A cross-linguistic study of the effect of early experience on vocal development

Kasia Hitczenko (kasia.hitczenko@gmail.com)[1,*], Elika Bergelson (elika.bergelson@duke.edu)[2], Marisa Casillas (mcasillas@uchicago.edu)[3], Heidi Colleran (heidi_colleran@eva.mpg.de)[4], Margaret Cychosz (mcychosz@umd.edu)[5], Pauline Grosjean (p.grosjean@unsw.edu.au)[6], Lisa R. Hamrick (lrague@purdue.edu)[7], Bridgette L. Kelleher (bkelleher@purdue.edu)[7], Camila Scaff (camila.scaff@iem.uzh.ch)[1,8], Amanda Seidl (aseidl@purdue.edu)[9], Sarah Walker (s.walker@unsw.edu.au)[6], Alejandrina Cristia (alecristia@gmail.com)[1]

1. Laboratoire de Sciences Cognitives et Psycholinguistique, Département d'Etudes Cognitives, ENS, EHESS, CNRS, PSL University, 2. Department of Psychology and Neuroscience, Duke University, 3. Comparative Human Development Department, University of Chicago, 4. BirthRites Independent Max Planck Research Group, Department of Human Behavior, Ecology and Culture, Max Planck Institute for Evolutionary Anthropology, 5. Department of Hearing and Speech Sciences & Center for Comparative and Evolutionary Biology of Hearing, University of Maryland, College Park, 6. School of Economics, University of New South Wales, CEPR, 7. Department of Psychological Sciences, Purdue University, 8. Human Ecology Group, Institute of Evolutionary Medicine, University of Zurich, 9. Department of Speech, Language, and Hearing Sciences, Purdue University

Acknowledgments: *Funding Statement:* Funding for this work comes from the Agence Nationale de la Recherche, the J. S. McDonnell Foundation Understanding Human Cognition Scholar Award, and the European Research Council under the European Union's Horizon 2020 research and innovation programme. *Conflict of Interest Statement:* The authors have no conflict of interests to declare. *Data Availability Statement:* Upon completion of the proposed analyses, all code and data will be made available on the first author's Github/OSF page: https://github.com/khitczenko and https://osf.io/tjdg7/. *Ethics Approval Statement:* The necessary funding, facilities, and approvals are in place for the proposed research and the proposed research could commence immediately after an In Principle Acceptance. We agree to register the approved protocol on the Open Science Framework following a Stage 1 acceptance. We confirm that if we later withdraw our paper, we will publish a short 500-word summary that describes the hypotheses and intended methods of the pre-registered studying, including the reasons for manuscript withdrawal.

*Correspondence concerning this article should be addressed to Kasia Hitczenko, 29 rue d'Ulm, Paris 75005, France. E-mail: kasia.hitczenko@ens.psl.edu

Proposal Research Highlights

- We study one aspect of vocal development by examining the prevalence of consonant-vowel/vowel-consonant transitions in child speech via a measure known as canonical proportion that can be extracted from naturalistic recordings of a large cross-cultural and cross-linguistic sample.
- We ask how a child's experience - namely, speech input quantity, contingent adult vocalizations, and child vocalization practice - relates to this measure of vocal development.
- [Highlight will be inserted based on results of study]
- [Highlight will be inserted based on results of study]

Abstract

Children exhibit substantial vocal development in their first years of life, laying the foundation for later language. This paper asks how key aspects of their linguistic experience relates to one aspect of this development. We focus on three factors that have been argued to influence vocal development: the amount of language input that infants hear, the likelihood of hearing adult vocalizations contingent on infants' productions, and the amount of time children spend vocalizing. We test the relationship between these three factors and one measure of vocal development (canonical proportion, which estimates the prevalence of transitions between consonants and vowels in spontaneous speech) using long-form recordings in a large sample. The children in the dataset (N = 204, aged 2-18 months) were growing up in diverse cross-cultural urban and rural settings, and were learning one or more typologically-varied languages. With this well-powered, diverse sample, we predict that we will see significant positive relationships between each of the factors explored and this measure of vocal development. [Description of results will be inserted here]

Keywords: vocal development; canonical proportion; input quantity; social feedback; vocalization practice; citizen science; long-form recordings

A cross-linguistic study of the effect of early experience on vocal development

1. Introduction

Infants exhibit substantial phonological development in their first year of life. For example, their speech perception becomes attuned to the language(s) they are learning (Werker & Tees, 1984) and their vocal productions mature and become more speech-like, forming the basis for meaningful language (Oller, 2000). While the trajectory of early phonological learning is well-established within certain cultural and linguistic contexts (Vihman, 2014), it is less clear how much variability children exhibit, and which experiential factors, if any, reliably promote or delay different aspects of early vocal development. For instance, Cristia (2020)'s review highlights that different theories make widely diverse predictions in terms of the extent of individual and group variation in phonological development, and concludes that evidence suggested group variation was small. This conclusion was based on work that may not have been in the best position to detect differences, because it studied development in a small and homogeneous group of children (e.g., primarily children in US or English-speaking households: Eilers et al., 1993). Indeed, recent research studying a cross-culturally and cross-linguistically diverse sample suggests there may be much more variability in some aspects of early phonological development across individual children, languages, and cultures than previously thought (Hitczenko et al., 2023). Moreover, it appears to us that Cristia's assumption that phonological development is monolithic was incorrect. Using more precise definitions of phonological development, other work predicts that some milestones of phonological development (e.g., age of onset of canonical babbling, which reflects articulatory/motor limitations of the developing vocal tract) may be robust to

experience while others will be experience-dependent (e.g., frequency of babbling, which is less affected by motor limitations; Oller, 2000). Additional theorizing has helped tease apart different dimensions of infants' experiences, distinguishing at least the effects of speech by others, the child's own vocal exploration, and the contingency of others' reactions to the infant's vocalization (e.g., Warlaumont et al., 2014; Ritwika et al., 2020). The resulting picture is complex, with various aspects of children's phonological development and experiences represented in a variety of proxies, which, together with limited sample sizes, means that much research remains to be done to draw more nuanced generalizations with respect to how certain experiences shape specific aspects of infants' vocal development.

In this paper, we focus on one measure of children's phonological development, namely, "canonical proportion", which estimates the proportion of a child's speech-like vocalizations that contain consonant-vowel/vowel-consonant transitions, and is thought to capture an important aspect of children's *vocal development*. Specifically, we ask how this measure is affected by three proximate aspects of experience - the quantity of speech input, contingent adult vocalizations, and the child's own vocalizations. We study this using naturalistic recordings collected from a large, diverse cross-cultural sample, which is critical for measuring how stable or variable this development is across different learning environments.

1.1. Definition of canonical proportion and current evidence on its development

Vocal development encompasses a broad set of skills, including the development of a stable inventory of sounds (e.g., vocal motor schemes, e.g., McCune & Vihman,

2001), the development of motor routines allowing fast transitions between sounds, and many other abilities. Here, we focus specifically on the development of canonical (consonant-vowel/vowel-consonant) transitions, one key aspect of vocal development that has been extensively studied in previous work, as summarized next.

While speech-like vocalizations are dominant even at birth (they outnumber cries 5:1; Oller et al., 2021), they exhibit important changes over the first years of life. Early on, both consonants and vowels appear in infants' productions but the timing of phonation and speech articulators' movement is such that most syllables are what Oller et al. (1994) described as marginal, to distinguish them from canonical syllables, which have adult-like, fast and smooth transitions -- which we will call "canonical" transitions. By about 7 months, infants' productions begin to include canonical transitions in Consonant-Vowel: [bɑ] “ba”, Vowel-Consonant: [ʌp] “up”, and other syllable shapes (Eilers et al., 1993; Morgan & Wren, 2018; Oller, 2000).

A number of different metrics have been used to study the emergence and prevalence of canonical transitions (a discussion in Lang et al., 2019). *Canonical babbling onset* is defined as the age at which infants begin to produce canonical syllables. *Canonical babbling rate/ratio* measures the rate of syllables with well-timed consonant-vowel transitions (Oller et al., 1994), and has been found to reach 15% by 10 months (Nyman et al., 2021). The focus of our study is on a metric proposed by Cychosz et al. (2021a; building on work such as Oller et al., 1994), which is called *canonical proportion*, and estimates the proportion of vocalizations that contain consonant-vowel transitions. This metric is highly-related to canonical babbling rate/ratio, with the slight difference that the prevalence of canonical transitions is

estimated over vocalization clips rather than syllables. We choose to adopt this metric because it can be easily measured from long-form recordings (i.e., without having to identify the onset and offset of each syllable) and thus greatly increases the number and diversity of children we can study.

Despite the fact that Cychosz et al. (2021a) used a different definition and process than Oller et al. (1994) and Nyman et al. (2021), they nonetheless found their canonical proportion metric increased from an average 7% at 1-6 months to 15% at 7-12 months, which aligns with results from Nyman et al. (2021) and many others. Less is known about how the other metrics change beyond 12 months, but two studies suggest canonical proportion continues to increase beyond this age (Cychosz et al., 2021a; Hitczenko et al., 2023).

