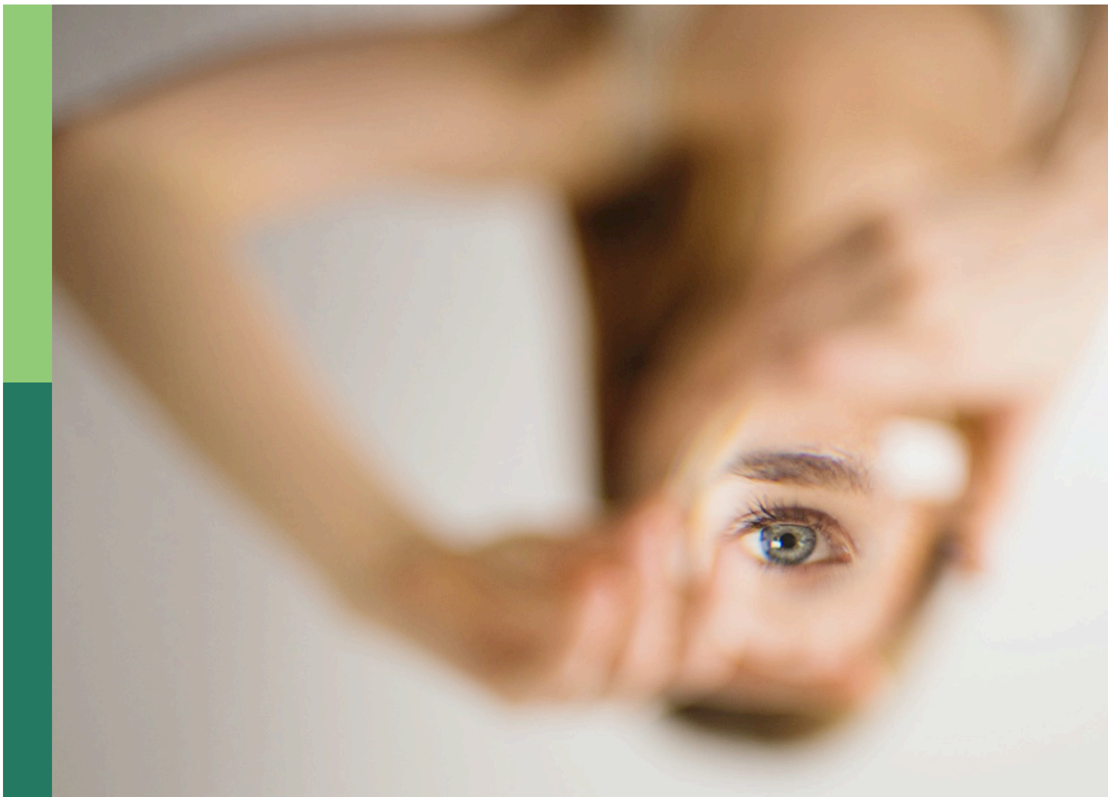
Research Topics

# Are there different types of child-directed speech? Dynamic variations according to individual and contextual factors

**Edited by**
Maria Spinelli, Chiara Suttora, Mirco Fasolo, Fabia Franco,
Adrian Garcia-Sierra and Francesca Lionetti

**Published in**
Frontiers in Psychology



ISSN 1664-8714
ISBN 978-2-83251-622-5
DOI 10.3389/978-2-83251-622-5

## About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

# Are there different types of child-directed speech? Dynamic variations according to individual and contextual factors

**Topic editors**
Maria Spinelli — University of Studies G. d'Annunzio Chieti and Pescara, Italy
Chiara Suttora — University of Bologna, Italy
Mirco Fasolo — University of Studies G. d'Annunzio Chieti and Pescara, Italy
Fabia Franco — Middlesex University, United Kingdom
Adrian Garcia-Sierra — University of Connecticut, United States
Francesca Lionetti — University of Studies G. d'Annunzio Chieti and Pescara, Italy

**Citation**
Spinelli, M., Suttora, C., Fasolo, M., Franco, F., Garcia-Sierra, A., Lionetti, F., eds. (2023). *Are there different types of child-directed speech? Dynamic variations according to individual and contextual factors*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-83251-622-5

# Table of contents

frontiers | Frontiers in Psychology

TYPE Editorial
PUBLISHED 24 January 2023
DOI 10.3389/fpsyg.2022.1056816



OPEN ACCESS

EDITED AND REVIEWED BY
Peter Klaver,
Interkantonale Hochschule für Heilpädagogik (HfH), Switzerland

*CORRESPONDENCE
Maria Spinelli
✉ maria.spinelli@unich.it

SPECIALTY SECTION
This article was submitted to Developmental Psychology, a section of the journal Frontiers in Psychology

RECEIVED 29 September 2022
ACCEPTED 03 October 2022
PUBLISHED 24 January 2023

CITATION
Spinelli M, Suttora C, Garcia-Sierra A, Franco F, Lionetti F and Fasolo M (2023) Editorial: Are there different types of child-directed speech? Dynamic variations according to individual and contextual factors. *Front. Psychol.* 13:1056816. doi: 10.3389/fpsyg.2022.1056816



# Editorial: Are there different types of child-directed speech? Dynamic variations according to individual and contextual factors

Maria Spinelli[1]*, Chiara Suttora[2], Adrian Garcia-Sierra[3], Fabia Franco[4], Francesca Lionetti[1] and Mirco Fasolo[1]

[1]Department of Neurosciences, Imaging and Clinical Sciences, University of Studies G. d'Annunzio Chieti and Pescara, Chieti, Italy, [2]Department of Psychology, University of Bologna, Bologna, Italy, [3]Department of Speech, Language and Hearing Sciences, University of Connecticut, Storrs, CT, United States, [4]Department of Psychology, Middlesex University, London, United Kingdom

KEYWORDS

**child directed speech, caregiver-child interaction, child development, language development, cognitive development, emotional development**

Editorial on the Research Topic
Are there different types of child-directed speech? Dynamic variations according to individual and contextual factors

Child-directed speech (CDS) is the particular voice register observed in the majority of caregivers during interaction with their infants and children. CDS represents a crucial part of the dyadic caregiver-child interaction, and its prosodic, lexical, syntactic, and functional characteristics are associated with several outcomes such as infant attention, engagement, linguistic acquisition, and affect transmission and sharing (Fernald and Kuhl, 1987; Hoff and Naigles, 2002; Saint-Georges et al., 2013; Rowe and Snow, 2020).

From a dynamic perspective, the specific features and modifications over time of CDS may be considered a way caregivers adjust their input to children's development and achievements during interactions (Soderstrom, 2007; Golinkoff et al., 2015). It is through these social exchanges that, in turn, CDS stimulates children's socio-cognitive development. However, individual fluctuations in CDS characteristics are documented (D'Odorico and Jacob, 2006). These variations—which are not necessarily adaptive or non-adaptive—could be determined by the dynamic interaction among contextual and individual factors.

