



## From data to discovery: Technology propels speech-language research and theory-building in developmental science

Zuzanna Laudańska <sup>a,b,1,\*</sup>, Anna Caunt <sup>a,c,1</sup>, Alejandrina Cristia <sup>d</sup>, Anne Warlaumont <sup>e</sup>, Katerina Patsis <sup>a</sup>, Przemysław Tomalski <sup>b</sup>, Petra Warreyn <sup>f</sup>, Drew H. Abney <sup>g</sup>, Jeremy I. Borjon <sup>h,i,j</sup>, Manu Airaksinen <sup>k</sup>, Emily JH Jones <sup>l</sup>, Sven Bölte <sup>m,n,o</sup>, Magdalena Dall <sup>p</sup>, Daniel Holzinger <sup>p</sup>, Luise Poustka <sup>a</sup>, Herbert Roeyers <sup>f</sup>, Sam Wass <sup>q</sup>, Dajie Zhang <sup>a,r,2</sup>, Peter B. Marschik <sup>a,m,r,s,\*2</sup>

<sup>a</sup> Department of Child and Adolescent Psychiatry, Heidelberg University Hospital, Heidelberg University, and German Center for Mental Health (DZPG), Heidelberg, Germany

<sup>b</sup> Institute of Psychology, Polish Academy of Sciences, Warsaw, Poland

<sup>c</sup> School of Psychology, University of Plymouth, UK

<sup>d</sup> Laboratoire de Sciences Cognitives et de Psycholinguistique, Département d'études cognitives, ENS, EHESS, CNRS, PSL University, France

<sup>e</sup> Departments of Communication and Psychology, University of California, Los Angeles, USA

<sup>f</sup> Department of Experimental Clinical and Health Psychology, Ghent University, Belgium

<sup>g</sup> Department of Psychology, University of Georgia, Athens, USA

<sup>h</sup> Department of Psychology, University of Houston, Houston, USA

<sup>i</sup> Texas Institute for Measurement, Evaluation, and Statistics, University of Houston, Houston, USA

<sup>j</sup> Texas Center for Learning Disorders, University of Houston, Houston, USA

<sup>k</sup> BABA Center, Department of Physiology, University of Helsinki, Finland

<sup>l</sup> Department of Psychology, Birkbeck, University of London, and Centre for Developmental Neurobiology and Department of Child and Adolescent Psychiatry, IoPPN, King's College London, United Kingdom

<sup>m</sup> Center of Neurodevelopmental Disorders (KIND), Department of Women's and Children's Health, Centre for Psychiatry Research, Karolinska Institutet & Region Stockholm, Stockholm, Sweden

<sup>n</sup> Child and Adolescent Psychiatry, Stockholm Health Care Services, Region Stockholm, Stockholm, Sweden

<sup>o</sup> Curtin Autism Research Group, Curtin School of Allied Health, Curtin University, Perth, Australia

<sup>p</sup> Research Institute for Developmental Medicine, Johannes Kepler University Linz, Linz, Austria

<sup>q</sup> Institute for the Science of Early Years, University of East London, London, UK

<sup>r</sup> iDN – interdisciplinary Developmental Neuroscience, Division of Phoniatrics, Medical University of Graz, Austria

<sup>s</sup> Child and Adolescent Psychiatry and Psychotherapy, University Medical Center Göttingen, Leibniz ScienceCampus Primate Cognition and German Center for Child and Adolescent Health (DZKJ), Göttingen, Germany

### ARTICLE INFO

#### Keywords:

Infant  
Child  
Vocalizations  
Speech-Language Development  
LENA  
Audio Analysis  
Automatic Speech Recognition  
Acoustic Analysis  
Day-long Recordings  
Long-form Recordings

### ABSTRACT

Research on speech and language development has a long history, but in the past decade, it has been transformed by advances in recording technologies, analysis and classification tools, and AI-based language models. We conducted a systematic literature review to identify recently developed (semi-)automatic tools for studying speech-language development and learners' environments in infants and children under the age of 5 years. The Language ENvironment Analysis (LENA) system has been the most widely used tool, with more and more alternative free- and/or open-source tools emerging more recently. Most studies were conducted in naturalistic settings, mostly recording longer time periods (daylong recordings). In the context of vulnerable and clinical populations, most research so far has focused on children with hearing loss or autism. Our review revealed notable gaps in the literature regarding cultural, linguistic, geographic, clinical, and social diversity. Additionally, we identified limitations in current technology—particularly on the software side—that restrict researchers

\* Corresponding authors at: Department of Child and Adolescent Psychiatry, Heidelberg University Hospital, Heidelberg University, and German Center for Mental Health (DZPG), Heidelberg, Germany.

E-mail addresses: [Zuzanna.Laudanska@med.uni-heidelberg.de](mailto:Zuzanna.Laudanska@med.uni-heidelberg.de) (Z. Laudańska), [Peter.Marschik@med.uni-heidelberg.de](mailto:Peter.Marschik@med.uni-heidelberg.de) (P.B. Marschik).

<sup>1</sup> shared first authors

<sup>2</sup> senior authors

from fully leveraging real-world audio data. Achieving global applicability and accessibility in daylong recordings will require a comprehensive approach that combines technological innovation, methodological rigour, and ethical responsibility. Enhancing inclusivity in participant samples, simplifying tool access, addressing data privacy, and broadening clinical applications can pave the way for a more complete and equitable understanding of early speech and language development. Automatic tools that offer greater efficiency and lower cost have the potential to make science in this research area more geographically and culturally diverse, leading to more representative theories about language development.

## 1. Introduction

Speech and language development is fundamental to human interaction, serving as the foundation for social bonding, communication and cognitive growth (e.g., Bruner, 1985; Tomasello, 2005; Vygotsky, 1978). Research on speech and language development has a long history, with studies dating back over centuries. The earliest systematic observations were conducted on individual children, often the researchers' own offspring (e.g., Darwin, 1877; Stern and Stern, 1907; Piaget, 1952). Apart from unstructured parental reports, standardised vocabulary checklists have been widely used to measure language development since the mid-1990s (e.g., MacArthur-Bates Communicative Development Inventories/MCDI; Fenson et al., 1994; Fenson et al., 2006). Parental questionnaires, specifically MCDIs, have been adapted into many different languages (e.g., Fenson et al., 1994; Eriksson et al., 2012; <https://mb-cdi.stanford.edu/adaptations.html>) and have more recently been adapted for computerized-adaptive testing (e.g., Makranksy et al., 2016; Mayor and Mani, 2019; Kachergis et al., 2022) and mobile app data collection (e.g., Mieszkowska et al., 2022; Muszyńska et al., 2025). Frequently, parental reports are also complemented by systematic manual annotations and analyses of audio-video recordings of children's behaviours in experimental settings (e.g., preverbal data collection pathways as outlined in Pokorny et al., 2020) and naturalistic home recordings (e.g., for recordings, see HomeBank, VanDam et al., 2016 and Databrary, Dressler, 2015; for transcriptions of child productions see CHILDES database, MacWhinney, 2000). Manual annotation techniques, such as behavioural micro-coding and phonetic or phonological transcriptions, have been widely applied, enabling detailed analyses of expressive language functions. They require specialist training to achieve a high level of inter-rater reliability, are time- and labour-intensive, and thus, they are limited in scalability (e.g., Oller et al., 2010). When analysing speech-language functions on signal level, similar limitations apply with respect to specialised software (e.g., Computerized Speech Lab by KayPENTAX/Pentax Medical; Praat, Boersma and Weenink, 2025 or WaveSurfer, Liu, 2021). Most solutions allow for accurate visualisation of sound waveforms but do not provide a fully automated analysis process.

In the past decade, advances in recording technologies, analysis and classification tools, and AI-based language models have transformed research on speech-language development. These tools allowed researchers to advance their study of speech and language development as well as the acoustic and socio-communicative environment. To date, we have tools at hand that allow for short-term or continuous recordings across diverse contexts such as homes, nurseries, schools, and neonatal intensive care units (NICUs), offering more objective ways to record speech-language development than checklists and questionnaires that may be affected by certain bias. In recent years, we have witnessed an increase in the popularity of a variety of technical solutions to record, assess, and analyse infant speech-language development. Some of these tools are proprietary (e.g., LENA, Language ENvironment Analysis; Greenwood et al., 2011; see also Ganek and Eriks-Brophy, 2018a, 2018b for a review), others are open-source and/or free algorithms for speech diarisation (e.g., ALICE; Räsänen et al., 2021), vocal type classification (e.g., VTC; Lavechin et al., 2020) and fully automated transcription (e.g., Whisper by OpenAI; Radford et al., 2023). These tools enable scalable analyses of early vocal behaviours, expressive language development,

linguistic and acoustic environments. Moreover, automated (or semi-automated) diarisation and transcription tools open new possibilities for continuous measurement of real-world behaviours that could lead to new insights into how infants interact with caregivers and siblings and how these interactions shape speech-learning processes in daily life. Overall, applying automatic tools for recording and analysing the development of child vocalisations and language opens new avenues for researchers to uncover phenomena that were previously difficult to detect using traditional methods.

To date, these tools have been used in a variety of settings and have helped to shed light on different aspects of language and communicative development, such as how the timing and frequency of caregiver responses (Weisleder and Fernald, 2013) or interaction with siblings (Laing and Bergelson, 2024) is associated with language learning trajectories (Weisleder and Fernald, 2013), and how specific environmental factors, such as household noise levels (Simon et al., 2022), background sounds (Suarez-Rivera et al., 2024), music input (Hippe et al., 2024) or digital exposure (Ferjan Ramírez et al., 2021; Brushe et al., 2024) contribute to variability in speech and language development. These tools also enable the inclusion of larger datasets collected across countries (e.g., Bergelson et al., 2023), which is important for diversifying research and asking research questions about cross-cultural differences. Moreover, they can be used to analyse how atypical developmental patterns, such as those in children with hearing loss (e.g., Altman et al., 2020; Aragon and Yoshinaga-Itano, 2012; Josvassen et al., 2024; Lee and Ha, 2024), genetic conditions (e.g., Marschik et al., 2017; 2022; Pokorny et al., 2016; 2022), autism (e.g., Oller et al., 2010; Warlaumont et al., 2014, see review in Putnam et al., 2024), and developmental delays (e.g., Oller et al., 2010), impact early language learning, providing critical insights for intervention strategies. However, the introduction of these tools also brings with it new challenges. Many current tools or algorithms are insufficiently flexible to account for diverse situational, linguistic and cultural settings, particularly in low-resource environments (Cristia et al., 2024). Moreover, speech recognition tools still struggle with infant vocalisations and child speech, which include a mix of prelinguistic vocalisations, proto-words and target-language utterances. To enhance the utility of modern technology in the field, it is essential to examine the scope of its existing application—including the populations, cultural contexts, and environments studied, as well as its limitations. This will help identify opportunities for improvement.

