Daily auditory environments in French-speaking infants: A longitudinal dataset

Estelle Hervé
CNRS & Aix Marseille Univ
Laboratoire Parole et Langage
Aix-en-Provence, France
estelle.herve@univ-amu.fr

Clément François
CNRS & Aix Marseille Univ
Laboratoire Parole et Langage
Aix-en-Provence, France
clement.francois@cnrs.fr

Laurent Prévot
CNRS & MEAE
CEFC
Taipei, Taiwan
laurent.prevot@cnrs.fr

Abstract

Babies' daily auditory environment plays a crucial role in language development. Most previous research estimating the quantitative and qualitative aspects of early speech inputs has predominantly focused on English- and Spanish-speaking families. In addition, validation studies for daylong recordings' analysis tools are scarce on French data sets. In this paper, we present a French corpus of daylong audio recordings longitudinally collected with the LENA (Language ENvironment Analysis) system from infants aged 3 to 24 months. We conduct a thorough exploration of this data set, which serves as a quality check for both the data and the analysis tools. We evaluate the reliability of LENA metrics by systematically comparing them with those obtained from the ChildProject set of tools and checking the known dynamics of the metrics with age. These metrics are also used to replicate, on our data set, findings from Warlaumont et al. (2014) about the increase of infants' speech vocalizations and temporal contingencies between infants and caregivers with age.

1 Introduction

Infants rely on their daily auditory environment to develop language and other cognitive skills. Pioneering studies interested in these early auditory inputs used observatory experiments in laboratory settings or short recordings that were manually annotated (Hart and Risley, 1992; Keller et al., 2004). In the last decades, technological advances brought new tools that allowed the collection and analysis of more considerable and ecological datasets. Daylong recordings are now increasingly used in developmental studies (Ganek and Eriks-Brophy, 2018; Bergelson et al., 2023), especially since the release of the Language Environment Analysis (LENA) system in 2004.

Daily auditory environments have been described in a variety of populations (Christakis et al., 2009; Aragon and Yoshinaga-Itano, 2012; Caskey et al., 2014; Warren et al., 2010), highlighting the positive effects of early caregiver-infant interactions on language development (Warlaumont et al., 2014; Gilkerson and Richards, 2008; Bergelson and Aslin, 2017). Nonetheless, only two datasets were collected in French-speaking households (Canault et al., 2016; Orena et al., 2019). Here, we expand the literature by describing an original dataset of infants' daylong audio recordings gathered in twenty French-speaking families.

We focused on 3-to-24-month-old babies for several reasons: (1) it allows a direct comparison with Canault et al. (2016)'s and Warlaumont et al. (2014) results ; (2) it includes crucial steps for language development, including the emergence of phonemic categories between 6 and 10 months (Werker and Tees, 1984; Cheour et al., 1998), and the vocabulary spurt between 18 and 24 months (Benedict, 1979; Goldfield and Reznick, 1990; Nazzi and Bertoncini, 2003) ; (3) the two first years of life constitute a critical period where caregiver-infant interactions are early precursors of later language outcomes and cognitive skills (Warlaumont et al., 2022; Gilkerson and Richards, 2009; Weisleder and Fernald, 2013; Bergelson and Aslin, 2017).

LENA's output correlation with human annotations has been assessed in several languages, suggesting good reliability (Xu et al., 2008b; Weisleder and Fernald, 2013; Gilkerson et al., 2015; Busch et al., 2018; Pae et al., 2016; Ganek and Eriks-Brophy, 2017). However, only one study provided evidence for LENA system reliability in European French, yielding relatively good results (Canault et al., 2016). Moreover, LENA validation studies implied listening to the continuous raw audio recordings. In addition to being highly time-consuming, this approach raises critical ethical issues associated with data privacy (Casillas and Cristia, 2019; Cychosz et al., 2020). Here, we override these difficulties by comparing the LENA metrics outputs with other

Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics, pages 132–151

annotation systems.
The paper has the following contributions: (i) describe a new French corpus of auditory LENA-recorded data, (ii) compare different automatic annotation tools, (iii) provide a picture of the daily auditory environment in French-speaking families in 3-to-24-months infants, (iv) show the potential of the data set by replicating daylong recordings-based results on developmental trajectories.

2 Related Work

This section is a brief overview of the existing literature regarding daylong recording studies in developmental populations. We identified two main types of studies: (i) experimental studies that used daylong recordings as a tool to answer a specific research question and (ii) validation studies that focused on assessing the reliability of the recording and analysis tools themselves.

2.1 Experimental studies

Daylong recording studies in infants often involved the LENA system. Four years after its release, the first LENA normative study, the "Natural Language Study (NLS)" was conducted by the LENA Research Foundation (Gilkerson and Richards, 2008). This report relied on the three main LENA metrics (Adult Word Count: AWC; Child Vocalization Count: CVC; Conversational Turn Counts: CTC) to describe daily auditory environments in 329 English-speaking infants aged 2 to 48 months. Then, more experimental works involving daylong recordings in infants began to emerge (see Ganek and Eriks-Brophy (2018) for a review).
Studies that focused on the characteristics of the daily auditory environment in typically developing infants revealed that children's vocalizations and child-caregiver interactions increased with age within the first two years of life (Gilkerson and Richards, 2008; Pae et al., 2016). Warlaumont et al. (2014) proposed a "social feedback loop" in which contingencies between adult-child and child-adult speech-like vocalizations contribute to increasing interactions between infants and adults through age. Additionally, a higher proportion of adult-child interactions has been associated with larger vocabulary size (Weisleder and Fernald, 2013). The impact of various factors like multilingualism (Oller et al., 2010; Orena et al., 2019; Ramírez and Hippe, 2024), socio-economic status (Bergelson et al., 2023), exposure to TV Christakis et al. (2009); Zimmerman et al. (2009), musical inputs (Mendoza and Fausey, 2021), activity during the day (Soderstrom and Wittebolle, 2013) and temporal dynamics of the surrounding sounds (Warlaumont et al., 2022) have been investigated as well. Daylong recording studies in clinical populations showed the importance of understanding infants' daily soundscape for early language intervention (Caskey et al., 2011; Warren et al., 2010; Warlaumont et al., 2014; Aragon and Yoshinaga-Itano, 2012).

Overall, age ranges, sample sizes, and recording spans varied across studies. Some infants were included as early as 2 months of age (Aragon and Yoshinaga-Itano, 2012; Bergelson and Aslin, 2017; Zimmerman et al., 2009), while others started after 12 months of age (Oller et al., 2010; Warren et al., 2010; Weisleder and Fernald, 2013). Children could be followed longitudinally within various periods (Gilkerson et al., 2018; Sy et al., 2023) but not systematically (Weisleder and Fernald, 2013; Bergelson and Aslin, 2017).
Although most studies relied on the LENA system, methodological choices regarding data collection and analysis were various. For example, some authors chose to rely on preexisting datasets that already fitted their research questions (Christakis et al., 2009; Aragon and Yoshinaga-Itano, 2012; Warren et al., 2010). For data analysis, the LENA metrics were mostly used although some preferred to develop their own tools (MacWhinney, 2000; Al Futaisi et al., 2019; Lavechin et al., 2020; Räsänen et al., 2021).

2.2 Validation studies

The first LENA validation study was led in 2008 on American English, as part of the NLS (Xu et al., 2008b). Human annotations were compared with automatic outputs provided by the LENA software to determine agreement scores, measured with Pearson's correlations. LENA's AWC and CVC reached $r = 0.82$ and $r = 0.76$ respectively, indicating reliable LENA annotations for subsequent English-speaking environment studies (Christakis et al., 2009; Warren et al., 2010; Xu et al., 2008a; Zimmerman et al., 2009; Gilkerson et al., 2017).
Later, the same validation procedure was applied to other languages, focusing on the three main LENA metrics (AWC, CVC, and CTC). Overall, the AWC metric was the most reliable, although several authors reported that, on average,

the LENA's estimations were lower than the human counts (Xu et al., 2009; Canault et al., 2016). Agreement scores for AWC were reported in Spanish ($r = 0.80$, Weisleder and Fernald (2013)), Mandarin ($r = 0.72$, Gilkerson et al. (2015)), Korean ($r = 0.72$, Pae et al. (2016)), and Dutch ($r = 0.87$, Busch et al. (2018)). CVC and CTC's reliability were not systematically assessed and yielded variable results, ranging from $r = 0.52$ (Busch et al., 2018) to $r = 0.84$ (Gilkerson and Richards, 2008) (see Table 3 in Appendices). In French, we found Canault et al. (2016)'s report as the only existing validation study so far. They manually annotated and transcribed 324 ten-minute samples recorded in 3-to-48-month-olds: Pearson's correlation scores were $r = 0.64$ for AWC and $r = 0.71$ for CVC. These results suggest good reliability for LENA metrics in French, although slightly below the abovementioned languages.

Cristia et al. (2021)'s comprehensive validation study in three different linguistic and socio-cultural environments calls for more validation studies with more detailed and systematic methods. However, the concurrent emergence of annotation tools (MacWhinney, 2000; Al Futaisi et al., 2019; Lavechin et al., 2020; Räsänen et al., 2021) tends to increase methodological variability. To converge toward a standardized pipeline for daylong data management, Gautheron et al. (2023) developed the ChildProject package. It is compatible with many existing annotation formats and allows annotation systems comparisons. Here, we relied on these tools to compare LENA's metrics with measures extracted from the `Voice Type Classifier (VTC)` from Lavechin et al. (2020) and the `VoCalisation Maturity analysis (VCM)` from Al Futaisi et al. (2019).