Intriguingly, the age of onset and prevalence of canonical transitions have been found to predict later language outcomes (e.g., Chapman et al., 2003; Lang et al., 2019; McDaniel et al., 2019; Oller et al., 1998, 1999).[1] For example, Oller et al. (1999) found that infants with delayed canonical babbling onset had smaller expressive vocabularies at 2-3 years of age, and McDaniel et al. (2019) and Yankowitz et al. (2022) both found a very strong relationship between canonical babbling ratio and expressive language measures one year later in children with autism spectrum disorder (but see Lang et al., 2021 and Fagan, 2009 who, respectively, found that canonical babbling onset and reduplicated babbling onset - i.e., producing utterances with *multiple* canonical syllables - did not predict age of word onset within children without delays). In addition, many studies have shown that infants with delayed or reduced canonical productions are

[1] Other aspects of early vocal development also predict later language, in literature too extensive to summarize here (e.g., vocal motor schemes, or the size of the stable consonant inventory: e.g., McCune & Vihman, 2001; Majorano et al., 2014; McGillion et al., 2017).

more likely to have a speech, language, developmental, and/or genetic disorder (Lohmander et al., 2017; Oller et al., 1998). In particular, delayed or deviant canonical transitions have been reported in Williams syndrome (e.g., Masataka, 2001), autism spectrum disorder (e.g., Patten et al., 2014; Yankowitz et al., 2022; Lang et al., 2019), Rett syndrome (e.g., Marschik et al., 2012, 2013; Roche et al., 2018; Lang et al., 2019), Fragile X syndrome (e.g., Belardi et al., 2017; Lang et al., 2019; Marschik et al., 2014), Angelman syndrome (e.g., Semenzin et al., 2021), and many other conditions (e.g., Chapman et al., 2009; Eilers & Oller, 1994; Lynch et al., 1995; Moeller et al., 2007; Overby & Caspari, 2015; Sohner & Mitchell, 1991).

1.2. Factors hypothesized to predict canonical proportion

The upward trajectory of children's canonical proportion over the first years of life is well-documented (Cychosz et al., 2021a; Hitczenko et al., 2023; Peute & Casillas, 2022) and consistent with research on related metrics (Fagan, 2009; Oller et al., 1997; Warlaumont & Ramsdell, 2016), but it is less clear what factors, if any, are associated with its development cross-linguistically or cross-culturally. We focus on three candidate factors that existing theories of vocal and language development suggest are likely to play a role[2]: quantity of language input (e.g., Marklund et al., 2019; Weisleder & Fernald, 2013), social feedback from caregivers, specifically how often the child hears adult vocalizations contingent on their productions (e.g., Goldstein et al., 2003; Gros-Louis et al., 2006; Warlaumont et al., 2014), and child vocalization

[2] We are inspired by work which shows that certain aspects of both developmental pathways and outcomes of the human-universal skills of walking may vary as a function of experience (Karasik et al., 2010; Karasik & Kuchirko, 2022). In language as well, there is ongoing discussion on the ways in which development of various linguistic skills varies as a function of experiences (e.g., Goldin-Meadow, 2014; Kidd & Garcia, 2022; Lopez et al., 2020; Oller et al., 1985; Vihman et al., 1985; Weisleder & Fernald, 2013).

experience/practice (e.g., Long et al., 2020; Ritwika et al., 2020), each detailed below. These three factors are not mutually-exclusive and, indeed, have often been studied together (e.g., Long et al., 2022). Nonetheless, they do represent three conceptually distinct contributions, which we aim to dissociate in this work. The degree to which they can be empirically separated or measured is a question we will return to in the results/discussion.

1.2.1. Input quantity

The first factor that we hypothesize relates to vocal development is the amount of language input that children hear. There is ample evidence from other areas of language development research that input quantity is correlated with language learning (e.g., word comprehension and processing speed: Shneidman & Goldin-Meadow, 2012; Weisleder & Fernald, 2013; grammatical complexity: Huttenlocher et al., 2002). Consequently, it is reasonable to predict that input quantity would also relate to early vocal development. At the same time, phonological[3] development could be less "input-hungry" than e.g. lexical development for a few reasons, including that there are fewer "bits" of information to be learned (Mollica & Piantadosi, 2019; see Cristia, 2020 for a detailed discussion).

Indeed, prior studies suggest that phonological development is less sensitive to input experience than other language domains (summarized in Cristia, 2020, but see

[3] We use the term "phonological development" to refer to, broadly, how children learn about the sounds of their language(s), including what they are, how they are organized, how to perceive them, and how they are produced. We use the term "vocal development" to refer more specifically to how children learn to produce the sounds of their language(s). In this Introduction, we consider vocal development to be an instance of phonological development and we use previous work studying the relationship between input quantity and all forms of phonological development to inform our predictions (though, of course, it need not be the case that input quantity affects all forms of phonological development in the same way).

Marklund et al., 2019 and Cychosz et al., 2021b). Within the subdomain of vocal development, this relative resilience has been argued for through proxies of input quantity (e.g., socioeconomic status, which is correlated with input quantity; Dailey & Bergelson, 2021), rather than by directly studying the correlation between input quantity and vocal development. For example, the onset of canonical babbling was found to be similar in children from low vs. high socioeconomic status backgrounds (Eilers et al., 1993; N=49 split between high versus low socioeconomic status crossed with preterm vs. full-term) and children living in extreme poverty (Oller et al., 1995; N=41). More recently, Cychosz et al. (2021a) found similar canonical proportions in a cross-cultural sample of 49 children, including rural and urban populations, which have been reported to differ in child-directed speech input quantity (Cristia, 2022)[4]. Similarly, Peute & Casillas (2022) found that Tseltal-learning children (N=20) and Yélî Dnye-learning children (N=12) had similar canonical proportions and similar vocal motor schemes (i.e., sizes of their stable consonant inventories) both as compared to each other and as compared to English-learning children from previous studies, despite reported differences in child-directed input quantities between all three communities.

Nonetheless, the conclusion that canonical transitions are robust to input quantity variation may be premature, particularly given the relatively small sample sizes in prior work. Indeed, if input effects on canonical transitions are similar to those in other areas of language (r=.21 in a meta-analysis of input quantity and standardized language measures, Wang et al., 2020), a well-powered study would need to include nearly 200 participants which none, to our knowledge, have done.

[4] Throughout the paper, we use the short-hand term “rural” to refer to rural, small-scale, subsistence-level populations and the term “urban” to urban or suburban, industrialized, or post-industrial populations. We refer the reader to Cristia (2022) for more discussion on this distinction and terminology.

In addition, the proxies for input quantity have been called into question since the publication of those studies. For instance, when Cychosz et al., (2021a) was published, evidence suggested that the included populations differed in how much people talked to children: the American English infant learners in that study were thought to hear several magnitudes more infant-directed speech than the Tsimane', Yélî, and Tseltal learners (Casillas et al., 2020, 2021; Cristia et al., 2019), so the absence of an effect of culture was interpreted as evidence that input quantity does not play a large role in vocal development. However, more recent work has suggested overlap in total input quantities across American English, Yélî, and Tseltal children when measured from long-form recordings (Bunce et al., 2021), calling the initial design and conclusion into question. Indeed, when the link between input quantity and other aspects of phonology has been directly studied, researchers have found significant effects between the two (e.g., between input quantity and development of phonological working memory in Cychosz et al., 2021b).

Taken together, while past work generally argues that the development of canonical transitions proceeds similarly regardless of input quantity, few studies have actually directly studied this link in a way that would allow researchers to detect the effect given the sample sizes. Consequently, the relationship between input quantity and canonical proportion merits dedicated and direct examination, using a large corpus with sufficient population and input variability.

1.2.2. Social feedback: Likelihood of adult vocalizations contingent on child productions

A second factor that could predict children's vocal development is social feedback – specifically, how often children hear contingent adult vocalizations following their productions. The idea, as proposed for example by Bloom et al. (1987) and explored by a host of work states that adults may respond preferentially to more mature child vocalizations (e.g., speech-like over non-speech-like, or canonical over non-canonical; Albert et al., 2018; Gros-Louis et al., 2006; Warlaumont et al., 2014, see also McGillion et al., 2013). This may, in turn, lead infants to produce more advanced vocalizations, either via reinforcement or in seeking interactions with their caregivers, which would further reinforce caregiver behaviors in a social feedback loop. This view predicts that children whose caregivers more frequently reward advanced vocalizations with infant-directed input should have a higher canonical proportion than caregivers whose feedback does not distinguish between vocalization types or does so less frequently or less systematically.