The current Research Topic brings together researchers working on exploring CDS from this perspective by analyzing specific linguistic and prosodic characteristics of CDS, how these are affected by individual, dyadic, and contextual factors, and the role of those variations in child development.

The importance of considering CDS by focusing on its role in the dyadic exchanges is transversal to all the studies collected in this Research Topic. When exploring CDS universality across cultures and contexts (see Soderstrom et al. and Sarvasy et al.), the findings evidence that even if some elements of CDS appear similar across different languages and cultural contexts, each language has its specific characteristics. These peculiarities appear to be related to the peculiarities of the caregiver-child interaction, which can be influenced by the caregiver's cultural background and practices.

Other studies underlined the importance of considering the role of the child in dyadic exchanges. Fetus-directed speech is influenced by fetal movements, interpreted by mothers as the participation of the fetus in the social interaction (see Parlato-Oliveira et al.). This reciprocal influence is not generic, but specific to the different characteristics of CDS. At 12 months, during interactive turns, the caregiver's IDS phonetic complexity negatively influences the infant's vocalization (see Marklund et al.). The age of the child is a relevant variable in this process. The 10 studies covered different child developmental ages, from fetal to school age, showing how the important characteristics of CDS to explore are different, as are the different competencies and needs of the child. Studies with different time points confirmed that CDS varied over infancy and childhood (see Gram Garmann et al.), adapting to the infant's and children's communication abilities. These variations play an important role in later language development in children (see Cychosz et al.).

Child language development was shown to be influenced by the ability of the caregiver to adapt CDS to the contingent needs of the child over time. For example, maternal circumstances (i.e., parenting stress), dyad aspects (i.e., quality of co-regulation) as well as infant specific circumstances (i.e., preterm birth; see Spinelli et al.). Again, CDS varies according to the specific characteristics of the actors in the dyadic interaction. These variabilities are not always interpreted as different from normative CDS, but as specific adaptations of maternal speech to the infant's needs and communication abilities. Adaptations can be more or less appropriate according to the characteristics of the dyad. It is within environmental and individual risk contexts that the last two studies in this Research Topic illustrate the characteristics and effectiveness of interventions aimed at improving caregivers' CDS to promote, in turn, children's linguistic development (see Hindman et al. and Suttora et al.).

These 10 studies together positively answer the main question of this Research Topic: Are there different types of CDS? CDS was proved to be a dynamic construct, not stable, with variations in the function of the dynamic needs of the dyadic interaction. From this point of view, when studying CDS, the individual role of the child as an interactive partner should be always considered, from the first fetus-caregiver to the teacher/caregiver in interactions with school-aged children. To this extent, going beyond the communicative role of the child (i.e., considering linguistic comprehension and production abilities), researchers should take into account the complexity of the child as an individual with all his/her emotional, interactive, and cognitive abilities. At the same time, the adult should be considered in their complexity as an interactive partner, who uses CDS as an interactive modality with the aim not only to transmit language but also to communicate emotions, affect, and knowledge. Through this lens, considering the caregiver's individual characteristics, such as wellbeing and culture becomes necessary. All these elements dynamically influence each other in the dyadic interactive experience.

This point of view also affects how we should study the impact of CDS on child development. The multifarious peculiarities of CDS imply that CDS might have multifarious purposes, presuming it affects not only a child's linguistic competencies but also several other developmental aspects. We highly encourage more longitudinal studies on this point. It is with such studies that scientific knowledge could gain more information to develop suitable intervention programs to promote the best child development.

In conclusion, the present Research Topic successfully collected contributions from researchers and clinicians, evidencing the importance of looking at CDS from a dynamic perspective and considering the interaction of individual and contextual factors. This is possible only if there is a virtuous integration of expertise between linguistic, emotional, cognitive, and clinical frameworks. Planning studies within this perspective is a future challenge that will advance scientific knowledge in directions that could strongly affect our ability to provide novel knowledge on CDS and hence shape efficient preventive and intervention strategies to promote its quality and its impact on child development. The present Research Topic is the first step forward in this direction.

## Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.


## Acknowledgments

In memory of Laura D'Odorico whose view on child direct speech complexity encouraged us to organize this Research Topic.



## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## References

D'Odorico, L., and Jacob, V. (2006). Prosodic and lexical aspects of maternal linguistic input to late-talking toddlers. *Int. J. Lang. Commun. Disord.* 41, 293–311. doi: 10.1080/13682820500342976

Fernald, A., and Kuhl, P. K. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behav. Dev.* 10, 279–293. doi: 10.1016/0163-6383(87)90017-8

Golinkoff, R. M., Can, D. D., Soderstrom, M., and Hirsh-Pasek, K. (2015). (Baby)talk to me: the social context of infant-directed speech and its effects on early language acquisition. *Curr. Direct. Psychol. Sci.* 24, 339–344. doi: 10.1177/0963721415595345

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

Rowe, M. L., and Snow, C. E. (2020). Analyzing input quality along three dimensions: interactive, linguistic, and conceptual. *J. Child Lang.* 47, 5–21. doi: 10.1017/S0305000919000655

Saint-Georges, C., Chetouani, M., Cassel, R., Apicella, F., Mahdhaoui, A., Muratori, F., et al. (2013). Motherese in interaction: at the cross-road of emotion and cognition? (a systematic review). *PLoS ONE* 8, e78103. doi: 10.1371/journal.pone.0078103

Soderstrom, M. (2007). Beyond babytalk: re-evaluating the nature and content of speech input to preverbal infants. *Dev. Rev.* 27, 501–532. doi: 10.1016/j.dr.2007.06.002

frontiers in Psychology

ORIGINAL RESEARCH
published: 09 March 2021
doi: 10.3389/fpsyg.2021.646170

Check for updates

# "Motherese" Prosody in Fetal-Directed Speech: An Exploratory Study Using Automatic Social Signal Processing

*Erika Parlato-Oliveira[1,2,3], Catherine Saint-Georges[3,4], David Cohen[3,4], Hugues Pellerin[3,4], Isabella Marques Pereira[1], Catherine Fouillet[4], Mohamed Chetouani[3], Marc Dommergues[5] and Sylvie Viaux-Savelon[2,4]**

*[1] School of Medicine, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, [2] CRPMS, Université de Paris, Paris, France, [3] Institut des Systèmes Intelligents et de Robotiques (ISIR), équipe Perception, Interaction, et Robotiques Sociales (PIRoS), Sorbonne Université, Paris, France, [4] Service de Psychiatrie de l'Enfant et de l'Adolescent, APHP Pitié Salpêtrière, Sorbonne Université, Paris, France, [5] Service de Gynécologie Obstétrique, APHP Pitié Salpêtrière, Sorbonne Université, Paris, France*