3 Rationale

3.1 Participants

Infants were recruited between 3 and 18 months of age in three daycare centers in south-east France. An official collaboration between our team and the daycare centers was established to facilitate both participants' recruitment and data collection. We met parents in person to communicate the project and obtain their informed consent. The French Ethics Committee Review Board approved the study (Agreement 2022-A02281-42) which was conducted according to the guidelines of the Declaration of Helsinki (World Health Organisation, 2008). Parents filled out a questionnaire to ensure that infants did not have any hearing, cognitive, or developmental disorders and that they were raised in a dominant French-speaking environment. Other metadata were gathered through this questionnaire: number of caregivers, musical practice of the caregiver(s), linguistic environment (which language(s) spoken around the child), and socio-economic status (SES) assessed via profession.

Independently of their age at the inclusion date, we followed infants until 24 months of age when possible, or as long as possible otherwise. Twenty infants were involved, with a mean age at inclusion of 12 months (m = 360 days, sd = 132.5). Six additional babies were recruited but excluded from the analysis because parents did not provide enough recordings (<5).

3.2 Procedure

As mentioned above, we collaborated with three daycare centers that became a hub for data collection. Once parents had given their informed consent, they were provided with the LENA materials: a recorder and a t-shirt with a frontal pocket. Each infant had one unique recorder that they kept until the end of data collection. Clothing size was adapted to infants and changed throughout the months when needed. To help parents get used to the LENA system, we gave them some oral instructions when possible so they could ask questions and we could make sure they understood everything. Additionally, all families were given an instruction sheet that was taken from LENA's support materials and adapted to our study. Instructions comprised information about when and how often to record, how to use the device, various recommendations, and the procedure for device deposit and pickup at daycare. Parents also had our contact information and could reach us whenever they needed.

Parents were asked to have their child wear the recorder once a week for a full day, preferentially at home or during the weekends. To limit attrition, we accepted recordings at daycare occasionally, when they could not record another day or if they forgot. The frequency of the recordings was hard to maintain for some families, so we had to send them kind reminders sometimes. But overall, all families were very involved and consistent. We recommended that during a recording day, they never turn the recorder off, limit noisy environ-

ments, and let the recorder nearby while showering and during bedtimes. Once the recording was completed, they were asked to bring the LENA recorders back to the daycare center once a week. Then, the investigator could transfer the data to the database the same day, so parents could get the recorder back and start over for a new week. At the end of data collection (when the child reached 24 months or when families decided to stop), infants were given a "baby researcher" diploma and a customized t-shirt as a reward.

4 Tools and Methods

Daylong recordings present a set of challenges in terms of processing. The first constraint is the data set size: we gathered 10^4 hours of highly heterogeneous recordings (both across and within recordings) that need to be sampled. Existing literature has used the notion of *hot spots* (areas in the recordings with a high density of speech events) as well as a method consisting of human labeling of extremely short sound events (Semenzin et al., 2021). Due to the required infrastructure, the latter approach was not considered for our work. Instead, we applied state-of-the-art computational techniques and packages to perform step-by-step reliability tests and calculate agreement scores between them. The automatic tools we used for our analyses are the LENA suite (Gilkerson and Richards, 2008) and a set of tools developed or adapted whithin the framework of ChildProject (Lavechin et al., 2020).

4.1 LENA

We used the LENA system for both data collection and analysis. For data collection, LENA provides a small digital language processor (DLP) that is easily held in a child's hand and can be directly inserted into child-adapted clothing equipped with a specific pocket on the front. The DLP can save up to 16 hours of auditory input. Recordings are then processed with the LENA software, which provides automatic annotations and quantification reports.
The annotation process starts by segmenting the continuous audio recordings based on acoustic features such as intensity and pitch. The segments are then compared to general models of eight categories (Christakis et al., 2009) to be labeled as target child (CHN), adult male (MAN), adult female (FAN), other child (CXN), TV/electronic sounds (TVN), noise (NOI), silence (SIL), or overlapping sounds (OVL). Next, the four categories CHN, MAN, FAN, and CXN are further analyzed to differentiate speech-related from non-speech vocalizations (see Figure 17). The LENA software provides estimations of the number of words produced by adults (AWC) and infants' speech-related vocalizations (CVC). In this study, we only used the raw sound event segmentation (timestamps) and labeling.

4.2 VTC and VCM

The ChildProject suite starts processing recordings with the *voice-type-classifier (VTC)* (Lavechin et al., 2020), which relies on the state-of-the-art speech diarization tool, *pyannote* (Bredin et al., 2020). *VTC* identifies sound activity segments that can be mapped to some of LENA's categories: target key child (KCHI), other children (CHI), female (FEM), and male (MAL). Another tool, *Vocalisation Maturity analysis (VCM)* (Al Futaisi et al., 2019), refines the output of *VTC*. *VCM* is grounded on the state-of-the-art signal processing and emotion recognition tool, *SMILE* (Eyben et al., 2010), and more precisely on the Geneva Minimal Acoustic Parameter Set (GeMAPS) (Eyben et al., 2015). It adds information to the labeled categories (e.g., speech from the target child) by determining whether the targeted speech is Canonical (CNS), Non-canonical (NCS), cries (CRY), or other sounds (noise, laughter). Such classification has been used in (Casillas et al., 2017), for example.

5 Data set

The corpus currently consists of 8286 hours of LENA daylong recordings. Table 1 indicates the mean, minimum, and maximum values for the recording period (age span), number of recording sessions, and length of the recordings.

	Avg	Min	Max	Sum
Age span (months)	9.85	3	18	-
# sessions	27.0	6	66	540
Duration (hours)	414	87	1022	8286

Table 1: The data set. N = 20 children. Age span: number of months between the first and the last recording.

Table 1 reflects a high variability in parents' use of the LENA device. As mentioned earlier, we asked them to turn the DLP on in the morning and leave it until it automatically turns off after 16 hours of recording. However, some families turned

the device on and off multiple times during the day or stopped the recording before reaching 16 hours. Thus, we observed variability across participants in recording length and number. Additionally, there was variability in the recording span: not all children started the recordings at the same age, and not all were followed until 24 months of age. Given these observations, we selected a sample of children that 1) had at least 9 months of recording span and 2) provided at least 10 recordings. These thresholds allowed us to focus on more representative datasets while maintaining a sufficient number of data points to observe developmental trajectories. A sample of 10 children met these two criteria and were selected for complementary analyses (see Table 2).

	Avg	Min	Max	Sum
Age span (months)	14.2	10	18	-
# sessions	40.9	12	66	409
Duration (hours)	637	192	1022	6366

Table 2: Selected children for individual longitudinal metrics and plots. N = 10 children (age recording span $\geq$ 9 months; number of recordings $\geq$ 10).

6 Investigating the data set

6.1 Testing age

Our first goal was to test the reliability of the metrics extracted with the three targeted tools (LENA, VTC, VCM) on our data. A crucial check for our dataset consisted of testing whether children's production evolved with age. We expected an increase in children's speech-related metrics (such as speaking time and ratio, vocalization counts, etc.), while adults' metrics would remain stable. Finally, the voices of other children present in the recordings were expected to increase as the siblings of the target child followed their own development. We began by examining the production time ratio (the percentage of the recording time occupied by a given category), for example, for the target child as shown in Figure 1 (See Appendix D for other categories).

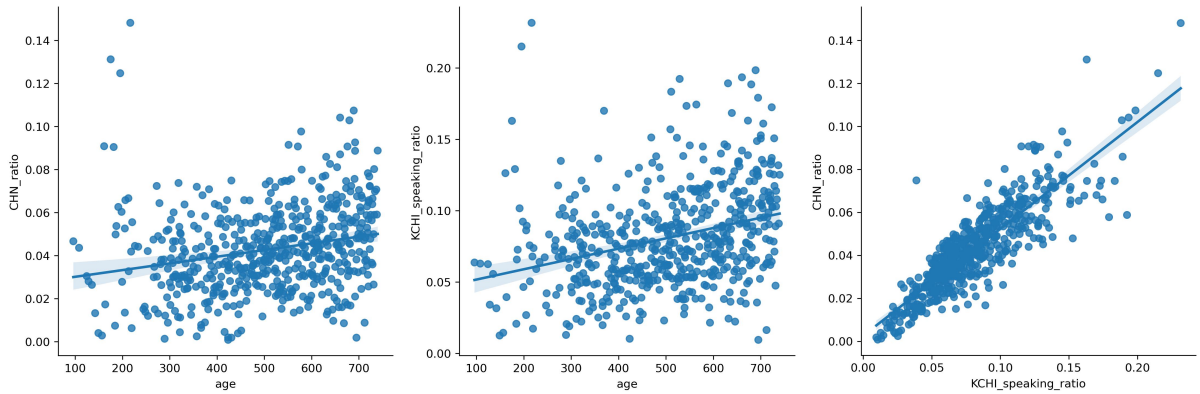
More precisely, we examined the production ratio calculated from LENA (sum of the duration of intervals labeled with *CHN* as the speaker, divided by recording duration), VTC (using the same approach with the label *KCHI*), as well as the correlation between the two measures. Importantly, the ratios obtained for the different voices from these two tools were highly correlated. More generally, all the metrics extracted with both approaches were highly correlated for all comparable categories.