There is evidence supporting this "social feedback" hypothesis in children as young as 5 months (Elmlinger et al., 2022). Warlaumont et al. (2014) found that caregivers preferentially responded to speech-like vocalizations and that infants were more likely to vocalize when their previous vocalization was met with a contingent, adult response. Similarly, Gros-Louis & Miller (2018) found that children were more likely to produce Consonant-Vowel vocalizations, rather than Vowel vocalizations, if their caregiver had just given them contingent feedback on previous Consonant-Vowel vocalizations (though this was only true of 12-month-olds, not 10-month-olds). Bloom et al. (1987) and Hsu & Fogel (2001) found that infants were more likely to produce speech-like syllabic vocalizations (relative to non-speech-like vocalic vocalizations)

when engaged in give-and-take turn-taking interactions that involved symmetric engagement from infants and caregivers. Finally, in controlled lab experiments, Goldstein et al. (2003) and Goldstein & Schwade (2008) found that 8-10 month-old children produced more and more advanced vocalizations when parents were instructed to reward child vocalizations (either by touching them or responding vocally), relative to a control condition where parents engaged with children the same amount, but without regard to their vocalizations.

That being said, Fagan & Doveikis (2017) failed to find evidence for this social feedback hypothesis in naturalistic mother-infant home interactions (in English). They found that while mothers preferentially responded to speech-like over non-speech-like child vocalizations, they were equally likely to respond to Consonant-Vowel vs. Vowel child vocalizations and that maternal responses did not affect children's subsequent vocalizations.

Taken together, however, this research suggests a possible influence of social feedback on vocal development, which we test in a large cross-linguistic/cultural corpus that directly links contingency rates and canonical proportion.

1.2.3. Child vocalization practice

We also study a third factor, namely child vocalization experience. The idea is that, like elsewhere, practice makes perfect: children's vocalizations become more advanced in proportion to how much they vocalize. This could lead to improvements for a number of reasons. First, vocalizing allows children to practice required speech motor coordination and adjust their vocalizations based on self-auditory feedback (Goffman,

1999; Vihman, 2013). Vocalizing also elicits caregiver productions which could serve as production models for future vocalizations (Moulin-Frier et al., 2014). When applied to canonical proportion, this theory would predict that the more children vocalize, the more advanced their vocalizations become and the higher their canonical proportion is (once other factors, such as age, are controlled for).

The primary evidence for the child vocalization experience hypothesis comes from the fact that there is a substantial non-social function to children's early vocalizations. Children vocalize at roughly the same rate when alone vs. with their caregiver (Oller et al., 2019) and, even when in the company of a caregiver, most child vocalizations are not social in nature (as judged by an experimenter on the basis of a number of factors e.g. gaze, body position, timing, etc.; Long et al., 2020; Oller et al., 2013), though this has only been studied in children growing up in the US. While these findings do not necessarily imply a practice effect, they do suggest that children have substantial intrinsic and non-social motivation to vocalize.

Another line of research suggests that children with more production experience have more advanced vocal/linguistic development (see Vihman, 2022 for a comprehensive overview of this line of research). For example, the more practice a child has with producing a particular sound, the more likely they are to use that sound to match objects in their environment (e.g., produce 'ba' near a ball) (Laing & Bergelson, 2020), the more likely it is that their first words will have that sound (Vihman et al., 1985), and the better their phonological memory is for that sound (Keren-Portnoy et al., 2010). Similarly, the more overall vocalization practice a child has, the more mature their speech tends to be (as measured by how much they coarticulate, relative to adults;

Cychosz et al., 2021c). Taken together, these findings suggest that infants exhibit significant internal motivation to vocalize and that their production experience significantly impacts later phonological development.

1.3. A cross-cultural study using long-form recordings

We aim to address limitations of previous work concerning the three hypotheses. To begin with, while some predictions from them have been born out (e.g., showing that social feedback affects children's behavior locally), no previous work has directly linked all three factors to overall vocal development. In addition, previous research has tended to focus on English-speaking children raised in urbanized societies. Children raised in different communities may differ in how much social feedback they are exposed to, how that feedback is conveyed (e.g., through speech, touch, gaze, etc.), as well as how much time they spend vocalizing on their own versus interacting with or hearing input from others, which could change the effects observed. For this reason, it will be important to test whether these effects are stable across cultures.

The importance of studying language acquisition as it happens in the real world has led to a rise in the use of long-form recordings (e.g., Bunce et al., 2020) and we adopt this approach here. Long-form recordings are collected by equipping children with non-invasive Language ENvironment Analysis (LENA) audio recorders (e.g., Oller et al., 2010; VanDam & Yoshinaga-Itano, 2019; Wang et al., 2020) or other small, lightweight audio recorders (Casillas et al., 2020, 2021; Cassar et al., 2021; Cristia & Colleran, 2018) that children wear over the course of a normal day. This approach has many benefits, especially for the study of vocal development. First, naturalistic recordings

offer researchers an ecologically valid way of measuring child vocalizations and caregiver input, without having to rely on parental reports, lab observations, or short recordings which have been shown to underestimate children's vocalization rates (Lewedag et al., 1994) and overestimate input (Bergelson et al., 2019). Second, automating and crowd-sourcing the annotation of the recordings reduces the time needed to analyze the large volumes of data collected and permits larger sample sizes than would be possible with traditional methods. Finally, due to the portability of the recorders, this method allows standardized data collection across a broad range of settings, enabling researchers to study cultures and populations where behavioral research would not otherwise be possible. Indeed, through adopting this approach, we are able to study a large sample of ~200 children, exhibiting significant cultural and linguistic diversity. The children in our sample speak 40 languages, differing in their phonological properties (e.g., whether they feature complex syllables), which affects children's early vocalizations (de Boysson-Bardies & Vihman, 1991). And they are raised in varied environments (e.g., rural vs. urban; across 7 countries and 4 continents), which can impact how much input and social interaction they experience (Cristia, 2022). By studying a culturally and linguistically diverse sample, we are better-equipped to test the role of experience in vocal development.

1.4. Testing the relationship between experience and the development of canonical proportion

In this paper, we ask how experience - in the form of input quantity, contingency of adult vocalizations, and child vocalization practice - relates to one aspect of vocal

development, by studying a large, cross-cultural sample of naturalistic, child-centered speech experiences.

We test the following hypotheses:

- Hypothesis 1 (Input quantity): Children who hear more adult speech input will have a higher canonical proportion.
- Hypothesis 2 (Social feedback): Two reasonable proxies of social feedback exist, leading to two variants of this hypothesis:
    - Hypothesis 2a (Social feedback specificity): Children who hear a relatively higher proportion of adult vocalizations contingent on their speech-like productions than their non-speech-like productions will have a higher canonical proportion.
    - Hypothesis 2b (Social feedback frequency): Children who hear a larger number of adult vocalizations contingent on their own speech-like productions will have a higher canonical proportion. In other words, this hypothesis differs from Hypothesis 2a, by stating that it is the overall frequency of contingent responses that promotes development, rather than proportional differences in what types of vocalizations are met with responses (e.g., the idea is that a large number of contingent responses is more helpful than a small number of contingent responses that preferentially reward advanced vocalizations to a large extent).
- Hypothesis 3 (Child vocalization experience): Children who practice vocalizing more (measured as the total duration of their speech-like vocalizations) will have a higher canonical proportion.

2. Methods

2.1. Data corpora

We will use nine corpora of long-form recordings from a total of 204 children who are reported to be typically-developing.

- English-Bergelson (N=44): We will include data from 44 children aged 6-18 months learning American English in the Rochester, NY area, USA, North America (Bergelson, 2017; Bergelson et al., 2019; Bergelson & Aslin, 2017).
- English-Seidl (N=10): We will include data from 10 children aged 4-18 months learning American English in Indiana, USA, North America (Semenzin et al., 2021).
- French-Canault (N=9): We will include data from 9 children aged 3-18 months growing up in a monolingual context in Lyon, France, Europe (Canault et al., 2016).
- French-Cristia (N=10): We will include data from 10 children aged 11-12 months growing up in Paris, France, Europe (Cristia, 2021). Six of these children are French monolinguals, 3 are bilinguals, and for one this information is missing.
- Quechua (N=7): We will include data from 7 bilingual children aged 5-13 months, learning Quechua and Spanish in the south Bolivian highlands of South America (Cychosz, 2018).
- Solomon Islands (N=43): We will include data from 43 children aged 4-18 months growing up on the Solomon Islands in Oceania (Cassar et al., 2021). These children are raised in multilingual environments, each learning a subset of

Roviana, Avaso, Babatana, Marco, Marovo, Pidjin, Senga, Simbo, Sisinga, Ughele, Vaghua, and Varisi.

