



HAL
open science

Vocal input and output among infants in a multilingual context: Evidence from long-form recordings in Vanuatu

Alejandrina Cristia, Lucas Gautheron, Heidi Colleran

► To cite this version:

Alejandrina Cristia, Lucas Gautheron, Heidi Colleran. Vocal input and output among infants in a multilingual context: Evidence from long-form recordings in Vanuatu. *Developmental Science*, 2023, 26 (4), pp.e13375. 10.1111/desc.13375 . hal-04352827

HAL Id: hal-04352827

<https://cnrs.hal.science/hal-04352827>

Submitted on 11 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Vocal input and output among infants in a multilingual context: Evidence from long-form recordings in Vanuatu

Alejandrina Cristia¹, Lucas Gautheron^{1,2}, & Heidi Colleran³

¹ Laboratoire de Sciences Cognitives et de Psycholinguistique, Département d'études cognitives, ENS, EHESS, CNRS, PSL University

² Interdisciplinary Centre for Science and Technology Studies (IZWT), University of Wuppertal

³ BirthRites Independent Max Planck Research Group, Department of Human Behavior, Ecology and Culture, Max Planck Institute for Evolutionary Anthropology

Word Count: 7,944

Data availability statement: The data and code necessary to reproduce the analyses presented here are publicly accessible from <https://osf.io/t8r5j/>

Acknowledgments: HC thanks The Vanuatu Cultural Centers in Port Vila and Malakula, especially directors Marcellin Ambong, Richard Shing, and Numa Fred Longga for their support. The late Chief Jimmysan Sanhapath, late Chief Gilbert Nihapath, Chief Joachim Kaibaba, and

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

Chief Arnold Niptik Virampath all provided crucial community-level support on Malakula. HC also thanks Jacklyn and Rona Willie, Kalmet and Sylvie Tulili, Olivier and Margaret Nihapath, Jeanette Sanhapath, Susan Arknaveh, and Laura Loveland for invaluable support and friendship in the field; and Erina William for help collecting some of the data. We are also grateful to Mustapha Mardi for the human annotations; Jonathan Stieglitz, Marisa Casillas, Gandhi Yetish, Camila Scaff, and Richard McElreath for helpful discussion.

Ethics approval statement: The project protocols were approved by the Jena Ethik-Kommission at Friedrich Schiller University of Jena, Germany, under the auspices of the “Vanuatu Languages and Lifeways” project co-directed by HC at the Max Planck Institute for the Science of Human History (2016-2019), with the amendment entitled “The effect of input on output: Language acquisition among ni-Vanuatu children in Malakula” to the project number 4818-06/16; and research agreements between the Vanuatu Cultural Council of the Vanuatu Cultural Center and HC. Following an explanation of the aims, methods, and expected outcomes of the project in local communities (in Bislama), individual families were invited to take part in recording their children. The majority (34 participants) signed a consent form, six others chose to provide their consent orally. The data were collected between 2016 and 2018 as part of a broader mixed-methods research project during which HC lived in the field, and which includes ongoing participant feedback through repeat visits to the communities. On returning to the field in 2023, HC plans to continue feedback with the study communities. The research results will be presented orally both in participating communities and at the Vanuatu Cultural Center in port Vila, and via written materials to be distributed to participating families. All feedback will take place in Bislama.

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

Funding statement: We acknowledge funding and institutional support from The Max Planck Society, the Agence Nationale pour la Recherche [ANR-14-CE30-0003 MechELex, ANR-17-EURE-0017, ANR-16-DATA-0004 ACLEW], the J. S. McDonnell Foundation, European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ExELang, Grant agreement No. 101001095).

Conflict of interest disclosure: The authors have no conflict of interest to disclose

Statement of contributions: Conceptualization AC HC; Data collection HC; Methodology AC HC LG; Software LG; Resources and Funding AC HC; Data curation AC LG HC; Writing AC HC LG; Visualization: AC; Funding acquisition: HC AC; Project administration: HC.

Corresponding author: Correspondence concerning this article should be addressed to Alejandrina Cristia, 29, rue d'Ulm, 75005 Paris, France. E-mail: alecristia@gmail.com
<http://orcid.org/0000-0003-2979-4556>

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

Research highlights

- Combining long-form recordings with automated analyses, we estimated infants potentially exposed to ~2.6 languages heard ~11 minutes of speech per hour.
- Infants' input was dominated by vocalizations from female adults and from other children, particularly for the oldest infants in our sample.
- The strongest association between children's own vocalization counts and input counts was with those of other children, and not those with adults.
- Results invite further research on individual, group, and population variability in input quantity and composition, and its potential effects on vocal development.

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

Abstract

What are the vocal experiences of children growing up on Malakula island, Vanuatu, where multilingualism is the norm? Long-form audio-recordings captured spontaneous speech behavior by, and around, 38 children (5-33 months, 23 girls) from 11 villages. Automated analyses revealed most children's vocal input came from female adults and other children's voices, with small contributions from male adult voices. The greatest changes with age involved an increase in the input vocalizations from other children. Total input (collapsing across child-directed and overheard speech, and across languages) was ~11 minutes per hour, which was at least 5 minutes (31%) lower than that found in other populations studied using comparable methods in previous literature, as well as in archival American data analyzed with the same algorithm. In contrast, children's own vocalization counts were two to four times higher than previous reports for North-American English-learning monolingual infants at matched ages, and comparable to estimates from archival American data, consistent with a resilient language-learning cognitive system for this aspect of vocal development. The strongest association between input and output was with vocalizations by other children, rather than those by adults, which is consistent with research in anthropology but less so with current theoretical trends in developmental psychology. These results invite further research in populations that are under-represented in developmental science.

Keywords: language acquisition; long-form recordings; infant-directed speech; multilingualism

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

Vocal input and output among infants in a multilingual context: Evidence from long-form recordings in Vanuatu

Despite decades of study of language acquisition, basic aspects of this process have not been well-documented, including who talks to children and for how long across diverse human populations (Bunce et al., 2020; Casillas, Brown, & Levinson, 2020; Casillas, Brown, & Levinson, 2021; Cristia, 2022; Scaff et al., 2022). The present study provides descriptive evidence on an understudied population. This Introduction motivates our aim (a child-centered view of children's vocal input) and methods (long-form audio-recordings) by connecting with key anthropological and psycholinguistic literature, with an exhaustive review being beyond the scope of this article.

Describing language experiences

Recent work describes how language exposure correlates with phonological (e.g., Garcia-Sierra et al., 2011) and lexical (e.g., Grüter, Hurtado, Marchman, & Fernald, 2014) processing among toddlers exposed to one or more languages. By and large, more input is associated with better performance in, for example, toddlers' word comprehension (Weisleder & Fernald, 2013). Despite the many conceptual and empirical advances this work brings, there is a strong focus on balanced bilinguals, who hear exactly two languages, each representing between 30 and 70% of language exposure (see Byers-Heinlein et al., 2020 for a discussion of exposure estimation). Consequently, it is difficult to extrapolate those ideas and results to children with more diverse language experiences, including children who hear more than two languages, or for whom exposure to different languages is not balanced (cf. Kremin & Byers-Heinlein, 2021).

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

In another line of work, researchers discuss how much input children are afforded, with a recent systematic review suggesting potentially large cross-population variation in the prevalence of infant-directed speech (Cristia, 2022). Large variation in experiences, paired with low variation in certain outcomes, may be consistent with specific aspects of language acquisition being resilient (Goldin-Meadow, 2015), helping us adjudicate among alternative language acquisition theories (Cristia, 2020). So how much input is afforded to young infants across populations? The most accurate estimates may come from research using a novel technique called long-form recordings, in which the infant wears a device that records audio continuously over a typical day. Long recordings with unobtrusive devices may reduce self-consciousness in the family and the child, resulting in more accurate estimations of input quantities (comparing audio versus video; Bergelson et al., 2019) and output quantities (in the home versus the lab; Lewedag, Oller, & Lynch, 1994) than shorter recordings.

The field has not yet standardized ways of measuring input and output in such recordings, with a current divide between two forms of analyses, one that relies on purely automated algorithms, and another that builds on human annotation. The advantage of an algorithmic approach is its coverage, but, until recently, the most commonly used algorithm (LENA, e.g., Gilkerson et al., 2017) only estimates accurately the child's own vocalizations and input from adult speakers (systematically reviewed in Cristia et al., 2020), ignoring children in the child's vocal environment as a potential source of input. According to two previous studies, children growing up in North America hear about 3.3-3.5 min/h of speech from male and female adults combined (Gilkerson et al., 2017; Weisleder & Fernald, 2013).