We tested whether age remained a dominant factor when controlling for the available metadata. In Figure 2, we plot the production ratio alongside gender, socio-economic status (high vs. low), and linguistic context (monolingual vs. plurilingual). From these figures, a general observation is that children's speaking ratio increased with age. This was tested by conducting a linear mixed model analysis using pymer4 (Jolly, 2018). We treated *'target child speaking ratio'* as the dependent variable and *'age'* as the fixed effect, with *'child ID'* as the random effect. Only *'age'* had a significant effect on the target child production ratio ($\beta = 0.261$, $SE = 0.028$, and $p < 0.001$), controlling for *'gender'*, *'linguistic environment'*, and *'socio-economic status'* (all not significant).

VCM metrics allowed us to refine our evaluation of child production with age. We applied it to our set of selected children and found (Figure 3) that the increase in speaking ratio was due to an increase in real child speech (both canonical and non-canonical) rather than to a variation in the proportion of cries, laughter, or noise. In addition, we replicated findings from Warlaumont et al. (2014), who tested the evolution of speech-related vocalization ratio (versus non-speech-related) produced by the target child. Using our own pipeline and analysis, we also found a significant increase in this ratio with age while controlling for all metadata ($\beta = 0.219$, $SE = 0.022$, $p < 0.001$) (See Appendix C.1).

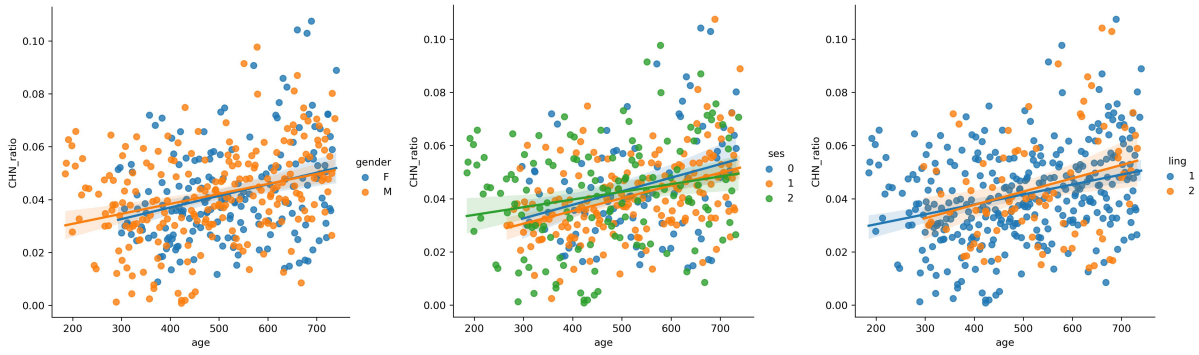
6.2 Interactional metrics

Conversations arguably constitute the most important aspect of the linguistic environment. First, we approached this by examining temporal contingencies (Bloom et al., 1987; Warlaumont et al., 2014) between the child and the other voices in the recordings. Specifically, we analyzed instances of target child productions followed by another participant, as well as target child productions preceded by other participants. Working with manually transcribed data, Nikolaus et al. (2022) explored the effect of time-contingent responses on children's intelligibility. These studies found that (1) caregivers provided more time-contingent responses to intelligible utterances from the child and (2) children



(a) LENA (b) VTC (c) Correlation ($r = 0.862$)

Figure 1: Target Child Speaking Time Ratio for (a) LENA (controlled for child id, $\beta = 0.159$; SE = 0.027; p < 0.001); (b) VTC ($\beta = 0.242$; SE = 0.029; p < 0.001) and (c) correlation plot between LENA and VTC. All children included (n=20). Age in days.



(a) Gender (b) Socio Economic Status (c) Linguistic Context

Figure 2: Target child Production Time (from LENA) differentiating for (a) gender ; (b) socio-economic status ; (c) linguistic background. Only age is significant. The three other variables are not. See Appendix D for details. Selected children. Age in days.

produced more intelligible utterances if their caregivers were responsive. Then, we investigated the "social feedback loop" as proposed by Warlaumont et al. (2014).

We employed a similar approach for both tools at our disposal. For *CHILD>ADULT* contingencies, we selected all target child productions and checked whether there was a production from an adult (MAL + FEM) participant **1 second after**. While LENA metrics extract similar information, we aimed to use the same method for both tools to enable a direct comparison. Figure 4 depicts the comparison for *CHILD>ADULT* contingencies. We also looked at ANY>CHILD contingencies by considering any activity occurring **2 seconds before**[1] before a child production. Figure 5 provides the VTC extraction for *ANY>CHILD* contingencies (additional combinations are included in Appendix E). We tested, for VTC data, the relationship between *age* and temporal contingencies for *CHILD>ADULT* ($\beta = 0.326$, $SE = 0.026$, $p < 0.001$) and *ANY>CHILD* ($\beta = 0.158$, $SE = 0.030$, $p < 0.001$), while controlling for gender, linguistic environment, and socio-economic status. This result was further refined by replicating the second finding from Warlaumont et al. (2014) on our dataset (and with LENA metrics this time). Specifically, we confirmed that children's speech-related productions tend to elicit more feedback from adults (See Appendix C.2).

Finally, we also replicated Warlaumont's result about the social loop on our data. We tested whether initial speech-related CHILD productions

[1]We considered allowing for a longer gap for children's follow-up to be appropriate. To count "turns", LENA metrics use a 5-second threshold.

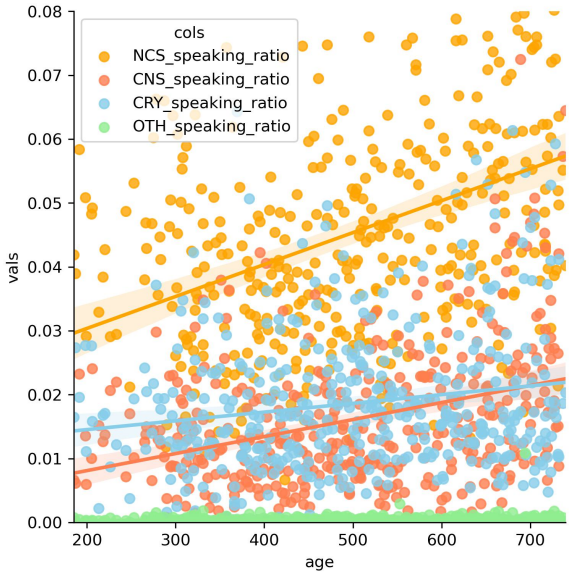


Figure 3: Speech vs. Non-speech VCM extraction for the selected children.

that were followed by an ADULT production 1 second after[2] were more likely to be followed by a speech-related CHILD production than a non-speech CHILD production within 3.5 seconds, compared to initial speech-related CHILD productions that did not receive an adult response.[3] This is illustrated in Figure 6, which shows the difference in speech-related / non-speech related ratio in child's production after an adult's response versus no response. This difference is positive, indicating that adult responses to children's speech-related productions tend to increase the proportion of child speech-related follow-ups.

7 Discussion

The collection and investigation of our original corpus in French households facilitated the comparison of different analysis tools and the replication of previous results within the same age ranges.

First, we identified a robust relationship between target child production metrics and age. As expected, we did not observe a similar significant change for adult voices. However, the evolution was also positive and significant for other children's voices ($\beta = 0.179$, $SE = 0.027$, $p < 0.001$). This can be attributed to the behavioral path of siblings as well as other children in kindergartens. All these results were obtained using both LENA and VTC pipelines. Finally, we found coherent results in line with existing literature and across the tools we used. These results strengthen our confidence in both our recording protocol and in the metrics extraction and analysis.

These comments hold for interactional metrics as well: the increases in temporal contingencies involving the children were consistent with findings from previous studies (Warlaumont et al., 2014; Nikolaus et al., 2022). We calculated temporal contingencies in a way that ensured these increases were not influenced by the overall amount of child productions.

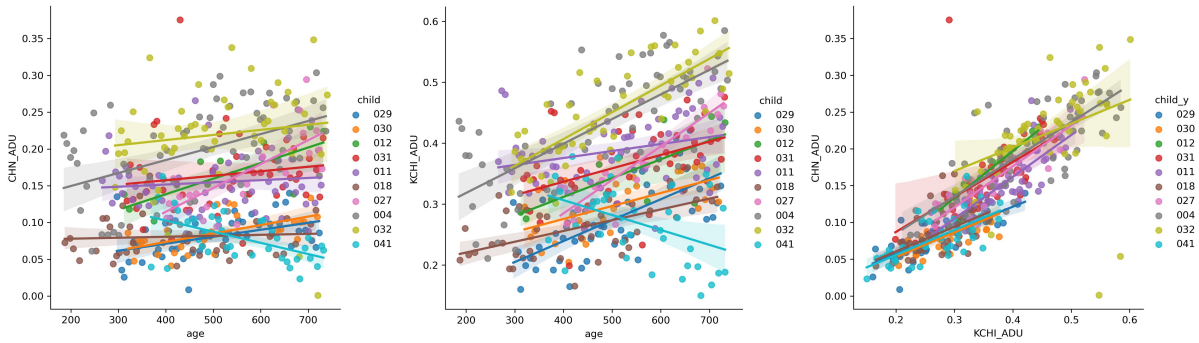
Furthermore, our more detailed analysis (depicted in Figure 3) and the replication of Warlaumont's first results, showed that the increase in child's productions with age was due to speech-related productions and not to vegetative sounds or noise. In summary, our data show that children do produce more speech while growing up in their first two years of life. These increased productions are temporally contingent on other speakers, regardless of the initiator of the interaction (target child or other speaker).