- Tseltal (N=37). We will include data from 37 children aged 2-18 months learning Tseltal monolingually. These data were collected in a rural subsistence farming community in the Chiapas highlands of southern Mexico in North America (Casillas et al., 2017, 2020).
- Tsimane' (N=21): We will include data from 21 children aged 5-18 months learning Tsimane' in two villages in the lowlands of Northern Bolivia, South America (Scaff et al., 2018, 2019).
- Vanuatu (N=12): We will include data from 12 children aged 5-16 months growing up in a multilingual environment in Malekula, Vanuatu in Oceania (Cristia & Colleran, 2018). The represented languages are Bislama, Venen Taut, Petarmul, Neverver, Uripiv, Vinmavis, Francis, Novol, Epi, Nah'ai, Paama, Ninde, Tautu, French, Pinalum, Malo, Rano,Tauta, Santo Language, Ambae, Maevo, South, Atchin, and Tempun.
- Yélî Dnye (N=20): We will include data from 20 children aged 3-16 months learning Yélî Dnye on the Rossel Islands in Papua New Guinea (Cristia & Casillas, 2019).

From this set, we will only include data from one child per family to guarantee independence between samples, chosen randomly but ensuring balanced coverage across ages/sexes. Many of the children have multiple recordings from different days/ages. For each child, we will randomly select two annotated recordings (only one

for those who do not have multiple annotated recordings), again ensuring balanced coverage across ages. The first set of recordings will constitute our main dataset from which we will draw conclusions, while the second set of recordings will be used as an "exploration set" to help finalize our model structure, before analyzing our main dataset, as will be described in the Methods section. In the process of randomly choosing which child recordings to analyze, any recording with fewer than 4 hours between 6am-9pm will be excluded from consideration to ensure that all measure estimates are based on a sufficiently long sample. We chose 4 hours both because all researchers intended to record at least four hours (the latest recording start times were 5pm), so shorter recording times indicate a malfunction of some sort. We refer the reader to Supplementary Materials[5] SM4 for a compiled list of exclusionary criteria.

2.2. Pre-processing and calculation of canonical proportion (outcome measure)

Figure 1 details our full data-processing pipeline. To simplify exposition, we do not go into details of the data collection (which is described in the papers cited for each included corpus) nor of the pre-processing, since it was done in previous work and is not central to the current study (but see Supplementary Materials, SM2 and SM3 for the most relevant information).

To calculate canonical proportion, we will rely on vocalization type labels obtained in previous work for a subset of each key child's vocalizations (key child = the child wearing the recorder; Cychosz et al., 2021a; Semenzin et al., 2021). That work relied on citizen science, a growing crowd-sourced approach, in which volunteers from the general public assist with research tasks online. Specifically, at least three citizen

[5] Supplementary Materials: https://osf.io/9e6ch/?view_only=a4bce608850648a9a3d17c6b800d087e

scientists labeled child vocalizations as: (i) a canonical vocalization, (ii) a non-canonical vocalization (a speech-like vocalization that does not have both a consonant and vowel), (iii) laughing, (iv) crying, or (v) junk (a clip without a child vocalization), based on instructions provided in SM2. Not all instructions included reference to fast adult-like transitions, but they all required the presence of both a consonant and a vowel within the ~500ms clip. See Figure 1 and Supplementary Materials for more details.

In the current work, these three (or more) labels will be combined to arrive at one final vocalization type per clip. Each clip will receive the label chosen by the majority of annotators, where a majority is defined as at least 50%, with all other choices receiving less than 50% of the votes. We use the 50% cut-off because of recommendations from the citizen science platforms used, which is thus informed by extensive experience. Moreover, this cut-off was used in the previous study that established a very high correlation between the derived canonical proportion metric used here and laboratory annotations (Semenzin et al., 2021). This would suggest that a higher level of agreement may not be needed when using this method, perhaps because there are more individuals independently making decisions about each clip (often 5, whereas in many lab studies there is only 1, with agreement being decided on a subset of the data, rather than each individual clip); and because we can include more data per child than typical laboratory annotations studies (including vocalizations sampled throughout the whole day, versus those collected in 1-2h observations). We will exclude any clips without majority agreement.

Each child's canonical proportion will be calculated from the citizen scientists' labels as the proportion of their speech-like vocalizations that have both a consonant and vowel, as follows:

$$Canonical\ proportion\ = \frac{Number\ of\ canonical\ vocalizations}{Number\ of\ canonical\ vocalizations + Number\ of\ non{-}canonical\ vocalizations}$$

This value represents one aspect of the child's vocal development, with higher values indicating a greater proportion of syllables with a mature shape.

2.3. Pre-processing and calculation of input quantity, social feedback, and vocalization practice measures (predictor variables)

To calculate our predictor variables, we will first automatically identify the portions of each long-form recording that contain key child or adult vocalizations. We will use Voice Type Classifier (VTC), a neural network model specifically designed to classify speaker type in child-centered long-form recordings (for VTC details and validation, see Lavechin et al., 2020). VTC returns a log with the start and end time of stretches of the audio attributed to the different talker types, and which covers the entire audio length. We will calculate all of our predictor variables from the period between 6am-9pm to maximize the chance that children are awake. We will normalize all measures by the total duration of the recording to arrive at an hourly rate for each metric of interest:

- **Input quantity:** For each recording, we will extract the total number of vocalizations produced by adults (female or male) from 6am-9pm. Prior work

indicates high accuracy in detecting adult speech, particularly for female adults (F-score = 82% in a held out test set), which make up a large majority in recordings across many cultures and languages (Lavechin et al., 2020).

- **Social feedback: contingent adult vocalizations (predictors):** We will measure the contingency of adult vocalizations on child productions in two ways (see hypotheses above). Both use the chattr package (Casillas & Scaff, 2021), where we operationalize contingent responses as any adult vocalization that occurs within 1-second of a child vocalization, but does not overlap with it. We choose this 1-second threshold based on a recent meta-analysis showing infant-adult turn-taking latencies of 1-second (Nguyen et al., 2022) and following past work studying this social feedback loop (Warlaumont et al., 2014).

    - **Specificity of adult contingent vocalizations to speech-like child vocalizations (socialfeedback$_{prop}$):** Following Warlaumont et al. (2014), we will calculate the difference between the proportion of child speech-like vocalizations (as determined by citizen science labels) that receive a response and the proportion of child non-speech-like vocalizations (as determined by citizen science labels) that receive a response:

$$\text{socialfeedback}_{prop} = \frac{N_{child\ speech-like\ vocs\ that\ receive\ a\ response}}{N_{child\ speech-like\ vocs}} - \frac{N_{child\ non-speech-like\ vocs\ \ that\ receive\ a\ response}}{N_{child\ non-speech-like\ vocs}}$$

This measure can vary between -1 (adults only verbally respond to non-speech-like vocalizations) and 1 (adults only verbally respond to speech-like vocalizations), and is positive if adults preferentially respond to speech-like vocalizations.

- **Total frequency of adult contingent vocalizations (socialfeedback$_N$):** Rather than the proportional measure above, this metric provides a count i.e. the total number of adult vocalizations that occur within 1s of a child speech-like vocalization.

- **Child vocalization experience:** For each key child, we will estimate the total duration of speech-like vocalizations they produced between 6am-9pm. We first sum the duration of all stretches of audio that VTC has attributed to the key child, which gives the total duration of key child vocalizations. To arrive at an estimate of total duration of *speech-like* key child vocalizations, we then multiply the total duration of each key child's vocalizations by the percentage of their vocalizations that were labeled as speech-like by citizen scientists. Reliability for this measure comes from VTC evaluation (Lavechin et al., 2020), where accuracy for the key child category is high (77% F-score). The 77% F-score is based on the proportion of 100ms frames that VTC correctly categorized, indicating high agreement even at this low level of resolution. We favored using total duration rather than number of vocalizations in order to capture the idea that longer vocalizations allow infants to experiment more (i.e., a child that produces 100

vocalizations each 2 seconds long has had more "practice" than one that produced 100 1-second vocalizations).

2.5. Analyses

2.5.1. Power analysis

In what follows, we propose a series of logistic mixed-effects models to test our hypotheses. Because previous work has not performed analyses of this type using logistic mixed-effects models, we cannot estimate the size of the effects involved in the models and thus cannot conduct an appropriate power analysis for the model structure we propose. To achieve a more meaningful sample estimate, we conduct our power analysis based on bivariate correlations, using effect sizes reported in a recent meta-analysis by Wang et al. (2020). Note that because bivariate correlations are less powerful than mixed effect models and because Wang et al. (2020) found no evidence of publication bias, this approach, if anything, should lead to overestimations of the needed sample size.

Wang et al. (2020) found an r = 0.21 correlation between LENA's adult word count measure (similar to input quantity) and language outcomes (e.g., MCDI, Fenson, 2007), an r = 0.31 correlation between LENA's conversational turn count measure (similar to social feedback) and language outcomes, and an r = 0.32 correlation between child vocalization counts (similar to child vocalization durations) and language outcomes. To detect similarly large relationships in our analyses (with 90% power and an alpha of .05), we would require sample sizes of N = 191, N = 86, and N = 80, for the

three respective factors we study. With our sample size of N = 204, we are well-powered to conduct the basic bivariate analyses.