**Edited by:**
*Chiara Suttora,*
*University of Bologna, Italy*

**Reviewed by:**
*Paola Zanchi,*
*University of Milano-Bicocca, Italy*
*Eon-Suk Ko,*
*Chosun University, South Korea*

***Correspondence:**
*Sylvie Viaux-Savelon*
*sylvie.viaux@icloud.com*

**Specialty section:**
*This article was submitted to Developmental Psychology, a section of the journal Frontiers in Psychology*

**Received:** *25 December 2020*
**Accepted:** *15 February 2021*
**Published:** *09 March 2021*

**Citation:**
*Parlato-Oliveira E, Saint-Georges C, Cohen D, Pellerin H, Pereira IM, Fouillet C, Chetouani M, Dommergues M and Viaux-Savelon S (2021) "Motherese" Prosody in Fetal-Directed Speech: An Exploratory Study Using Automatic Social Signal Processing. Front. Psychol. 12:646170. doi: 10.3389/fpsyg.2021.646170*

**Introduction:** Motherese, or emotional infant directed speech (IDS), is the specific form of speech used by parents to address their infants. The prosody of IDS has affective properties, expresses caregiver involvement, is a marker of caregiver-infant interaction quality. IDS prosodic characteristics can be detected with automatic analysis. We aimed to explore whether pregnant women "speak" to their unborn baby, whether they use motherese while speaking and whether anxio-depressive or obstetrical status impacts speaking to the fetus.

**Participants and Methods:** We conducted an observational study of pregnant women with gestational ages from 26 to 38 weeks. Women were recruited in a university hospital department of obstetrics. Thirty-five women agreed to participate in the study, and 26 audio records were exploitable. We collected obstetrical and sociodemographic data, pregnancy outcomes, anxiety and depressive status using the Covy and Raskin Scales, and life events using the Sensations During Pregnancy and Life Event Questionnaire. Each participant was left alone with an audio recorder with a recommendation to feel free to speak to her fetus as she would have done at home. The recording was stopped after 3 min. Audio recordings were analyzed by two methods: psycholinguist experts' annotation and computational objective automatic analyses.

**Results:** Most mothers (89%) reported speaking to their fetuses. We found a correlation between maternal first perceptions of fetal movements and the start of mother's speaking to fetus. Motherese prosody was detected with both annotation and automatic analysis with a significant correlation between the two methods. In this exploratory study, motherese use was not associated with maternal anxiodepressive or obstetrical status. However, the more future mothers were depressed, the less they spoke with their fetuses during the recording.

**Conclusion:** Fetal directed speech (FDS) can be detected during pregnancy, and it contains a period of prosody that shares the same characteristics of motherese that can be described as prenatal motherese or emotional fetal-directed speech (e-FDS). This means that pregnant women start using motherese much earlier than expected. FDS seems to be correlated with maternal first perceptions of fetal movements and depression scores. However, more research is needed to confirm these exploratory results.

**Keywords: motherese, prenatal, mother-fetus interaction, fetal-directed speech, machine learning, social signal processing**

## INTRODUCTION

Infant-directed speech (IDS) or motherese is a specific register, which includes peculiar prosodic characteristics, that parents or caregivers often use when speaking to infants (Fernald and Simon, 1984; Fisher and Tokura, 1995; Spinelli et al., 2017). The use of motherese progressively increases as the baby grows and then usually decreases and disappears when the child becomes able to communicate verbally (Saint-Georges et al., 2013). IDS has been studied extensively across a number of interactive situations and contexts, especially by researchers interested in understanding language acquisition. IDS is also a marker of the parent infant interaction quality. Motherese characteristics have been shown to be linked with emotional prosody characteristics (Trainor et al., 2000). Behavioral studies have shown that infants prefer and respond better to motherese than to regular prosody, typical of adult directed speech (Fernald and Kuhl, 1987; Dupoux and Mehler, 1990; Saint-Georges et al., 2013; Outters et al., 2020). IDS has affective properties, expresses parental involvement, and contributes to regulating caregiver-infant interactions (Cohen et al., 2013). Thus, IDS is part of an interactive loop that may play an important role in infants' cognitive and social development (Saint-Georges et al., 2013). Experimental data suggest that very young infants in their first month of life (Cooper and Aslin, 1990; Cooper, 1993) or in their first week (Ramus, 1999) and even neonates (Saito et al., 2007) are sensitive to this prosody.

The *in utero* period has been less explored. A recent study (Bartha-Doering et al., 2019) suggested that neural discrimination of speech begins *in utero*. Some reports also show that future mothers speak to their fetus (DeCasper et al., 1994). In addition, parents observing their fetuses during ultrasound prenatal screening were shown to present mirroring movement activities (Ammaniti et al., 2010). These studies suggest that motherese may already be present in the prenatal period and may be associated with mother involvement and emotional tone regarding prenatal interactions. However, motherese has not been clearly demonstrated during the prenatal period.

With the development of automatized methods of social signal processing, machine-learning methods can now detect the acoustic characteristics of emotional speech, such as motherese, in the human voice. It can distinguish motherese sequences from adult-directed speech (Mahdhaoui et al., 2011). Traditionally, the design of computerized classifiers aims to capture supra-segmental features like pitch (fundamental frequency), duration, energy of vocalizations as well as global dynamics of spectrum (Williams and Stevens, 1972; Sherer, 1986; Chetouani et al., 2014). Mel frequency cepstral coefficients (MFCC) capture short-term dynamics of spectrum and are termed segmental features. The motherese detection algorithm system exploits the combination of two classifiers, segmental and supra-segmental, that are weighted and fused to reach best classification rates (Mahdhaoui et al., 2011). Previous works have shown that these methods can contribute to exploring parent-infant interactions in video/audio recordings in natural or experimental settings (Cohen et al., 2013; Weisman et al., 2016; Bourvis et al., 2018). Moreover, these automatic analyses of motherese have contributed to state the universality of the emotional prosodic characteristics of IDS across languages (Parlato et al., 2020).

Here, we describe an exploratory observational study based on interviews and audio recording of volunteer pregnant women with the following aims: (1.1) to determine whether pregnant women speak to their unborn baby, (1.2) if so, to determine, if they speak using motherese prosody or not, with two methods (1.3) with prosody analyzed by clinical experts, and (1.4) using computational analysis of speech with machine learning method. In addition, (2) we will assess if prenatal stress, obstetrical or fetal complications, and future mother emotional state would influence the quantity and characteristics of mother's prosody (Watson et al., 2002; Viaux-Savelon et al., 2012).