This is further elaborated by the replication of Warlaumont's third result about the social loop that showed the benefit of follow-up productions of adult feedback on children's speech. Contrary to Warlaumont et al. (2014) (but in line with Bergelson et al. (2023) we did not find any effect of parental SES, gender, or linguistic environment on children's productions. This was the same for temporal contingencies used as a proxy for measuring linguistic interaction with the child.

Other studies have attempted to dive further into linguistic metrics from large child-caregiver datasets. Some refined the speech contingencies to distinguish corrective feedback or to approach the grammaticality evaluation of the productions in these datasets (Nikolaus et al., 2023). However, those datasets have been manually transcribed, while ours did not (and will not) receive such a transcription. A second major difference is the age range of the children. Most of the existing studies involved children from 12-24 months up to

[2]Following Warlaumont et al. (2014) and Nikolaus et al. (2022), who followed Oller et al. (2010), which used a 1-second window to extract relevant vocal activity to investigate the "social feedback loop" for children between 8 to 48 months.

[3]This 3.5 second is between the end of initial CHILD production and the start of the following one. The delay is to allow for a potential ADULT response to occur in between. It is not possible to use the ADULT production for proposing a simpler time threshold for the following utterance since there is not always an ADULT production in-between.



(a) LENA (b) VTC (c) Correlation ($r = 0.831$

Figure 4: CHILD>ADULT contingencies for (a) LENA ($\beta = 0.131$; SE = 0.024; p < 0.001), (b) VTC ($\beta = 0.130$, $SE = 0.024$, $p < 0.001$) and (c) correlation plot. Selected children. Age in days.

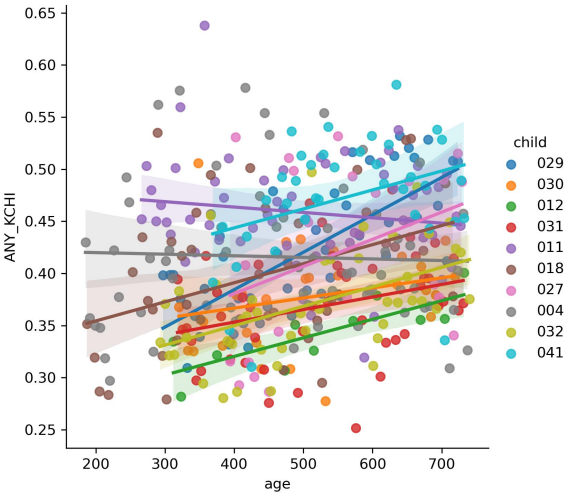


Figure 5: ANY>CHILD contingencies from VTC speech extraction. ($\beta = 0.158$, $SE = 0.030$, $p < 0.001$)

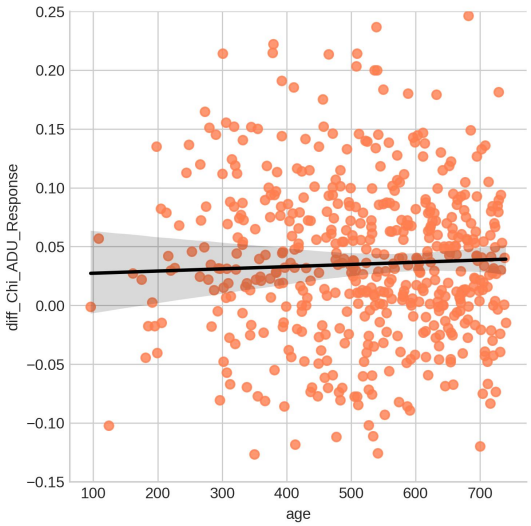


Figure 6: Child ratio difference between Child speech and non-speech productions depending on whether or not an initial child speech-related utterance was responded by an adult or not (positive mean : indicating a tendency for more speech-related responses), replication of Warlaumont et al. (2014).

later ages. This inevitably changes the nature and characteristics of the metrics that can be extracted. Another trend of work focuses on phonetic learning based on daylong recordings, such as Lavechin et al. (2024) on perceptual attunement. Finally, in a more cultural and typological direction, Bergelson et al. (2023) studied a global sample of 1,001 child-centered audio capturing 2- to 48-months-olds from many countries and various cultural backgrounds.

8 Conclusion

This work first constitutes a replication of earlier results on daylong recordings, on a completely new and independent data set, using two different tools. This contributes to answer Cruz Blandón et al. (2023)'s call for more and better meta-studies on long recordings. Indeed, despite their creation cost, daylong recordings' significance is growing in cognitive science. Showing that these data sets, despite their noisy nature, do present enough reliability to gain insights about the children's language and communicative skills development is crucial.

The second contribution consists of the characterization of our data set itself. It is a large data set (>8000 hours of recordings) that is still growing at the time of writing. It is unique by being truly longitudinal with some children's environments being recorded over a two-year span. Finally, other experimental data were collected longitudinally in the same sample of infants. These are beyond the scope of this paper but open up the possibilities of cross-analyses between the characterization of

the linguistic environment and other experimental results regarding language development. The present study, therefore, constitutes a crucial first step in this direction through the thorough exploration of the data set, and by considering individual variability.

From here, we now plan to refine the analyses by entering into more linguistic characterizations of these productions in terms of richness. We will consider tools that allow for more phonological measures such as Räsänen et al. (2021) and more content-based metrics (Nikolaus et al., 2023, 2024) that have been used so far only on transcript-based corpora from CHILDES (MacWhinney, 2000) and that can be now explored on daylong recordings thanks to the improvement of automatic speech recognition engines.

9 Limitations

One major limitation of this work is the absence of manual annotation. For legal and ethical reasons, we are not in position to perform extensive manual annotations of raw audio data, as well as sharing raw audio. We needed to find other ways to check the reliability of our dataset. By replicating previous results from the literature and comparing different computational tools, we reinforced our trust in our dataset and overcame this constraint. All metrics derived from the corpus related to this paper as well as for future work will be made available in the LLDC public repository on Ortolang institutional platform https://www.ortolang.fr/. Moreover the code for producing the analyses presented in this paper is available at : https://github.com/prevotlaurent/LENA_CMCL.

Acknowledgments

We thank all the babies and their families for participating in the study. We also thank the directors and the staff of the Babilou daycare centers where the babies were evaluated. We thank Mathilde Gaujard for her help in the recruitment and data collection. This research has been supported by an ANR grant ("BabyLang project", ANR-20-CE28-0017) to CF and a CNRS 80PRIME grant to CF and LP ("LangDev project"). This work, carried out within the Institut Convergence ILCB (ANR-16-CONV-0002), has benefited from support from the French government, managed by the French National Agency for Research (ANR) and the Excellence Initiative of Aix-Marseille University (A*MIDEX).

References

Najla Al Futaisi, Zixing Zhang, Alejandrina Cristia, Anne Warlaumont, and Bjorn Schuller. 2019. Vcmnet: Weakly supervised learning for automatic infant vocalisation maturity analysis. In *2019 International Conference on Multimodal Interaction*, pages 205–209.

Miranda Aragon and Christine Yoshinaga-Itano. 2012. Using language environment analysis to improve outcomes for children who are deaf or hard of hearing. In *Seminars in Speech and Language*, volume 33, pages 340–353. Thieme Medical Publishers.

Helen Benedict. 1979. Early lexical development: Comprehension and production. *Journal of child language*, 6(2):183–200.

Elika Bergelson and Richard N Aslin. 2017. Nature and origins of the lexicon in 6-mo-olds. *Proceedings of the National Academy of Sciences*, 114(49):12916–12921.

Elika Bergelson, Melanie Soderstrom, Iris-Corinna Schwarz, Caroline F Rowland, Nairán Ramírez-Esparza, Lisa R. Hamrick, Ellen Marklund, Marina Kalashnikova, Ava Guez, Marisa Casillas, et al. 2023. Everyday language input and production in 1,001 children from six continents. *Proceedings of the National Academy of Sciences*, 120(52):e2300671120.

Kathleen Bloom, Ann Russell, and Karen Wassenberg. 1987. Turn taking affects the quality of infant vocalizations. *Journal of child language*, 14(2):211–227.

Hervé Bredin, Ruiqing Yin, Juan Manuel Coria, Gregory Gelly, Pavel Korshunov, Marvin Lavechin, Diego Fustes, Hadrien Titeux, Wassim Bouaziz, and Marie-Philippe Gill. 2020. Pyannote. audio: neural building blocks for speaker diarization. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7124–7128. IEEE.

Tobias Busch, Anouk Sangen, Filiep Vanpoucke, and Astrid van Wieringen. 2018. Correlation and agreement between language environment analysis (lena™) and manual transcription for dutch natural language recordings. *Behavior research methods*, 50:1921–1932.

Mélanie Canault, Marie-Thérèse Le Normand, Samy Foudil, Natalie Loundon, and Hung Thai-Van. 2016. Reliability of the language environment analysis system (lena™) in european french. *Behavior research methods*, 48:1109–1124.

Marisa Casillas, John Bunce, Melanie Soderstrom, Celia Rosemberg, Maia Migdalek, Florencia Alam, Alejandra Stein, and Hallie Garrison. 2017. Introduction: the aclew das template. https://osf.io/aknjv.

Marisa Casillas and Alejandrina Cristia. 2019. A step-by-step guide to collecting and analyzing long-format speech environment (lfse) recordings. *Collabra: Psychology*, 5(1):24.