2.5.2. Quality controls

Throughout our workflow, we will take precautions to make sure that our data analyses are valid. We will report the final average agreement (ignoring clips that have <50% agreement, since these are not included in our analyses). We will also inspect the distribution of our variables and the presence of outliers or overly influential individual data points (using Cook's distance) to identify potential errors. We will inspect all of our logistic regressions to make sure all assumptions are met. In particular, we will first check that there is no multicollinearity among the independent variables, by calculating Variance Inflation Factors (VIF) scores and determining that a variable is problematic if the VIF score is greater than 10. Second, we will test whether there is a linear relationship between each independent variable and the logit, using the Box-Tidwell test. That is, we will add log-transformed interactions between each independent variable and its log into the regression model. If any of the log interaction terms are significant, we will infer that the assumption of linearity of the logit has been violated. In our regressions, we intend to weigh data points to give more importance to individual children whose canonical proportions are based on a greater number of vocalizations, because of previous methodological work suggesting canonical proportions are more reliable when they are based on more data (Semenzin et al., 2021). In case anything looks amiss, we will use our exploration set (second recordings for each child) to transform variables (e.g., by standardizing them) until the fit models respect the

assumptions, before running the models on our main dataset. This process would be documented via updated pre-registrations.

2.5.3. Analysis 1: Testing the relationship between input quantity and canonical proportion

Analysis 1 will bear on our first stated hypothesis:

> *Hypothesis 1: Children who hear more input will have a higher canonical proportion.*

To test this hypothesis, we will fit a logistic mixed-effects model, predicting canonical proportion (cp) from input quantity. We fit a logistic mixed-effects model, rather than a linear mixed-effects model, because, canonical proportion, though continuous, is bounded between 0 and 1. To control for child age (in months, z-scored) and child sex, we will include them as fixed effects. In addition, we will include the interaction between child age and input quantity, child age^2 (z-scored), and the interaction between child age^2 and input quantity as predictors in the model. We will include the quadratic effect to account for potential floor and ceiling effects in our outcome measure (e.g., below or above a certain age). We will include the interaction terms because the effect of input quantity likely differs by age and may only appear within a certain developmental window. Finally, to account for potential differences between corpora (both cultural/age differences, but also more practical site differences), we will include corpus as a random effect. Following recommendations in Seedorff et

al., (2019), we will choose the best random effects structure by fitting the full space of models and selecting the (converging) model with the random effects structure that yields the lowest AIC. If corpus is found not to explain any variance, we will remove the random structure entirely and default to a simple regression. The most complex model we will consider is:

Regression 1.1

cp ~ input_quantity*child_age_z + input_quantity*child_age_z^2 + child_sex + (1 + input_quantity*child_age_z + input_quantity*child_age_z^2 + child_sex|corpus)

We will finalize the model structure on an "exploration set" (which will consist of a second recording for each child in the sample who has more than one recording, see §2.1), before running the model on the main held-out data set. Once we have arrived at the optimal model, we will remove any observations that have a Cook's distance greater than 3SDs from the mean and will refit the same model on the remaining observations (this will remove observations that have a particularly large effect on the fitted model).

If input quantity is related to canonical proportion, we would expect a significant main effect of input quantity, a significant input quantity*age interaction, or a significant input quantity*age^2 interaction, such that higher input quantities are related to higher canonical proportions. We will inspect betas to confirm that the pattern observed fits the prediction that more input is related to higher canonical proportion. Note this is not a causal interpretation: perhaps input quantity leads to higher canonical proportion, but it is also possible that a higher canonical proportion elicits more caregiver input (even

when controlling for age) and indeed this directionality has been documented as part of the social feedback loop in Warlaumont et al. (2014). If, on the other hand, input quantity is not significant in the model, then we will conclude that there is no evidence for an effect of input quantity on canonical proportion. We refer the reader to the Supplementary Materials, which provides more details on all of the analyses we describe here, as well as additional analyses that try to dissociate the three hypotheses, by controlling for potential confounds between them.

2.5.4. Analysis 2: Testing the relationship between contingent adult vocalization and canonical proportion

Analysis 2 will bear on our second set of hypotheses:

> *Hypothesis 2a: Children who hear relatively more adult vocalizations contingent on their speech-like vocalizations over their non-speech-like vocalizations will have a higher canonical proportion.*
> *Hypothesis 2b: Children who hear a larger number of adult vocalizations contingent on their own productions will have a higher canonical proportion.*

To test this set of hypotheses, we will fit two base logistic mixed-effects models predicting canonical proportion (cp) from contingent adult vocalization, one for each of the two social feedback measures we introduced in §2.3 ($\text{socialfeedback}_{\text{prop}}$ and $\text{socialfeedback}_{\text{N}}$). All modeling decisions will be identical to those in Analysis 1. The most complex base models we will consider are:

Regression 2.1a

$cp \sim socialfeedback_{prop}*child\_age\_z + socialfeedback_{prop}*child\_age\_z^2 + child\_sex + (1 + socialfeedback_{prop}*child\_age\_z + socialfeedback_{prop}*child\_age\_z^2 + child\_sex|corpus)$

Regression 2.1b

$cp \sim socialfeedback_{N}*child\_age\_z + socialfeedback_{N}*child\_age\_z^2 + child\_sex + (1 + socialfeedback_{N}*child\_age\_z + socialfeedback_{N}*child\_age\_z^2 + child\_sex|corpus)$

If contingent adult vocalization is related to canonical proportion, we would expect a significant main effect of social feedback, a significant social feedback*age interaction or a significant social feedback*age^2 interaction. We will inspect the directionality of the beta coefficients to confirm that the pattern observed fits the prediction that more or more systematic contingent adult vocalizations are related to higher canonical proportion. If social feedback is not significant in either model, then we will conclude that there is no evidence for an effect of social feedback on canonical proportion.

2.5.5. Analysis 3: Testing the relationship between child vocalizations and canonical proportion

Analysis 3 will bear on our final stated hypothesis:

Hypothesis 3: Children who vocalize more will have higher canonical proportions.

To test this hypothesis, we fit a logistic mixed effects model, predicting canonical proportion from child vocalization duration (child_voc), controlling for child age and child sex, and including corpus in the random effects structure. All decision-making will be identical to that in Analysis 1.

Regression 3.1

cp ~ child_voc*child_age_z + child_voc*child_age_z^2 + child_sex + (1 + child_voc*child_age_z + child_voc*child_age_z^2 + child_sex|corpus)

If child vocalization practice predicts canonical proportion, we would expect a significant main effect of total child vocalization duration, a significant child vocalization duration*age interaction, or a significant child vocalization duration*age^2 interaction such that longer total child vocalization durations are related to higher canonical proportion. We will inspect the directionality of the beta coefficients to confirm that the pattern observed fits the prediction that more child vocalization is related to higher canonical proportion. If we do not find this effect, we will conclude that there is no evidence for an effect of child vocalization experience.

2.6. Timeline:

- Prior to acceptance:
  - Compile all existing data needed for analyses.
- Within 3 months of acceptance:
  - Extract input quantity, social feedback, and vocalization experience measures from each key child's long-form recording
  - Merge these measures with citizen-science vocalization type annotations which have already been completed in the context of previous studies
  - Run pre-registered analyses as well as exploratory analyses
- Within 5 months of acceptance:
  - Write up results and submit paper draft to journal

References

Adda-Decker, M., Boula de Mareüil, P., Adda, G., & Lamel, L. (2005). Investigating syllabic structures and their variation in spontaneous French. *Speech Communication*, *46*(2), 119–139. https://doi.org/10.1016/j.specom.2005.03.006

Al Futaisi, N., Zhang, Z., Cristia, A., Warlaumont, A., & Schuller, B. (2019). VCMNet: Weakly Supervised Learning for Automatic Infant Vocalisation Maturity Analysis. *2019 International Conference on Multimodal Interaction*, 205–209. https://doi.org/10.1145/3340555.3353751

Albert, R. R., Schwade, J. A., & Goldstein, M. H. (2018). The social functions of babbling: Acoustic and contextual characteristics that facilitate maternal responsiveness. *Developmental Science*, *21*(5), e12641. https://doi.org/10.1111/desc.12641

Asteriou, D., & Hall, S. G. (2011). *Applied Econometrics* (2nd ed.). New York. Palgrave Macmillian.

Belardi, K., Watson, L. R., Faldowski, R. A., Hazlett, H., Crais, E., Baranek, G. T., ... & Oller, D. K. (2017). A retrospective video analysis of canonical babbling and volubility in infants with fragile X syndrome at 9–12 months of age. *Journal of autism and developmental disorders*, *47*, 1193-1206.

