *35*, 871–881. https://doi.org/10.1016/j.cub.2024.12.052

Farabolini, G., Rinaldi, P., Caselli, M. C., & Cristia, A. (2022). Nonword repetition in bilingual children: The role of language exposure, vocabulary scores and environmental factors. Speech. *Language and Hearing*, *25*, 283–298. https://doi.org/10.1080/2050571X.2021.1879609

Ferjan Ramírez, N., 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*, e12762. https://doi.org/10.1111/desc.12762

Ferjan Ramírez, N., Lytle, S. R., & Kuhl, P. K. (2020). Parent coaching increases conversational turns and advances infant language development. *Proceedings of the National Academy of Sciences*, *117*, 3484–3491. https://doi.org/10.1073/pnas.1921653117

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, *16*, 234–248. https://doi.org/10.1111/desc.12019

Foushee, R., & Srinivasan, M. (2024). Infants who are rarely spoken to nevertheless understand many words. *Proceedings of the National Academy of Sciences*, *121*, e2311425121. https://doi.org/10.1073/pnas.2311425121

Foushee, R., Srinivasan, M., & Xu, F. (2021). Self-directed learning by preschoolers in a naturalistic overhearing context. *Cognition*, *206*, 104415. https://doi.org/10.1016/j.cognition.2020.104415

Frye, H. (2022). *Child-directed speech in qaqet: A language of East New Britain, Papua New Guinea* (1st ed). ANU Press. https://doi.org/10.22459/cdsq.2022

García Coll, C., Lamberty, G., Jenkins, R., McAdoo, H. P., Crnic, K., Wasik, B. H., & & García, H. V. (1996). An integrative model for the study of developmental competencies in minority children. *Child Development*, *67*, 1891–1914. https://doi.org/10.1111/j.1467-8624.1996.tb01834.x

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. https://doi.org/10.4324/9781003135517-14

Gilkerson, J., Richards, J. A., Warren, S. F., Montgomery, J. K., Greenwood, C. R., Oller, D. K., & Paul, T. D. (2017). Mapping the early language environment using all-day recordings and automated analysis. *American Journal of Speech-Language Pathology*, *26*, 248–265. https://doi.org/10.1044/2016_AJSLP-15-0169

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, *19*, 515–523. https://doi.org/10.1111/j.1467-9280.2008.02117.x

Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby)Talk to me: The social context of infant-directed speech and its effects on early language acquisition. *Current Directions in Psychological Science*, *24*, 339–344. https://doi.org/10.1177/0963721415595345

Gordon, P., Li, Z., Medeiros, S., Tang, J. E., Bisbee, N., Kirby, E., & Everett, D. (June 2020). *Pirahã infants acquire language without direct input*. Presented at the Association for Psychological Science Virtual Conference, Virtual.

Hamrick, L. R., Seidl, A., & Kelleher, B. L. (2023). Semi-automatic assessment of vocalization quality for children with and without angelman syndrome. *American Journal on Intellectual and Developmental Disabilities*, *128*, 425–448. https://doi.org/10.1352/1944-7558-128.6.425

Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H. Brookes Publishing.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development

via maternal speech. *Child Development*, *74*, 1368–1378. https://doi.org/10.1111/1467-8624.00612

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development*, *73*, 418–433. https://doi.org/10.1111/1467-8624.00415

Hoff-Ginsberg, E. (1986). Function and structure in maternal speech: Their relation to the child's development of syntax. *Developmental Psychology*, *22*, 155–163. https://doi.org/10.1037/0012-1649.22.2.155

Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science*, *11*, F31–F39. https://doi.org/10.1111/j.1467-7687.2008.00768.x

Huttenlocher, J., Vasilyeva, M., Cymerman, E., & Levine, S. (2002). Language input and child syntax. *Cognitive Psychology*, *45*, 337–374. https://doi.org/10.1016/S0010-0285(02)00500-5

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

Keller, H. (2007). *Cultures of infancy*. Lawrence Erlbaum Associates. https://doi.org/10.4324/9781003284079

Keller, H., Borke, J., Staufenbiel, T., Yovsi, R. D., Abels, M., Papaligoura, Z., Jensen H., Lohaus A., Chaudhary N., Lo W., & Su, Y. (2009). Distal and proximal parenting as alternative parenting strategies during infants' early months of life: A cross-cultural study. *International Journal of Behavioral Development*, *33*, 412–420. https://doi.org/10.1177/0165025409338441