Melinda Caskey, Bonnie Stephens, Richard Tucker, and Betty Vohr. 2011. Importance of parent talk on the development of preterm infant vocalizations. *Pediatrics*, 128(5):910–916.

Melinda Caskey, Bonnie Stephens, Richard Tucker, and Betty Vohr. 2014. Adult talk in the nicu with preterm infants and developmental outcomes. *Pediatrics*, 133(3):e578–e584.

Marie Cheour, Rita Ceponiene, Anne Lehtokoski, Aavo Luuk, Jüri Allik, Kimmo Alho, and Risto Näätänen. 1998. Development of language-specific phoneme representations in the infant brain. *Nature neuroscience*, 1(5):351–353.

Dimitri A Christakis, Jill Gilkerson, Jeffrey A Richards, Frederick J Zimmerman, Michelle M Garrison, Dongxin Xu, Sharmistha Gray, and Umit Yapanel. 2009. Audible television and decreased adult words, infant vocalizations, and conversational turns: a population-based study. *Archives of pediatrics & adolescent medicine*, 163(6):554–558.

Alejandrina Cristia, Marvin Lavechin, Camila Scaff, Melanie Soderstrom, Caroline Rowland, Okko Räsänen, John Bunce, and Elika Bergelson. 2021. A thorough evaluation of the language environment analysis (lena) system. *Behavior research methods*, 53:467–486.

María Andrea Cruz Blandón, Alejandrina Cristia, and Okko Räsänen. 2023. Introducing meta-analysis in the evaluation of computational models of infant language development. *Cognitive Science*, 47(7):e13307.

Margaret Cychosz, Rachel Romeo, Melanie Soderstrom, Camila Scaff, Hillary Ganek, Alejandrina Cristia, Marisa Casillas, Kaya De Barbaro, Janet Y Bang, and Adriana Weisleder. 2020. Longform recordings of everyday life: Ethics for best practices. *Behavior research methods*, 52:1951–1969.

Florian Eyben, Klaus R Scherer, Björn W Schuller, Johan Sundberg, Elisabeth André, Carlos Busso, Laurence Y Devillers, Julien Epps, Petri Laukka, Shrikanth S Narayanan, et al. 2015. The geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing. *IEEE transactions on affective computing*, 7(2):190–202.

Florian Eyben, Martin Wöllmer, and Björn Schuller. 2010. Opensmile: the munich versatile and fast open-source audio feature extractor. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 1459–1462.

Hillary Ganek and Alice Eriks-Brophy. 2018. Language environment analysis (lena) system investigation of day long recordings in children: A literature review. *Journal of Communication Disorders*, 72:77–85.

Hillary V Ganek and Alice Eriks-Brophy. 2017. A concise protocol for the validation of language environment analysis (lena) conversational turn counts in vietnamese. *Communication Disorders Quarterly*, 39(2):371–380.

Lucas Gautheron, Nicolas Rochat, and Alejandrina Cristia. 2023. Managing, storing, and sharing long-form recordings and their annotations. *Language Resources and Evaluation*, 57(1):343–375.

Jill Gilkerson and Jeffrey A Richards. 2008. The lena natural language study. *Boulder, CO: LENA Foundation. Retrieved March*, 3(2009):15–17.

Jill Gilkerson and Jeffrey A Richards. 2009. The power of talk. *Impact of adult talk, conversational turns and TV during the critical 0-4 years of child development: Boulder, CO: LENA Foundation*.

Jill Gilkerson, Jeffrey A Richards, Steven F Warren, Judith K Montgomery, Charles R Greenwood, D Kimbrough Oller, John HL Hansen, and Terrance D Paul. 2017. Mapping the early language environment using all-day recordings and automated analysis. *American journal of speech-language pathology*, 26(2):248–265.

Jill Gilkerson, Jeffrey A Richards, Steven F Warren, D Kimbrough Oller, Rosemary Russo, and Betty Vohr. 2018. Language experience in the second year of life and language outcomes in late childhood. *Pediatrics*, 142(4).

Jill Gilkerson, Yiwen Zhang, Dongxin Xu, Jeffrey A Richards, Xiaojuan Xu, Fan Jiang, James Harnsberger, and Keith Topping. 2015. Evaluating language environment analysis system performance for chinese: A pilot study in shanghai. *Journal of Speech, Language, and Hearing Research*, 58(2):445–452.

Beverly A Goldfield and J Steven Reznick. 1990. Early lexical acquisition: Rate, content, and the vocabulary spurt. *Journal of child language*, 17(1):171–183.

Betty Hart and Todd R Risley. 1992. American parenting of language-learning children: Persisting differences in family-child interactions observed in natural home environments. *Developmental psychology*, 28(6):1096.

Eshin Jolly. 2018. Pymer4: Connecting r and python for linear mixed modeling. *Journal of Open Source Software*, 3(31):862.

Heidi Keller, Elke Hentschel, Relindis Dzeaye Yovsi, Bettina Lamm, Monika Abels, and Verena Haas. 2004. The psycho-linguistic embodiment of parental ethnotheories: A new avenue to understanding cultural processes in parental reasoning. *Culture & Psychology*, 10(3):293–330.

Marvin Lavechin, Ruben Bousbib, Hervé Bredin, Emmanuel Dupoux, and Alejandrina Cristia. 2020. An open-source voice type classifier for child-centered

daylong recordings. In *Interspeech 2020 - Conference of the International Speech Communication Association*, Shanghai / Virtual, China.

Marvin Lavechin, Maureen de Seyssel, Marianne Métais, Florian Metze, Abdelrahman Mohamed, Hervé Bredin, Emmanuel Dupoux, and Alejandrina Cristia. 2024. Modeling early phonetic acquisition from child-centered audio data. *Cognition*, 245:105734.

Brian MacWhinney. 2000. The childes project. *Computational Linguistics*, 26(4):657–657.

Jennifer K Mendoza and Caitlin M Fausey. 2021. Everyday music in infancy. *Developmental Science*, 24(6):e13122.

Thierry Nazzi and Josiane Bertoncini. 2003. Before and after the vocabulary spurt: Two modes of word acquisition? *Developmental Science*, 6(2):136–142.

Mitja Nikolaus, Abhishek Agrawal, Petros Kaklamanis, Alex Warstadt, and Abdellah Fourtassi. 2024. Automatic annotation of grammaticality in child-caregiver conversations.

Mitja Nikolaus, Laurent Prévot, and Abdellah Fourtassi. 2022. Communicative feedback as a mechanism supporting the production of intelligible speech in early childhood. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 44.

Mitja Nikolaus, Laurent Prévot, and Abdellah Fourtassi. 2023. Communicative feedback in response to children's grammatical errors. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 45.

D Kimbrough Oller, Partha Niyogi, Sharmistha Gray, Jeffrey A Richards, Jill Gilkerson, Dongxin Xu, Umit Yapanel, and Steven F Warren. 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.

Adriel John Orena, Krista Byers-Heinlein, and Linda Polka. 2019. Reliability of the language environment analysis recording system in analyzing french–english bilingual speech. *Journal of Speech, Language, and Hearing Research*, 62(7):2491–2500.

Soyeong Pae, Hyojin Yoon, Ahyoung Seol, Jill Gilkerson, Jeffrey A Richards, Lin Ma, and Keith Topping. 2016. Effects of feedback on parent–child language with infants and toddlers in korea. *First Language*, 36(6):549–569.

Naja Ferjan Ramírez and Daniel S Hippe. 2024. Estimating infants' language exposure: A comparison of random and volume sampling from daylong recordings collected in a bilingual community. *Infant Behavior and Development*, 75:101943.

Okko Räsänen, Shreyas Seshadri, Marvin Lavechin, Alejandrina Cristia, and Marisa Casillas. 2021. Alice: An open-source tool for automatic measurement of phoneme, syllable, and word counts from child-centered daylong recordings. *Behavior Research Methods*, 53:818–835.

Chiara Semenzin, Lisa Hamrick, Amanda Seidl, Bridgette Kelleher, and Alejandrina Cristia. 2021. Towards large-scale data annotation of audio from wearables: validating zooniverse annotations of infant vocalization types. In *2021 IEEE Spoken Language Technology Workshop (SLT)*, pages 1079–1085. IEEE.

Melanie Soderstrom and Kelsey Wittebolle. 2013. When do caregivers talk? the influences of activity and time of day on caregiver speech and child vocalizations in two childcare environments. *PloS one*, 8(11):e80646.

Yaya Sy, William Havard, Marvin Lavechin, Emmanuel Dupoux, and Alejandrina Cristia. 2023. Measuring language development from child-centered recordings. In *Interspeech 2023*, pages 4618–4622. ISCA.

Anne S Warlaumont, Kunmi Sobowale, and Caitlin M Fausey. 2022. Daylong mobile audio recordings reveal multitimescale dynamics in infants' vocal productions and auditory experiences. *Current directions in psychological science*, 31(1):12–19.

Anne S. Warlaumont, Anne S. Warlaumont, Jeffrey A. Richards, Jeffrey A. Richards, Jill Gilkerson, Jill Gilkerson, D. Kimbrough Oller, and D. Kimbrough Oller. 2014. A social feedback loop for speech development and its reduction in autism. *Psychological Science*.