Bennett, R. (2016). Mayan phonology. *Language and Linguistics Compass*, *10*(10), 469–514. https://doi.org/10.1111/lnc3.12148

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

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

Bergelson, E., & Aslin, R. N. (2017). Nature and origins of the lexicon in 6-mo-olds. *Proceedings*

of the National Academy of Sciences, *114*(49), 12916–12921. https://doi.org/10.1073/pnas.1712966114

Bloom, K., Russell, A., & Wassenberg, K. (1987). Turn taking affects the quality of infant vocalizations. *Journal of child language*, *14*(2), 211-227.

Bunce, J., Soderstrom, M., Bergelson, E., Rosemberg, C., Stein, A., Alam, F., Migdalek, M., & Casillas, M. (2020). *A cross-cultural examination of young children's everyday language experiences*. PsyArXiv. https://doi.org/10.31234/osf.io/723pr

Canault, M., Le Normand, M. T., Foudil, S., Loundon, N., & Thai-Van, H. (2016). Reliability of the language environment analysis system (LENA™) in European French. *Behavior research methods*, *48*, 1109-1124.

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

Casillas, M., Brown, P., & Levinson, S. C. (2020). Early Language Experience in a Tseltal Mayan Village. *Child Development*, *91*(5), 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*(4), 792–814. https://doi.org/10.1017/S0305000920000549

Casillas, M., & Scaff, C. (2021). Analyzing contingent interactions in R with `chattr`. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*(43). https://escholarship.org/uc/item/4rr848x0

Cassar, A., Cristia, A., Grosjean, P., & Walker, S. (2021). *Long-form recordings in the Solomon Islands*.

Chapman, K. L., Hardin-Jones, M., & Halter, K. A. (2003). The relationship between early speech and later speech and language performance for children with cleft lip and palate. *Clinical Linguistics & Phonetics*, *17*(3), 173–197. https://doi.org/10.1080/0269920021000047864

Christiansen, M. H., Kallens, P. C., & Trecca, F. (2022). We need a comparative approach to language acquisition: A commentary on Kidd and Garcia (2022). *First Language*, 01427237221093847. https://doi.org/10.1177/01427237221093847

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

Cristia, A. (2021). *PhonSES: A pilot study to measure socioeconomic status association with infants' word and sound processing*. https://gin.g-node.org/LAAC-LSCP/phonSES-public

Cristia, A. (2022). A systematic review suggests marked differences in the prevalence of infant-directed vocalization across groups of populations. *Developmental Science*, *n/a*, e13265. https://doi.org/10.1111/desc.13265

Cristia, A., & Casillas, M. (2019). *LENA recordings in Rossel Island*.

Cristia, A., & Casillas, M. (2022). Non-word repetition in children learning Yélî Dnye. *Language Development Research*, *2*(1). https://doi.org/10.34842/ZR2Q-1X28

Cristia, A., & Colleran, H. (2018). *Long-form, child-centered recordings collected in Malekula in 2016-2018.*

Cristia, A., Dupoux, E., Gurven, M., & Stieglitz, J. (2019). Child-Directed Speech Is Infrequent in a Forager-Farmer Population: A Time Allocation Study. *Child Development*, *90*(3), 759–773. https://doi.org/10.1111/cdev.12974

Cristia, A., Farabolini, G., Scaff, C., Havron, N., & Stieglitz, J. (2020). Infant-directed input and literacy effects on phonological processing: Non-word repetition scores among the Tsimane'. *PLOS ONE*, *15*(9), e0237702. https://doi.org/10.1371/journal.pone.0237702

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

Cychosz, M., Cristia, A., Bergelson, E., Casillas, M., Baudet, G., Warlaumont, A. S., Scaff, C., Yankowitz, L., & Seidl, A. (2021a). Vocal development in a large-scale crosslinguistic corpus. *Developmental Science*, *24*(5), e13090. https://doi.org/10.1111/desc.13090

Cychosz, M., Edwards, J. R., Bernstein Ratner, N., Torrington Eaton, C., & Newman, R. S. (2021b). Acoustic-lexical characteristics of child-directed speech between 7 and 24 months and their impact on toddlers' phonological processing. *Frontiers in Psychology*, 3186.

Cychosz, M., Munson, B., & Edwards, J. R. (2021c). Practice and Experience Predict Coarticulation in Child Speech. *Language Learning and Development*, *17*(4), 366–396. https://doi.org/10.1080/15475441.2021.1890080

Dailey, S., & Bergelson, E. (2022). Language input to infants of different socioeconomic statuses: A quantitative meta‐analysis. *Developmental Science*, *25*(3), e13192.

de Boysson-Bardies, B., & Vihman, M. M. (1991). Adaptation to Language: Evidence from Babbling and First Words in Four Languages. *Language*, *67*(2), 297–319. https://doi.org/10.2307/415108

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

Eilers, R. E., & Oller, D. K. (1994). Infant vocalizations and the early diagnosis of severe hearing impairment. *The Journal of pediatrics*, *124*(2), 199-203.

Elmlinger, S., Goldstein, M., & Casillas, M. (2022). Immature vocalizations simplify the speech of Tseltal Mayan and US caregivers. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*(44). https://escholarship.org/uc/item/8qk657c3

Elmlinger, S. L., Schwade, J. A., Vollmer, L., & Goldstein, M. H. (2022). Learning how to learn from social feedback: The origins of early vocal development. *Developmental Science*, *n/a*, e13296. https://doi.org/10.1111/desc.13296

Fagan, M. K. (2009). Mean length of utterance before words and grammar: Longitudinal trends and developmental implications of infant vocalizations. *Journal of child language*, *36*(3),

495-527.

Fagan, M. K., & Doveikis, K. N. (2017). Ordinary interactions challenge proposals that maternal verbal responses shape infant vocal development. *Journal of Speech, Language, and Hearing Research*, *60*(10), 2819-2827.

Faraway, J. J. (2015). *Linear Models with R* (2nd ed.). Boca Raton. Chapman & Hall, CRC Press.

Fenson, L. (2007). MacArthur-Bates communicative development inventories.

Goffman, L. (1999). Prosodic Influences on Speech Production in Children With Specific Language Impairment and Speech Deficits: Kinematic, Acoustic, and Transcription Evidence. *Journal of Speech, Language, and Hearing Research*, *42*(6), 1499–1517. https://doi.org/10.1044/jslhr.4206.1499

Goldin-Meadow, S. (2014). In search of resilient and fragile properties of language. *Journal of Child Language, 41*(S1), 64–77. https://doi.org/10.1017/S030500091400021X

Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences*, *100*(13), 8030–8035. https://doi.org/10.1073/pnas.1332441100

Goldstein, M. H., & Schwade, J. A. (2008). Social Feedback to Infants' Babbling Facilitates Rapid Phonological Learning. *Psychological Science*, *19*(5), 515–523. https://doi.org/10.1111/j.1467-9280.2008.02117.x

Gros-Louis, J., & Miller, J. L. (2018). From 'ah' to 'bah': Social feedback loops for speech sounds at key points of developmental transition. *Journal of Child Language*, *45*(3), 807–825. https://doi.org/10.1017/S0305000917000472

Gros-Louis, J., West, M. J., Goldstein, M. H., & King, A. P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of Behavioral Development*, *30*(6), 509–516. https://doi.org/10/d2m833

Hitczenko, K., Bergelson, E., Casillas, M., Colleran, H., Cychosz, M., Grosjean, P., Hamrick, L.

R., Kelleher, B. L., Scaff, C., Seidl, A., Walker, S., & Cristia, A. (2023). The development of canonical proportion continues past toddlerhood. To appear in *Proceedings of the 20th International Congress of Phonetic Sciences, Prague, Czech Republic.*

Huttenlocher, J., Vasilyeva, M., Cymerman, E., & Levine, S. (2002). Language input and child syntax. *Cognitive psychology*, *45*(3), 337-374.

Hsu, H. C., & Fogel, A. (2001). Infant vocal development in a dynamic mother-infant communication system. *Infancy*, *2*(1), 87-109.

Karasik, L. B., Adolph, K. E., Tamis-LeMonda, C. S., & Bornstein, M. H. (2010). WEIRD walking: Cross-cultural research on motor development. *The Behavioral and brain sciences, 33*(2-3), 95-96. https://doi.org/10.1017%2FS0140525X10000117

Karasik, L. B., & Kuchirko, Y. A. (2022). Talk the talk and walk the walk: Diversity and culture impact all of development – A commentary on Kidd and Garcia (2022). *First Language*, 01427237221096508. https://doi.org/10.1177/01427237221096508

Keren-Portnoy, T., Vihman, M. M., DePaolis, R. A., Whitaker, C. J., & Williams, N. M. (2010). The Role of Vocal Practice in Constructing Phonological Working Memory. *Journal of Speech, Language, and Hearing Research*, *53*(5), 1280–1293. https://doi.org/10.1044/1092-4388(2009/09-0003)

Kidd, E., & Garcia, R. (2022). How diverse is child language acquisition research? *First Language*, 01427237211066405. https://doi.org/10.1177/01427237211066405

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

Lang, S., Bartl-Pokorny, K. D., Pokorny, F. B., Garrido, D., Mani, N., Fox-Boyer, A. V., Zhang, D., & Marschik, P. B. (2019). Canonical Babbling: A Marker for Earlier Identification of Late Detected Developmental Disorders? *Current Developmental Disorders Reports*, *6*(3), 111–118. https://doi.org/10.1007/s40474-019-00166-w

Lang, S., Willmes, K., Marschik, P. B., Zhang, D., & Fox-Boyer, A. (2021). Prelexical phonetic and early lexical development in German-acquiring infants: Canonical babbling and first spoken words. *Clinical linguistics & phonetics*, *35*(2), 185-200.