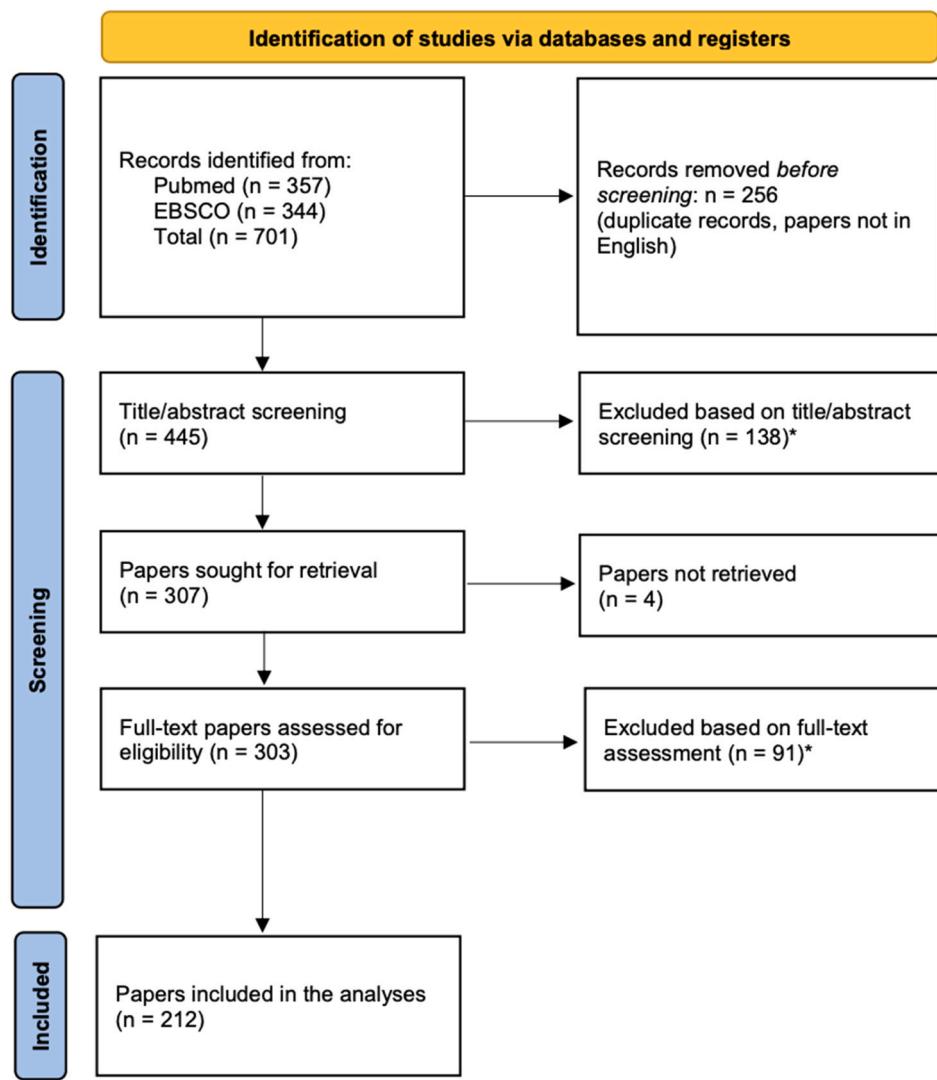
Here, we conduct a systematic literature review to identify recently developed (semi-)automatic tools for studying speech-language development and learners' environments in infants and children under the age of 5 years. We overview how these methods have been applied, including the country of publication, the spoken language studied, and the settings in which the methods were used. We review if participants were recruited from physiological/typical or clinical cohorts. Furthermore, we report what recording equipment and tools were used for data analysis.

In the narrative synthesis of results, we systematically evaluate the strengths, limitations, and potential of existing tools and discuss their impact on speech-language research in infancy and early childhood, while providing examples of studies. We begin with tools for acoustic analysis that commonly require labour-intensive manual pre-processing. We go on to examine the LENA system, the most used tool for both

recording and analysing speech-language development. This is followed by summarising alternative approaches for audio recording and analysis, considering their impact, potential and limitations for future applications. In the Discussion section, we embed findings from the systematic review in the broader context of theory and practice in developmental science, presenting newly emerging technological solutions and possible research directions that can advance our knowledge of speech-language development. Highlighting opportunities for innovation and expansion, we advocate for developing more inclusive, ecologically valid and reliable approaches for investigating language acquisition across diverse contexts, including various clinical conditions and cultural groups.

## 2. Methods

To systematically review the state-of-the-art methodologies for capturing and analysing speech and language, we conducted database searches following PRISMA guidelines (Page et al., 2021). The search strategy used the specific terms outlined below and was limited to peer-reviewed publications in English, excluding grey literature. The final search was completed on December 9, 2024, and included studies published from the year 2000 onward. The datasets generated by the survey research and analysed during the current study are available in OSF: [https://osf.io/xge3s/?view\\_only=e94b350bbe754a23841a10b1d7f8f887](https://osf.io/xge3s/?view_only=e94b350bbe754a23841a10b1d7f8f887). The code for generating figures is available on GitHub: [https://github.com/kpatsis97/Cost\\_review](https://github.com/kpatsis97/Cost_review).



### \*Reasons for exclusion:

- publication before the year 2000;
- systematic reviews, meta-analyses or dissertations;
- participants older than 5 years;
- no reference to the recording/analysis using semi-automatic or automatic methods

Fig. 1. PRISMA Flow diagram.

## 2.1. Search strategy

The search strategy was tailored to PubMed and EBSCOhost, using database-specific search terms:

- a) PubMed: (((infan\*[tiab]) OR (child\*[tiab])) AND ((vocal\*[tiab]) OR (speech[tiab]) OR (daylong audio recordings[tiab])) AND ("computational analysis"[tiab] OR "automatic classification"[tiab] OR "automatic measurement"[tiab] OR "acoustic analysis"[tiab] OR "audio analysis"[tiab] OR "automatic speech recognition"[tiab] OR "LENA"[tiab]));
- b) EBSCOhost: AB (((infan\*) OR (child\*)) AND ((vocal\*) OR (speech) OR (daylong audio recordings)) AND ("computational analysis" OR "automatic classification" OR "automatic measurement" OR "acoustic analysis" OR "audio analysis" OR "automatic speech recognition" OR "LENA")).

## 2.2. Selection criteria

### 2.2.1. Title-abstract review

The exclusion criteria for this first selection phase were: (a) publication before the year 2000; (b) systematic reviews, meta-analyses or dissertations; (c) majority of participants older than 5 years; (d) no reference to the recording and/or analysis using semi-automatic or automatic methods. The first 100 abstracts (22.5 %) were independently double-screened by two authors (ZL, MD). The initial kappa value for interrater reliability was  $\kappa = .69$ . All disagreements were resolved by consensus. In the second step, 50 additional abstracts were independently screened by the same authors, improving the kappa to  $\kappa = .74$ . Following this substantial agreement, the remaining articles were split for review between the two authors.

### 2.2.2. Full-text review

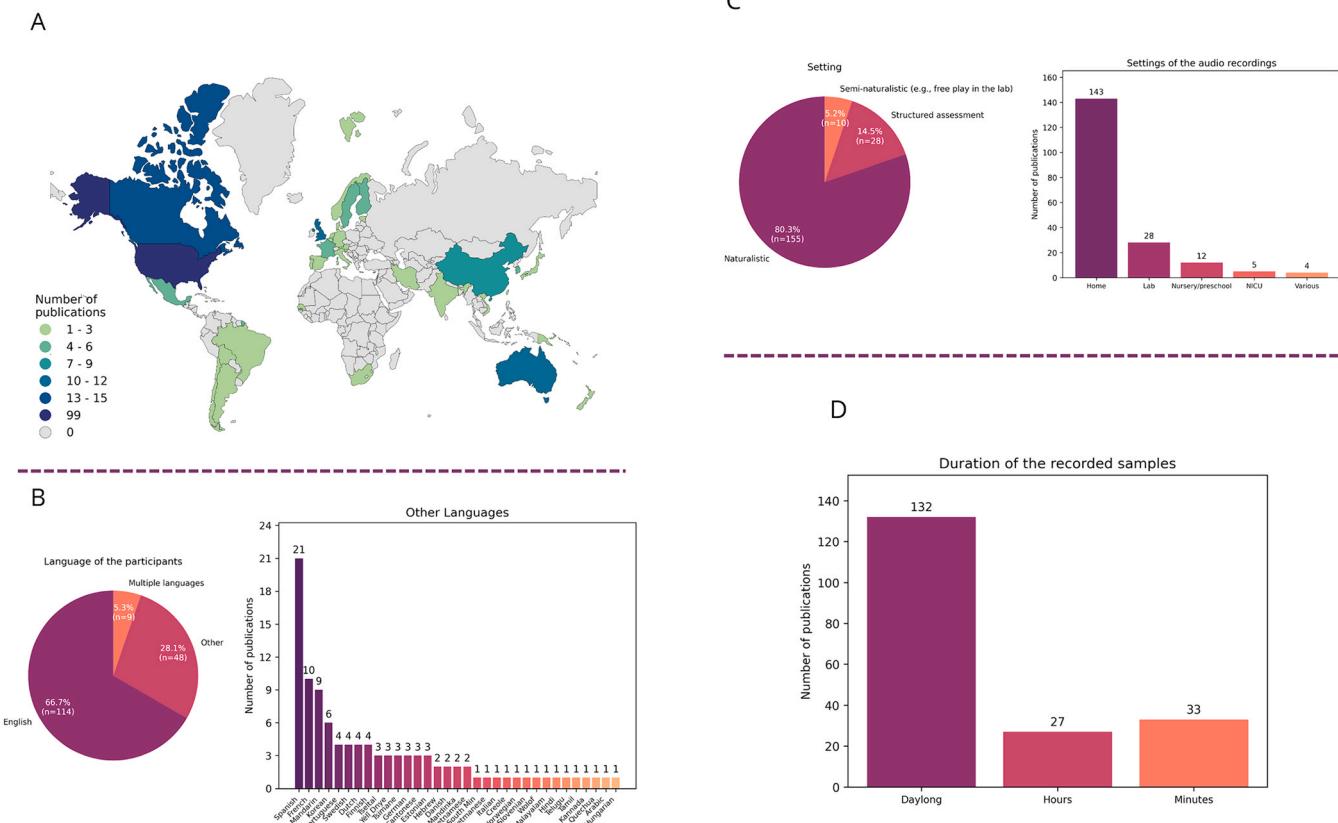
Full texts of all studies that passed the title-abstract review phase were retrieved and split between three authors (ZL, AC, MD), who applied the same exclusion criteria as in the previous article selection phase (A) (see Fig. 1).

### 2.2.3. Study selection

The systematic search strategy yielded a total of 701 publications. After removing duplicates and papers written in a language other than English, 445 publications proceeded to the first screening stage (A). The title-abstract review eliminated 138 publications, leaving 307 for the full-text review (B). Out of those 307, four papers could not be retrieved, leaving 303 publications retained for data extraction and in-depth review. During the full-text review, a further 91 publications were excluded, again applying the above-mentioned criteria, resulting in 212 publications included in the review (see Fig. 1).

## 3. Results

Our systematic literature review was focused on recently developed (semi-)automatic tools for studying speech-language development and learners' environments in infants and children under the age of 5 years. Regarding participants' geographic location and language, English-learning children in the U.S. were the most intensively studied (Fig. 2A, B). The majority of studies (80.3 %) were conducted in naturalistic settings such as participants' homes, nurseries and preschool classrooms (Fig. 2C), mostly recording longer time periods, even across the entire day (Fig. 2D). A fifth of the screened studies (19.7 %) were conducted in semi-naturalistic settings, recording free play in the lab or some form of structured assessment over shorter periods of time (hours or minutes in contrast to daylong recordings, Fig. 2C, D). In the context of more vulnerable and clinical populations, most research so far has