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

Different estimations emerge from studies using random or periodic sampling of small audio sections to submit them to human annotation, which has primarily been employed with rural communities. Three studies are most relevant because they provide input estimates for each speaker type and include all vocalizations (and not just those directed to the child): Casillas, Brown, and Levinson (2021) on 10 infants learning Yéî Dnye; Casillas, Brown, and Levinson (2020) on 10 children learning Tseltal; and Scaff et al. (in prep) on 15 children learning Tsimane'; all three studies consider their participants to be largely monolingual. Estimates are much higher in this work than in Gilkerson's, and range from 16.9 min/h (Scaff et al., 2022) to 39 min/h (Casillas, Brown, and Levinson, 2019) total speech, collapsing across all speaker types. In all three populations, male adults provided little input, and instead, a majority of the input came from female adults, although a sizable proportion, increasing with age, came from other children (Bunce et al., 2020). In the research on toddler bilingualism, the question is generally posed in terms of the *proportion* of input afforded in each of the languages (e.g., Marchman et al., 2017; Montant, Byers-Heinlein, & Perez, 2022). As for sources of input, our unsystematic review of this work suggests a focus on adult primary caregivers, potentially because most of their toddler participants are growing up in urban, industrialized societies, where other children do not feature saliently before attending pre-school, and where families are small.

One of the motivations to look at input quantity and input sources is the hypothesis that variation in these dimensions may relate to variation in language skills (Golinkoff, Hoff, Rowe, Tamis-LeMonda, & Hirsh-Pasek, 2019; Ramírez-Esparza et al., 2014; Shneidman & Goldin-Meadow, 2012; Weisleder & Fernald, 2013; for contrasting views see Sperry, Sperry, & Miller, 2019). Although the impression that emerges from this work is that *adult* input,

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

particularly in one-on-one conversations, is most useful for language learning (e.g., Ramírez-Esparza et al., 2014), such evidence is generally scant and debates are ongoing. One area of disagreement of particular relevance to the present work pertains to the *sources* of input, and particularly how much infants can learn from adults as opposed to children. Shneidman and Goldin-Meadow (2012) studied individual variation in vocabulary among 15 Yucatec Mayan children, and found that this was better predicted by lexical diversity in child-directed speech from adults, and not child-directed speech from other children. In a study of 30 toddlers in Chicago, the number of tokens addressed to the child from all household members explained more variance in key children's vocabulary than input from primary caregivers (in this sample, mothers) alone (Shneidman, Arroyo, Levine, & Goldin-Meadow, 2013). Little is known about the potential relevance of male, as opposed to female, adult speech for language development, although one study using long-form recordings among bilingual Spanish-English North American families found that input quantity from fathers accounted for variance in children's own vocal production above and beyond input from mothers (Ferjan Ramirez et al., 2022).

In sum, the literature on how much input infants need to develop functional language skills is still relatively small and it is unclear whether some types of speech matter more than others (such as speech coming from adults as opposed to other children). It is therefore important to describe experiences afforded to children along both of these dimensions.

Towards increasing cultural and linguistic diversity in acquisition via long-form recordings

Previous work on language acquisition relies predominantly on samples that could be described as WEIRD-EM (Kidd & Garcia, 2022). The WEIRD acronym refers to the fact that study samples drawn from Western, Educated, Industrialized, Rich, and Democratic populations

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

in psychological (Henrich, Heine, & Norenzayan, 2010) and developmental research (Nielsen, Haun, Kärtner, & Legare, 2017) are unusual in terms of both today's human populations, and the history of humankind. WEIRD infants' life experiences probably differ from human infants' life experiences in most of the world and throughout human history along many dimensions, including the number of people they interact with, their family sizes, and the age distribution of the local population – factors that may impact the quantity of child-directed speech coming from adults that infants could experience (Cristia, 2022) as well as the frequency of exposure to other children. A classical line of research called *language socialization* highlights potential cultural diversity that directly impacts young children's language use (e.g., Ochs & Schieffelin, 2001). These researchers study interactions between children and many others in a variety of societies. The focus is often on individual or group interactions, and their rich pragmatics, and it is crucial in reminding us of the diversity of ways in which childcare and language acquisition can be conceived across cultures (e.g., Keller et al., 2007; Rogoff, 2003).

We added the “-EM” to “WEIRD-EM” to highlight that, in general terms, language acquisition research is biased towards English, and, and to monolingual (or at most bilingual) children (Kidd & Garcia, 2022). This is not to deny that researchers in language acquisition have made strong efforts to study children from diverse language backgrounds – but the nature of the methods typically used, relying on manual annotation and transcription, necessarily curtailed both the number of children they could include in their samples, and the publishable progress that could be made. Therefore, this psycholinguistic literature has not fully addressed the many ways in which children's input and outcomes may relate to variation at population and individual levels. For example, at the population level, variation in the prevalence of allocare, particularly

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

from older children, has been tied to variation in input composition (with less speech from mothers and more from other children in Loukatou et al., 2022's study of three children; cf. Kondaurova et al., 2022); at the individual level, variation in terms of birth order has been linked to language outcomes (e.g., Havron et al., 2021, based on an epidemiological cohort); and although effects are small and variable, some suggestions have been made for variation as a function of children's reported sex (e.g., Oller et al., 2020).

To base our developmental theories on a broader range of populations and individuals, we need to study development at scale, which requires zooming out from a rich situated understanding of specific interactions, as well as from the highly intensive nature of manual annotation and transcription. We propose to study *vocal behavior*, without describing its intent or content, by analyzing vocalizations that occur in children's daily experiences as captured by long-form recordings. As summarized above, this method is already allowing diversification of participant samples, with our best estimates of input emerging from studies of children growing up in rural settings (Casillas et al., 2020, 2021; Scaff et al., 2022). Of course, we cannot use this approach to learn e.g. how children acquire nouns versus verbs, appropriate code-switching, their own place in a stratified society, or when they use one or another language they are exposed to, and many other topics that are crucial to study. However, we *can* investigate broader aspects of language use as summarized in the previous section, contributing a data point that falls outside the WEIRD-EM majority sample (SM1.1). We hope one day research like ours will help us understand how broad cultural variation, such as household structure and the family's subsistence economy, may affect children's input composition (how much they get talked to and by whom), and how this may in turn relate to the children's own linguistic outcomes. Importantly, long-form

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

recordings allow us to capture the language ecology that children are embedded in with minimal distortion (Lavechin et al., 2022).

The precise approach we take relies on an open-source algorithm that has been found to outperform the more widely used commercial one (see Methods for validation details), and which allows us to extract estimates of both children's experiences (their input) and children's own vocalizations (their output), attributing each discovered vocalization to one of four different classes: female adults, male adults, other children, and key child (the infant who is wearing the recording device). Therefore, we also have a first, broad impression of children's own vocal productions. Although children's production is not the main focus of the present study, it is relevant to mention that automated child vocalization counts have been found to be highly reliable between human and automated algorithmic methods (Cristia et al., 2020). In addition, a systematic review finds moderate but significant correlations ($r \sim .3$) between this measure and standardized tests of children's language (Wang et al., 2020). Although the large majority of studies included in that review were on monolingual English learners (and all were on urban samples), one did bear on children learning multiple languages (Ramírez-Esparza et al., 2017), with a correlation $r = .83$ between children's vocalization counts and their language outcomes. Based on results such as these, the language acquisition field has come to view this metric as a potentially useful proxy for early development (Gilkerson et al., 2017). Children's vocalization counts increase with age: In the age range covered in the present study, North American English learners produced an average of 75 vocalizations per hour at about 5 months of age, which increased to about 200 per hour by about 33 months (Gilkerson et al., 2017). Participants in the current study were in a multilingual context, but we do not know when (or even whether) they

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

use one or another language in their output. Our hypothesis is that, even if exposed to multiple languages, these children are learning some language or language combinations, and thus develop some kind of language skill. That is, to a certain extent, the cognitive system underlying *the process of language development* is shared across children regardless of what language, or how many languages, they are exposed to or will come to develop, in the broad behavioral sense defined above, whereby they will use information in their experiences (their language input) and that this may shape their language outcomes (including their vocal production). Leaning on such an assumption, we follow previous research in using child vocalization counts as an unspecific proxy for language development. Note specifically that we do not take a stance in the ongoing debates on the cognitive structures underlying the use of multiple languages (e.g., MacSwan, 2017), since we do not study children's lexicon or grammar, but only their verbal behavior at a broader level of description and thus simply cannot contribute to such debates.

The present work

We used long-form recordings to capture spontaneous linguistic behavior of infants growing up in 11 villages on Malakula, the second largest island in the Republic of Vanuatu. In addition to three official languages (French, English and the English-lexified creole Bislama), around 138 indigenous languages of the Oceanic subgroup of the Austronesian language family are attested (François et al 2015) by a population of approximately 300,000 inhabitants (Vanuatu National Statistics Office, 2020). Malakula is the most linguistically dense of Vanuatu's islands, with up to 42 attested language varieties (François et al 2015), a population of about 32,300 people (Vanuatu National Statistics Office, 2020), and marked socio-cultural and political variation. Outside of the major towns Port Vila (the capital, located on Efate island) and

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

Luganville (located on Espiritu Santo), the vast majority of the population of Vanuatu practices subsistence horticulture combined with fishing, hunting and cash-cropping. Our study sample is predominantly from more rural and remote regions of the archipelago.