Lavechin, M., Bousbib, R., Bredin, H., Dupoux, E., & Cristia, A. (2020). *An open-source voice type classifier for child-centered daylong recordings* (arXiv:2005.12656). arXiv. https://doi.org/10.48550/arXiv.2005.12656

Lavechin, M., de Seyssel, M., Gautheron, L., Dupoux, E., & Cristia, A. (2022). Reverse Engineering Language Acquisition with Child-Centered Long-Form Recordings. *Annual Review of Linguistics*, *8*, 389–407.

Lewedag, V. L., Oller, D. K., & Lynch, M. P. (1994). Infants' vocalization patterns across home and laboratory environments. *First Language*, *14*(42–43), 049–065. https://doi.org/10.1177/014272379401404204

Lohmander, A., Holm, K., Eriksson, S., & Lieberman, M. (2017). Observation method identifies that a lack of canonical babbling can indicate future speech and language problems. *Acta Paediatrica*, *106*(6), 935-943.

Long, H. L., Bowman, D. D., Yoo, H., Burkhardt-Reed, M. M., Bene, E. R., & Oller, D. K. (2020). Social and endogenous infant vocalizations. *PLOS ONE*, *15*(8), e0224956. https://doi.org/10.1371/journal.pone.0224956

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*(12), e0279395.

Lopez, L. D., Walle, E. A., Pretzer, G. M., & Warlaumont, A. S. (2020). Adult responses to infant prelinguistic vocalizations are associated with infant vocabulary: A home observation study. *PLOS ONE*, *15*(11), e0242232. https://doi.org/10.1371/journal.pone.0242232

Lynch, M. P., Oller, D. K., Steffens, M. L., & Levine, S. L. (1995). Onset of speech-like

vocalizations in infants with Down syndrome. *American Journal on Mental Retardation*.

Majorano, M., Vihman, M. M., & DePaolis, R. A. (2014). The relationship between infants' production experience and their processing of speech. *Language Learning and Development*, *10*(2), 179-204.

Marklund, E., Schwarz, I.-C., & Lacerda, F. (2019). Amount of speech exposure predicts vowel perception in four- to eight-month-olds. *Developmental Cognitive Neuroscience*, *36*, 100622. https://doi.org/10.1016/j.dcn.2019.100622

Marschik, P. B., Pini, G., Bartl‐Pokorny, K. D., Duckworth, M., Gugatschka, M., Vollmann, R., ... & Einspieler, C. (2012). Early speech–language development in females with Rett syndrome: focusing on the preserved speech variant. *Developmental Medicine & Child Neurology*, *54*(5), 451-456.

Marschik, P. B., Bartl-Pokorny, K. D., Sigafoos, J., Urlesberger, L., Pokorny, F., Didden, R., ... & Kaufmann, W. E. (2014). Development of socio-communicative skills in 9-to 12-month-old individuals with fragile X syndrome. *Research in Developmental Disabilities*, *35*(3), 597-602.

Marschik, P. B., Kaufmann, W. E., Sigafoos, J., Wolin, T., Zhang, D., Bartl-Pokorny, K. D., ... & Johnston, M. V. (2013). Changing the perspective on early development of Rett syndrome. *Research in developmental disabilities*, *34*(4), 1236-1239.

Masataka, N. (2001). Why early linguistic milestones are delayed in children with Williams syndrome: late onset of hand banging as a possible rate–limiting constraint on the emergence of canonical babbling. *Developmental Science*, *4*(2), 158-164.

McCune, L., & Vihman, M. M. (2001). Early phonetic and lexical development: A productivity approach. *Journal of Speech, Language, and Hearing Research, 44*(3), 670-684.

McDaniel, J., Woynaroski, T., Keceli-Kaysili, B., Watson, L. R., & Yoder, P. (2019). Vocal Communication With Canonical Syllables Predicts Later Expressive Language Skills in Preschool-Aged Children With Autism Spectrum Disorder. *Journal of Speech, Language,*

and Hearing Research, *62*(10), 3826–3833. https://doi.org/10.1044/2019_JSLHR-L-19-0162

McGillion, M. L., Herbert, J. S., Pine, J. M., Keren-Portnoy, T., Vihman, M. M., & Matthews, D. E. (2013). Supporting early vocabulary development: What sort of responsiveness matters? *IEEE Transactions on Autonomous Mental Development, 5*(3), 240-248.

McGillion, M., Herbert, J. S., Pine, J., Vihman, M., dePaolis, R., Keren-Portnoy, T., & Matthews, D. (2017). What Paves the Way to Conventional Language? The Predictive Value of Babble, Pointing, and Socioeconomic Status. *Child Development*, *88*(1), 156–166. https://doi.org/10.1111/cdev.12671

Melvin, S. A., Brito, N. H., Mack, L. J., Engelhardt, L. E., Fifer, W. P., Elliott, A. J., & Noble, K. G. (2017). Home Environment, But Not Socioeconomic Status, is Linked to Differences in Early Phonetic Perception Ability. *Infancy*, *22*(1), 42–55. https://doi.org/10.1111/infa.12145

Moeller, M. P., Hoover, B., Putman, C., Arbataitis, K., Bohnenkamp, G., Peterson, B., ... & Stelmachowicz, P. (2007). Vocalizations of infants with hearing loss compared with infants with normal hearing: Part I–phonetic development. *Ear and hearing*, *28*(5), 605-627.

Mollica, F., & Piantadosi, S. T. (2019). Humans store about 1.5 megabytes of information during language acquisition. *Royal Society Open Science*, *6*(3), 181393. https://doi.org/10.1098/rsos.181393

Morgan, L., & Wren, Y. E. (2018). *A Systematic Review of the Literature on Early vocalizations and Babbling Patterns in Young Children*. https://doi.org/10/gfgdbk

Moulin-Frier, C., Nguyen, S. M., & Oudeyer, P.-Y. (2014). Self-organization of early vocal development in infants and machines: The role of intrinsic motivation. *Frontiers in Psychology*, *4*. https://www.frontiersin.org/article/10.3389/fpsyg.2013.01006

Nguyen, V., Versyp, O., Cox, C., & Fusaroli, R. (2022). A systematic review and Bayesian

meta‐analysis of the development of turn taking in adult–child vocal interactions. *Child Development*, *93*(4), 1181-1200.

Nyman, A., Strömbergsson, S., & Lohmander, A. (2021). Canonical babbling ratio – Concurrent and predictive evaluation of the 0.15 criterion. *Journal of Communication Disorders*, *94*, 106164. https://doi.org/10.1016/j.jcomdis.2021.106164

Oller, D. K. (2000). *The Emergence of the Speech Capacity*. New York. Psychology Press. https://doi.org/10.4324/9781410602565

Oller, D. K., Buder, E. H., Ramsdell, H. L., Warlaumont, A. S., Chorna, L., & Bakeman, R. (2013). Functional flexibility of infant vocalization and the emergence of language. *Proceedings of the National Academy of Sciences*, *110*(16), 6318–6323. https://doi.org/10.1073/pnas.1300337110

Oller, D. K., Caskey, M., Yoo, H., Bene, E. R., Jhang, Y., Lee, C.-C., Bowman, D. D., Long, H. L., Buder, E. H., & Vohr, B. (2019). Preterm and full term infant vocalization and the origin of language. *Scientific Reports*, *9*(1), 14734. https://doi.org/10.1038/s41598-019-51352-0

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*(44), 167–187. https://doi.org/10.1177/014272379501504403

Oller, D. K., Eilers, R. E., Bull, D. H., & Carney, A. E. (1985). Prespeech vocalizations of a deaf infant: A comparison with normal metaphonological development. *Journal of Speech, Language, and Hearing Research, 28*(1), 47-63. https://doi.org/10.1044/jshr.2801.47

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*(4), 223–245. https://doi.org/10.1016/S0021-9924(99)00013-1

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

Oller, D. K., Eilers, R., Neal-Beevers, A., & Cobo-Lewis, A. (1998). Late Onset Canonical

Babbling: A Possible Early Marker of Abnormal Development. *American Journal of Mental Retardation : AJMR*, *103*, 249–263. https://doi.org/10.1352/0895-8017(1998)103<0249:LOCBAP>2.0.CO;2

Oller, D.K., Eilers, R. E., Urbano, R., & Cobo-Lewis, A. B. (1997). Development of precursors to speech in infants exposed to two languages. *Journal of Child Language*, *24*(2), 407–425. https://doi.org/10/bhzqxh

Oller, D. K., Niyogi, P., Gray, S., Richards, J. A., Gilkerson, J., Xu, D., Yapanel, U., & Warren, S. F. (2010). Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. *Proceedings of the National Academy of Sciences*, *107*(30), 13354–13359. https://doi.org/10.1073/pnas.1003882107

Oller, D. K., Ramsay, G., Bene, E., Long, H. L., & Griebel, U. (2021). Protophones, the precursors to speech, dominate the human infant vocal landscape. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1836), 20200255. https://doi.org/10.1098/rstb.2020.0255

Overby, M., & Caspari, S. S. (2015). Volubility, consonant, and syllable characteristics in infants and toddlers later diagnosed with childhood apraxia of speech: A pilot study. *Journal of Communication Disorders*, *55*, 44-62.