## METHODS

### Participants and Ethics

From September 2013 to January 2014, we proposed to pregnant women attending the prenatal clinic of the Pitié-Salpêtrière University Hospital in Paris, France, to participate in a survey on maternal speech. They received oral and written information explaining that their participation would require filling out self-questionnaires, answering questions related to their emotional status, and being audiotaped when speaking to their future baby. The study was approved by the local Ethical Committee (CPPIDF6) under the number n°09012014. All participants gave a written consent. The inclusion criteria were mothers aged 18 or above, with a gestational age of 26–38 weeks, and able to understand the protocol. Indeed, during this period of pregnancy, future mothers are less concerned by the fetus viability. They begin preparing their meeting and relationship with the future child with more dreams and more detailed representations about the future baby (Ammaniti et al., 2000). Provided women were fluent in French, they could be included even if French was not their native language. Exclusion criteria were mental disorders and absence of health coverage according to the French ethical rules that require that studies be carried out only on people with health coverage. The mental disorder information were extracted from the obstetric record. Mental disorders were considered to be present if the pregnant woman was cared or treated for mental disorder by specialist before the pregnancy.

## Data Collection

### Clinical Data Collection

We collected several clinical variables. Social and demographic characteristics included age, parity, marital status, native language, education level, and occupation. We also assessed life events using the Sensations During Pregnancy and Life Event Questionnaire (Tordjman et al., 2004). Fetal-oriented interaction variables included gestational age when the pregnant women reported spontaneous moment of speaking to their infant to come and the gestational age they started perceiving fetal movement. To assess maternal anxiety-depression status, we used two specific scales. Maternal anxiety was assessed using the Covi anxiety scale, a questionnaire completed by the investigator (EP), based on clinical assessment. This score ranges from 0 (no anxiety) to 12 (high anxiety), with a threshold of six defining clinically relevant anxiety (Covi, 1986). Depression was assessed using the RASKIN score based on clinical assessment. This score ranges from 0 (no depression) to 12 (high depression), with a threshold of six defining clinically relevant depression (Raskin and Crook, 1976). In case of clinically relevant depression or anxiety, we planned that the investigator would warn the doctor or midwife in charge of the patient to organize adequate follow-up.

Finally, we retrospectively recorded medical history and obstetrical outcomes after birth based on obstetrical and neonatal records by professionals who were blinded to the audio analysis. Variables are listed in **Table 1** (pregnancy and delivery sections). Breastfeeding initiation was collected as some studies have pointed out that stress events during pregnancy may influence breastfeeding initiation or duration (Evers et al., 1998; Figueiredo et al., 2013).

**TABLE 1** | Description of study participants $N = 35$.

| | |
|---|---|
| ***Sociodemographics*** | |
| Age: mean (SD) [range] in years | 32.34 (6.4) [18.9–42.66] |
| Marital status: single/in couple | 7 (20%)/28 (80%) |
| Working status: No/Working/Student | 6 (17.1%)/27 (77.1%)/2 (5.7%) |
| Years of education: ≤12/>12 years | 8 (22.9%)/27 (77.1%) |
| Mother tongue: French/other | 30 (85.7%)/5 (14.3%) |
| ***Life events*** | |
| Number of life events: mean (SD) [range] | 7.69 (5.95) [0–24] |
| Significant obstetrical history: n (%) | 21 (60%) |
| Significant medical history: n (%) | 23 (65.7%) |
| ***Pregnancy*** | |
| Gestational age at study recruitment: mean (SD) [range] in weeks | 32.45 (3.69) [22–38] |
| At least one fetal risk: n (%) | 17 (49%) |
| Ultrasound soft marker: n (%) | 2 (6%) |
| At least one maternal risk: n (%) | 21 (60%) |
| Fetus gender: Female/Male | 22 (63%)/13 (37%) |
| Complication during pregnancy: n (%) | 13 (37%) |
| Global risk: No/Fetus only/Mother only/Both | 7 (20%)/4 (11.4%)/11 (31.4%)/13 (37.1%) |
| ***Psychopathology*** | |
| Covy anxiety score: mean (SD) [range] | 1.7 (2.15) [0–6] |
| Raskin depression score: mean (SD) [range] | 1.67 (1.84) [0–6] |
| ***Fetal oriented interaction variables*** | |
| Mother gestational age when first sentences were addressed to the fetus: mean (SD) [range] in months | 3.63 (1.64) [0–6] |
| Mother declare spontaneously speaking to fetus: never or rarely/frequently/missing data | 9 (25.7%)/22 (62.9%)/4 (11.4%) |
| Mother gestational age when they first perceived their fetus moving: mean (SD) [range] in months | 2.83 (1.89) [1–7] |
| ***Delivery*** | |
| Gestation duration: mean (SD) [range] in weeks | 38.82 (2.23) [29.1–41.5] |
| Baby weight at birth: mean (SD) [range] in g | 3047.6 (684.32) [2140–3850] |
| 5 min APGAR: mean (SD) [range] | 9.91 (0.37) [8–10] |
| Breast feeding: n (%) | 29 (83%) |
| Mode of delivery: Basse/VBI/Anticipated Caesarian/Emergency Caesarian/Missing data | 17 (48.6%)/5 (14.3%)/9 (25.7%)/3 (8.6%)/1 (2.9%) |

### Audio Data Collection and Analyses

After the questionnaires were completed, the investigator invited the participant to sit in a quiet room, independent of the prenatal clinic suite. The participant was left alone with a recorder (Zoom recorder AT170 PRO-Sony) lying on a nearby table. She was asked to feel free to speak or not to her baby, as she would have done at home. The mother was taken to a quiet room with a comfortable chair and invited to speak with her fetus, if only she wanted to. The interviewer would turn on the recorder and leave the room, so as not to intimidate the mother and allow the environment to be as close as possible to the mother's usual situation with her fetus. The recording was stopped after 3 min. Audio recordings were analyzed at two levels: (i) maternal vocalization characteristics (low-level features) and (ii) affective speech analysis (high-level audio features).

### Maternal Vocalization Characteristic

During a dialogue or a monolog, vocalization can be characterized by a series of quantitative parameters that allow describing the features and their dynamics. For our survey, we adapted this method to the monolog uttered by the mother, making the hypothesis that what we recorded was the equivalent of a dialogue between the mother and her unborn child. We first segmented and annotated the mothers' vocalization based on the Weisman et al. (2016) method. Two experts (EP and IR), one linguist and one speech therapist, listened to every 3-min recording. Using the ELAN EUDICO Linguistic Annotator (Institut Max-Planck, Nimègue, Nederland), they worked together to split them into segments of vocalization defined as continuous streams of speech with <150 ms of silence. Then, they labeled each segment as vocalization, laughing, singing, crying or other sounds. The maternal vocalizations consisted in the mother's recorded sound production directed or not to the fetus. For examples: "I'm very tired," "I can't wait to see you," "My baby, your room is ready, we await you with love," "I am afraid about childbirth." Maternal vocalization, maternal pause, and silence were extracted using an automated algorithm (for details see Bourvis et al., 2018). It calculated the duration of each segment and the amount of pause time within each 3-min recording, corresponding to the sum of silences >150 ms between two segments. Thus, we obtained the *maternal vocalization mean duration*, the *vocalization number during the 3-min window*, the *maternal pause mean duration* and the *vocalization ratio of time during the 3 min*.