Data were collected by the last author in 11 villages on Malakula between 2016 and 2018, with a local field assistant helping in one village. Villages were located in the Northeast (3 villages; 3 dominant language varieties), Northwest (7 villages; 2 dominant language varieties) and Southwest (1 village; 1 dominant language variety) of the island. Children were recruited opportunistically, following both formal presentations of the study aims in local communities and informal conversations with interested mothers. In many of the villages, most eligible children participated, allowing us a near-complete coverage of the small available sample populations.

Brief demographic interviews were conducted with the mothers of all children who participated in the study, following their oral or written consent to take part (see Ethics statement [REMOVED FOR ANONYMIZATION BY JOURNAL POLICY] for further details). We describe the number of languages present in the child's environment at the 'kitchen' level, i.e. the cluster of sleeping houses around one 'kitchen' and which is the primary residence of the focal child within a particular patrilineal compound (a collection of kitchens). A single kitchen can include multiple adults beyond the parents and also adopted or fostered (related) children, which is common in Vanuatu. We asked mothers to report all languages spoken in the kitchen that defines where the child lives. Bislama was reported for all kitchens; for one child included in the study, it was the only language spoken. Therefore, while all villages are multilingual, not all kitchens are. French was reported for six kitchens; English was not reported for any kitchen. The mean number of languages reported in a kitchen was 2.6 (range 1-8). A total of 22 languages are

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

represented in the dataset: 14 attested varieties from Malakula, six from other neighboring islands, alongside Bislama and French. Before proceeding, we stress that we will not classify children based on the number of languages they are exposed to, nor will we measure their acquisition of individual languages (SM1.1).

The village we will call "Village A" is on the Northeast of the island. We collected the majority of data there (from 20 out of 38 included children). It is a short distance from Lakatoro, the main town on Malakula, and may differ from all other sampled villages along dimensions that could affect child care and potentially the prevalence of child-directed speech. These include higher levels of (female) education and employment, greater access to market goods and transportation, better internet and phone coverage, and exposure to foreign tourists. All the other 10 villages (jointly represented by the remaining 18 included children) require combinations of 4x4 truck and boat transportation to reach Lakatoro and are therefore less easily connected. In our data, infants sampled from Village A did not differ from the others in terms of the number of siblings they had, the average number of languages spoken in the kitchen, or in the number of boys versus girls. However, Village A had significantly larger households (mean 6.9) than the other villages considered together (mean 5), which could reflect a higher rate of child fostering.

Importantly, we want to stress that there is a great deal more variability on Malakula than can be captured by our data. Nonetheless, our sample suffices to provide a broad strokes picture of early vocal development in a cultural setting among the many that are under-represented in the developmental literature.

We wanted to know, as a function of children's ages and other descriptors which may be relevant moderators (e.g., child sex):

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

- input: how much talk do the children hear, and who is it coming from?
- output: how much do the children themselves vocalize?
- input-output association: (how) are input and output correlated?

Regarding input, the reports summarized before suggest considerable population variation in this estimate, with the current low-bound estimate at about 16 min/h. Infants in our sample can contribute particularly interesting data to this research area because they are growing up in a multilingual setting, which means that they could potentially provide a novel bound in terms of the minimum amount of language experience that is necessary to become functional language users in the presence of multiple languages. Although there are important limitations to our approach that we turn to in the Discussion, this paper takes a first step to answering that important question by providing an estimate for *total* vocal input. In addition, exploratory analyses document whether input quantity and composition may vary as a function of child and village characteristics.

As for output, we describe children's number of vocalizations produced as the proxy of their early vocal development proposed by previous similar work (Wang et al., 2020), and used also in non-monolingual samples (Ramírez-Esparza et al., 2017), needed to address our third research question. In addition, we were motivated to provide this information so that future studies can use it as an additional comparison point, since there is remarkably little work on this beyond North American samples. It is beyond the scope of the present report to study how these children learn any of the languages they are exposed to.

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

Finally, we ask how input and output estimates relate to each other in order to inform the literature attempting to assess whether input may play a causal role in language development, and if so, whether it is only input from adults or whether other children's speech may also be helpful. The current participant sample is important for such a question since these infants will have been surrounded by a wide diversity of individuals, and are thus in a potentially better position to benefit from their speech than samples typically included in previous research, in which infants grow up in small, nuclear families.

Methods

Participants

See the Ethics approval statement for information about our ethics approval and consent. We excluded data from one child who did not have any data within the 9am-6pm recording period, and one child who was much older (4 years, 5 months at the time of the recordings) than all other children. After these exclusions, we include data from 38 children (15 boys, 23 girls, as reported by parents), belonging to 35 families (three children were siblings). At the time of the recordings, the children averaged 20.11 months of age ($SD = 8.70$; range 5-33). As noted above, 20 of the children were growing up in one village, and the 18 others in 10 other villages. On average, children's birth order was 2 (range 1-5; this variable was strictly equivalent to the number of siblings a child had). The size of the household averaged 6 people (range 3-15). Ethnicity/race is not relevant in this population since all participants are ni-Vanuatu.

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

Procedure and equipment

Infants wore sleeveless cotton t-shirts with a custom-made pocket sewn onto the front, in which a USB-sized audiorecorder was placed (Etekcity©, 2016 model; see Figure 1). These devices can record up to 15 hours with a full charge, and some children were recorded more than once (and therefore given multiple t-shirts). USB-recorders were waterproofed with a small amount of clear cellophane wrap, so that children could move around unimpeded, and the t-shirts could get wet.

----- INSERT FIGURE 1 HERE -----

Either (ANONYMIZED) or a local research assistant brought the pre-prepared equipment to the families' home, or gave them to mothers, showing them how it was used. If this visit occurred early in the day, then the child was fitted with the t-shirt and the recording device was turned on by the researcher. Otherwise, we asked parents to launch the recording the following day. We then picked up the device and t-shirt on our next visit. As a result of this inherent variability in procedure, especially where mothers changed the t-shirts mid-way during recording sessions, we do not have precise start times for all of the recordings, with accuracies probably being at the hourly (not minute) level.

Infants were then free to go about their typical days. Detailed ethnographic information on how young children in these villages spend their time is missing for this population. According to an analysis of 158 societies, mothers were the principal caretakers in only 57% of them; and the number of people as well as their relationship to the child may vary (Barry & Paxson, 1971). By putting the child at the center of the recording, rather than some chosen individual, we decrease

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

the chances of biasing our sample of children's experiences. That is, had we asked mothers to e.g. report on what the child was doing and with whom throughout the day, this would have obliged mothers to follow the child around, which may not be what she typically does, and would have resulted in our overestimating the amount of her speech captured in the recording; and if her presence changed the way siblings or peers acted (e.g., they talk less around her, or they are around the child less since she is already looking after the child), then this may have resulted in our underestimating the prevalence of other child input.

Due to logistical and technical issues, the amount of data varied across children (SM2.2). On average, we could include data from 9.47 hours (SD = 3.93; range 1.50-17.43) recordings occurring between 9am and 6pm. Data points are weighted by length of recording.

Automatic annotations

We used Lavechin, Bousbib, Bredin, Dupoux, and Cristia (2020)'s Voice Type Classifier (VTC) to automatically assign portions of the audio to the key child, other children, female adults, and male adults. A technical description is in SM2.4. VTC was trained, developed, and tested with held-out subsets of a corpus containing over 200 hours of audio from a variety of languages (including 2h29m from the present corpus). F-score performance on the test set from this multilingual corpus (which contained data from both urban samples with few other children present, and rural samples with many other children present) was 77% for key child, 82% for female adult, 42% for male adult, and 26% for other children. Comparison of performance with the state-of-the-art LENA algorithm on a different, completely held-out, dataset revealed that the VTC outperformed LENA on all classes (in all cases, F-scores for VTC versus LENA, key child: 69 versus 55%; female adult: 63 versus 43%; male adult 43 versus 37; other child 33 versus 29%;

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

see Lavechin et al., 2020). In addition, we assessed VTC's performance on a section annotated manually (SM2.5), and found the Voice Type Classifier algorithm performed quite well, with the following F-scores: 73% for key child, 60% for female adult, 56% for male adult, and 46% for other children.

Analyses

Corpus organization, processing, and preliminary analyses were done with ChildProject (Gautheron, Rochat, & Cristia, 2022). Additionally, transparency was ensured by publicly posting all of our materials (https://osf.io/t8r5j/?view_only=4e6f8a3b37f84da681b414bc058deca4, Anonymized, 2022), including code to reproduce results thanks to RMarkdown (Baumer & Udwin, 2015) on R (R Consortium Team, 2013), as well as DataLad (Wagner et al., 2020) and GIN (<https://gin.g-node.org/>).

Given the relatively small sample size, and the number of descriptors extracted from the data set, we use statistical significance descriptively, to discuss whether the effects are large given the variance in the data, and not as hypothesis tests (Amrhein, Trafimow, & Greenland, 2019).