Patten, E., Belardi, K., Baranek, G. T., Watson, L. R., Labban, J. D., & Oller, D. K. (2014). Vocal patterns in infants with autism spectrum disorder: Canonical babbling status and vocalization frequency. *Journal of autism and developmental disorders*, *44*, 2413-2428.

Peute, B., & Casillas, M. (2022). (Non-) effects of linguistic environment on early stable consonant production: A cross cultural case study. *Glossa: a journal of general linguistics*, *7*(1).

Ritwika, V. P. S., Pretzer, G. M., Mendoza, S., Shedd, C., Kello, C. T., Gopinathan, A., & Warlaumont, A. S. (2020). Exploratory dynamics of vocal foraging during infant-caregiver communication. *Scientific Reports*, *10*(1), 10469.

https://doi.org/10.1038/s41598-020-66778-0

Roche, L., Zhang, D., Bartl-Pokorny, K. D., Pokorny, F. B., Schuller, B. W., Esposito, G., ... & Marschik, P. B. (2018). Early vocal development in autism spectrum disorder, Rett syndrome, and fragile X syndrome: Insights from studies using retrospective video analysis. *Advances in neurodevelopmental disorders*, *2*, 49-61.

Scaff, C., Stieglitz, J., & Cristia, A. (2018). *Tsimane' daylong recordings collected with LENA in 2017-2018*. https://doi.org/DOI 10.17605/OSF.IO/6NEZA

Scaff, C., Stieglitz, J., & Cristia, A. (2019). *Excerpts from daylong recordings of young children learning Tsimane' in Bolivia*. https://doi.org/DOI 10.17605/OSF.IO/5869Q

Seedorff, M., Oleson, J., & McMurray, B. (2019). *Maybe maximal: Good enough mixed models optimize power while controlling Type I erro*. https://doi.org/10.31234/osf.io/xmhfr

Semenzin, C., Hamrick, L., Seidl, A., Kelleher, B. L., & Cristia, A. (2021). Describing Vocalizations in Young Children: A Big Data Approach Through Citizen Science Annotation. *Journal of Speech, Language, and Hearing Research*, *64*(7), 2401–2416. https://doi.org/10.1044/2021_JSLHR-20-00661

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

Sohner, L., & Mitchell, P. (1991). Phonatory and phonetic characteristics of prelinguistic vocal development in cri du chat syndrome. *Journal of communication disorders*, *24*(1), 13-20.

Sultana, A. (2022). The missing majority in child language research: A commentary on Kidd and Garcia (2022). *First Language*, 01427237221094738. https://doi.org/10.1177/01427237221094738

VanDam, M., & Yoshinaga-Itano, C. (2019). Use of the LENA Autism Screen with Children who are Deaf or Hard of Hearing. *Medicina*, *55*(8), 495. https://doi.org/10.3390/medicina55080495

Vihman, M. M. (2014). *Phonological development: The first two years (2nd ed)*. Wiley-Blackwell: Oxford, UK.

Vihman, M. M. (2013). *Phonological Development: The First Two Years*. John Wiley & Sons.

Vihman, M. M. (2022). The developmental origins of phonological memory. *Psychological Review*.

Vihman, M. M., Macken, M. A., Miller, R., Simmons, H., & Miller, J. (1985). From Babbling to Speech: A Re-Assessment of the Continuity Issue. *Language*, *61*(2), 397–445. https://doi.org/10.2307/414151

Wang, Y., Williams, R., Dilley, L., & Houston, D. M. (2020). A meta-analysis of the predictability of LENA™ automated measures for child language development. *Developmental Review*, *57*, 100921. https://doi.org/10.1016/j.dr.2020.100921

Warlaumont, A. S., & Ramsdell, H. (2016). *Detection of Total Syllables and Canonical Syllables in Infant Vocalizations* (p. 2680). https://doi.org/10.21437/Interspeech.2016-1518

Warlaumont, A. S., Richards, J. A., Gilkerson, J., & Oller, D. K. (2014). A Social Feedback Loop for Speech Development and Its Reduction in Autism. *Psychological Science*, *25*(7), 1314–1324. https://doi.org/10.1177/0956797614531023

Weisleder, A., & Fernald, A. (2013). Talking to Children Matters: Early Language Experience Strengthens Processing and Builds Vocabulary. *Psychological Science*, *24*(11), 2143–2152. https://doi.org/10/f5gdg2

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3

Yankowitz, L. D., Petrulla, V., Plate, S., Tunc, B., Guthrie, W., Meera, S. S., ... & Parish-Morris, J. (2022). Infants later diagnosed with autism have lower canonical babbling ratios in the first year of life. *Molecular Autism*, *13*(1), 1-16.

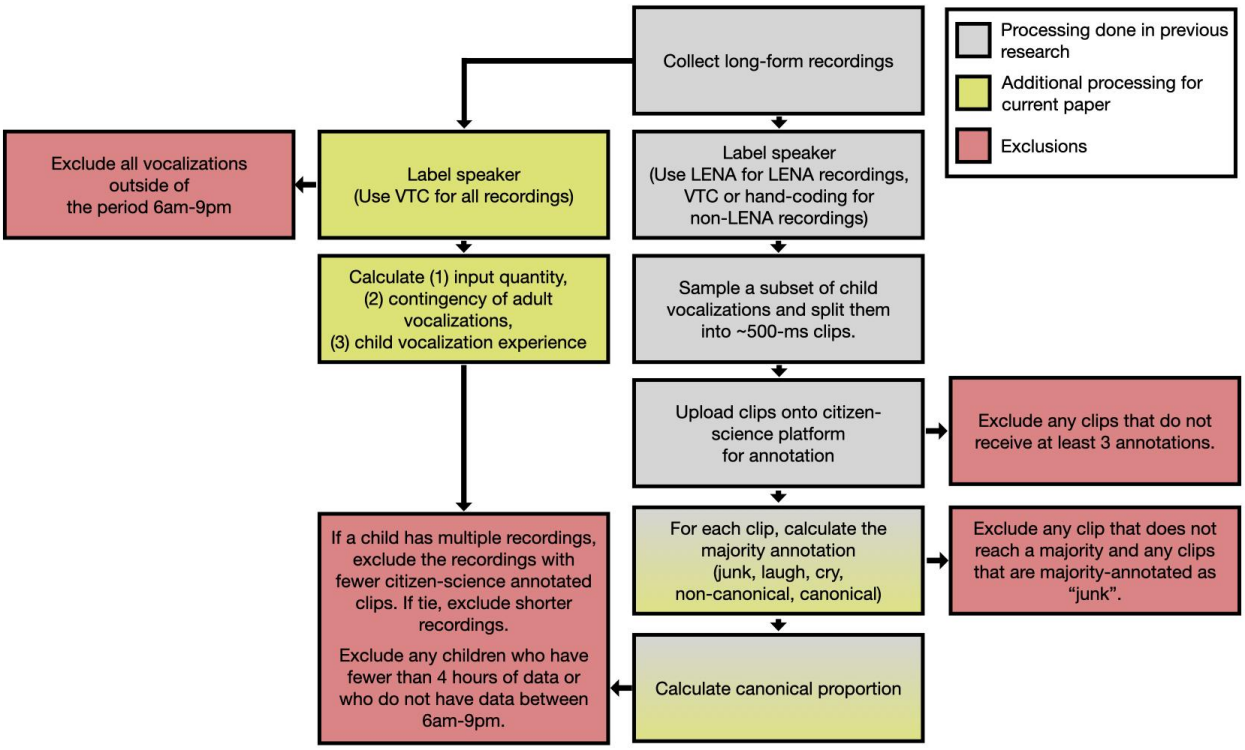

Figure 1. *Flowchart of data-processing steps.* Two boxes contain both gray and yellow sections to indicate that some of these steps have been done in previous work for some corpora, and for the rest of the corpora they were done for the current paper.