### Affective Speech Analysis

Each speech segment labeled "vocalization" by the experts was extracted as a digital audio sample, stored and submitted to affective speech analysis based on high-level audio features. The goal of this analysis was to categorize each vocalization as "motherese," based on the presence of the emotional component of IDS, vs. "non-motherese," which refers to prosody more typical of adult directed speech. This was achieved by two methods: expert evaluation by listening to the audio samples of the segments labeled "vocalization" and computational automatic assessment of the same digital samples.

For the manual qualitative annotation, the two experts (EPO and IMP) worked independently to assess the presence of motherese characteristics in each vocalization segment. Interrater agreement between the two independent raters was calculated on the whole sample of vocalizations and was found to equal 80%. In case of disagreement, they listened again together to the remaining segments with no agreement (20%) and reached a consensus. This method allowed us to obtain a unique manual label for each vocalization segment.

Automated labeling for motherese or non-motherese was performed using an *ad hoc* algorithm developed in the ISIR *(Institut des Systèmes Intelligents et de Robotiques)* laboratory in Paris. This motherese classifier, based on machine learning methods, uses both segmental (mel-frequency cepstrum coefficients, MFCCs) and suprasegmental (e.g., statistics with regard to fundamental frequency, energy, and duration) acoustic characteristics of speech and SVM (support vector machine) classifiers. The algorithm classifier was trained on a data set of both motherese and non-motherese. It can distinguish emotional sequences of motherese from normal speech (Mahdhaoui et al., 2011). In previous studies, it was able to identify motherese during early interaction in both experimental (Weisman et al., 2016; Bourvis et al., 2018) and natural settings (Cohen et al., 2013), in both mothers and fathers (Cohen et al., 2013; Weisman et al., 2016; Parlato et al., 2020), in various languages (Parlato et al., 2020), and in parents speaking to infants with later psychopathology (e.g., autism, Cohen et al., 2013).

Both motherese detection methods created two subclass labels of maternal vocalization: "motherese" labeled Emotional Fetal-Directed Speech (e-FDS) vs. "non-motherese" (non-e-FDS). Three variables were derived: e-FDS ratio during the 3 min, non-e-FDS ratio during the 3 min, and *e-FDS/vocalization ratio* (duration of "motherese" vocalization/duration of maternal vocalization).

## Statistical Analysis

Statistical analyses were performed using R Software, Version 2.12.2. For all tests, the level of significance alpha was fixed at 5%. Given the sample size and the exploratory nature of the study, we used univariate analysis only. Quantitative variables were presented as the mean, standard deviation, and range. Qualitative variables were presented as frequencies.

First, we explored the correlation between maternal first perception of fetal movements and first vocalizations to the fetus.

We then successively estimated the relationship between:

(i) Variable "*Vocalization ratio of time during the 3 min*" and the following variables: "depression score" (score Raskin), "anxiety score" (score Covy), "life events" and "fetal risk".
(ii) *Variable ratio Emotional-Fetal Directed Speech (e-FDS)/vocalization according to psycholinguist expert* and the following variables: "depression score" (score Raskin), "anxiety score" (score Covy), "life event" and "fetal risk".

The relationship between two continuous variables was either tested using Pearson r or Spearman rho, depending on the validity of the assumptions. The relationship between a continuous and a binary variable (fetal risk) was either tested using the Welch *t*-test or Wilcoxon rank sum test, depending on the validity of the assumptions. Finally, we estimated the agreement between our experts' measures and the algorithm's measures using ICC (single random raters, ICC2) and calculated the 95% confidence interval (R psych package).

# RESULTS

## Flow Chart

One hundred forty-five pregnant women were considered eligible from September 2013 to January 2014. Thirty-five of them agreed to participate in the study and were enrolled. Four women participated in the clinical part of the study but did not record audio data; five audio records were not exploitable because of technical difficulties. Thus, 26 audio records were used for analysis. The most frequent reason declared by the invited mothers to decline the participation to the research was the lack of available time, considering that the interview needed at least 1 h and 30 min. Indeed in order to avoid displacements, we proposed the study to mothers who were already present at the Hospital for a pregnancy consultation.

## Description of the Sample (*N* = 35)

**Table 1** summarizes the study sample in terms of socio-demographics, life events and pregnancy risk factors, delivery, psychopathology and fetal-oriented interaction variables. Mothers presented a high number of life events and significant obstetrical (60%) and medical (65.7%) history: endocrinologic conditions ($N = 7$), multisystemic pathologies ($N = 3$), neurologic disorders ($N = 4$), psychiatric disorders ($N = 2$), uterine anomaly ($N = 2$), social precariousness ($N = 2$), and cardiac anomaly ($N = 1$). Regarding fetal risk, we found trisomy 21 risk ($N = 3$), intrauterine growth retardation ($N = 2$), premature delivery threats ($N = 2$), cardiovascular anomalies ($N = 1$), drug exposure ($N = 1$), and prior history of neonatal death ($N = 1$). This could be related to recruitment inside a free public university hospital in a maternity unit specializing in complex cases.

Nevertheless, our sample presented a low mean level of anxiety and depression scores and a high percentage of breastfeeding compared to the French general population (68.1–70.5% in the general population with 59% exclusive breastfeeding https://drees.solidarites-sante.gouv.fr/IMG/pdf/dt68-sources_et_methodes.pdf; Kersuzan et al., 2014).

## Analysis of Fetal-Oriented Interaction Variables ($N = 35$)

Mothers declared to start speaking or vocalizing to their fetus on average at 3.63 (±1.64) months during pregnancy. Additionally, they started perceiving fetal movement on average at 2.83 (±1.89) months during pregnancy. We found a significant correlation between speaking to the fetus and perceiving fetal movements (**Figure 2**, right).

Regarding audio analyses, only 26 mothers were included because of technical issues (see **Figure 1** flow chart). For low-level audio analysis (quantitative speech analysis), a total of 856 vocalization segments (mean vocalization number = 32.92) were detected. The duration of the vocalizations ranged from 0 to 3.95 s during the 3-min audio record (**Table 2**). The vocalization ratio (vocalization time during the 3 min) ranged from 0 to 65%. Indeed, two mothers did not speak during the 3-min recording.