Steven F Warren, Jill Gilkerson, Jeffrey A Richards, D Kimbrough Oller, Dongxin Xu, Umit Yapanel, and Sharmistha Gray. 2010. What automated vocal analysis reveals about the vocal production and language learning environment of young children with autism. *Journal of autism and developmental disorders*, 40:555–569.

Adriana Weisleder and Anne Fernald. 2013. Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological science*, 24(11):2143–2152.

Janet F Werker and Richard C Tees. 1984. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant behavior and development*, 7(1):49–63.

Dongxin Xu, Umit Yapanel, and Sharmi Gray. 2009. Reliability of the lena language environment analysis system in young children's natural home environment. *Boulder, CO: Lena Foundation*, pages 1–16.

Dongxin Xu, Umit Yapanel, Sharmi Gray, and Charles T Baer. 2008a. The lena language environment analysis system: The interpreted time segments (its) file. *Boulder, CO: Lena Foundation*, pages 1–7.

Dongxin Xu, Umit Yapanel, Sharmi Gray, Jill Gilkerson, Jeff Richards, and John Hansen. 2008b. Signal processing for young child speech language development. In *First Workshop on Child, Computer and Interaction*.

Frederick J Zimmerman, Jill Gilkerson, Jeffrey A Richards, Dimitri A Christakis, Dongxin Xu, Sharmistha Gray, and Umit Yapanel. 2009. Teaching by listening: The importance of adult-child conversations to language development. *Pediatrics*, 124(1):342–349.

A Validation studies

Reference	Language(s)	r(AWC)	r(CVC)	r(CTC)
Busch et al. (2018)	Dutch	0.87	0.77	0.52
Canault et al. (2016)	European French	0.64	0.71	NI
Caskey et al. (2014)	American English and Spanish	0.93	NI	NI
Cristia et al. (2021)	American English	0.76	0.76	0.57
Ganek and Eriks-Brophy (2017)	Vietnamese	NI	NI	0.70*
Gilkerson et al. (2015)	Mandarin	0.72	0.84; 0.70**	0.72
Pae et al. (2016)	Korean	0.72	NI	0.67
Weisleder and Fernald (2013)	Spanish	0.80	NI	NI
Xu et al. (2008b)	American English	0.82	0.76	NI

Table 3: Pearson's r correlation scores between human and LENA automatic annotations for AWC, CVC and CTC in various languages. NI = No Information.*agreement score assessed via a Spearman rank correlation test. **0.84 for speech-like vocalizations, 0.70 for non-speech-like vocalizations.

B Meta-data

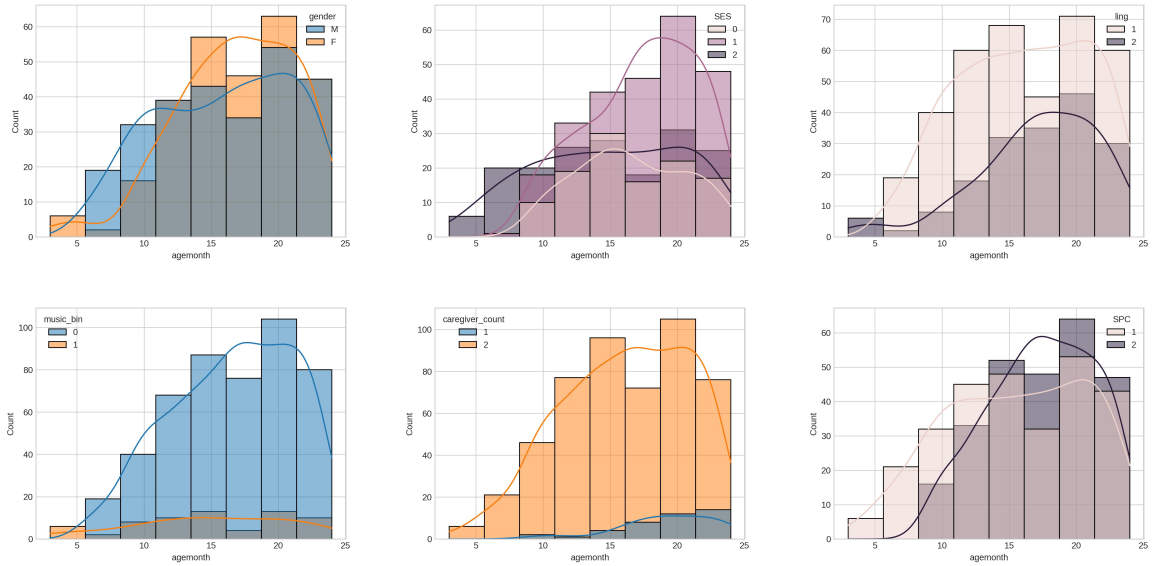
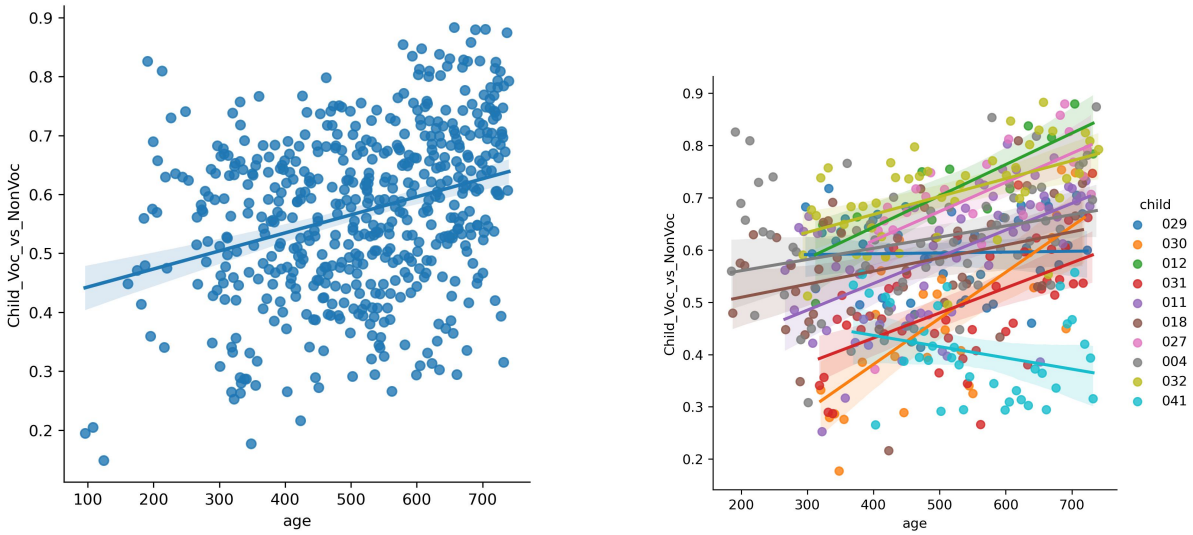


Figure 7: Distribution of session through age, depending on various demographic information: from left to right, top to bottom : gender, socio-economic status (SES: low,mid,high), linguistic environment (1:monolingual, 2: plurilingual), music practice at home(yes/no), number of cargiver, socio-professional category of the parents.

C Replication (Warlaumont et al., 2014)

C.1 Ratio of Speech Related Vocalisations



(a) All children

(b) Selected Children

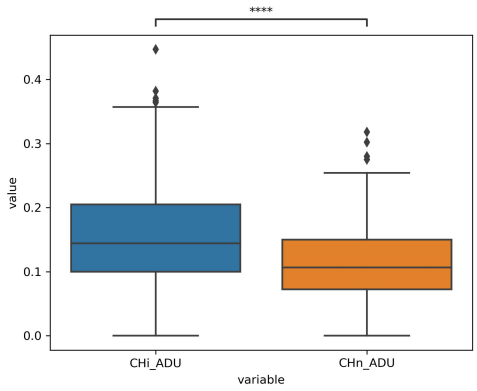
Figure 8: Ratio between the child total amount of speech related productions (`ChildVocDuration`) and all production (including Non speech related) (`ChildNonVocDuration`)(age significant $\beta = 0.219$, $p < 0.001$; `child_id` and all metadata variables not significant). Age in days.

The Figure 8 presents the first result of (Warlaumont et al., 2014), which is that the speech related percentage productions of children increase with age. See the figure caption and below for statistics.

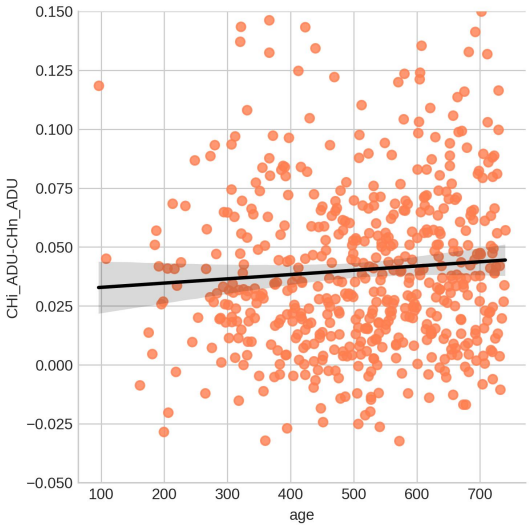
```
Formula:
Child_Voc_vs_NonVoc~age_rs+gender+ses_bin+ling_bin+gender*ses_bin
                          +gender*ling_bin+ling_bin*ses_bin+(1|child)}

Number of observations: 540  Groups: {'child': 20.0}

Random effects:
                 Name    Var    Std
child     (Intercept)  0.008  0.090
Residual               0.010  0.098

Fixed effects:
                   Estimate  2.5_ci  97.5_ci     SE       DF  T-stat  P-val  Sig
(Intercept)           0.443   0.356    0.530  0.044   17.986   9.964  0.000  ***
age_rs                0.219   0.175    0.262  0.022  532.437   9.910  0.000  ***
genderM               0.001  -0.114    0.117  0.059   13.872   0.021  0.984
ses_binL              0.028  -0.113    0.169  0.072   14.613   0.385  0.706
ling_binP            -0.111  -0.245    0.024  0.069   14.548  -1.611  0.129
genderM:ses_binL      0.104  -0.143    0.351  0.126   14.711   0.824  0.423
genderM:ling_binP     0.096  -0.109    0.301  0.104   14.535   0.917  0.374
```

C.2 Adult response to Child speech vs. non-speech productions

Figure 9 illustrates that children's speech-related productions tend to elicit more feedback from adults.