**Fig. 2.** Overview of the results based on the number of publications in a given category.

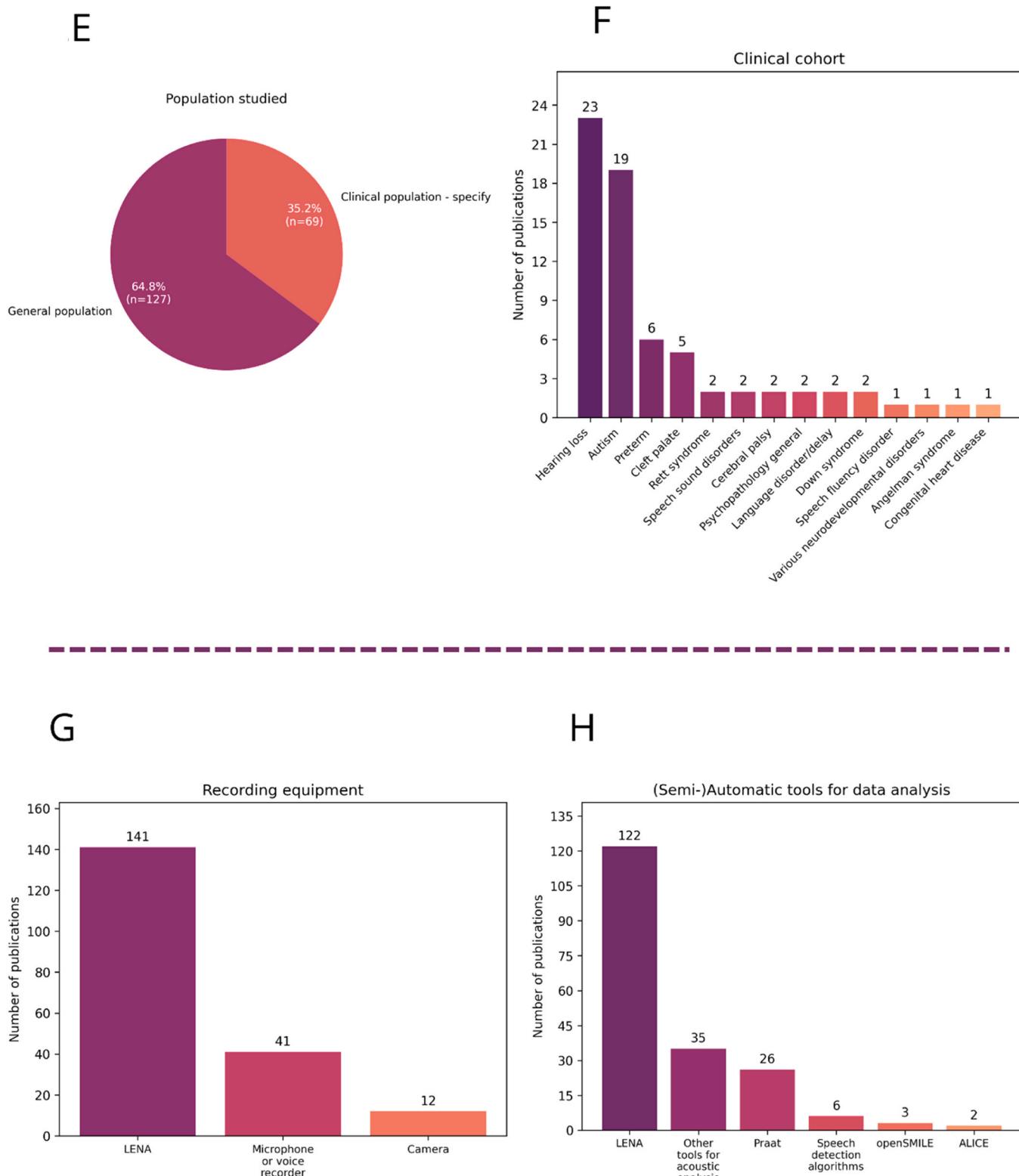


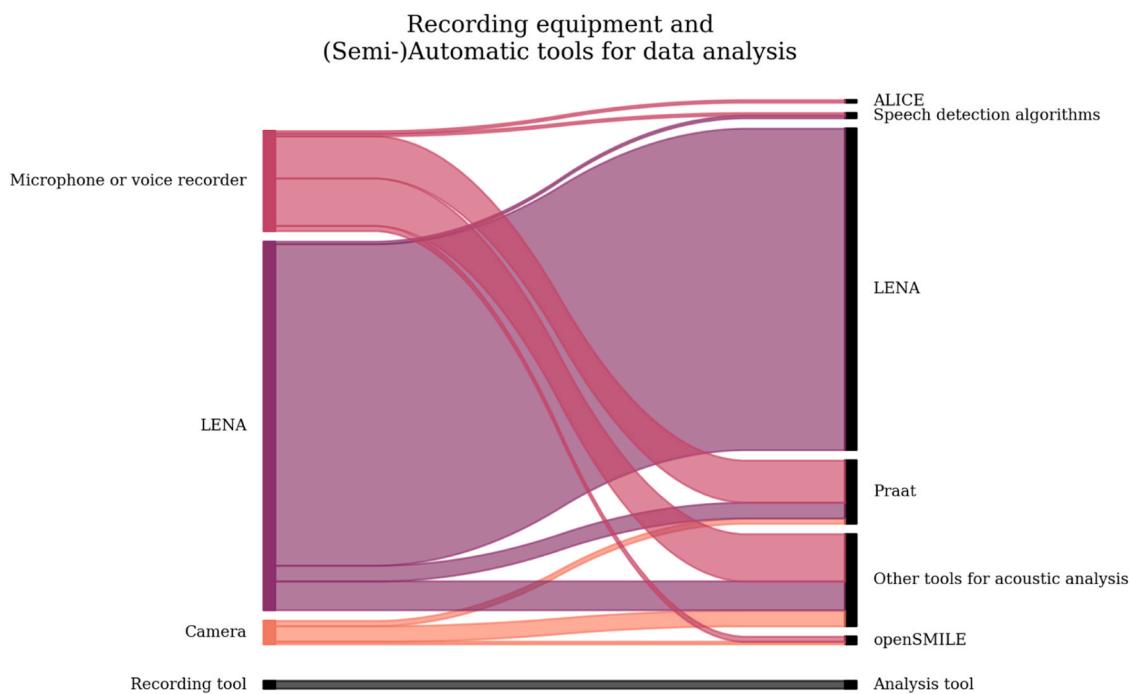
Fig. 2. (continued).

focused on children with hearing loss and children with (elevated likelihood of) autism (Fig. 2E, F). The search revealed a few major tools for studying early vocal/speech production and linguistic environments. LENA has been the most used solution for recording, with microphones/voice recorders and cameras being less frequent tools (Fig. 2G). LENA has also been the most widely used tool for analysing (pre)linguistic development (Fig. 2H). Other approaches for analysis were primarily

non-commercial and custom-made by research groups. Fig. 3 presents the combination of tools used for recording and analysing data and Fig. 4 shows an overview of the recording and analysis process and tools.

### 3.1. Narrative synthesis of results

Here, we summarise key findings and examples of studies that



**Fig. 3.** The Sankey diagram shows the tools used for recording (left) and the corresponding tools used for data analysis (right). The width of the band indicates the number of publications with each combination.

applied (semi-)automatic tools in studying speech-language development in infancy and early childhood, starting with tools for acoustic analysis that were some of the first semi-automatic approaches to analyse speech-language development, still requiring significant efforts related to manual preprocessing of recordings. Then, we describe findings related to LENA-based studies and the limitations of this approach. Finally, we present recording and analysis alternatives to LENA.

### 3.2. Acoustic analysis

Some of the first initiatives to automate the analysis of speech-language development were focused on developing specialised software for acoustic analysis. Acoustic analysis of audio signals and extraction of key features, such as speech sounds and words, is necessary to conduct any higher-level analyses of speech production. Such programs provide visual displays of audio signals such as waveforms, amplitude spectra and spectrograms. They also provide quantitative analysis derived from these acoustic waveforms and spectrograms, which can be valuable for assessing speech and language development and are widely used in research and clinical practice in the fields of speech pathology and audiology for diagnosis and monitoring treatment progress (Segura-Hernández et al., 2019; Coelho et al., 2015; 2016; Knight et al., 2016; Glaspey and Macleod, 2010; Chang et al., 2002). Acoustic analysis was used, for example, to investigate the impact of age at cochlear implantation on vocal development in children (e.g., Knight et al., 2016), the effects of vocal rehabilitation on voice acoustics in children with cleft lip and palate (Segura-Hernández et al., 2019) and the changes in voice physiology after surgical correction in infants with congenital heart disease (Joo et al., 2015). Some studies used proprietary software-hardware solutions such as the Computerized Speech Lab (KayPENTAX/Pentax Medical, e.g., Smith et al., 2008; Goldfield, 2000) that can include software modules such as Multidimensional Voice Program (MDVP, for detailed analysis of voice quality by measuring parameters like jitter, shimmer and harmonic-to-noise ratio, Coelho et al., 2016; Knight et al., 2016; Joo et al., 2015), Multi-Speech (for acoustic analysis and speech signal processing, Niwano and Sugai, 2002) and Real Time Pitch (focuses on pitch tracking and provides visual

feedback for pitch variation, Coelho et al., 2016).

Other studies relied on free and open-source software such as Praat (Boersma and Weenink, 2025), which supports pitch tracking, spectrogram analysis, and formant measurement. It can also be used for manual annotation of vocalisations (León et al., 2023; Mahmoudian et al., 2019; Mealigns and Demuth, 2014). Other examples include TF32 (Milenkovic, 2018) for time-frequency analysis, which displays the acoustic waveform along with pitch and sound spectrogram (e.g., Yoo et al., 2019) and WaveSurfer (Liu, 2021). This methodological approach allows for in-depth phenotyping of acoustic properties and voice quality in typically and atypically developing children. However, these tools have been minimally tested using large datasets or daylong recordings. While some studies have applied tools like PRAAT for large-scale acoustic analyses (Ritwika et al., 2020), systematic validation of pitch estimates in long-form recordings remains limited. Further, because many acoustic features, such as F0, are meaningful only within voiced speech, researchers must carefully consider the theoretical foundations of speech signals to avoid misinterpreting acoustic analyses, especially in long-form recordings where a diverse range of vocalisation contexts arise. Additionally, while automated tools could potentially facilitate acoustic analysis (see Pokorny et al., 2022, for an example of combining manual annotation of vocalisations with automatic classification based on acoustic features), they do not replace the need for human oversight in selecting and segmenting relevant portions of audio, a crucial step before meaningful interpretation can occur. This limitation constrains the potential for large-scale applications.

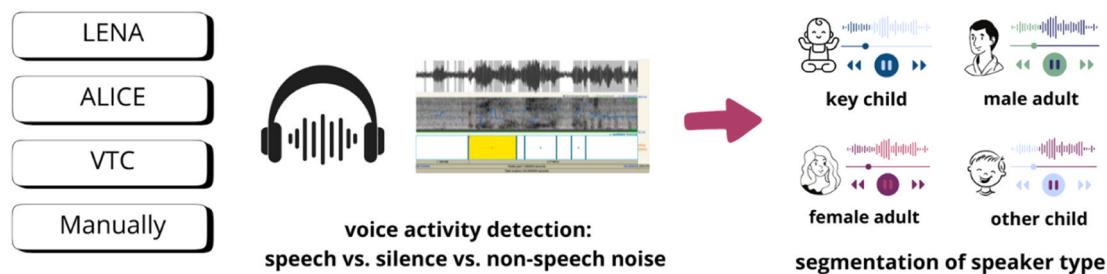
### 3.3. LENA – currently the most common tool for recording and analysing speech-language development

An important milestone in automating the measurement of child speech and language environment was the development of the proprietary system Language Environment Analysis (LENA) (Greenwood et al., 2011). To date, LENA is the most commonly used tool (Fig. 2G, H), so we begin with a more detailed overview of its features, which also illustrates the requirements of researchers in the field and the challenges faced by new tool developers. LENA is a commercial tool for audio

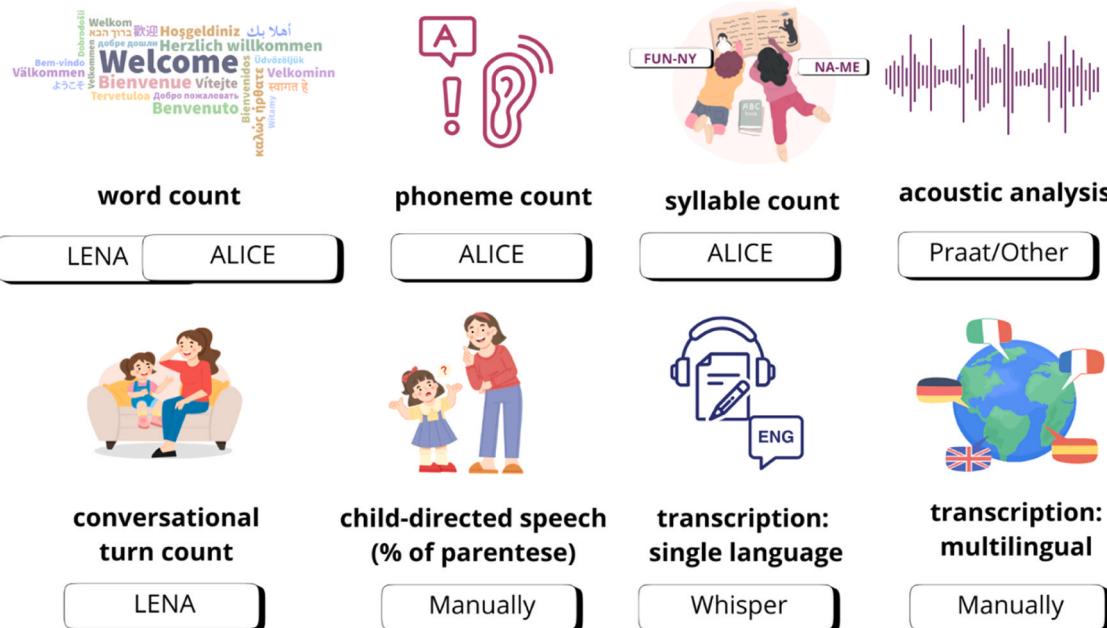
## Recording



## Voice activity detection & Segmentation of speaker type



## Feature extraction



**Fig. 4.** Overview of the recording and analysis process and tools. LENA: Language Environment Analysis; ALICE: Automatic Linguistic unit Count Estimator; VTC: Voice Type Classifier; Whisper: a neural net by OpenAI; Praat: open source software.