Regarding high-level audio analysis (qualitative affective speech analysis), two complementary methods were performed: a qualitative manual annotation of maternal vocalization to the fetus by two expert psycholinguists and an automatic classification. In both clinical expert and automatized classifications, we found that pregnant women when speaking to their fetus (FDS) used sometimes a specific prosody that usually characterized motherese (or emotional IDS), which we called emotional fetal directed speech (e-FDS). We called the FDS without motherese characteristics "non-e-FDS." The automatic classification yielded a mean e-FDS ratio during the 3 min of 0.12, whereas the expert classification found a mean e-FDS ratio during the 3 min of 0.14 (**Table 2**). **Figure 2** (left) shows the strong and significant correlation between expert and automatic classification on e-FDS recognition. We also calculated the intraclass correlation (ICC) between the "two" raters (the expert

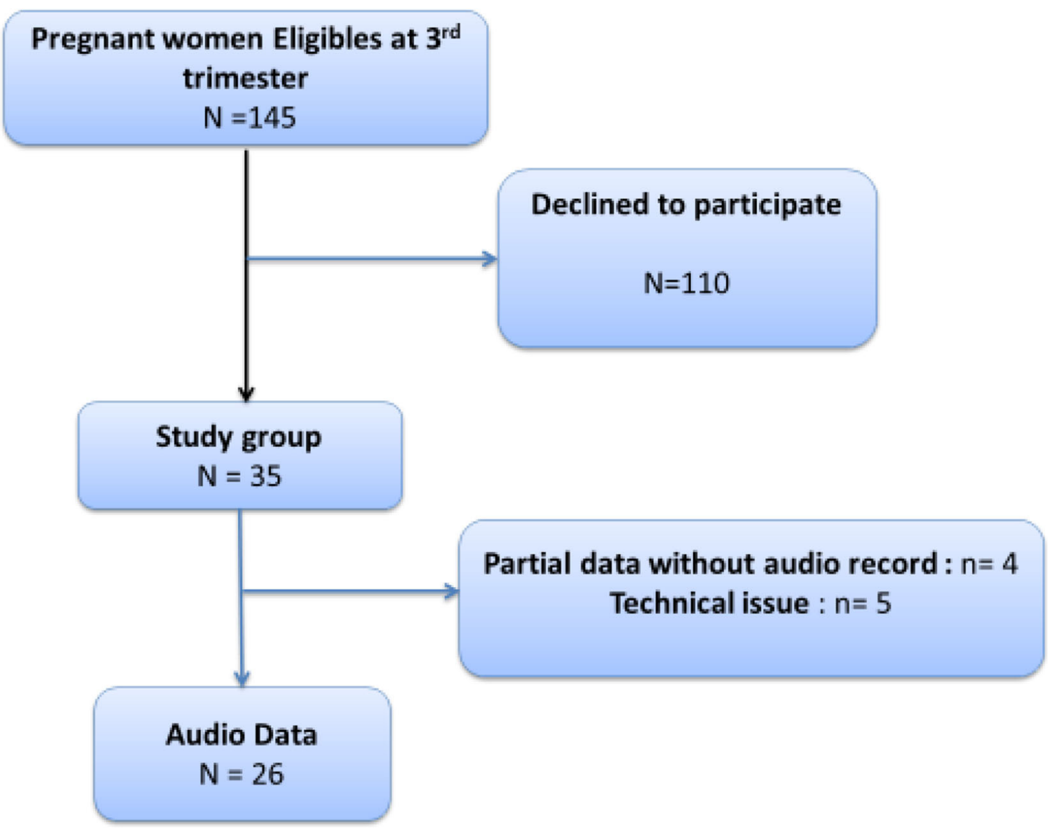



**FIGURE 1 |** Flow chart of recruitment.

and the algorithm) and found a good and very significant ICC (ICC = 0.79 (95% CI: 0.59–0.90), $p < 0.001$).

## Correlation of Maternal Audio Data With Maternofetal Characteristics and Anxiety Depression Status (*N* = 26)

Given the limited sample size, we used only exploratory univariate analysis to address whether some stress or psychopathological variables could influence the ability to produce e-FDS. We found no association between speaking to the fetus (whether prosody had characteristics of e-FDS or not) and being a fetus at risk during pregnancy (correlation ratio = 0.36 (±0.2) and 0.33 (±0.2), respectively, $t$-test, $p = 0.69$).

**TABLE 2 |** Maternal vocalization characteristics during the experiment ($N = 26$).

| | |
|---|---|
| Vocalization mean duration: mean (SD) [range] | 1.89 (0.92) [0–3.95] |
| Vocalization number during the 3 min window: mean (SD) [range] | 32.92 (16.66) [0–59] |
| Maternal Pause mean duration: mean (SD) [range] | 4.69 (6.3) [0–28.38] |
| Vocalization ratio of time during the 3 min: mean (SD) [range] | 0.35 (0.19) [0–0.65] |
| ***Emotional-Fetal Directed Speech according to automatic classification*** | |
| e-FDS ratio during the 3 min: mean (SD) [range] | 0.12 (0.17) [0–0.65] |
| Non e-FDS ratio during the 3 min: mean (SD) [range] | 0.23 (0.18) [0–0.64] |
| e-FDS/vocalization ratio: mean (SD) [range] | 0.29 (0.3) [0–1] |
| ***Emotional-Fetal Directed Speech according to psycholinguist expert*** | |
| e-FDS ratio during the 3 min: mean (SD) [range] | 0.14 (0.16) [0–0.51] |
| Non e-FDS ratio during the 3 min: mean (SD) [range] | 0.21 (0.16) [0–0.62] |
| e-FDS/vocalization ratio: mean (SD) [range] | 0.34 (0.34) [0–1] |

**Figure 3** shows the correlation between the speaking-to-fetus ratio and the Covy anxiety score, Raskin depression score and number of maternal stressful events. As shown, we found no correlation with the Covy anxiety score, a tendential negative correlation with the number of maternal stressful events, and a significant negative correlation with the Raskin depression score ($\rho = -0.4$, $p = 0.046$), meaning that the more the future mothers were depressed during pregnancy, the less they spoke to their fetuses during the experiment.

We performed the same analyses using only the e-FDS/vocalization ratio. None of the variables modulated the e-FDS/vocalization ratio, meaning that the impact of the Raskin depression score was on speaking to the fetus as a whole whether the pregnant women had e-FDS prosody or not. However, there was a trend toward a negative correlation between the e-FDS/vocalization ratio and the number of maternal stressful events (rho = −0.421, $p = 0.072$).

# DISCUSSION

## Can We Define Emotional Fetal Directed Speech (e-FDS)?

To answer this question, we proposed to address two different issues: (1) Is the mother speaking to the fetus during the experiment truly oriented toward the fetus? And (2) does FDS include some sequences that share the same prosodic characteristics of postnatal IDS (motherese)?

To address this first question (Is the mother speaking to the fetus during the experiment truly oriented toward the fetus?), we explored whether the pregnant women reported spontaneous moments of speaking and vocalization with their infant to come.