Keller, H., Otto, H., Lamm, B., Yovsi, R. D., & Kärtner, J. (2008). The timing of verbal/vocal communications between mothers and their infants: A longitudinal cross-cultural comparison. *Infant Behavior & Development*, *31*, 217–226. https://doi.org/10.1016/j.infbeh.2007.10.001

Kitamura, C., & Burnham, D. (2003). Pitch and communicative intent in mother's speech: Adjustments for age and sex in the first year. *Infancy: The Official Journal of the International Society on Infant Studies*, *4*, 85–110. https://doi.org/10.1207/S15327078IN0401_5

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

Larson, A. L., Barrett, T. S., & McConnell, S. R. (2020). Exploring early childhood language environments: A comparison of language use, exposure, and interactions in the home and childcare settings. *Language, Speech, and Hearing Services in Schools*, *51*, 706–719. https://doi.org/10.1044/2019_LSHSS-19-00066

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

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.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle, & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403–439). Springer Netherlands. https://doi.org/10.1007/978-94-009-2037-8_16

Long, H. L., Ramsay, G., Griebel, U., Bene, E. R., Bowman, D. D., Burkhardt-Reed, M. M., & Oller, D. K. (2022). Perspectives on the origin of language: Infants vocalize most during independent vocal play but produce their most speech-like vocalizations during turn taking. *PLoS One*, *17*, e0279395. https://doi.org/10.1371/journal.pone.0279395

Loukatou, G., Scaff, C., Demuth, K., Cristia, A., & Havron, N. (2022). Child-directed and overheard input from different speakers in two distinct cultures. *Journal of Child Language*, *49*, 1173–1192. https://doi.org/10.1017/S0305000921000623

Luo, R., Masek, L. R., Alper, R. M., & Hirsh-Pasek, K. (2022). Maternal question use and child language outcomes: The moderating role of children's vocabulary skills and socioeconomic status. *Early Childhood Research Quarterly*, *59*, 109–120. https://doi.org/10.1016/j.ecresq.2021.11.007

Ma, W., Golinkoff, R. M., Houston, D. M., & Hirsh-Pasek, K. (2011). Word learning in infant- and adult-directed speech. *Language Learning and Development: The Official Journal of the Society for Language Development*, *7*, 185–201. https://doi.org/10.1080/15475441.2011.579839

Mahr, T., & Edwards, J. (2018). Using language input and lexical processing to predict vocabulary size. *Developmental Science*, *21*, 1–14. https://doi.org/10.1111/desc.12685

Majorano, M., Vihman, M. M., & DePaolis, R. A. (2014). The relationship between infants' production experience and their processing of speech. *Language Learning and Development: The Official Journal of the Society for Language Development*, *10*, 179–204. https://doi.org/10.1080/15475441.2013.829740

ManyBabies. (2020). Quantifying sources of variability in infancy research using the infant-directed-speech preference. *Advances in Methods and Practices in Psychological Science*, *3*, 24–52. https://doi.org/10.1177/2515245919900809

Mastin, J. D., & Vogt, P. (2016). Infant engagement and early vocabulary development: A naturalistic observation study of Mozambican infants from 1;1 to 2;1. *Journal of Child Language*, *43*, 235–264. https://doi.org/10.1017/S0305000915000148

McCune, L., & Vihman, M. M. (2001). Early phonetic and lexical development: A productivity approach. *Journal of Speech Language and Hearing Research*, *44*, 670–684. https://doi.org/10.1044/1092-4388(2001/054)

Mendelsohn, A. L., Cates, C. B., Weisleder, A., Berkule, S. B., & Dreyer, B. P. (2013). Promotion of early school readiness using pediatric primary care as an innovative platform. *Zero to Three*, *34*, 29–40.