recording, language measurement, and basic analysis designed for long-form recordings. The system consists of a small audio recorder with a single microphone, which is made to be placed in a chest pocket sewn into children's clothing (vest, t-shirt, overalls) made by LENA for this

purpose. Once the audio has been collected, the software analyses the recordings and provides a numerical and graphical data output. First, the LENA algorithm classifies speech into speaker type (female adult, male adult, key child, and other child) and non-speaker classes (noise,

television, and silence). For speech recorded from the key child (the child wearing the audio recorder), LENA uses algorithms to identify and exclude crying or vegetative sounds. This ensures that only linguistically relevant vocalisations from the child are included in the output. The final output provides researchers with numerical data on various features captured from the linguistic input recorded by the LENA device, including adult word count, conversational turn count, and child vocalisation count. Conversational turns are defined as pairs of adult-child or child-adult vocalisations separated by no more than five seconds. Research using LENA has successfully contributed to a substantial body of work on infant and child language development, both for exploratory research purposes (e.g., Laing and Bergelson, 2024; Orena et al., 2020; Weisleder and Fernald, 2013) and intervention programs for children (e.g., Cunha et al., 2024; Joseph et al., 2022; Rowe et al., 2023; Suskind et al., 2016). Exploratory research has used LENA to investigate a variety of linguistic features of children's environments, for example, the language input to preterm infants (Caskey et al., 2011) and to children with diverse abilities and neurodivergent profiles (e.g., Aragon and Yoshinaga-Itano, 2012; Dykstra et al., 2013; Irvin et al., 2013; Oller et al., 2010; Thiemann-Bourque et al., 2014). Research with LENA has also been used to measure the amount of language exposure children receive in monolingual English-speaking homes (e.g., Bergelson et al., 2019) and bilingual homes (e.g., Marchman et al., 2017; Orena et al., 2020) in North America. Researchers have also implemented parent-focused interventions using LENA, such as parental interventions in low SES samples (e.g., Cunha et al., 2024; Joseph et al., 2022; Rowe et al., 2023; Suskind et al., 2013). In addition to studies examining the speech children hear from their environment, researchers have used LENA to explore the interaction between the language input children receive and the language output they produce. For example, studies have investigated the relationship between the number of conversational turns between adults and children and the child's overall word count (Weisleder and Fernald, 2013), as well as adult response rates to both speech-like and non-speech vocalisations (Warlaumont et al., 2014). Researchers have also used LENA to focus specifically on infants' vocalisations, such as comparing the amount of speech-like vocalisations to crying vocalisations (Oller et al., 2021).

#### 3.4. LENA research in languages other than English

LENA has been predominantly used in monolingual English-speaking communities in North America. The original dataset used to train LENA's algorithms included over 65,000 hours of recordings from over 300 monolingual English-speaking families raising infants between 1 and 42 months in North America (Gilkerson and Richards, 2008a). Since then, LENA has been used with a wide range of languages, and numerous papers report on the accuracy of their algorithms: Hebrew and Arabic (Levin-Asher et al., 2023), Mandarin Chinese (Zhang, X. et al., 2024), Danish (Jovsassen et al., 2024), Slovenian (Ferjan Ramírez et al., 2024), Shanghainese-Mandarin (Gilkerson et al., 2015), Vietnamese (Ganek and Eriks-Brophy, 2018a, 2018b), French (Canault et al., 2016), Dutch (Bruyneel et al., 2021), Swedish (Schwarz et al., 2017), Spanish (Weisleder and Fernald, 2013) and Italian (Bastianello et al., 2024). A recent systematic review of studies focusing on the validation of LENA across various languages found that, while some LENA outputs had moderate to high accuracy across languages (e.g. child vocalisation counts), others did not (conversational turn counts) (Cristia et al., 2020). Moreover, previous validation studies have typically focused on a specific corpus where participants share similar age ranges and languages. This makes it challenging to determine whether any differences in results are due to variations in how the corpus was annotated by researchers, or if LENA's accuracy is influenced by the specific characteristics of the population, such as age range and language. For instance, accuracy assessments of LENA vary widely across studies. Canault et al. (2016) considered LENA to be sufficiently reliable for adult speakers according to their reliability score ( $r = .64$ ), whereas

Meera et al., 2025 showed similar reliability for adult speakers ( $r = .62$ ) but concluded there was still scope for further improvement (Meera et al., 2025). These discrepancies highlight the subjectivity in determining whether a given level of accuracy is 'good enough'. To validate LENA reliably for use in multiple languages, it is essential for researchers to develop a standardised validation process applicable to diverse datasets. Beyond accuracy metrics, validation can also be approached at the level of findings rather than individual classifications. Some errors may not systematically bias results, meaning that even an imperfect classifier may still be useful for capturing broader patterns (e.g., Ritwika et al., 2020). A robust validation framework should, therefore, consider both quantitative accuracy and the broader implications of classification errors on research conclusions. Such an approach to validation could also be applicable to other tools designed to analyse daylong recordings.

#### 3.5. Limitations of LENA

While studies using LENA have contributed to a substantial body of research that has advanced early language development research, LENA can be costly for research labs. To start a project using two recorders over 6 months, the costs could be up to \$20,000, including equipment and subscription costs to LENA ((LENA Research Foundation, n.d) prices as of February 2025). Therefore, research institutions with fewer financial resources may lack the funds to conduct such research, limiting its applicability. Furthermore, new versions of the LENA analysis component require researchers to upload participants' recordings to servers based in the U.S. In some countries, particularly in the European Union, legal restrictions, such as those under GDPR, pose challenges for data transfers to non-European sites, sometimes leading to the rejection of research proposals based on legal rather than ethical considerations. Navigating these legal requirements involves substantial administrative effort, which can limit the feasibility of such research.

The level of performance for some of LENA's outputs relating to infant vocalisations has been critically discussed, e.g. highlighting poor performance for the key child's vocalisation recall, with one study finding LENA captured only 50 % of the key child's vocalisations identified by human annotators (Cristia et al., 2020). In addition, while LENA can distinguish between speech-like vocalisations and non-speech vocalisations (e.g., crying), its algorithms cannot provide more detailed classifications, such as distinguishing between canonical and non-canonical babbling or between different levels of crying (e.g., crying for different emotional needs). Thus, researchers using LENA to collect and identify instances of infant vocalisation often need to analyse those vocalisations in further detail. This can be achieved either manually (e.g., Soderstrom et al., 2021) or by extracting detailed acoustic features using OpenSMILE (Eyben et al., 2010). While these features provide valuable insight, they are typically only a starting point for a deeper analysis. Advanced algorithms, such as models designed to detect infants' emotional responses (e.g., ZhuParris et al., 2021), are often required to interpret the data for the studied research questions adequately. Between the limits to the accuracy and detail of the LENA system's automatically generated labels, the high cost of the system, and the proprietary algorithms, researchers are increasingly turning to alternative audio recording devices and analysis software.

#### 3.6. Recording and analysis alternatives to LENA

In recent years, alternative audio recording and analysis methods have been employed (e.g., Casillas et al., 2020; Cristia et al., 2023, see also Lavechin et al., 2025). Devices for recording purposes include USB "spy" recording devices (e.g., Cristia et al., 2023; Scaff et al., 2024; Caunt and Abu-Zhaya, 2024) and Olympus recorders (e.g., Casillas et al., 2020; Scaff et al., 2024) and the average prices are in the low-to-moderate range (starting from \$20–70 for a basic voice recorder). While most of these devices do not come with their own software for speech analysis, researchers engaged in this line of work have been investing in building

open-source alternatives to the closed-source LENA speech processing algorithms. Such open-source alternatives can provide researchers with similar outputs to those provided by LENA.

Firstly, the voice type classifier (VTC; [Lavechin et al., 2020](#)) was created with the purpose of classifying audio segments into speaker categories similar to LENA, for example, *female adult*, *male adult*, *key child* (the child wearing the recording device), and *other child*. Building upon this, the Automatic Linguistic unit Count Estimator (ALICE; [Räsänen, 2021](#)) was created using the VTC, allowing researchers to measure additional metrics in their data, such as the number of words, syllables and phonemes produced by adult speakers. These additional features and the ability to fine-tune them are essential for enhancing the algorithm's usability across different languages. For example, estimating word count can be challenging in languages where acoustic patterns and language-specific lexical entries do not follow the same sentence structure or acoustic patterns. By counting phonemes and syllables in parallel with words, ALICE allows researchers to use the algorithm across different languages. Both the VTC and ALICE have been shown to either outperform LENA or achieve similar performance, although the different systems make different trade-offs in terms of precision versus recall ([Lavechin et al., 2020](#); [Räsänen, 2021](#)). However, researchers need programming proficiency in Python to use ALICE and the VTC effectively. Without sufficient experience in Python or technical support, researchers may encounter challenges, which could create barriers and limit labs' ability to automate the processing of audio data.

Further challenges arise in analysing infant vocalisations. Currently, no widely available algorithms provide detailed automated speech analysis of day-long audio files (e.g., 16-hour recordings) beyond simple key-child segmentation of vocalisations. While algorithms like ALICE and VTC show promise for day-long recordings through speaker diarization, they were not trained to categorise features of infant vocalisations. Although researchers have employed feature extraction tools such as OpenSMILE ([Eyben et al., 2010](#)) on shorter audio segments (up to 10 seconds) to generate statistics on infants' emotional states, such as cry detection ([ZhuParris et al., 2021](#); [Micheletti et al., 2023](#)) and the classification of babbles (e.g., canonical vs non-canonical, [Fell et al., 2003](#); [Yeh et al., 2019](#); vocant vs squeal vs vowel, [Warlaumont et al., 2010](#)), there are no open-source tools that provide comprehensive labelling of infant vocalisations across different stages of speech-language development (cooing, babbling, proto-words etc.). Further, advancing current algorithms to classify infant vocalisations – e.g. distinguishing canonical from non-canonical babbles – within a single program rather than requiring multiple different tools (e.g., one software for audio preprocessing and another for classification) would be a significant improvement. These challenges highlight the need for continued development in algorithms for recognising infant vocalisations, driven by several factors, including the highly variable and context-dependent nature of infant vocalisations, as well as the need to create flexible tools that can analyse a broad range of vocalisations across age groups and developmental stages.

There is significant potential for advancements in this area, which could be achieved through increased collaborations between speech technologists and researchers collecting data with infants. This collaboration would be crucial in developing more effective algorithms and tools. A key factor in driving these advancements would be the establishment of a culture of data sharing, which will enable better training of classifiers and the creation of open-source tools that are tailored for the analysis of infant vocalisation ([VanDam et al., 2016](#)). To facilitate this, developing multi-site recording efforts and ensuring open data sharing across research groups will be crucial for building a robust foundation for the analysis of infant speech.

Other important limitations of the currently-available automatic tools that need to be addressed include the issues with the correct classification of various child speakers present in the recording. This is crucial for research conducted in more noisy environments with multiple speakers present at the same time, such as in a nursery or preschool.

Similar concerns are related to capturing the simultaneous speech of multiple speakers that consists of many overlapping vocalisations. Additionally, these tools must be adapted to ensure accessibility for low-income, linguistically and culturally diverse, and neurodiverse populations, making the methodology more inclusive and widely applicable.

#### 4. Discussion

In this review, we aimed to identify key tech-enhanced tools for studying speech-language development in infants and children under 5 years of age. We present an overview of methods that allow for investigating early speech-language skills as well as learners' environments. We review studies focusing on different language acquisitional aspects such as acoustic analysis of audio signals, frequency of vocalisations, or conversational turns. In our discussion, we highlight the existing sampling bias regarding geographical regions and linguistic groups. We point out the underrepresentation of neurodiverse and clinical groups in the research landscape and discuss how technological and utility limitations could hamper broader and more inclusive research practices.