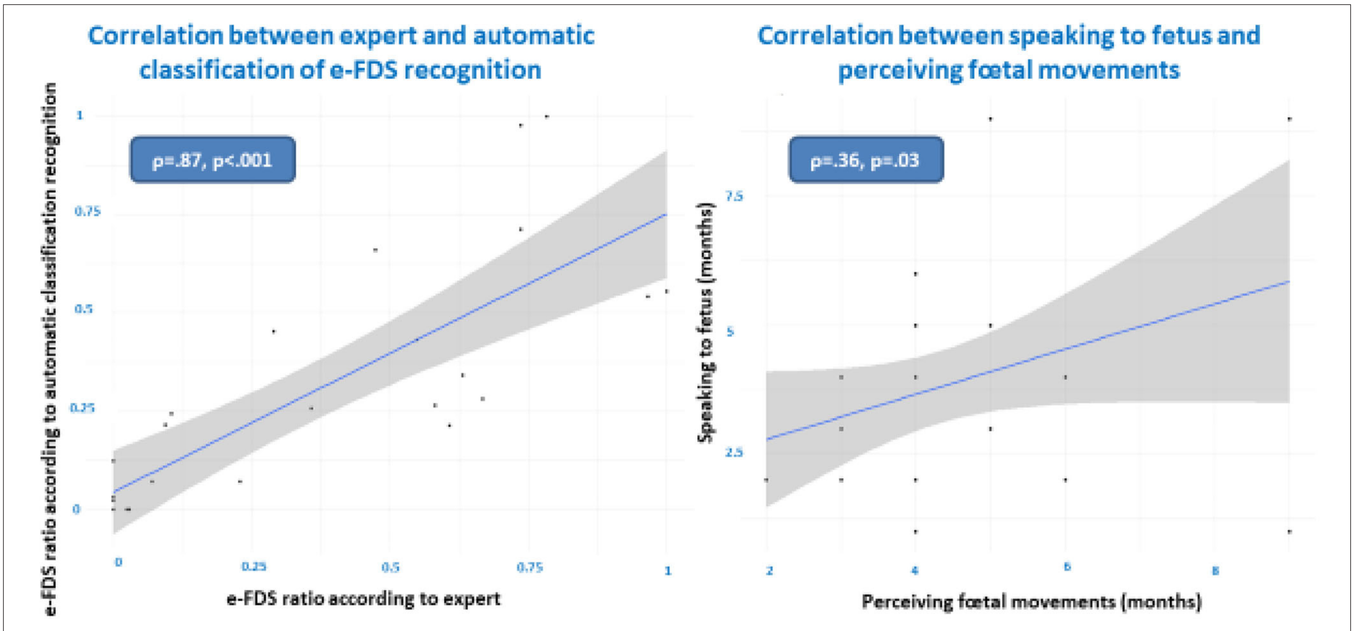



**FIGURE 2 |** Correlations between expert and automatic classification of emotional fetal directed speech (e-FDS, left) and between speaking to fetus and perceiving fetal movements (right).

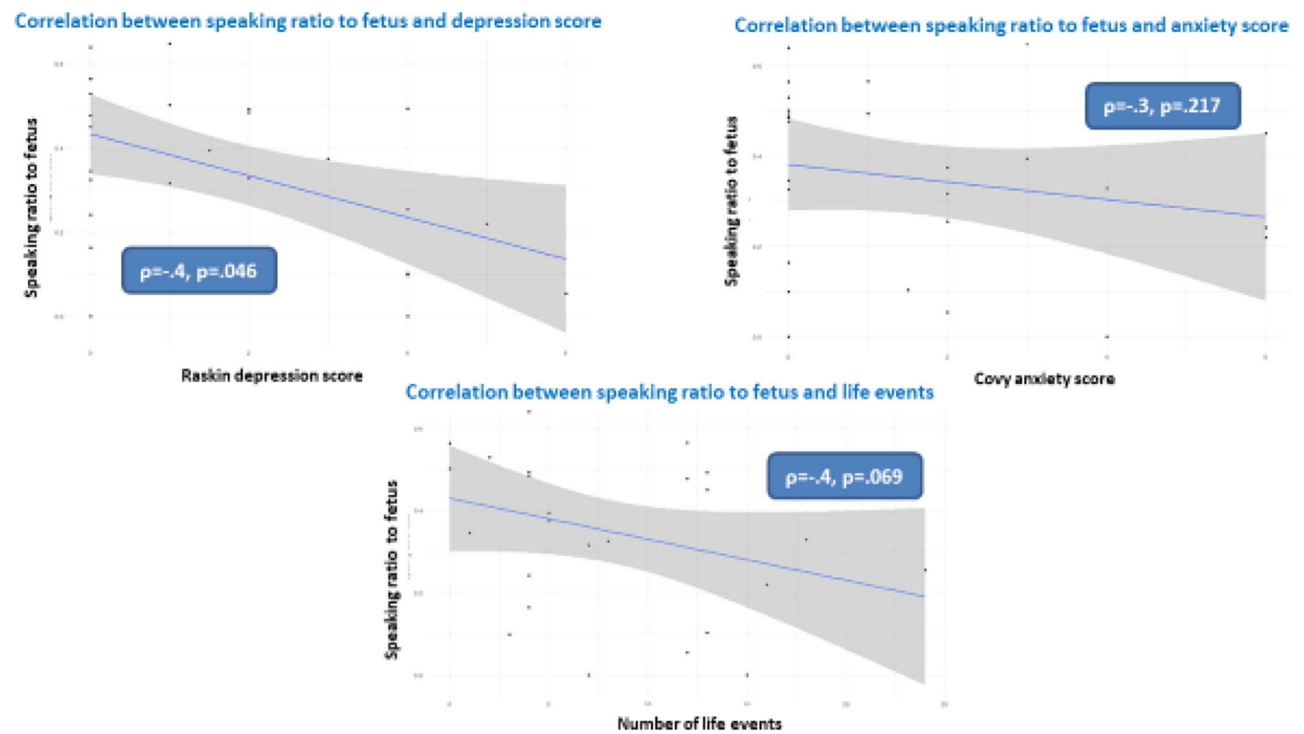



**FIGURE 3 |** Correlations between speaking to fetus and Raskin depression score (upper left), Covy anxiety ratio (upper right) and number of mother stressful life events.

For the mothers who reported doing so ($n = 26$), mothers started speaking or vocalizing with their fetus on average at 3.63 months during pregnancy (**Table 1**). Additionally, they started perceiving fetal movement on average at 2.83 months during pregnancy. This means that they could feel physically the existence of their fetus before they reported speaking to their fetus. Given the significant correlation between speaking to the fetus and perceiving fetal movement gestational ages (**Figure 2**, right), we can hypothesize that speaking to the fetus was indeed oriented toward the fetus. This result supports the hypothesis of a preliminary dialogue between future mothers and their fetus, as shown when mothers observed fetal movements during ultrasound scans that were interpreted by mothers as a response or solicitation from the fetus. Mirroring movements were seen as motor turn taking (Ammaniti et al., 2010). We believe that the current results on e-FDS are in the same vein and support the idea that prenatal development influences maternal infant attachment (Ammaniti et al., 2014; Feldman, 2016; Malm et al., 2016) and maternal representations of her future child (Viaux-Savelon et al., 2012, 2020).