Molemans, I., Van Den Berg, R., Van Severen, L., & Gillis, S. (2012). How to measure the onset of babbling reliably? *Journal of Child Language*, *39*, 523–552. https://doi.org/10.1017/S0305000911000171

Morgan, L., & Wren, Y. E. (2018). A systematic review of the literature on early vocalizations and babbling patterns in young children. *Communication Disorders Quarterly*, *40*, 3–14. https://doi.org/10.1177/1525740118760215

Nathani, S., Ertmer, D. J., & Stark, R. E. (2006). Assessing vocal development in infants and toddlers. *Clinical Linguistics & Phonetics*, *20*, 351–369. https://doi.org/10.1080/02699200500211451

Ochs, E. (1988). *Culture and language development: Language acquisition and language socialization in a Samoan village*. Cambridge University Press. https://doi.org/10.2307/414735

Ochs, E., & Kremer-Sadl, T. (2020). Ethical blind spots in ethnographic and developmental approaches to the language gap

debate. *Langage et société*, *170*, 39–67. https://doi.org/10.3917/ls.170.0039

Ochs, E., & Kremer-Sadlik, T. (2015). How postindustrial families talk. *Annual Review of Anthropology*, *44*, 87–103. https://doi.org/10.1146/annurev-anthro-102214-014027

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

Oller, D. K., Eilers, R. E., Basinger, D., Steffens, M. L., & Urbano, R. (1995). Extreme poverty and the development of precursors to the speech capacity. *First Language*, *15*, 167–187. https://doi.org/10.1177/014272379501504403

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. https://doi.org/10.1016/S0021-9924(99)00013-1

Padilla-Iglesias, C., Woodward, A. L., Goldin-Meadow, S., & Shneidman, L. (2021). Changing language input following market integration in a Yucatec Mayan community. *PLoS One*, *16*, e0252926. https://doi.org/10.1371/journal.pone.0252926

Parra, M., Hoff, E., & Core, C. (2011). Relations among language exposure, phonological memory, and language development in Spanish-English bilingually-developing two-year-olds. *Journal of Experimental Child Psychology*, *108*, 113–125. https://doi.org/10.1016/j.jecp.2010.07.011

Place, S., & Hoff, E. (2016). Effects and noneffects of input in bilingual environments on dual language skills in 2 /-year-olds. *Bilingualism: Language and Cognition*, *19*, 1023–1041. https://doi.org/10.1017/S1366728915000322

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2014). 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*, 880–891. https://doi.org/10.1111/desc.12172

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2017). The impact of early social interactions on later language development in Spanish-English bilingual infants. *Child Development*, *88*, 1216–1234. https://doi.org/10.1111/cdev.12648

Richman, A. L., Miller, P. M., & LeVine, R. A. (1992). Cultural and educational variations in maternal responsiveness. *Developmental Psychology*, *28*, 614–621. https://doi.org/10.1037/0012-1649.28.4.614

Rowe, M. L. (2008). Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language*, *35*, 185–205. https://doi.org/10.1017/S0305000907008343

Rowe, M. L., & Snow, C. E. (2020). Analyzing input quality along three dimensions: Interactive, linguistic, and conceptual. *Journal of Child Language*, *47*, 5–21. https://doi.org/10.1017/S0305000919000655

Scaff, C., Casillas, M., Stieglitz, J., & Cristia, A. (2024). Characterization of children's verbal input in a forager-farmer population using long-form audio recordings and diverse input definitions. *Infancy: The Official Journal of the International Society on Infant Studies*, *29*, 196–215. https://doi.org/10.1111/infa.12568

Schick, J., Daum, M. M., & Stoll, S. (2025). Input to the language learning infant: The impact of other children. *Developmental Science*, *28*, e70045. https://doi.org/10.1111/desc.70045

Schieffelin, B. B. (1990). *The give and take of everyday life: Language socialization of Kaluli children*. Cambridge University Press. https://doi.org/10.1525/jlin.1992.2.1.122

Schieffelin, B. B., & Ochs, E. (1986). Language socialization. *Annual Review of Anthropology*, *15*, 163–191. https://doi.org/10.1146/annurev.an.15.100186.001115

Shneidman, L., Arroyo, M. E., Levine, S. C., & Goldin-Meadow, S. (2013). What counts as effective input for word learning? *Journal of Child Language*, *40*, 672–686. https://doi.org/10.1017/S0305000912000141

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

Shneidman, L., & Woodward, A. L. (2016). Are child-directed interactions the cradle of social learning? *Psychological Bulletin*, *142*, 1–17. https://doi.org/10.1037/bul0000023

Singh, L., Cheng, Q., & Yeung, W. J. J. (2023). Effects of socio-economic status on infant native and non-native phoneme discrimination. *Developmental Science*, *26*, e13351. https://doi.org/10.1111/desc.13351