##### 4.1. Sampling bias and consequences for understanding language acquisition

The majority of research on speech-language development has been conducted in the Global North, with participants from the U.S. and Western Europe. This geographical bias has systematically influenced theories on language development. Some theories have already been challenged by data from daylong recordings collected in non-WEIRD populations ([Henrich et al., 2010](#)). For example, research on children living in the U.S. has suggested that child-directed speech produced by adults is a key type of vocal input that significantly impacts children's language development and children who are exposed to more child-directed speech are developing faster-growing linguistic capacities (e.g., [Weisleder and Fernald, 2013](#)). However, recent findings point out that in some cultures, children are infrequently spoken to ([Casillas et al., 2020](#)), the quantity of child-directed speech is relatively low ([Bunce et al., 2024](#)), and the association between children's vocal output and input of vocalisations produced by other children is stronger than the association with vocalisations produced by adults ([Cristia et al., 2023](#)). Thus, the structure of home life in various cultural contexts relates to exposure to child- and adult-directed speech on a daily basis (e.g., [Bunce et al., 2024](#); [Casillas et al., 2020](#)). Another prominent theory based on data from children in the U.S. suggests that children of families from lower socioeconomic status (SES) backgrounds are exposed to less language input, which impacts their speech-language development (e.g., [Hart and Risley, 2003](#)). Yet, a recent study conducted across six continents, which also used daylong recordings to examine language input and children's vocal output, found no significant association between SES and children's input quantity or vocal production ([Bergelson et al., 2023](#)), showcasing how the use of novel tools can challenge monocultural theories on speech and language learning.

Automatic tools for processing (daylong) recordings and speech recognition in real-world settings can also help broaden existing theories to better account for factors that contribute to speech acquisition, such as, for instance, capturing the influence of a wider diversity of caregivers ([Katus et al., 2024](#)), siblings (e.g., [Laing and Bergelson, 2024](#)) and other peers (e.g., [Perry et al., 2018](#)) on the development of early conversational turns. Automatic tools that offer greater efficiency and decreased cost compared to manual annotation and transcription of recordings have the potential to make science in this research area more diverse. Moreover, they facilitate tracking infants' daily electronic media exposure as a component of their language environment and its impact on their vocalisations (e.g., [Ferjan Ramírez et al., 2022](#); [Brushe et al., 2024](#)). Taken together, these recent findings on language acquisition highlight the value of research that spans not only different cultures, but

also different contexts and household structures.

#### 4.2. Future challenges and opportunities

Making research on speech-language development more accessible for global applications and clinical populations is a multifaceted challenge requiring technological innovation, methodological adaptation, and ethical vigilance. Below, we discuss potential pathways and considerations organised by key challenges and opportunities.

A significant limitation in current research is the overrepresentation of studies focused on typically developing English-learning children in the U.S. Findings from these studies, while informative, do not capture the diversity of early language development on a global scale (see [Singh et al., 2025](#) for a broader discussion). Expanding research to include populations from diverse linguistic, cultural, geographical and socio-economic contexts is critical. However, practical barriers such as funding, training, and infrastructure often impede these efforts, particularly in low- and middle-income countries (LMICs). Establishing respectful and collaborative partnerships with local institutions and investing in capacity-building initiatives can help overcome these challenges (e.g., [Léon et al., 2024](#)). Similarly, the efforts to develop free and open-source tools can also result in broader use across different research sites. Furthermore, the dominance of English-speaking participants in language development research ([Fig. 2A](#)) creates a feedback loop that perpetuates the scarcity of data from other languages. Insufficient data hinders the development of tools for under-represented languages. Further, the limited applicability of these tools in non-English languages can discourage researchers from collecting new data. To break this cycle, strategic efforts to collect and share recordings in diverse languages are essential (see the excellent example in [Bergelson et al., 2023](#)). Collaborative cross-linguistic research efforts and shared repositories of annotated data are essential steps towards building a more comprehensive dataset that reflects the diversity of global languages. Another step is training initiatives accessible to diverse groups of researchers – with a particular emphasis on involving research groups from LMICs – focused on cross-linguistic validation of various algorithms dedicated to infant/toddler data and specialised programming workshops. Finally, new funding opportunities supporting multi-site collaborations would allow for minimising the practical barriers.

In clinical contexts, most research has focused on populations such as children with hearing loss and those with (elevated likelihood of) autism ([Fig. 2E, F](#), see also [Putnam et al., 2024](#) for a review of LENA-based studies focused on autism). Broadening the scope to include other clinical and/or vulnerable populations, such as children with developmental language disorder and Down syndrome, is necessary to understand the full spectrum of challenges related to early language development. Automatised AI-supported speech analysis might augment the clinical diagnosis and thus support earlier and more reliable identification and differentiation of different types of speech disorders (e.g., phonological, dysarthria, childhood apraxia of speech). However, obtaining patient data for algorithm-training purposes poses unique challenges due to the sensitive nature of medical data. For this reason, more consortia-based collaborations with predefined rules of data sharing across sites are necessary to develop automatic tools to accommodate the needs of specific clinical populations.

Technological limitations also restrict the accessibility of daylong recording studies. For example, the Language Environment Analysis (LENA) system, widely used in the field, faces several challenges in adapting to new languages and contexts. LENA's algorithms were trained on English recordings, and validation efforts in other languages have shown mixed results, underscoring its limited generalisability. Open-source alternatives offer a promising avenue but often require programming expertise, creating barriers for researchers and clinicians without technical training. Additionally, cloud-based systems cannot be easily used in many European countries due to data privacy regulations – this issue is especially pressing for clinical research, where

safeguarding sensitive data is paramount. Decentralised or localised data-processing solutions could mitigate these challenges and ensure compliance with regional regulations. Simplifying the setup and usability of open-source tools through community-driven platforms and shared resources could further enhance their adoption.

While partial solutions exist for sharing long-form audio data as well as annotations and tools (e.g., HomeBank, [VanDam et al., 2016](#)), data sharing remains a major challenge in the field, particularly when balancing open science principles with ethical and legal obligations to protect participants' privacy. Sharing annotated recordings on such platforms, along with organising thematic engineering challenges such as ComParE (Computational Paralinguistic Challenge; [Schuller et al., 2020](#)), can accelerate algorithm development, but this must be done in a way that respects participants' rights and adheres to ethical guidelines, as emphasised by [Cychosz et al. \(2020\)](#). Innovative approaches such as federated learning – when the algorithm travels, not the data (see example for pediatric care in [Rb-Silva et al., 2023](#)), allow data analysis without the transfer of raw data and could be a promising direction for certain research scenarios. In addition, anonymisation techniques (such as sharing annotations, transcripts and output files rather than raw audio data) and secure data-sharing protocols are essential to ensure ethical and responsible research practices.

Data coming from daylong recordings can facilitate the development of computational models of early language acquisition ([Dupoux, 2018](#); [Lavechin et al., 2022](#)). Daylong recordings provide realistic input data in ecologically valid conditions that can be used to train AI algorithms. The approach of reverse engineering language development, i.e. building computational systems that can be trained on realistic input data to mimic the process of infant language acquisition, can provide us with scientific insights into human language learning ([Dale et al., 2022](#)). In addition, it allows for the creation of better artificial language learners ([Dupoux, 2018](#); [Lavechin et al., 2022](#)).

Whilst naturalistic recordings offer valuable insights into real-world language exposure, it is important to consider whether they are sufficient on their own or if experimental approaches remain necessary. Naturalistic recordings can capture authentic interactions between children and others and provide data to input into computational models, yet the data can be limited in recognising causal relationships or systematically manipulating variables. Thus, researchers could integrate experimental methods to test specific hypotheses, using a mixed-methods approach where naturalistic data informs experimental design and vice versa. This combination could lead to a more comprehensive understanding of early language acquisition.

A recent breakthrough in minimising time- and labour-intensive tasks in speech development research comes from developments in open-source automatic speech recognition (ASR) models, such as the Whisper algorithm developed by OpenAI ([Radford et al., 2023](#)). It is designed to effectively handle transcription in many languages (although not when multiple languages are spoken in a single recording) and translation tasks, using a transformer-based architecture trained on diverse datasets of speech recordings and their transcriptions. Whisper emphasises robustness over a wide range of accents, background noise and languages, making it suitable for real-world applications. However, children's speech is characterised by high variability in pitch, articulation, and speech patterns due to developmental differences (e.g., [Li, 2012](#); [Redford, 2014](#)), so it can present unique challenges to automatic speech recognition algorithms, which may be built with limited infant/child-centric training data. Nevertheless, preliminary reports suggest that Whisper is a promising tool for recognising children's speech recorded in challenging environments – for example, in noisy school environments ([Southwell et al., 2024](#)) or spoken by children who are not native English speakers ([Jain et al., 2024](#)).

Finally, improvements in data collection and analysis techniques are critical to capturing the complexity of language environments. Emerging research highlights the importance of contextual factors, such as exposure to electronic media (Ferjan Ramírez et al., 2022; [Brushe et al.,](#)

2024), household noise levels (Simon et al., 2022), background sounds (Suarez-Rivera et al., 2024), music input and episodes (Hippe et al., 2024; Lerma-Arregocés and Pérez-Moreno, 2024; Mendoza and Fausey, 2022) or sibling interactions (Laing and Bergelson, 2024). Recording tools must be capable of encoding these influences while addressing technical challenges such as speaker differentiation and detection of overlapping speech. Beyond the complexity of the environment around the infant, producing speech requires the coordination and activity of multiple biological processes, from the brain to the orofacial articulators, the body, and the autonomic nervous system. Measuring these indices presents unique challenges which multiple open source or non-commercial devices (Maitha et al., 2020; Geangu et al., 2023; Islam et al., 2024), software (Borjon et al., 2024), and algorithms (Xu et al., 2020; Zhang Y. et al., 2024; Mason et al., 2024) have been developed to address. Continued efforts in linking multiple modalities of infant behaviour during language production and perception, particularly in diverse naturalistic settings, will benefit from consortium efforts in agreed-upon construct definitions, synchronisation protocols, and analysis pipelines. Furthermore, methodological advances allow for including multimodal data, such as video, accelerometry, or physiological measures (e.g., Abney et al., 2021; Madden-Rusnak et al., 2024; Wass et al., 2022; Smith et al., 2023; Borjon et al., 2024; Sullivan et al., 2021), alongside audio recordings, could provide rich insights into language environments and developmental trajectories.

## 5. Conclusions

We performed a systematic review of the literature on technology for automated analysis of audio data for studying children's speech-language experiences. The review identified a large body of studies utilising the LENA system to study language development in a variety of contexts, with alternative free- and/or open-source tools emerging more recently and offering new possibilities for multi-site collaborations. Our review identified gaps in the diversity of cultural, linguistic, geographic, clinical, and social contexts represented in the literature. We also observed limits in the currently available technology, especially on the software end, that, in turn, limit how much researchers can capitalise on their ability to capture real-world audio from children. Achieving global applicability and accessibility requires a holistic approach that integrates technological innovation, methodological rigour, and ethical responsibility. By fostering inclusivity in participant samples, simplifying access to tools, addressing data privacy concerns, and expanding clinical applications, we will move toward a more comprehensive, ecologically valid and equitable understanding of early speech-language development. Capturing the full range of its variability across populations and contexts is key to informing theory-building for typical and atypical language acquisition.

## Funding

This work was supported by the Cost Action CA22111 (SW), DFG grant 454648639, SFB 1528, Cognition of Interaction C03 to PM with DZ, ZL; Volkswagen Foundation Project IDENTIFIED to PM; National Science Center of Poland grant 2022/45/N/HS6/02085 to ZL and 2018/30/E/HS6/00214 to PT; Rett Deutschland e.V. to PM and DZ; National Institutes of Health grant R00HD105920 to JIB. A.Cristia acknowledges the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ExELang, Grant agreement No. 101001095). ASW received funding from the National Science Foundation (grants 1529127 and 1539129/1827744) and the James S. McDonnell Foundation for related work. This study bridges the German Center for Mental Health (DZPG VISIONS25-MA2 FIND & ACT grant to ZL) with the German Center for Child and Adolescent Health (DZKJ).