We control for variation in recording length by normalizing to minutes spoken and vocalization counts per hour. Specifically, we sum vocalization numbers for each speaker type (key child, female adult, male adult, other child), and divide these numbers by the duration of the included data, to extract a mean number or duration per hour. The majority of our analyses are based on vocalization counts, rather than vocalization duration, because vocalization counts are more frequently used for children's output, and it seemed to us desirable to use the same metric

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

for input and output. For comparability with previous work on input, however, we additionally report on total vocalization duration.

Results

Children's input

Collapsing across ages, there were significantly more input vocalizations from female adults than from male adults (paired, $t(37) = 12.14$, $p < .001$, $d = 1.969$), and significantly more input vocalizations from other children than female adults (paired, $t(37) = 2.07$, $p = 0.045$, $d = 0.336$) or male adults (paired, $t(37) = 8.33$, $p < .001$, $d = 1.352$). Taking into account both vocalization count and duration, this represented an average of 5.09 min/h of female adult, 1.24 min/h of male adult, and 4.89 min/h of other child vocalizations, totaling about 11.22 min/h (SD = 4.22) of input vocalization. Notice that there are fewer female adult vocalizations than other child vocalizations but similar total quantities from these two speaker types because the latter produce vocalizations that are shorter on average.

Further inspection revealed that the sources of input varied with child age. Figure 2 shows the average number of vocalizations per hour that children were exposed to, as a function of speaker and child age. Mixed regression models controlling for mother ID and village ID (i.e., $lmer(count \sim age + (1|mother_ID) + (1|village_ID), weights=Duration)$) suggested that, with age, neither male nor female vocalization counts varied, but there was a significant increase in speech from other children ($\beta = 7.53$, $t = 4.53$, $p < .001$). This increase is sizable: Assuming 10 waking hours in a day, this translates to ~2,260 more vocalizations with each monthly increase in age. The mixed regression models fit to the male adults' data was a generalized one, using a

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

gamma link, because of the non-normal distribution of this variable (i.e., $voc_mal_ph \sim Age + (1|mother_id) + (1|location_id)$, $data = vtc_md$, $weights = Duration$, $family = Gamma(link = log)$, $control = glmerControl(optimizer = "bobyqa")$). We observed a significant, albeit very small, decrease in male adult vocalizations with child age ($\beta = -0.01$, $t = -4.23$, $p < .001$), which translates to ~3 fewer vocalizations with each monthly increase in age.

----- INSERT FIGURE 2 HERE -----

Exploratory analyses were consistent with the idea that there may be significant variation that can be attributed to child-level or village-level variables (Figure 3 and SM3.5-3.7). A mixed model where speaker type and birth order were allowed to interact (while controlling for age, mother ID and village ID) suggested that, compared to earlier-born infants, later-born infants heard significantly fewer vocalizations from female adults ($\beta = -37.92$, $t = -3.02$, $p = .003$), marginally fewer vocalizations from male adults ($\beta = -24.63$, $t = -1.96$, $p = .054$), and marginally more vocalizations from other children ($\beta = 16.86$, $t = 1.71$, $p = .091$). Another model suggested that, compared to Village A, children growing up in other villages heard significantly more vocalizations from female adults ($\beta = 67.29$, $t = 2.19$, $p = .032$) and male adults ($\beta = 119.10$, $t = 3.87$, $p < .001$). This potentially points to the importance of the age structure of a population (Village A had larger kitchen sizes). Child sex, however, did not explain significant variance in overall input quantity, nor in the amount of input afforded by different speaker types.

----- INSERT FIGURE 3 HERE -----

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

Children's output

In this cross-sectional sample, the number of vocalizations children produced per hour increased with child age ($\beta = 5.97$, $t = 3.18$, $p = 0.003$), a pattern obvious in Figure 4. Once age was controlled for, we did not find significant effects of birth order, child sex, or village (SM3.11).

----- INSERT FIGURE 4 HERE -----

Input and output

Figure 5 shows the key children's vocalization count as a function of the number of vocalizations found for each speaker type separately. Having observed positive associations between, on the one hand, key children's output and, on the other, female adults' and other children's vocalizations, as well as a negative correlation with male adults' input, we followed up with two analyses. First, we assessed whether these associations were significant when controlling for child age, mother ID and village ID ($lmer(count_key_child \sim count*age + (1|mother_ID) + (1|village_ID))$), and found that only the association between other child vocalization counts and key child vocalization counts was significant (see Table 1).

----- INSERT FIGURE 5 HERE -----

----- INSERT TABLE 1 HERE -----

Second, we checked whether the correlations observed could be due to algorithmic confusion. Indeed, correlations between input and output vocalizations can be spuriously inflated due to speaker misclassifications introduced by the diarization algorithm. For instance,

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

vocalizations from the key child may be erroneously attributed to the other children instead, or vice versa. As a result, an actual increase in the quantity of, say, speech from other children around the key child, will spuriously lead to an increase in the amount of speech attributed to the key child by the algorithm. Such spurious correlations may occur among any couple of speakers, to various extents, depending on how good the algorithm is at telling them apart.

In order to assess the significance of the observed correlations in spite of this confusion, we simulated 100,000 realistic datasets in which the null hypothesis was true, based on our study of a large collection of manually annotated data from several corpora (some collected in urban conditions, other in rural conditions) available to our team (SM3.14). This simulation allowed us to measure the correlation coefficients that could be observed due purely to confusion between the speakers (i.e., when the null hypothesis of no association between the speakers was true). In simulations (controlling for potential corpus bias), 27% of the simulated datasets yielded a correlation between key child vocalization count and that of female adults that was as large or larger than the correlation observed (partialling out age, $r_p(37) = .335$), suggesting that the level of association observed between the key child data and that of female adults could be due merely to the algorithm confusing these two speakers sometimes. In contrast, the level of association observed between other child vocalizations and those by the key child were higher than predicted: about 0.5% of the simulations yielded a correlation as high or higher than that observed ($r_p(37) = .859$), which is consistent with an interpretation where key child and other child vocalizations are correlated significantly and positively. As for male adults, the correlation observed when partialling out age is very close to zero ($r_p(37) = -.008$). In sum, these analyses suggest that children's vocalization counts were significantly and positively associated with the

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

number of vocalizations by other children, while association with the number of vocalizations by adults remains inconclusive.

Discussion

We sought out to address three research questions, concerning infants input, output, and the relationship between the two. We contextualize key findings for each question in turn, before turning to a broader discussion.

For clarity, we discuss precise input estimates, although we discuss limitations of this approach below. The total input rate from all speakers combined averaged about 11 min/h (SD = 4.22), which is lower than has been previously observed through the three studies that had the most similar methods and populations (long-form recordings of children growing up in small-scale societies, with estimates considering all speech, and not just adults'; Casillas et al., 2020, 2021; Scaff et al., 2022). Since those papers relied on human-annotated data, we reanalyzed, with the exact same algorithms and procedure used here, the Cougar corpus of long-form recordings gathered from American English-learning infants (VanDam, 2018; VanDam & Silbert, 2016), available from Homebank (VanDam et al., 2016). This dataset was chosen because it is available to Homebank members without restriction; it is particularly large; and includes children within the age range studied here (SM4.1-3). Our analyses suggest these American English monolingual learners are exposed to a total of 16.44 min/h, which is 47% more than the input estimate we draw from our own participants' data. Strikingly, such a low input occurs in a highly multilingual context, with on average 2.6 languages spoken around the child. However, it is possible that not all input counts equally for children's language learning, but instead vocalizations that are directed to the key child are more useful than merely overheard

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

ones (e.g., Weisleder & Fernald, 2013; Shneidman & Goldin-Meadow, 2013). Future work should assess what proportion of those 11 min/h are directed to the key child, in each language, and compare that against input estimates provided in other work. Across five urban and rural monolingual populations studied through representative long-form recordings, child-directed speech has been estimated at around 3-4 min/h (Bunce et al., 2020). The proportion of total input that is directed to the key child varies widely across populations: Casillas and colleagues found that 8% of total input was directed to Yéli Dnye-learning infants (Casillas et al., 2021) and this estimate was 14% for Tseltal Mayans (Casillas et al., 2020). Using short video-recordings, Shneidman and Goldin-Meadow (2013) report that the proportion of utterances that were directed, as opposed to overheard, averaged ~20% among Yucatec Mayan and ~70% among Chicagoans. What proportion of those 11 min/h will infants in our sample hear as directed to them? Assuming the highest estimate among all of these, children in our sample would have received about 7.7 min/h of directed speech, to be distributed among the 2.6 languages they were exposed to (on average, with a range of 1-8 languages). This does not entail that this amount is sufficient to become proficient in all linguistic domains. We are only considering vocalizations and thus we are not including in our estimates non-vocal (visual, tactile, and all other sensory) information, which are relevant for certain aspects of acquisition (e.g., Landau, Gleitman, & Landau, 2009). Moreover, even for something like learning the native language's sound system, it is likely that non-vocal input (via lexical knowledge; e.g., Swingley, 2009) as well as the children's own vocal output (e.g., Oller & Eilers, 1988) matter.