Snow, C. E. (1972). Mothers' speech to children learning language. *Child Development*, *43*, 549–565. https://doi.org/10.2307/1127555

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review: DR*, *27*, 501–532. https://doi.org/10.1016/j.dr.2007.06.002

Soderstrom, M., Grauer, E., Dufault, B., & McDivitt, K. (2018). Influences of number of adults and adult: Child ratios on the quantity of adult language input across childcare settings. *First Language*, *38*, 563–581. https://doi.org/10.1177/0142723718785013

Sperry, D. E., Sperry, L. L., & Miller, P. J. (2019a). Language does matter: But there is more to language than vocabulary and directed speech. *Child Development*, *90*, 993–997. https://doi.org/10.1111/cdev.13125

Sperry, D. E., Sperry, L. L., & Miller, P. J. (2019b). Reexamining the verbal environments of children from different socioeconomic backgrounds. *Child Development*, *90*, 1303–1318. https://doi.org/10.1111/cdev.13072

Tamis-LeMonda, C. S., Kuchirko, Y., & Song, L. (2014). Why is infant language learning facilitated by parental responsiveness? *Current Directions in Psychological Science*, *23*, 121–126. https://doi.org/10.1177/0963721414522813

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy: The Official Journal of the International Society on Infant Studies*, *7*, 53–71. https://doi.org/10.1207/s15327078in0701_5

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, *57*, 1454–1463. https://doi.org/10.2307/1130423

Tronick, E. Z., Morelli, G. A., & Winn, S. (1987). Multiple caretaking of Efe (Pygmy) infants. *American Anthropologist*, *89*, 96–106. https://doi.org/10.1525/aa.1987.89.1.02a00050

Van Kleeck, A. (1994). Potential cultural bias in training parents as conversational partners with their children who have delays in language development. *American Journal of Speech-Language Pathology*, *3*, 67–78. https://doi.org/10.1044/1058-0360.0301.67

Vanormelingen, L., Faes, J., & Gillis, S. (2020). Language development in children from different SES backgrounds: Babbling onset and consonant characteristics. *Dutch Journal of Applied*

Linguistics, *9*, 132–161. https://doi.org/10.1075/dujal.19032.van

Vogt, P., Mastin, J. D., & Schots, D. M. A. (2015). Communicative intentions of child-directed speech in three different learning environments: Observations from The Netherlands, and rural and urban Mozambique. *First Language*, *35*, 341–358. https://doi.org/10.1177/0142723715596647

Vorperian, H. K., Kent, R., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., & Yandell, B. S. (2005). Development of vocal tract length during early childhood: A magnetic resonance imaging study. *The Journal of the Acoustical Society of America*, *117*, 338–350. https://doi.org/10.1121/1.1835958

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*, 1314–1324. https://doi.org/10.1177/0956797614531023

Weber, A., Fernald, A., & Diop, Y. (2017). When cultural norms discourage talking to babies: Effectiveness of a parenting program in rural Senegal. *Child Development*, *88*, 1513–1526. https://doi.org/10.1111/cdev.12882

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

Weisleder, A., & Mendelsohn, A. (2019). Daylong recordings of 2-12 month-old infants from Spanish-speaking homes in the U. S. https://doi.org/10.17605/OSF.IO/JBTNC

Weisner, T. S. (1987). Socialization for parenthood in sibling caretaking societies. In J. B. Lancaster, J. Altmann, A. S. Rossi, & L. Sherrod (Eds.), *Parenting across the life span* (pp. 237–270). Routledge. https://doi.org/10.4324/9781315126005-12

Weisner, T. S., & Gallimore, R. (1977). My brother's keeper: Child and sibling caretaking. *Current Anthropology*, *18*, 169–190. https://doi.org/10.1086/201883

Xu, D., Yapanel, U., & Gray, S. (2009). *Reliability of the LENA language environment analysis system in young children's Natural home environment* (technical report lTR-05-2). LENA Research Foundation. https://www.lena.org/wp-content/uploads/2016/07/LTR-05-2_Reliability.pdf

Yurovsky, D. (2018). A communicative approach to early word learning. *New Ideas in Psychology*, *50*, 73–79. https://doi.org/10.1016/j.newideapsych.2017.09.001

Zentella, A. C. (1998). *Growing up bilingual: Puerto rican children in New York*. Blackwell.