## Acknowledgements

We would like to thank Prof. Jukka Leppänen for his role as a critical friend in the study conceptualisation and writing process. Your input is truly appreciated, thank you very much indeed.

## References

Abney, D.H., daSilva, E.B., Lewis, G.F., Bertenthal, B.I., 2021. A method for measuring dynamic respiratory sinus arrhythmia (RSA) in infants and mothers. *Infant Behav. Dev.* 63, 101569. <https://doi.org/10.1016/j.infbeh.2021.101569>.

Altman, R.L., Laursen, B., Perry, L.K., Messinger, D.S., 2020. Validation of continuous measures of peer social interaction with self- and teacher-reports of friendship and social engagement. *Eur. J. Dev. Psychol.* 17 (5), 773–785. <https://doi.org/10.1080/17405629.2020.1716724>.

Aragon, M., Yoshinaga-Itano, C., 2012. Using Language ENvironment Analysis to improve outcomes for children who are deaf or hard of hearing (Thieme Medical Publishers). *Semin. Speech Lang.* 33 (4), 340–353. <https://doi.org/10.1055/s-0032-1326912>.

Bastianello, T., Lorenzini, I., Nazzi, T., Majorano, M., 2024. The Language ENvironment Analysis system (LENA): A validation study with Italian-learning children. *J. Child Lang.* 51 (5), 1172–1192. <https://doi.org/10.1017/S0305000923000326>.

Bergelson, E., Casillas, M., Soderstrom, M., Seidl, A., Warlaumont, A.S., Amatuni, A., 2019. What do North American babies hear? A large-scale cross-corpus analysis. *Dev. Sci.* 22 (1), e12724.

Bergelson, E., Soderstrom, M., Schwarz, I.C., Rowland, C.F., Ramírez-Esparza, N., R. Hamrick, L., Cristia, A., 2023. Everyday language input and production in 1,001 children from six continents. *Proc. Natl. Acad. Sci.* 120 (52), e2300671120.

Boersma, P., Weenink, D. (2025). Praat: doing phonetics by computer [Computer program]. Version 6.4.26, retrieved 8 January 2025 from (<http://www.praat.org/>).

Borjon, J.I., Sahoo, M.K., Rhodes, K.D., Lipschutz, R., Bick, J.R., 2024. Recognizability and timing of infant vocalizations relate to fluctuations in heart rate. *Proc. Natl. Acad. Sci. USA* 121 (52), e2419650121. <https://doi.org/10.1073/pnas.2419650121>.

Bruner, J., 1985. Child's talk: Learning to use language. *Child Lang. Teach. Ther.* 1 (1), 111–114.

Brushe, M.E., Haag, D.G., Melhuish, E.C., Reilly, S., Gregory, T., 2024. Screen Time and Parent-Child Talk When Children Are Aged 12 to 36 Months. *JAMA Pediatr.* 178 (4), 369–375. <https://doi.org/10.1001/jamapediatrics.2023.6790>.

Bruyneel, E., Demurie, E., Boterberg, S., Warreyn, P., Roeyers, H., 2021. Validation of the Language ENvironment Analysis (LENA) system for Dutch. *J. Child Lang.* 48 (4), 765–791.

Bunce, J., Soderstrom, M., Bergelson, E., Rosemberg, C., Stein, A., Alam, F., Casillas, M., 2024. A cross-linguistic examination of young children's everyday language experiences. *J. Child Lang.* 1–29. <https://doi.org/10.1017/S030500092400028X>.

Canault, M., Le Normand, M.T., Foudil, S., Loudon, N., Thai-Van, H., 2016. Reliability of the language environment analysis system (LENA™) in European French. *Behav. Res. Methods* 48, 1109–1124.

Casillas, M., Brown, P., Levinson, S.C., 2020. Early language experience in a Tzeltal Mayan village. *Child Dev.* 91 (5), 1819–1835.

Caskey, M., Stephens, B., Tucker, R., Vohr, B., 2011. Importance of parent talk on the development of preterm infant vocalizations. *Pediatrics* 128 (5), 910–916.

Caunt, A., & Abu-Zhaya, R. (2024). Language Mixing Patterns in Multilingual Homes: Evidence from Daylong Recordings. [Conference Presentation] *Workshop for Infant Language Development 2024, Lisbon, Portugal*.

Chang, S.E., Ohde, R.N., Conture, E.G., 2002. Coarticulation and formant transition rate in young children who stutter. *J. Speech, Lang., Hear. Res.* 45 (4), 676–688. [https://doi.org/10.1044/1092-4388\(2002/054\)](https://doi.org/10.1044/1092-4388(2002/054)).

Coelho, A.C., Brasolotto, A.G., Bevilacqua, M.C., 2015. An initial study of voice characteristics of children using two different sound coding strategies in comparison to normal hearing children. *Int. J. Audiol.* 54 (6), 417–423. <https://doi.org/10.3109/14992027.2014.998784>.

Coelho, A.C., Brasolotto, A.G., Bevilacqua, M.C., Moret, A.L., Bahmad Júnior, F., 2016. Hearing performance and voice acoustics of cochlear implanted children. *Braz. J. Otorhinolaryngol.* 82 (1), 70–75. <https://doi.org/10.1016/j.bjorl.2015.11.002>.

Cristia, A., Bulgarelli, F., Bergelson, E., 2020. Accuracy of the language environment analysis system segmentation and metrics: A systematic review. *J. Speech, Lang., Hear. Res.* 63 (4), 1093–1105.

Cristia, A., Gautheron, L., Colleran, H., 2023. Vocal input and output among infants in a multilingual context: Evidence from long-form recordings in Vanuatu. *Dev. Sci.* 26 (4), e13375.

Cristia, A., Gautheron, L., Zhang, Z., Schuller, B., Soderstrom, M., 2024. Establishing the reliability of metrics extracted from long-form recordings using LENA and the ACLEW pipeline. *Behav. Res. Methods* 1–20.

Cunha, F., Gerdes, M., Hu, Q., Nihitianova, S., 2024. Language environment and maternal expectations: An evaluation of the LENA start program. *J. Hum. Cap.* 18 (1), 105–139.

Cychosz, M., Romeo, R., Soderstrom, M., et al., 2020. Longform recordings of everyday life: Ethics for best practices. *Behav. Res. Methods* 52 (5), 1951–1969.

Dale, R., Warlaumont, A.S., Johnson, K.L., 2022. The fundamental importance of method to theory. *Nat. Rev. Psychol.* 2, 55–66.

Darwin, C., 1877. A biographical sketch of an infant. *Mind* 2 (7), 285–294. <https://doi.org/10.1093/mind/os-2.7.285>.

Dressler, V., 2015. Databrary. *Ref. Rev.* 29 (4), 32–39.

Dupoux, E., 2018. Cognitive science in the era of artificial intelligence: A roadmap for reverse-engineering the infant language-learner. *Cognition* 173, 43–59. <https://doi.org/10.1016/j.cognition.2017.11.008>.

Dykstra, J.R., Sabatos-DeVito, M.G., Irvin, D.W., Boyd, B.A., Hume, K.A., Odom, S.L., 2013. Using the Language Environment Analysis (LENA) system in preschool classrooms with children with autism spectrum disorders. *Autism* 17 (5), 582–594.

Eriksson, M., Westerlund, M., Berglund, E., 2012. A cross-linguistic perspective on the development of communicative skills in toddlers: The Swedish Early Communicative Development Inventory. In: Eriksson, M. (Ed.), *Cross-linguistic adaptations of the MacArthur-Bates Communicative Development Inventories*. Singular Publishing Group, San Diego, CA, pp. 7–23.

Eyben, F., Wöllmer, M., & Schuller, B. (2010, October). Opensmile: the munich versatile and fast open-source audio feature extractor. In *Proceedings of the 18th ACM international conference on Multimedia* (pp. 1459–1462).

Fell, H., Cress, C., MacAuslan, J., & Ferrier, L. (2003). VisiBabble for reinforcement of early vocalization. *Assets '04: Proceedings of the 6th International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 161–168). <https://doi.org/10.1145/1028630.1028659>.

Fenson, L., Dale, P.S., Reznick, J.S., Thal, D., Bates, E., Hartung, J.P., Pethick, S., Reilly, J.S., 1994. Variability in early communicative development. *Monogr. Soc. Res. Child Dev.* 59 (5), i–185. <https://doi.org/10.2307/1166093>.

Fenson, L., Marchman, V.A., Thal, D.J., Dale, P.S., Reznick, J.S., & Bates, E. (2006). *MacArthur-Bates Communicative Development Inventories, Second Edition (CDIs)* [Database record]. APA PsycTests. <https://doi.org/10.1037/t11538-000>.

Ferjan Ramírez, N., Hippé, D.S., Shapiro, N.T., 2021. Exposure to electronic media between 6 and 24 months of age: An exploratory study. *Infant Behav. Dev.* 63, 101549. <https://doi.org/10.1016/j.infbeh.2021.101549>.

Ferjan Ramírez, N., Marjanović Umek, L., Fekonja, U., 2024. Language environment and early language production in Slovenian infants: An exploratory study using daylong recordings. *Infancy* 29 (5), 811–837.

Ganek, H., Eriks-Brophy, A., 2018b. Language ENvironment analysis (LENA) system investigation of day long recordings in children: A literature review. *J. Commun. Disord.* 72, 77–85. <https://doi.org/10.1016/j.jcomdis.2017.12.005>.

Ganek, H.V., Eriks-Brophy, A., 2018a. A concise protocol for the validation of Language ENvironment Analysis (LENA) conversational turn counts in Vietnamese. *Commun. Disord. Q.* 39 (2), 371–380.

Geangu, E., Smith, W.A.P., Mason, H.T., Martinez-Cedillo, A.P., Hunter, D., Knight, M.I., Liang, H., Del Carmen Garcia de Soria Bazan, M., Tse, Z.T.H., Rowland, T., Corpuz, D., Hunter, J., Singh, N., Vuong, Q.C., Abdelpayed, M.R.S., Mullineaux, D.R., Smith, S., Müller, B.R., 2023. EgoActive: Integrated Wireless Wearable Sensors for Capturing Infant Egocentric Auditory-Visual Statistics and Autonomic Nervous System Function 'in the Wild. *Sens. (Basel, Switz.)* 23 (18), 7930. <https://doi.org/10.3390/s23187930>.

Gilkerson, J., Richards, J.A., 2008a. The LENA Natural Language Study. *LENA Foundation*.

Gilkerson, J., Zhang, Y., Xu, D., Richards, J.A., Xu, X., Jiang, F., Topping, K., 2015. Evaluating language environment analysis system performance for Chinese: A pilot study in Shanghai. *J. Speech, Lang., Hear. Res.* 58 (2), 445–452.

Glasper, A.M., Macleod, A.A., 2010. A multi-dimensional approach to gradient change in phonological acquisition: a case study of disordered speech development. *Clin. Linguist. Phon.* 24 (4–5), 283–299. <https://doi.org/10.3109/02699200903581091>.

Goldfield, E.C., 2000. Exploration of vocal tract properties during serial production of vowels by full-term and preterm infants. *Infant Behav. Dev.* 23 (3–4), 421–439. [https://doi.org/10.1016/S0163-6383\(01\)00053-4](https://doi.org/10.1016/S0163-6383(01)00053-4).

Greenwood, C.R., Thiemann-Bourque, K., Walker, D., Buzhardt, J., Gilkerson, J., 2011. Assessing children's home language environments using automatic speech recognition technology. *Commun. Disord. Q.* 32 (2), 83–92.

Hart, B., Risley, T.R., 2003. The early catastrophe: The 30 million word gap by age 3. *Am. Educ.* 27 (1), 4–9.

Henrich, J., Heine, S.J., Norenzayan, A., 2010. The weirdest people in the world? *Behav. brain Sci.* 33 (2–3), 61–135. <https://doi.org/10.1017/S0140525X0999152X>.

Hippe, L., Hennessey, V., Ferjan Ramírez, N., Zhao, T.C., 2024. Comparison of speech and music input in North American infants' home environment over the first 2 years of life. *Dev. Sci.* 27 (5), e13528. <https://doi.org/10.1111/desc.13528>.