Turning now to the composition of the input, we found the majority of vocalizations came from other children, with female adults being a close second, and both of these contributing

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

significantly more to infants' input than male adults. For comparison, we found that 42% of vocalizations were attributed to other children, 46% to female adults, and 12% to male adults. In the American English Cougar corpus, these percentages were 27%, 48%, and 25% respectively. The fact that more of the input comes from female than male adults has been observed not only in research on rural communities but also in urban settings (see Bunce et al., 2020 for convergent results). In contrast, the proportion of input from other children varies widely not only between Cougar and our data, but also across Tseltal, Tsimane', and Yéli data reported on in previous work. For instance, only 20% of Tseltal infants' input came from other children (Casillas et al., 2020); whereas other children contributed about as much to the input as female adults did among the Tsimane' (Scaff et al., 2022). Increases with age were only significant in two of these three studies (Casillas et al., 2021; Scaff et al., 2022), but in the remaining case, this may be due to limited power as only 10 infants could be included (Casillas et al., 2020). Taking all of these results together, the picture is one of considerable diversity across populations that are otherwise similar in many respects, including the fact that young children have been described as integrating peer groups from a young age (as in many other societies; Lew-Levy et al 2020; Hewlett & Cavalli-Sforza 1986).

These general trends should not detract from a consideration of individual variation in input quantity and composition, some of which was statistically traced back to child-level variables (birth order) and potential village differences. As these analyses are purely exploratory, we do not want to dwell on these results. However, we would like to point out connections with other research. For birth order, our input results are only partially consistent with the "resource dilution" theory underlying recent epidemiological studies on urban cohorts (e.g., Havron et al.,

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

2021). These studies observe lower language skills among later-born than earlier-born children, and attribute this to differences in linguistic stimulation, as siblings would compete for parental attention and stimulation. While we find decreases in adult speech in the input as a function of birth order, we do not observe significant differences in key children's vocalization counts (perhaps because of compensation by the marginal increase in other children's vocalizations). For potential village differences, Village A differed from other villages along many dimensions, but a salient one is the level of market integration. A recent report by Padilla et al. (2021) discusses differences in input quantity and composition in a bilingual (Mayan-Spanish) village that experienced fast market integration, by collecting infant-centered short video-recordings in 2007 and 2013. The authors find a decrease in speech afforded to young children by other children, which they reason may result from increased school attendance and/or reduced infant caregiving by older children (although neither explanation was directly measured in their study). In contrast, we found significant differences in terms of adult input, with fewer male and female vocalizations in the more market-exposed Village A than the combined data for other villages. Given that ours is only the second paper to report on differences in input composition that are potentially associated with market integration, more work is needed before we can understand the complex pathways through which integration may relate to children's language environments.

Without negating the importance of individual and group variability, the general trend observed was for relatively low levels of input, particularly if only adult speech is considered. Based on previous research extolling the virtues of adult speech (e.g., Ramírez-Esparza et al., 2014), one may have expected slower language development. Our data clearly suggest this is not the case, with about 4 times as many hourly vocalizations for the youngest age represented in this

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

corpus (5 months) compared to Gilkerson et al. (2017)'s large sample of North American English learners recorded at the same age. The difference across the two samples narrows over age, with child vocalization counts in Vanuatu that are only 50% higher among our oldest infants (33 months) when compared with age-matched American ones. This result should be interpreted with care, since our algorithm performs better than Gilkerson et al. (2017)'s, and in particular, it has a lower tendency to miss key children's vocalizations. For a closer comparison, we extracted predictions for amount of vocalizations for 5- and 33-month-old typically developing children in the Cougar dataset (SM4.7), analyzed with the same procedure as our data, and found estimates much closer to ours: At 5 months, we estimate about 159 vocalizations per hour from our corpus versus 153 for Cougar; and at 33 months, we estimate around 326 vocalizations per hour from our data versus 306 for Cougar. Although the meta-analysis summarized in the Introduction found significant correlations between children's vocalization counts and standardized tests (Wang et al., 2020), these correlations were only moderate in size, based exclusively on data from urbanites, and, in any case, the number of vocalizations infants spontaneously produce is not a direct proxy of their grammatical or vocabulary skills. For the time being, however, our method may be one of the few ones that can be employed in such a highly multilingual setting, where it will be basically impossible to develop formal grammatical and vocabulary tests that can be administered to children aged 5-33 months, across the 22 languages the children in this sample were exposed to. In the future, it may be interesting to adapt the descriptive method used by Casillas et al. (2020, 2021) to try to infer potential differences in complexity from those spontaneous vocalizations, or other techniques that allow us to more finely measure individual and group vocal development (e.g., Cychosz et al., 2021).

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

Our last research question pertained to the potential association between input and output. Broadly, we expected positive associations with all speaker types, perhaps more strongly for (female) adult speech --- the focus of attention of much language acquisition work (e.g., Ramírez-Esparza et al., 2014). We found that infants who heard more speech by other children also vocalized more. For male and female adults, simulations suggested that the associations observed were also compatible with a null hypothesis of no association. The association we do observe (between key children and other children) could arise via a range of causal processes, all of which could be jointly true, including: 1) infants who are spoken to more vocalize more; 2) infants who vocalize more elicit more vocalizations; 3) through a confound (e.g., motorically developed infants vocalize more and are better able to follow their peer group around). Future work may also be in a better position to study the precise ways in which infants interact with other children (including both younger and older siblings), and what contexts they vocalize in, for instance by asking families to interpret clips extracted from the long-form recording.

The present study is one of a handful using long-form recordings as a way to more accurately and naturalistically represent language experiences in a more varied set of populations (cf. Soderstrom et al., 2021). This move is important as researchers in child development become increasingly aware of the low levels of linguistic and cultural diversity represented in our mainstream journals (Heinlein et al., 2017; Kidd & Garcia, 2022). Long-form recordings hold many promises in this regard, including capturing children's spontaneous language use with minimal discomfort and intrusion for participants, particularly when paired with automatized algorithms like those used here, which were trained with multi-cultural data allowing relatively high levels of accuracy even in multilingual corpora like ours (Cychosz & Cristia, 2022).

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

Despite its many strengths, this method has several drawbacks. Some limitations in our data are inherent to our equipment and field conditions, including the fact that there is wide variability in the amount of audio available from each child. Others are more general to the field: There are relatively few reports on the accuracy of automated algorithms applied to long-form recordings (see Cristia et al., 2020, for a meta-analysis of LENA), and none for VTC in particular (other than the evaluation the algorithm authors themselves have carried out; Lavechin et al., 2020). To our knowledge, no study has reported the absolute error rates in number of vocalizations or amount of speech as estimated algorithmically from the full recordings versus estimated through human annotation, which means that we do not know the extent to which our estimates (11 min/h for our data, 16.4 min/h for Cougar) are precise. The most salient limitation is the fact that the different languages cannot be identified algorithmically, and it will be challenging to find research associates who are able to identify each of the 22 indigenous languages represented in our dataset, many of which are endangered or have not yet been systematically documented. Until this is done, we will not be able to tell how much each of the languages was spoken. By the same token, it would be ideal to have an algorithmic approach to distinguishing speech that was directed to the key child from that merely overheard. Unfortunately, this has proven challenging in long-form recordings even in monolingual North American data (Schüller et al., 2017). Moreover, evidence is accumulating on the cross-cultural and cross-linguistic diversity in phonetic cues used by male and female adult speakers in their infant-directed speech (e.g., McClay et al., 2022) -- and little is known about these cues in the speech of other children. Given that, at least in our data, a sizable proportion of input afforded to infants came from other children, solving this technical and theoretical challenge will be required before we can gain a full picture of input composition. Additionally, as mentioned above, our

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

measure of child output is also a coarse approximation to their vocal development, and cannot tell us about their phonological, morphosyntactic, or pragmatic acquisition in each of the languages reportedly used in their village. Thus, we still have a way to go before being able to address key issues regarding the mechanistic causal links between input and outcomes, including cases of reverse causality in which the child's vocalizations shape their input. Finally, these broad quantitative analyses do not consider the child in their holistic, multimodal language socialization context. Research with qualitative or mixed methods is necessary to better understand the sociological and cultural contexts in which the different languages are used, and how children and those around them co-construct meaning.

Conclusion

Using an innovative technique combining long-form recordings with automated analyses, we were able to provide the first estimate of the amount of vocal input available to young children growing up in multilingual villages. Although additional methodological work is needed to measure the error term in this estimate, our results hint at an even lower bound in terms of the amount of experience required to gain (presumed) functional use of language (in the broad behavioral sense discussed in this paper). We do not have data on the prevalence in the input, or in the child's production, of the multiple languages reported at the household level, so assessment of multilingual proficiency remains for future work. Children's input was dominated by vocalizations from female adults and other children, particularly for the oldest infants in our sample. The strongest association between children's own vocalization counts was with those of other children, and not those with adults.