(a) Adult response to Child speech vs. non-speech productions (Mann-Whitney-Wilcoxon test two-sided, ****: p <= 1.00e-04)

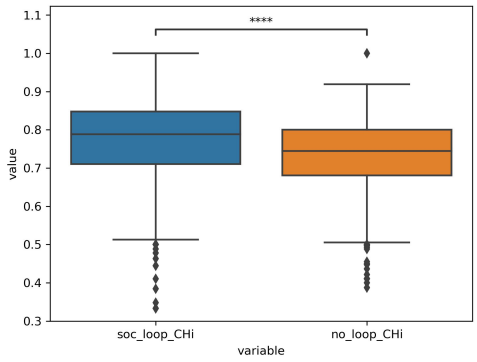


(b) Adult responses ratio difference between Child speech and non-speech productions (positive mean indicating a tendency for more responses to speech-related productions.

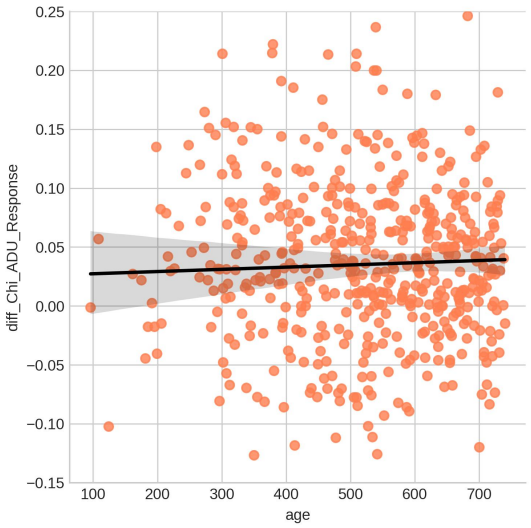
Figure 9: Adult response to Child speech vs. non-speech productions, replication of the second result of (Warlaumont et al., 2014)

C.3 Social Loop

The Figure 6 present the replication of the result on the social loop from (Warlaumont et al., 2014). More precisely it shows that given a children speech-related utterance, adult providing a response increase the proportion of speech-related (instead of non-speech related) follow-up utterance from the child.



(a) Child speech ratio (vs. non-speech) follow-up depending on whether or not an initial child speech-related utterance was responded by an adult or not (Mann-Whitney-Wilcoxon test two-sided, ****: p <= 1.00e-04)

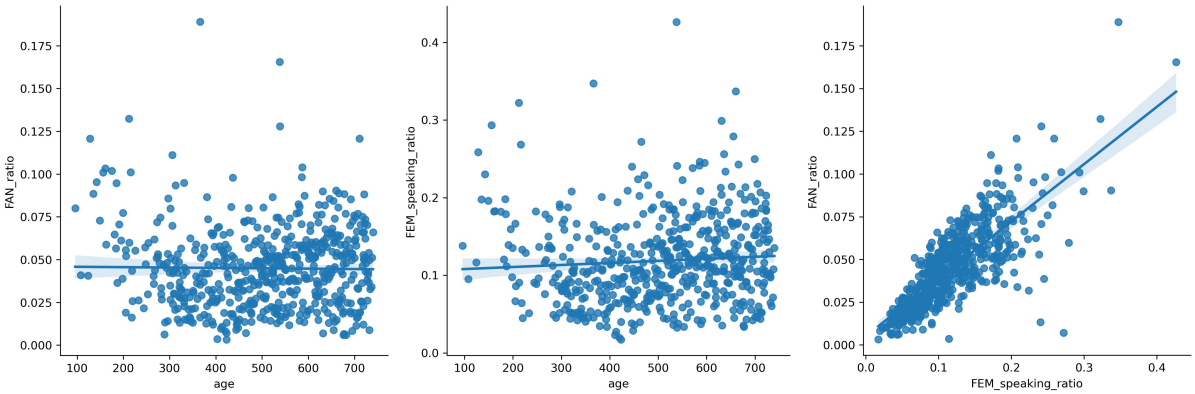


(b) Child ratio difference between Child speech and non-speech productions depending on whether or not an initial child speech-related utterance was responded by an adult or not (positive mean : indicating a tendency for more speech-related responses)

Figure 10: Adult response to Child speech vs. non-speech productions, replication of the second result of (Warlaumont et al., 2014) (left repeated from 6 in main text)

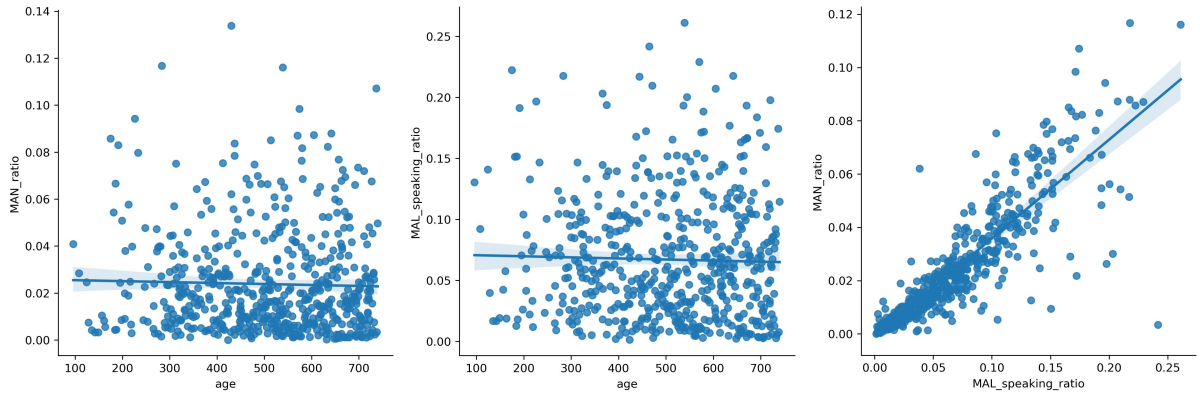
D Producing time ratios

The Figures 11-13 are showing the ration of the time occupied by a category when available for both LENA and VTC as well as the correlation metrics. The Figure 14 focuses on individual children recordings from the selected data set.



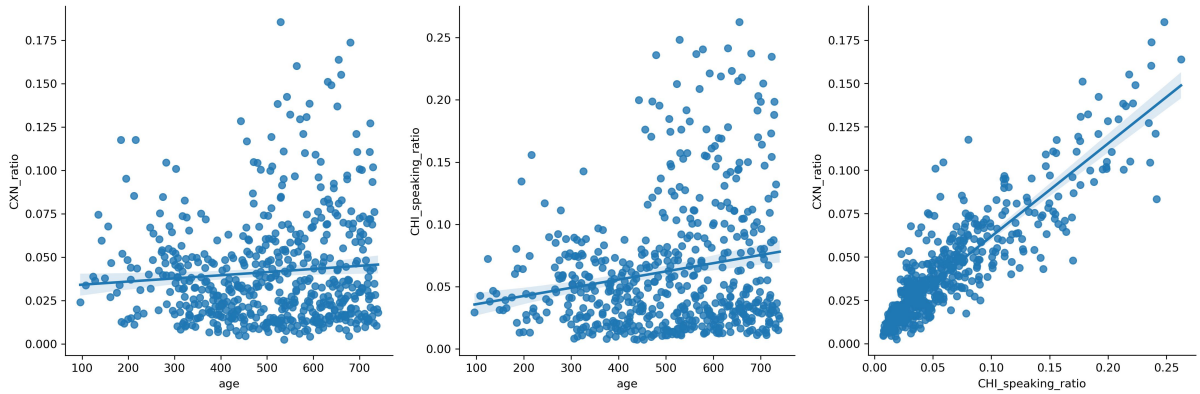
(a) LENA (b) VTC (c) Correlation ($r = 0.773$)

Figure 11: Female Speaking Time Ratio for LENA (a), VTC (b) and correlation plot between LENA and VTC (c). All children included (n = 20). Age in days.