Irvin, D.W., Hume, K., Boyd, B.A., McBee, M.T., Odom, S.L., 2013. Child and classroom characteristics associated with the adult language provided to preschoolers with autism spectrum disorder. *Res. Autism Spectr. Disord.* 7 (8), 947–955.

Islam, B., McElwain, N.L., Li, J., Davila, M.I., Hu, Y., Hu, K., Bodway, J.M., Dhekne, A., Roy Choudhury, R., Hasegawa-Johnson, M., 2024. Preliminary Technical Validation of LittleBeats™: A Multimodal Sensing Platform to Capture Cardiac Physiology, Motion, and Vocalizations. *Sens. (Basel, Switz.)* 24 (3), 901. <https://doi.org/10.3390/s24030901>.

Jain, R., Barcovschi, A., Yiwere, M.Y., Corcoran, P., Cucu, H., 2024. Exploring native and non-native English child speech recognition with Whisper. *IEEE Access* 12, 41601–41610. <https://doi.org/10.1109/ACCESS.2024.3378738>.

Joo, C.U., Choi, Y.M., Kim, S.J., 2015. Acoustic improvements after surgical correction in congenital heart disease. *Pediatr. Int.: Off. J. Jpn. Pediatr. Soc.* 57 (2), 205–209. <https://doi.org/10.1111/ped.12551>.

Joseph, G.E., Soderberg, J., Abbott, R., Garzon, R., Scott, C., 2022. Improving language support for infants and toddlers: Results of FIND coaching in childcare. *Infants Young-.. Child.* 35 (2), 91–105.

Josvassen, J.L., Hedegaard, V.A.M., Jørgensen, M.L., Percy-Smith, L., 2024. The Effect of LENA (Language ENvironment Analysis) for Children with Hearing Loss in Denmark including a Pilot Validation for the Danish Language. *J. Clin. Med.* 13 (9), 2688.

Kachergis, G., Marchman, V.A., Dale, P.S., Mankowitz, J., Frank, M.C., 2022. Online Computerized Adaptive Tests of Children's Vocabulary Development in English and Mexican Spanish. *J. Speech, Lang., Hear. Res.: JSLHR* 65 (6), 2288–2308. [https://doi.org/10.1044/2022\\_JSLHR-21-00372](https://doi.org/10.1044/2022_JSLHR-21-00372).

Katus, L., Crespo-Llado, M.M., Milosavljevic, B., Saidykhan, M., Njie, O., Fadera, T., McCann, S., Acolatse, L., Perapoch Amadó, M., Rozhko, M., Moore, S.E., Elwell, C.E., Lloyd-Fox, S., BRIGHT Project Team, 2024. It takes a village: Caregiver diversity and language contingency in the UK and rural Gambia. *Infant Behav. Dev.* 74, 101913. <https://doi.org/10.1016/j.infbeh.2023.101913>.

Knight, K., Ducasse, S., Coetzee, A., van der Linde, J., Louw, A., 2016. The effect of age of cochlear implantation on vocal characteristics in children. *South Afr. J. Commun. Disord.* 63 (1), 142. <https://doi.org/10.4102/sajcd.v63i1.142>.

Laing, C., Bergelson, E., 2024. Analyzing the effect of sibling number on input and output in the first 18 months. *Infancy* 29 (2), 175–195.

Lavechin, M., Bousbib, R., Bredin, H., Dupoux, E., & Cristia, A. (2020). An open-source voice type classifier for child-centered daylong recordings. *arXiv preprint 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. *Annu. Rev. Linguist.* 8, 389–407. <https://doi.org/10.1146/annurev-linguistics-031120-122120>.

Lavechin, M., Hamrick, L.R., Kelleher, B., & Seidl, A. (2025, April 4). Performance and biases of the LENA® and ACLEW algorithms in analyzing language environments in Down, Fragile X, Angelman syndromes, and populations at elevated likelihood for autism. [https://doi.org/10.31234/osf.io/fd3cr\\_v1](https://doi.org/10.31234/osf.io/fd3cr_v1).

Lee, M., Ha, S., 2024. Vocal and early speech development in Korean-acquiring children with hearing loss and typical hearing (Advance online publication). *Clin. Linguist. Phon.* 1–20. <https://doi.org/10.1080/02699206.2024.2380442>.

LENA Research Foundation. (n.d.). *LENA device*. LENA Research Foundation. (<https://shop.lenafoundation.org/products/lenadevice>).

León, M., Washington, K.N., McKenna, V.S., Crowe, K., Fritz, K., Boyce, S., 2023. Characterizing Speech Sound Productions in Bilingual Speakers of Jamaican Creole and English: Application of Durational Acoustic Methods. *J. Speech, Lang., Hear. Res.: JSLHR* 66 (1), 61–83. [https://doi.org/10.1044/2022\\_JSLHR-22-00304](https://doi.org/10.1044/2022_JSLHR-22-00304).

León, M., Meera, S.S., Fiévet, A.C., Cristia, A., 2024. Long-form recordings in low- and middle-income countries: recommendations to achieve respectful research. *Res. Ethics* 20 (1), 96–111. <https://doi.org/10.1177/17470161231199382>.

Llerena-Arregocés, D., Pérez-Moreno, J., 2024. Musical communication among parents and their children: An analysis tool to study their interaction. *Int. J. Music Educ.* 42 (3), 409–424. <https://doi.org/10.1177/02557614231174033>.

Levin-Asher, B., Segal, O., Kishon-Rabin, L., 2023. The validity of LENA technology for assessing the linguistic environment and interactions of infants learning Hebrew and Arabic. *Behav. Res. Methods* 55 (3), 1480–1495.

Li, F., 2012. Language-specific developmental differences in speech production: a cross-language acoustic study. *Child Dev.* 83 (4), 1303–1315. <https://doi.org/10.1111/j.1467-8624.2012.01773.x>.

Liu, C.T., 2021. A First Step toward the Clinical Application of Landmark-Based Acoustic Analysis in Child Mandarin. *Child. (Basel, Switz.)* 8 (2), 159. <https://doi.org/10.3390/children8020159>.

MacWhinney, B., 2000. *The CHILDES project: Tools for analyzing talk*, 3rd ed. Lawrence Erlbaum Associates.

Madden-Rusnak, A., Micheletti, M., Bailey, L., de Barbaro, K., 2024. Soothing touch matters: Patterns of everyday mother-infant physical contact and their real-time physiological implications. In: *Infant behavior & development*, 78. Advance online publication. <https://doi.org/10.1016/j.infbeh.2024.102021>.

Mahmoudian, S., Aminrasouli, N., Ahmadi, Z.Z., Lenarz, T., Farhadi, M., 2019. Acoustic Analysis of Crying Signal in Infants with Disabling Hearing Impairment. *J. Voice: Off. J. Voice Found.* 33 (6), 946.e7–946.e13. <https://doi.org/10.1016/j.jvoice.2018.05.016>.

Maitha, C., Goode, J.C., Maulucci, D.P., Lasassmeh, S.M.S., Yu, C., Smith, L.B., Borjon, J. I., 2020. An open-source, wireless vest for measuring autonomic function in infants. *Behav. Res. Methods* 52 (6), 2324–2337. <https://doi.org/10.3758/s13428-020-01394-4>.

Makranksy, G., Dale, P.S., Havmose, P., Bleses, D., 2016. An Item Response Theory-Based, Computerized Adaptive Testing Version of the MacArthur-Bates Communicative Development Inventory: Words & Sentences (CDI:WS). *J. Speech, Lang., Hear. Res.: JSLHR* 59 (2), 281–289. [https://doi.org/10.1044/2015\\_JSLHR-L-15-0202](https://doi.org/10.1044/2015_JSLHR-L-15-0202).

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, 10.1111/desc.12425. *Dev. Sci.* 20 (1). <https://doi.org/10.1111/desc.12425>.

Marschik, P.B., Pokorny, F.B., Peharz, R., Zhang, D., O'Muircheartaigh, J., Roeyers, H., Bölte, S., Spittle, A.J., Urlesberger, B., Schuller, B., Poustka, L., Ozonoff, S., Pernkopf, F., Pock, T., Tammimies, K., Enzinger, C., Krieber, M., Tomantschger, I., Bartl-Pokorny, K.D., Sigafoos, J., BEE-PRI Study Group, 2017. A Novel Way to Measure and Predict Development: A Heuristic Approach to Facilitate the Early Detection of Neurodevelopmental Disorders. *Curr. Neurotol. Neurosci. Rep.* 17 (5), 43. <https://doi.org/10.1007/s11910-017-0748-8>.

Marschik, P.B., Widmann, C.A.A., Lang, S., Kulvicius, T., Boterberg, S., Nielsen-Saines, K., Bölte, S., Esposito, G., Nordahl-Hansen, A., Roeyers, H., Wörgötter, F., Einspieler, C., Poustka, L., Zhang, D., 2022. Emerging Verbal Functions in Early Infancy: Lessons from Observational and Computational Approaches on Typical Development and Neurodevelopmental Disorders. *Adv. Neurodev. Disord.* 6 (4), 369–388. <https://doi.org/10.1007/s41252-022-00300-7>.

Mason, H.T., Martínez-Cedillo, A.P., Vuong, Q.C., García-de-Soria, M.C., Smith, S., Geangu, E., Knight, M.I., 2024. A Complete Pipeline for Heart Rate Extraction from Infant ECGs. *Signals* 5 (1), 118–146. <https://doi.org/10.3390/signals5010007>.

Mayor, J., Mani, N., 2019. A short version of the MacArthur-Bates Communicative Development Inventories with high validity. *Behav. Res. Methods* 51 (5), 2248–2255. <https://doi.org/10.3758/s13428-018-1146-0>.

Mealings, K.T., Demuth, K., 2014. The role of utterance length and position in 3-year-olds' production of third person singular -s. *J. Speech, Lang., Hear. Res.: JSLHR* 57 (2), 484–494. [https://doi.org/10.1044/2013\\_JSLHR-L-12-0354](https://doi.org/10.1044/2013_JSLHR-L-12-0354).

Meera, S.S., Swaminathan, D., Venkata Murali, S.R., Raju, R., Srikanth, M., Shyam Sundar, S., Mysore, A., 2025. Validation of the Language ENvironment Analysis (LENA) Automated Speech Processing Algorithm Labels for Adult and Child Segments in a Sample of Families From India. *J. Speech, Lang., Hear. Res.* 68 (1), 40–53.

Mendoza, J.K., Fausey, C.M., 2022. Everyday Parameters for Episode-to-Episode Dynamics in the Daily Music of Infancy. *Cogn. Sci.* 46 (8), e13178. <https://doi.org/10.1111/cogs.13178>.

Micheletti, M., Yao, X., Johnson, M., de Barbaro, K., 2023. Validating a model to detect infant crying from naturalistic audio. *Behav. Res. Methods* 55 (6), 3187–3197. <https://doi.org/10.3758/s13428-022-01961-x>.

Mieszkowska, K., Krajewski, G., Sobota, K., Dynak, A., Kolak, J., Krysztofiak, M., Łukomska, B., Łuniewska, M., Garmann, N.G., Hansen, P., Romøren, A.S.H., Simonsen, H.G., Alcock, K., Katsos, N., Haman, E., 2022. Parental Report via a Mobile App in the Context of Early Language Trajectories: StarWords Study Protocol. *Int. J. Environ. Res. Public Health* 19 (5), 3067. <https://doi.org/10.3390/ijerph19053067>.

Milenkovic P. (2018). *TF32 (Computer Software)*. Madison, WI: University of Wisconsin-Madison.

Muszyńska, K., Krajewski, G., Dynak, A., Garmann, N.G., Romøren, A.S.H., Łuniewska, M., Alcock, K., Katsos, N., Kolak, J., Simonsen, H.G., Hansen, P., Krysztofiak, M., Sobota, K., Haman, E., 2025. Bilingual children reach early language milestones at the same age as monolingual peers (Advance online publication). *J. Child Lang.* 1–24. <https://doi.org/10.1017/S0305000924000655>.

Niwano, K., Sugai, K., 2002. Intonation contour of Japanese maternal infant-directed speech and infant vocal response. *Jpn. J. Spec. Educ.* 39 (6), 59–68. [https://doi.org/10.6033/tokkyou.39.59\\_2](https://doi.org/10.6033/tokkyou.39.59_2).