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

References

- Cristia, A., Gautheron, L., & Colleran, H. (2022). Supplementary materials to Vocal input and output among infants in a multilingual context: Evidence from long-form recordings in Vanuatu. <https://osf.io/t8r5j/>
- Amrhein, V., Trafimow, D., & Greenland, S. (2019). Inferential statistics as descriptive statistics: There is no replication crisis if we don't expect replication. *The American Statistician*, 73(sup1), 262–270. <https://doi.org/10.1080/00031305.2018.1543137>
- Baumer, B., & Udwin, D. (2015). R markdown. *Wiley Interdisciplinary Reviews: Computational Statistics*, 7(3), 167–177. <https://doi.org/10.1002/wics.1348>
- 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>
- Barry, H., & Paxson, L. M. (1971). Infancy and early childhood: Cross-cultural studies. *Ethnology*, 10, 467-508. <https://doi.org/10.2307/3773177>
- Bunce, J., Casillas, M., Bergelson, E., Rosemberg, C., Rowland, C., Warlaumont, A. S., & Soderstrom, M. (2020). A cross-cultural examination of child directed speech across development. <https://doi.org/10.31234/osf.io/723pr>
- Byers-Heinlein, K., Schott, E., Gonzalez-Barrero, A. M., Brouillard, M., Dubé, D., Jardak, A., ... & Tamayo, M. P. (2020). MAPLE: A multilingual approach to parent language estimates.

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

Bilingualism: Language and Cognition, 23(5), 951-957.

<https://doi.org/10.1017/S1366728919000282>

Casillas, M., Brown, P., & Levinson, S. C. (2020). Early language experience in a Tzeltal 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>

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. (2022). A systematic review suggests marked differences in the prevalence of infant-directed vocalizations across populations. *Developmental Science*.

<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. *Child Development*, 90, 759–773.

<https://doi.org/10.1111/cdev.12974>

Cristia, A., Bulgarelli, F., & Bergelson, E. (2020). Accuracy of the Language Environment Analysis (LENATM) system segmentation and metrics: A systematic review. *Journal of Speech, Language, and Hearing Research*, 63(4), 1093-1105.

https://doi.org/10.1044/2020_JSLHR-19-00017

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

- Cychosz, M., & Cristia, A. (2022). Using big data from long-form recordings to study development and optimize societal impact. *Advances in Child Development and Behavior*, 62. <https://doi.org/10.31219/osf.io/ybqfw>
- Cychosz, M., Cristia, A., Bergelson, E., Casillas, M., Baudet, G., Warlaumont, A. S., ... & Seidl, A. (2021). Vocal development in a large-scale crosslinguistic corpus. *Developmental Science*, 24(5), e13090. <https://doi.org/10.1111/desc.13090>
- Ferjan Ramírez, N., Hippe, D. S., Correa, L., Andert, J., & Baralt, M. (2022). Habla conmigo, daddy! Fathers' language input in North American bilingual Latinx families. *Infancy*, 27(2), 301-3023. <https://doi.org/10.1111/infa.12450>
- François, A., Lacrampe, S., Franjeh, M., and Schnell, S. (2015). The languages of Vanuatu: Unity and diversity. Canberra: Asia Pacific Linguistics. URI: <http://hdl.handle.net/1885/14819>
- Garcia-Sierra, A., Rivera-Gaxiola, M., Percaccio, C. R., Conboy, B. T., Romo, H., Klarman, L., ... Kuhl, P. K. (2011). Bilingual language learning: An ERP study relating early brain responses to speech, language input, and later word production. *Journal of Phonetics*, 39(4), 546–557. <https://doi.org/10.1016/j.wocn.2011.07.002>
- Gautheron, L., Rochat, N., & Cristia, A. (2022). Managing, storing, and sharing long-form recordings and their annotations. *Language Resources and Evaluation*, 1-33. <https://doi.org/10.1007/s10579-022-09579-3>
- Gilkerson, J., Richards, J. A., Warren, S. F., Montgomery, J. K., Greenwood, C. R., Kimbrough Oller, D., ... & Paul, T. D. (2017). Mapping the early language environment using all-day

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

- recordings and automated analysis. *American Journal of Speech-Language Pathology*, 26(2), 248-265. https://doi.org/10.1044/2016_AJSLP-15-0169
- Goldin-Meadow, S. (2015). Studying the mechanisms of language learning by varying the learning environment and the learner. *Language, Cognition and Neuroscience*, 30(8), 899-911. <https://doi.org/10.1080/23273798.2015.1016978>
- Golinkoff, R. M., Hoff, E., Rowe, M. L., Tamis-LeMonda, C. S., & Hirsh-Pasek, K. (2019). Language matters: Denying the existence of the 30-million-word gap has serious consequences. *Child Development*, 90(3), 985–992. <https://doi.org/10.1111/cdev.13128>
- Grüter, T., Hurtado, N., Marchman, V. A., & Fernald, A. (2014). Language exposure and online processing efficiency in bilingual development. *Input and Experience in Bilingual Development*, 13, 15–36. <http://digital.casalini.it/9789027269454>
- Havron, N., Lovcevic, I., Kee, M. Z. L., Chen, H., Chong, Y. S., Daniel, M., Broekman, B. F. P., & Tsuji, S. (2022). The effect of older sibling, postnatal maternal stress, and household factors on language development in two- to four-year-old children. *Developmental Psychology*, 58(11), 2096–2113. <https://doi.org/10.1037/dev0001417>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, 466(7302), 29–29. <https://doi.org/10.1038/466029a>
- Hewlett, B. S., & Cavalli-Sforza, L. L. (1986). Cultural transmission among Aka pygmies. *American Anthropologist*, 88(4), 922–934. <https://doi.org/10.1525/aa.1986.88.4.02a00100>.

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

- Keller, H., Abels, M., Borke, J., Lamm, B., Su, Y., Wang, Y., & Lo, W. (2007). Socialization environments of Chinese and Euro-American middle-class babies: Parenting behaviors, verbal discourses and ethnotheories. *International Journal of Behavioral Development*, 31(3), 210–217. <https://doi.org/10.1177/0165025407074633>
- Kidd, E., & Garcia, R. (2022). How diverse is child language acquisition research? *First Language*, 42(6), 703-735. <https://doi.org/10.1177/01427237211066405>
- Kondaurova, M. V., Zheng, Q., VanDam, M., & Kinney, K. (2022). Vocal turn-taking in families with children with and without hearing loss. *Ear and Hearing*, 43(3), 883-898. <https://doi.org/10.1097/AUD.0000000000001135>
- Kremin, L. V., & Byers-Heinlein, K. (2021). Why not both? Rethinking categorical and continuous approaches to bilingualism. *International Journal of Bilingualism*, 25(6), 1560-1575. <https://doi.org/10.1177/13670069211031986>
- Landau, B., Gleitman, L. R., & Landau, B. (2009). *Language and experience: Evidence from the blind child*. Cambridge MA: Harvard University Press.
- Lavechin, M., Bousbib, R., Bredin, H., Dupoux, E., & Cristia, A. (2020). An open-source voice type classifier for child-centered daylong recordings. *Interspeech*. <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.

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

Annual Review of Linguistics, 8, 389-407.

<https://doi.org/10.1146/annurev-linguistics-031120-122120>

Lew-Levy, S., Kissler, S. M., Boyette, A. H., Crittenden, A. N., Mabulla, I. A., & Hewlett, B. S. (2020). Who teaches children to forage? Exploring the primacy of child-to-child teaching among Hadza and BaYaka hunter-gatherers of Tanzania and Congo. *Evolution and Human Behavior*, 41(1), 12-22.

<https://doi.org/10.1016/j.evolhumbehav.2019.07.003>

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>

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(6), 1173-1192.

<https://doi.org/10.1017/S0305000921000623>

MacSwan, J. (2017). A Multilingual Perspective on Translanguaging. *American Educational Research Journal*, 54(1), 167-201. <https://doi.org/10.3102/0002831216683935>

Marchman, V. A., Martínez, L. Z., Hurtado, N., Grüter, T., & Fernald, A. (2017). Caregiver talk to young Spanish-English bilinguals: Comparing direct observation and parent-report measures of dual-language exposure. *Developmental Science*, 20(1), e12425.

<https://doi.org/10.1111/desc.12425>

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

McClay, E. K., Cebioglu, S., Broesch, T., & Yeung, H. H. (2022). Rethinking the phonetics of baby-talk: Differences across Canada and Vanuatu in the articulation of mothers' speech to infants. *Developmental Science*, 25(2), e13180.
<https://doi.org/10.1111/desc.13180>

Montant, A. S., Byers-Heinlein, K., & Perez, M. L. (2022). The more they hear the more they learn? Using data from bilinguals to test models of early lexical development.
<https://psyarxiv.com/zd3m8>.
<https://doi.org/10.31234/osf.io/zd3m8>

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

Ochs, E., & Schieffelin, B. (2001). Language acquisition and socialization: Three developmental stories and their implications. In A. Duranti (Ed.) *Linguistic Anthropology: A Reader* (pp. 263–301). New York: John Wiley & Sons.