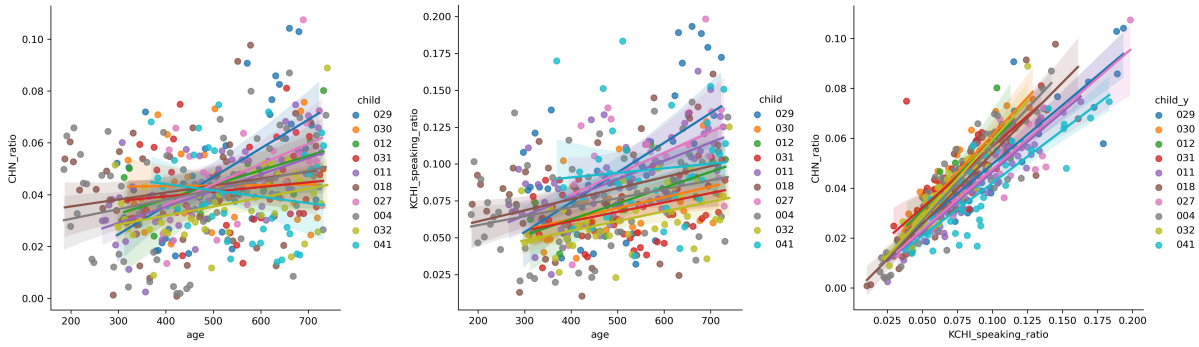
(a) LENA (b) VTC (c) Correlation ($r = 0.871$)

Figure 12: Male Speaking Time Ratio for LENA (a), VTC (b) and correlation plot between LENA and VTC (c). All children included (n = 20). Age in days.



(a) LENA (b) VTC (c) Correlation ($r = 0.917$)

Figure 13: Other child (than target child) Speaking Time Ratio for LENA (a), VTC (b) and correlation plot between LENA and VTC (c). All children included (n = 20). Age in days.



(a) LENA (b) VTC (c) Corelation

Figure 14: Target Child Speaking Time Ratio for LENA (a), VTC (b) and correlation plot (c). Selected children (n = 10, age span $\geq$ 9 months). Age in days.

```
Formula: chn_rs~age_rs+gender+ses_bin+ling_bin+gender*ses_bin
         +gender*ling_bin+ling_bin*ses_bin+(1|child)

Number of observations: 540  Groups: {'child': 20.0}
```

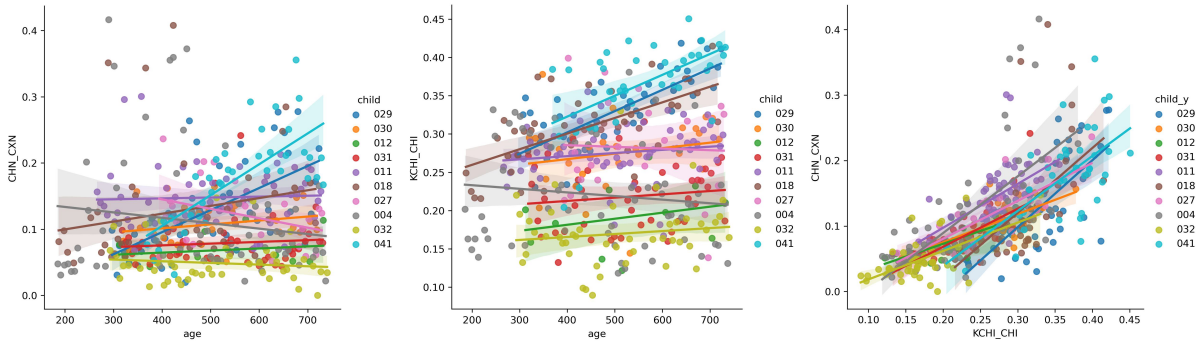
Random effects:

	Name	Var	Std
child	(Intercept)	0.012	0.109
Residual		0.014	0.119

Fixed effects:

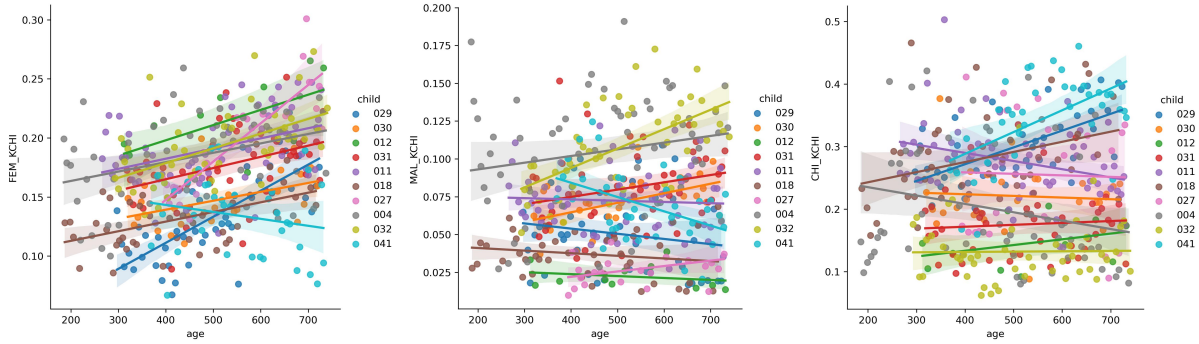
	Estimate	2.5_ci	97.5_ci	SE	DF	T-stat	P-val	Sig
(Intercept)	0.218	0.113	0.323	0.054	15.968	4.059	0.001	***
age_rs	0.163	0.110	0.215	0.027	532.415	6.063	0.000	***
genderM	-0.024	-0.163	0.115	0.071	12.270	-0.336	0.742	
ses_binL	-0.104	-0.275	0.066	0.087	12.936	-1.201	0.251	
ling_binP	0.046	-0.117	0.208	0.083	12.878	0.549	0.593	
genderM:ses_binL	0.104	-0.194	0.402	0.152	13.024	0.684	0.506	
genderM:ling_binP	-0.040	-0.288	0.207	0.126	12.866	-0.319	0.755	

E Temporal Contingencies Plots



(a) LENA (b) VTC (c) Correlation

Figure 15: CHILD>OTHER CHILD contingencies for LENA (a), VTC (b) and correlation plot (c). Selected children. Age in days.



(a) Female>Child (b) Male>Child (c) Other Child > Child

Figure 16: FEM>CHILD, MAL>CHILD, and OTHER CHILD > CHILD contingencies with VTC. Selected children. Age in days.

F Lena Annotation Process

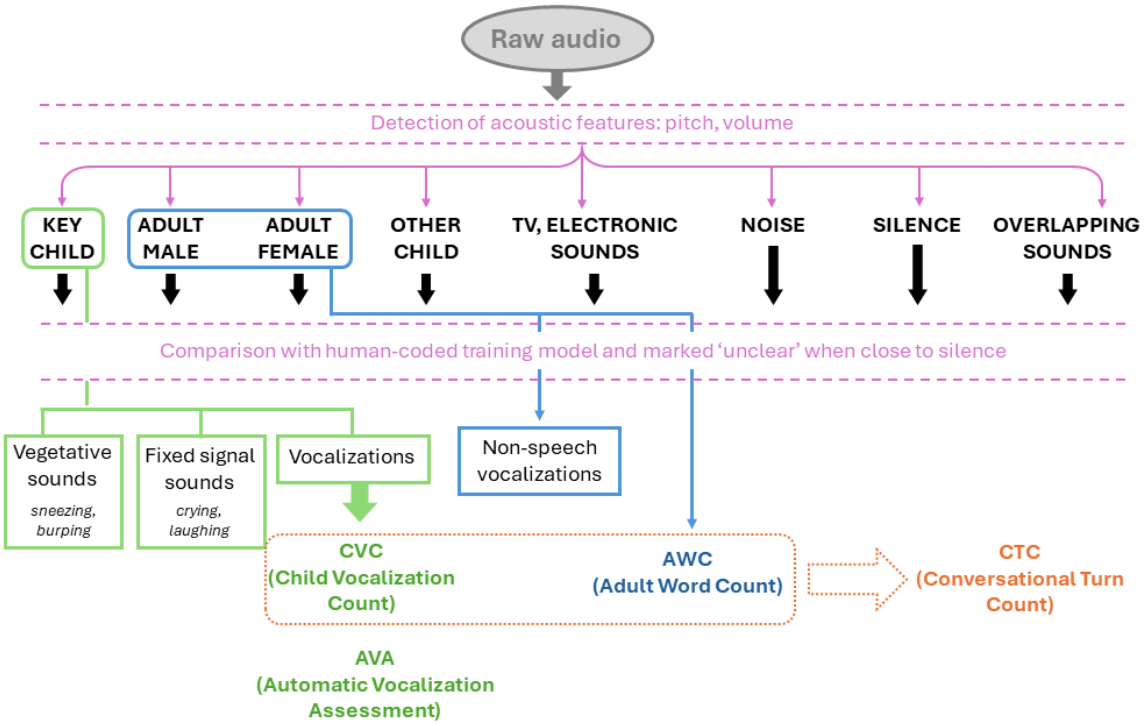


Figure 17: LENA annotation process.

G Instructions given to the families to use LENA

The instructions that were given to families were the following:

1. *When to record: once a week for a full day, until the child is 24 months old. Please prefer a day when you spend some time with your child (on the weekends for example). If you have no other choice, you can activate the device at daycare occasionally. We recommend that you record always on the same day of the week, to create a routine. Keep in mind that any day is a good day to record!*

2. *How to record: the instructions were kept identical to those provided by the LENA team. 1) Switch it on by pressing POWER; the screen should display "Paused". 2) Press RECORD for about 4 seconds; the screen should display "Recording". 3) Put the device in your child's shirt, the screen facing out, and close the pocket. 4) Leave it until the device turns off on its own at the end of the day.*

3. *Some various recommendations: never put the device out of the shirt; don't cover it with too many clothing layers; avoid noisy places as much as possible; remove the shirt (but leave the recorder inside) and keep it nearby during bath or nap times.*

4. *What to do after recording: bring the device back to the daycare center before a specific day of the week. The recorder will be ready for another week at the end of this day.*