Oller, D.K., Niyogi, P., Gray, S., Richards, J.A., Gilkerson, J., Xu, D., Warren, S.F., 2010. Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. *Proc. Natl. Acad. Sci.* 107 (30), 13354–13359.

Oller, D.K., Ramsay, G., Bene, E., Long, H.L., Griebel, U., 2021. Protophones, the precursors to speech, dominate the human infant vocal landscape. *Philos. Trans. R. Soc. B* 376 (1836), 20200255.

Orena, A.J., Byers-Heinlein, K., Polka, L., 2020. What do bilingual infants actually hear? Evaluating measures of language input to bilingual-learning 10-month-olds. *Dev. Sci.* 23 (2), e12901.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Moher, D., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ (Clin. Res. Ed.)* 372, n71. <https://doi.org/10.1136/bmj.n71>.

Perry, L.K., Prince, E.B., Valtierra, A.M., Rivero-Fernandez, C., Ullery, M.A., Katz, L.F., Laursen, B., Messinger, D.S., 2018. A year in words: The dynamics and consequences of language experiences in an intervention classroom. *PLoS One* 13 (7), e0199893. <https://doi.org/10.1371/journal.pone.0199893>.

Piaget, J., 1952. *The origins of intelligence in children* (M. Cook, Trans.). International Universities Press.

Pokorny, F.B., Marschik, P.B., Einspieler, C., Schuller, B., 2016. Does she speak RTT? Towards an earlier identification of Rett syndrome through intelligent pre-linguistic vocalisation analysis. *Proc. Inter. 2016*, 1953–1957. <https://doi.org/10.21437/Interspeech.2016-520>.

Pokorny, F.B., Bartl-Pokorny, K.D., Zhang, D., Marschik, P.B., Schuller, B., Schuller, B.W., 2020. Efficient Collection and Representation of Preverbal Data in Typical and Atypical Development. *J. Nonverbal Behav.* 44 (4), 419–436. <https://doi.org/10.1007/s10919-020-00332-4>.

Pokorny, F.B., Schmitt, M., Egger, M., Bartl-Pokorny, K.D., Zhang, D., Schuller, B.W., Marschik, P.B., 2022. Automatic vocalisation-based detection of fragile X syndrome and Rett syndrome. *Sci. Rep.* 12 (1), 13345. <https://doi.org/10.1038/s41598-022-17203-1>.

Putnam, O.C., Markfeld, J.E., Wright, S.T., Feldman, J.I., Goldblum, J., Karpinsky, M., Neal, A.J., Swanson, M.R., Harrop, C., 2024. The use of Language ENvironment Analysis in autism research: A systematic review, 13623613241290072. Advance online publication *Autism: Int. J. Res. Pract.* <https://doi.org/10.1177/13623613241290072>.

Radford, A., Kim, J.W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023, July). Robust speech recognition via large-scale weak supervision. In International conference on machine learning (pp. 28492–28518). PMLR.

Räsänen, O., (2021). *ALICE: Audio Language and Interspeech Characterization Environment*. GitHub. [https://github.com/orasanen/ALICE/blob/new\\_diarizer/docs/installation.md](https://github.com/orasanen/ALICE/blob/new_diarizer/docs/installation.md).

Räsänen, O., Seshadri, S., Lavechin, M., Cristia, A., Casillas, M., 2021. ALICE: An open-source tool for automatic measurement of phoneme, syllable, and word counts from child-centered daylong recordings. *Behav. Res. Methods* 53, 818–835.

Ribeiro, R., Ribeiro, X., Almeida, F., Ameijeiras-Rodriguez, C., Souza, J., Conceição, L., Taveira-Gomes, T., Marreiros, G., Freitas, A., 2023. Secur-e-Health Project: Towards Federated Learning for Smart Pediatric Care. *Stud. Health Technol. Inform.* 302, 516–520. <https://doi.org/10.3233/SHTI230196>.

Redford, M.A., 2014. The perceived clarity of children's speech varies as a function of their default articulation rate. *J. Acoust. Soc. Am.* 135 (5), 2952–2963. <https://doi.org/10.1121/1.4869820>.

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

Rowe, M.L., Romeo, R.R., Leech, K.A., 2023. Early Environmental Influences on Language. *Handb. Sci. Early Lit.* 23.

Scaff, C., Casillas, M., Stieglitz, J., Cristia, A., 2024. Characterization of children's verbal input in a forager-farmer population using long-form audio recordings and diverse input definitions. *Infancy* 29 (2), 196–215.

Schuller, B.W., Batliner, A., Bergler, C., Messner, E.-M., Hamilton, A., Amiriparian, S., Baird, A., Rizos, G., Schmitt, M., Stappen, L., Baumeister, H., MacIntyre, A.D., Hantke, S., 2020. The INTERSPEECH 2020 Computational Paralinguistics Challenge: Elderly Emotion, Breathing & Masks. In *Proceedings of the 21st Annual Conference of the International Speech Communication Association*. INTERSPEECH 2020, 2042–2046. <https://doi.org/10.21437/Interspeech.2020-32>.

Schwarz, I.C., Botros, N., Lord, A., Marcusson, A., Tidellius, H., & Marklund, E. (2017). The LENATM system applied to Swedish: Reliability of the Adult Word Count estimate. In *Interspeech 2017, Stockholm, Sweden, 20–24 August, 2017*(pp. 2088–2092). The International Speech Communication Association (ISCA).

Segura-Hernández, M., Valadez-Jiménez, V.M., Ysunza, P.A., Sánchez-Valerio, A.P., Arch-Tirado, E., Lino-González, A.L., Hernández-López, X., 2019. Acoustic analysis of voice in children with cleft lip and palate following vocal rehabilitation. Preliminary report. *Int. J. Pediatr. Otorhinolaryngol.* 126, 109618. <https://doi.org/10.1016/j.ijporl.2019.109618>.

Simon, K.R., Merz, E.C., He, X., Noble, K.G., 2022. Environmental noise, brain structure, and language development in children. *Brain Lang.* 229, 105112.

Singh, L., Basnight-Brown, D., Cheon, B.K., Garcia, R., Killen, M., Mazuka, R., 2025. Ethical and epistemic costs of a lack of geographical and cultural diversity in developmental science. *Dev. Psychol.* 61 (1), 1–18. <https://doi.org/10.1037/dev0001841>.

Smith, A.B., Lambrecht Smith, S., Locke, J.L., Bennett, J., 2008. A longitudinal study of speech timing in young children later found to have reading disability. *J. Speech, Lang., Hear. Res.: JSLHR* 51 (5), 1300–1314. <https://doi.org/10.1044/1092-4388-0806-0193>.

Smith, C.G., Jones, E.J.H., Charman, T., Clackson, K., Mirza, F.U., Wass, S.V., 2023. Vocalization and physiological hyperarousal in infant-caregiver dyads where the caregiver has elevated anxiety. *Dev. Psychopathol.* 35 (2), 459–470. <https://doi.org/10.1017/S095457942100153X>.

Soderstrom, M., Casillas, M., Bergelson, E., Rosenberg, C., Alam, F., Warlaumont, A.S., Bunce, J., 2021. Developing a cross-cultural annotation system and metacorpus for studying infants' real world language experience. *Collabra: Psychol.* 7 (1), 23445.

Southwell, R., et al. (2024). Automatic speech recognition tuned for child speech in the classroom. *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Seoul, Korea, Republic of, 12291–12295. <https://doi.org/10.1109/ICASSP48485.2024.10447428>.

Stern, W., Stern, C., 1907. *Die Kindersprache: Eine psychologische und sprachtheoretische Untersuchung*. Barth.

Suarez-Rivera, C., Fletcher, K.K., Tamis-LeMonda, C.S., 2024. Infants' home auditory environment: Background sounds shape language interactions. *Dev. Psychol.* 60 (12), 2274–2289. <https://doi.org/10.1037/dev0001762>.

Sullivan, J., Mei, M., Perfers, A., Wojcik, E., Frank, M.C., 2021. SAYCam: A Large, Longitudinal Audiovisual Dataset Recorded From the Infant's Perspective. *Open mind: Discov. Cogn. Sci.* 5, 20–29. [https://doi.org/10.1162/opmi\\_a\\_00039](https://doi.org/10.1162/opmi_a_00039).

Suskind, D.L., Leffel, K.R., Graf, E., Hernandez, M.W., Gunderson, E.A., Sapolich, S.G., Suskind, E., Leininger, L., Goldin-Meadow, S., Levine, S.C., 2016. A parent-directed language intervention for children of low socioeconomic status: a randomized controlled pilot study. *J. child lang.* 43 (2), 366–406. <https://doi.org/10.1017/S0305000915000033>.

Suskind, D., Leffel, K.R., Hernandez, M.W., Sapolich, S.G., Suskind, E., Kirkham, E., Meehan, P., 2013. An exploratory study of 'Quantitative Linguistic Feedback' effect of LENA feedback on adult language production. *Commun. Disord. Q.* 34 (4), 199–209.

Thiemann-Bourque, K.S., Warren, S.F., Brady, N., Gilkerson, J., Richards, J.A., 2014. Vocal interaction between children with Down syndrome and their parents. *Am. J. Speech-Lang. Pathol.* 23 (3), 474–485.

Tomasello, M., 2005. *Constructing a language: A usage-based theory of language acquisition*. Harvard university press.

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 audio recordings (May). In: *In Seminars in speech and language*, 37. Thieme Medical Publishers, pp. 128–142 (May).

Vygotsky, L.S., 1978. *Mind in society: The development of higher psychological processes*, 86. Harvard university press.

Warlaumont, A.S., Oller, D.K., Buder, E.H., Dale, R., Kozma, R., 2010. Data-driven automated acoustic analysis of human infant vocalizations using neural network tools. *J. Acoust. Soc. Am.* 127 (4), 2563–2577. <https://doi.org/10.1121/1.3327460>.

Warlaumont, A.S., Richards, J.A., Gilkerson, J., Oller, D.K., 2014. A social feedback loop for speech development and its reduction in autism. *Psychol. Sci.* 25 (7), 1314–1324.

Wass, S., Phillips, E., Smith, C., Fatimehin, E.O.O.B., Goupli, L., 2022. Vocal communication is tied to interpersonal arousal coupling in caregiver-infant dyads. *eLife* 11, e77399. <https://doi.org/10.7554/eLife.77399>.

Weisleder, A., Fernald, A., 2013. Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychol. Sci.* 24 (11), 2143–2152.

Xu, T.L., de Barbaro, K., Abney, D.H., Cox, R.F.A., 2020. Finding Structure in Time: Visualizing and Analyzing Behavioral Time Series. *Front. Psychol.* 11, 1457. <https://doi.org/10.3389/fpsyg.2020.01457>.

Yeh, S.L., Chao, G.Y., Su, B.H., Huang, Y.L., Lin, M.H., Tsai, Y.C., Lee, C.C., 2019. Using Attention Networks and Adversarial Augmentation for Styrian Dialect Continuous Sleepiness and Baby Sound Recognition (September). *Interspeech* 2398–2402.

Yoo, H., Buder, E.H., Bowman, D.D., Bidelman, G.M., Oller, D.K., 2019. Acoustic Correlates and Adult Perceptions of Distress in Infant Speech-Like Vocalizations and Cries. *Front. Psychol.* 10, 1154. <https://doi.org/10.3389/fpsyg.2019.01154>.

Zhang, X., Liu, D., Pappas, L., Dill, S.E., Feng, T., Zhang, Y., Ma, Y., 2024. The home language environment and early childhood development: A LENA study from rural and peri-urban China. *Appl. Dev. Sci.* 28 (4), 655–673.

Zhang, Y., Martinez-Cedillo, A.P., Mason, H.T., Vuong, Q.C., Garcia-de-Soria, M.C., Mullineaux, D., Knight, M., & Geangu, E. (2024). An automatic sustained attention prediction method for infants and toddlers using wearable device signals. *bioRxiv*. <https://doi.org/10.1101/2024.11.16.623211>.

ZhuParris, A., Kruizinga, M.D., Gent, M.V., Dassing, E., Exadaktylos, V., Doll, R.J., Cohen, A.F., 2021. Development and technical validation of a smartphone-based cry detection algorithm. *Front. Pediatr.* 9, 651356.