Oller, D. K., & Eilers, R. E. (1988). The role of audition in infant babbling. *Child Development*, 59(2), 441-449. <https://doi.org/10.2307/1130323>

Oller, D. K., Griebel, U., Bowman, D. D., Bene, E., Long, H. L., Yoo, H., & Ramsay, G. (2020). Infant boys are more vocal than infant girls. *Current Biology*, 30(10), R426-R42.
<https://doi.org/10.1016/j.cub.2020.03.049>

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

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(6), 880-89. <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(4), 1216-1234. <https://doi.org/10.1111/cdev.12648>

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

Scaff, C., Casillas, M., Stieglitz, J., & Cristia, A. (2022). Characterization of children's verbal input in a forager-farmer population using long-form audio recordings and diverse input definitions. <https://psyarxiv.com/mt6nz/>

Schüller, B., Steidl, S., Batliner, A., Bergelson, E., Krajewski, J., Janott, C., Zafeiriou, S., et al. (2017). The INTERSPEECH 2017 Computational Paralinguistics Challenge: Addressee, Cold & Snoring. *Proceedings of Interspeech 2017*. <https://doi.org/10.21437/Interspeech.2017-43>

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>

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

- Shneidman, L. A., Arroyo, M. E., Levine, S. C., & Goldin-Meadow, S. (2013). What counts as effective input for word learning? *Journal of Child Language*, *40*(3), 672-686.
<https://doi.org/10.1017/S0305000912000141>
- Soderstrom, M., Casillas, M., Bergelson, E., Rosemberg, C., Alam, F., Warlaumont, A. S., & Bunce, J. (2021). Developing a cross-cultural annotation system and metacorpora for studying infants' real world language experience. *Collabra: Psychology*, *7*(1), 23445.
<https://doi.org/10.1525/collabra.23445>
- Sperry, D. E., Sperry, L. L., & Miller, P. J. (2019). Reexamining the verbal environments of children from different socioeconomic backgrounds. *Child Development*, *90*(4), 1303–1318. <https://doi.org/10.1111/cdev.13072>
- Swingle, D. (2009). Contributions of infant word learning to language development. *Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3617-3632.
<https://doi.org/10.1098/rstb.2009.0107>
- VanDam, M. (2018). VanDam Cougar HomeBank Corpus. <https://doi.org/10.21415/T5WT25>
- VanDam, M., & Silbert, N. H. (2016). Fidelity of automatic speech processing for adult and child talker classifications. *PloS one*, *11*(8), e0160588.
<https://doi.org/10.1371/journal.pone.0160588>
- VanDam, M., Warlaumont, A.S., Bergelson, E., Cristia, A., Soderstrom, M., De Palma, P., & MacWhinney, B. (2016). HomeBank: An online repository of daylong child-centered

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

audio recordings. *Seminars in Speech and Language*, 37, 128-141.

<https://doi.org/10.1055/s-0036-1580745>

Vanuatu National Statistics Office (2020) National Population and Housing Census,

https://vnso.gov.vu/images/Pictures/Census/2020_census/Census_Volume_1/2020NPHC_Volume_1.pdf

Wagner, A. S., Waite, L. K., Meyer, K., Heckner, M. K., Kadelka, T., Reuter, N., ... Hanke, M.

(2020). *The DataLad handbook*. Zenodo. <https://doi.org/10.5281/ZENODO.3608612>

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>

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/10.1177/0956797613488145>

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT



Figure 1. Lewia Jimmysan from northwest Malakula was the first participant in our project, pictured here aged nearly 3 years. She is wearing one of our cotton t-shirts with a USB recorder in the custom-made breast pocket. (Photograph reproduced with permission from her parents Michael and Aline William).

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

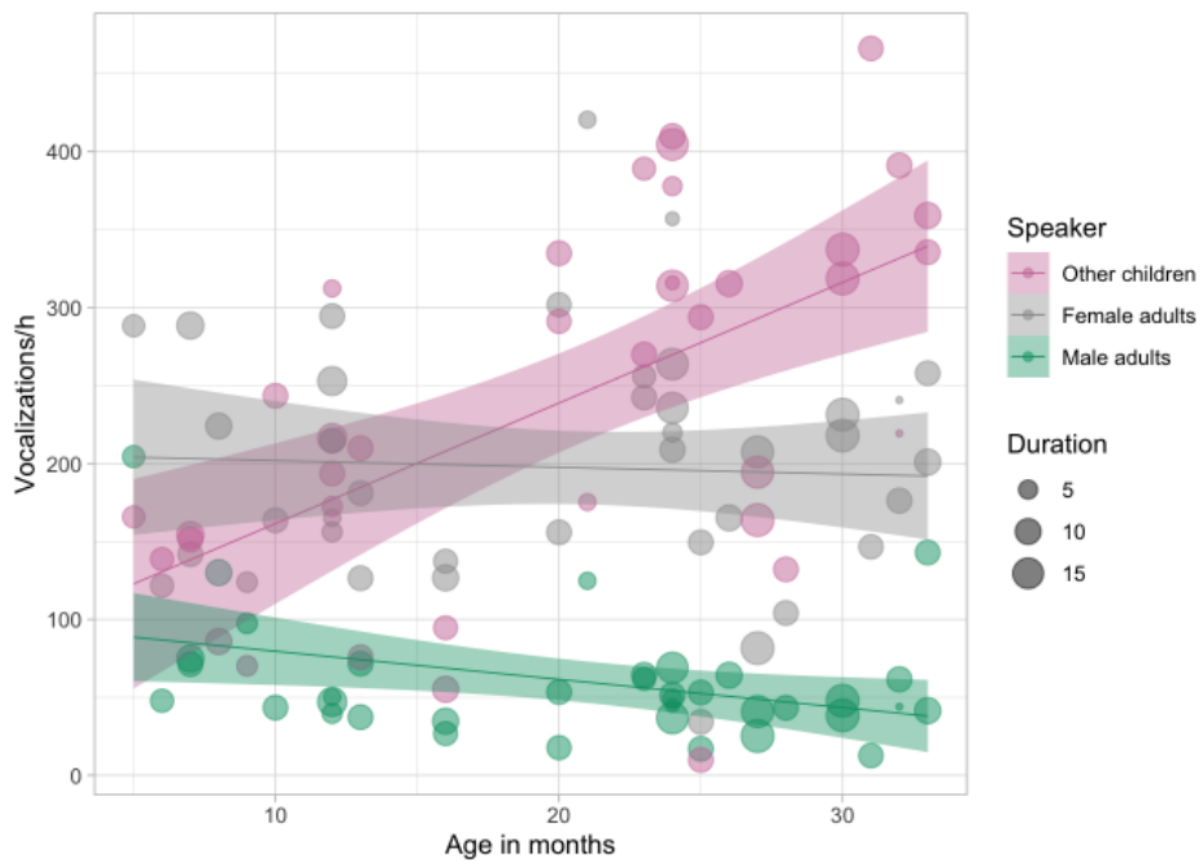


Figure 2. Number of vocalizations as a function of child age and vocalization speaker type. Each point corresponds to one child; the size of the point indicates how much audio data was available for that child. Lines are simple regression lines, fit to the three different speaker types. Duration is in hours.

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

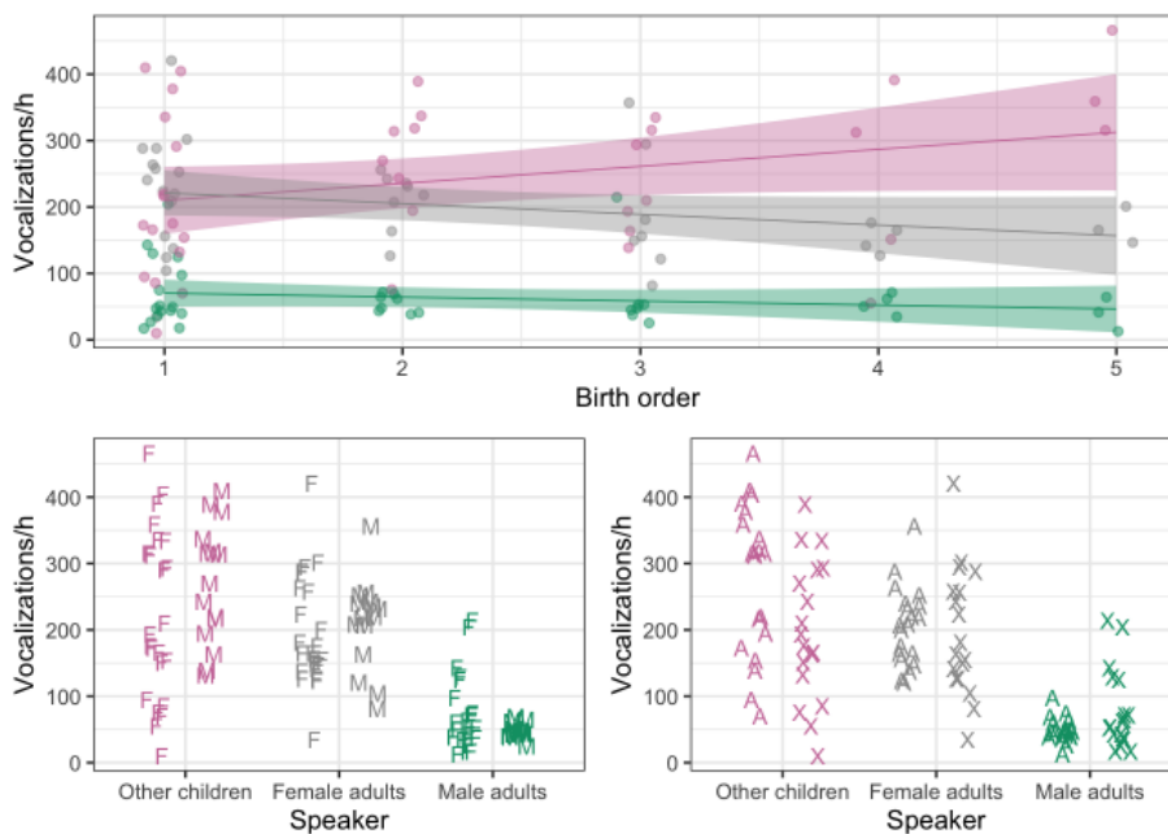


Figure 3. Input vocalization counts by speaker type (female adult: red, male adult: gray, other children: green), as a function of birth order (top panel), child sex (bottom left panel, F=female, M=male), and village (bottom right panel, A=Village A, X=Other villages). Each point corresponds to one child.

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

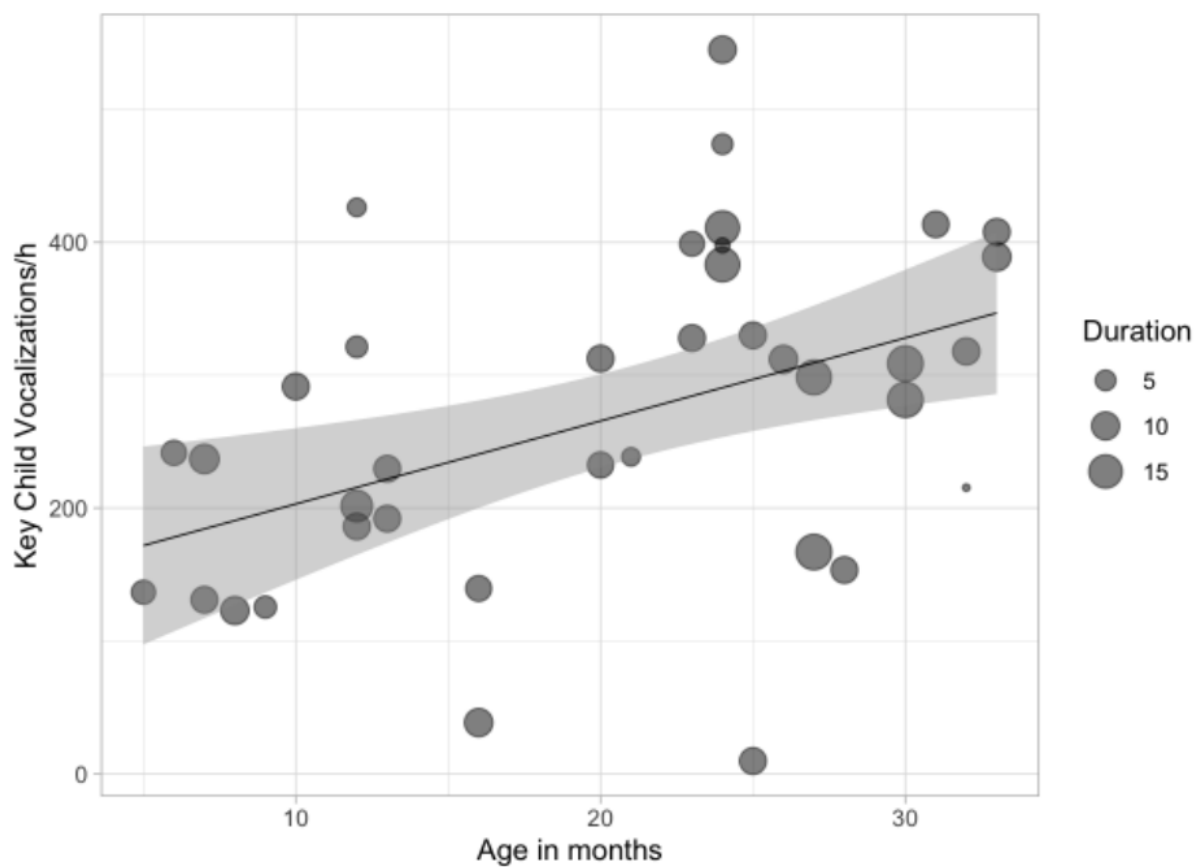


Figure 4. Number of vocalizations from the key children per hour, as a function of child age.

Each point corresponds to one child; the size of the point indicates how much audio data was available for that child (in hours). Lines are simple regression lines.

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

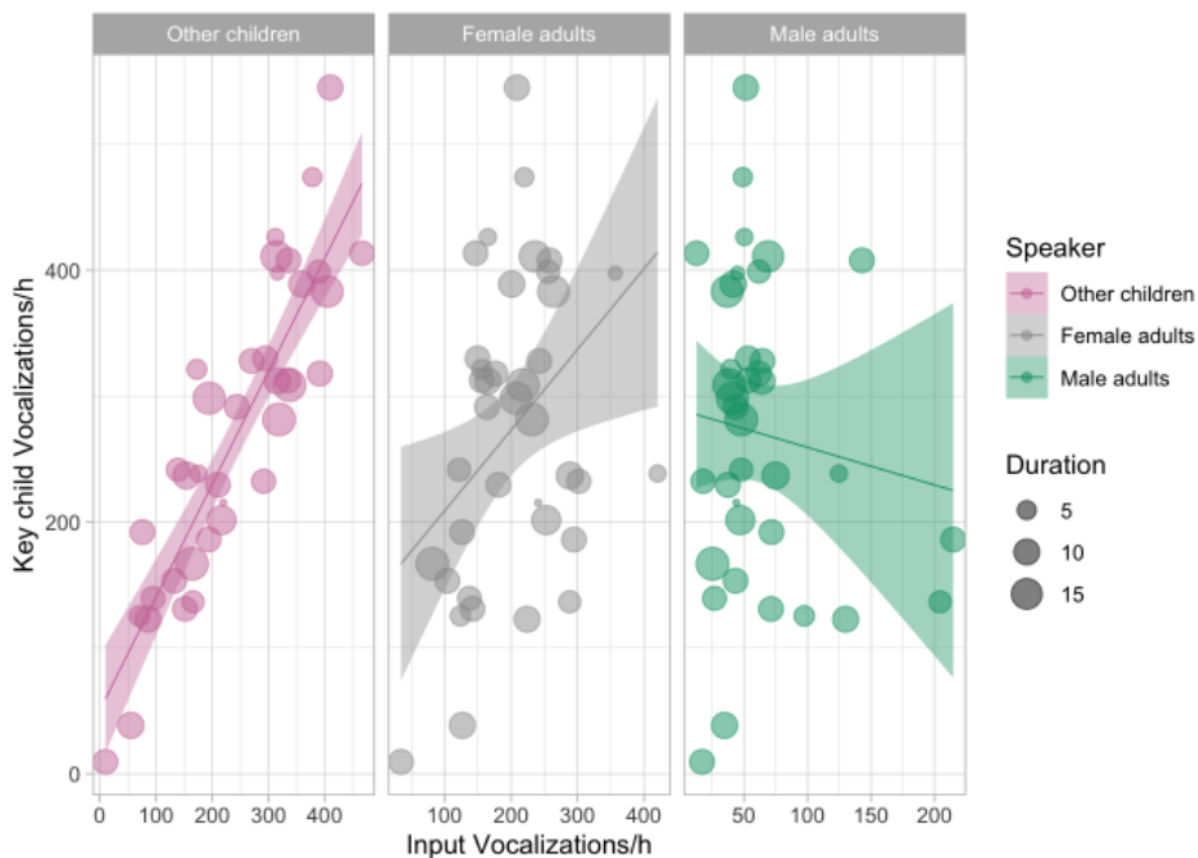


Figure 5. Key child vocalization counts as a function of the counts from the three speaker types contributing to children's input: female adults, male adults, and other children. Each point corresponds to one child; the size of the point indicates how much audio data was available for that child (between 9am and 6pm, in hours). Lines are weighted regression lines.

VOCAL INPUT AND OUTPUT IN A MULTILINGUAL CONTEXT

Table 1. Results of mixed regression models predicting key child vocalization counts from the input counts of each speaker type, in interaction with key child age. Both outcome and predictor variables were z-scored for this analysis. Each cell indicates the standardized β (standard error) and significance level: † < .1; * < .05; ** < .01; *** < .001.

Input	β Child Age	β Input Count	β Child Age*Count
Female adult	0.389 (.133)**	0.207 (.129)	0.184 (.165)
Male adult	0.355 (.134)*	0.176 (.179)	0.249 (.135)†
Other children	-.123 (.083)	0.924 (.095)***	-0.109 (.095)