Regarding the second question (do pregnant mothers sometimes use a motherese prosody (or here e-FDS) when speaking to their fetus?), our results show that futures mothers can use motherese prosody in their fetal-directed speech. The "manual" study of acoustic components of the voice takes a very long time and only allows the study of very short voice segments. The use of an automatic classifier allows extensive study of all vocalizations based on their acoustic characteristics and open perspectives for larger studies, and the machine learning classifier remains blind to the experiment or context. In this study, in addition to the expert "manual" categorization, the presence of e-FDS is confirmed by automatic measures that are strongly objective. Indeed, we found a strong and significant correlation between expert and automatic classification on e-FDS recognition ($\rho = 0.87$ $p < 0.01$) and a good and very significant ICC between expert and algorithm [ICC = 0.79 (95% CI: 0.59–0.90), $p < 0.001$]. This methodology of motherese detection using an algorithm has already shown robustness, as we have been able to distinguish motherese in early interaction with children with pathological outcome (Cohen et al., 2013), with both mothers and fathers (Weisman et al., 2016), and in five different languages (Parlato et al., 2020). Here, automatic annotation was useful to confirm that the prosody used during FDS shared the same characteristics of motherese. Manual and automatic labeling comparison realizes a validation of the two methods.

### Does Maternal Anxiety Depression Status Influence the Quality and Quantity of Fetal Directed Speech?

As expected, despite the limited sample size, the results show that the more the future mothers were depressed during pregnancy, the less they spoke to their fetuses during the experiment (**Figure 3**). We also found a tendency for a significant negative correlation between stressful life events and the speaking-to-fetus ratio ($p = 0.69$). These results are contingent with previous studies that have shown the impact of maternal prenatal states. Prenatal stress, particularly concerning prenatal diagnosis, increases the level of anxiety, disrupts the emotional investment of the parents toward the fetus (Watson et al., 2002; Petersen and Jahn, 2008; Kaasen et al., 2010) and disrupts parent-infant interactions after birth (Viaux-Savelon et al., 2012). In addition, pregnant women with depressive and anxiety symptoms talk and sing less to their fetuses (Hernandez-Reif et al., 2018).

Regarding fetal-directed speech *quality* (e-FDS or fetus-directed motherese), we found only a trend toward a negative correlation between the e-FDS/speaking to fetus ratio and the number of maternal stressful events. Thus, a high number of stressful events may reduce mothers' affective involvement with their future infant. We know that depressed mothers of young infants are not only less likely to speak with them (Herrera et al., 2004) but also more likely to display a reduced prosody of motherese with them (Bettes, 1988; Kaplan et al., 2001). Moreover, even when depressed mothers produce motherese, their infants fail to learn in response to their own-mother infant directed speech, despite normal competence (Kaplan et al., 1999, 2002). In our study with fetuses, we found that depressed mothers speak less to their fetus, but we did not find a correlation between depression score and e-FDS ratio. However, we cannot exclude that motherese quality could be poorer and less able to prepare language acquisition. As suggested in a recent study (Bartha-Doering et al., 2019), neural discrimination of speech could begin *in utero*. So we could expect that depression during the end of pregnancy may have repercussions on the first steps of language acquisition. However, given the small size of our sample and the exploratory nature of the study, we cannot conclude, and further studies with larger samples would be helpful.

Anxiety and depression status and a high level of stressful life events influence at least the quantity of fetal directed speech. Therefore, the quantity of fetal directed speech may be a sign to consider when detecting depression during pregnancy. Indeed, supporting these mothers in their investment toward the fetus and the future infant is compulsory for the prevention of later psychopathology (Mazzeschi et al., 2015; Røhder et al., 2020).

Finally, we found no significant association between speaking to the fetus (whether prosody had characteristics of e-FDS or not) and having a fetus at risk during pregnancy. This was not our hypothesis. However, the mothers' and fetuses' medical and obstetrical history of our population is very heterogeneous in this small sample size, and all risks may not be similar. In addition, the gestational age of the stressful event could also influence the impact on maternal representations and involvement. Again, a larger study would be necessary to better explore these factors with a comparison group according to the type of stress factor (e.g., mother complication/fetal complication/others) (Viaux-Savelon et al., 2012, 2020; Pisoni et al., 2016; Cuijlits et al., 2019).

### Study Limitations

As noted above, our sample was scarce ($N = 26$) and did not permit us to draw conclusions regarding the effects of various complex factors, such as maternal anxiety depression status or fetus risk. As many women declined participation in the study, we must discuss whether future mothers who agreed to participate could be more susceptible to speaking to their fetus than future mothers who declined participation. This study is only exploratory and used an experimental context. We also need to confirm that speaking to fetus also occurs spontaneously in more ecological contexts (e.g., at home). This might be achievable with automatic recording using portable devices for example. Additionally, we did not perform multivariate models to explore how relevant variables are robustly correlated or not. Nevertheless, it is important to note that one mother who declared before the audio recording she did not usually speak to her fetus actually spoke a lot to her during the recording. This may suggest that speaking to her fetus may be a widespread phenomenon.

## CONCLUSION

Fetal directed speech (FDS) can be detected during pregnancy, and it contains a period of prosody that shares the same characteristics of motherese that can be described as prenatal motherese or emotional fetus-directed speech (e-FDS). This means that pregnant women start using motherese much earlier than expected. FDS seems to be correlated with maternal first perceptions of fetal movements and depression scores. Although this study was exploratory, our results show that the more future mothers were depressed, the less they spoke to their fetuses during pregnancy. Therefore, the quantity of fetal directed speech may represent a useful sign for clinicians to detect prenatal depression and maternal involvement during pregnancy. However, more research (e.g., larger sample; prospective design with several timeline measures during pregnancy) is needed to confirm these exploratory results. Automatic audio detection and social signal processing should enable larger studies that explore prenatal emotional involvement with future infants.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the local Ethical Committee (CPPIDF6) under the number n°09012014. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

DC, SV-S, MD, and MC designed the study. CF, MD, EP-O, IP, and SV-S recruited the participants and assessed both obstetrical and psychological data. EP-O, SV-S, and CF performed the experiments. EP-O and IP assessed motherese prosody. CS-G, DC, and MC performed the automatic signal processing. DC and HP performed the statistical analysis. EP-O, DC, MD, and SV-S wrote the first draft of the manuscript. All authors contributed to the final version of the manuscript.

## FUNDING

This study was supported by the CAPES fund of Ministry of Education of Brazil and the Centre d'Activité et de Recherche en Psychiatrie Infanto-Juvénile (CARPIJ).

## REFERENCES

Ammaniti, A., Mazzoni, S., and Menozzi, F. (2010). Ecographia in gravidanza: studio della co-genitorialità/ultrasound during pregnancy: a study on co-parenting. *Infanz. Adolesc.* 9, 151–157. doi: 10.1710/535.6404

Ammaniti, M., Tambellini, R., and Perucchini, P. (2000). De la grossesse à la période post-accouchement: instabilité et évolution des représentations maternelles. *Devenir* 12, 57–74.

Ammaniti, M., Trentini, C., Menozzi, F., and Tambelli, R. (2014). "Transition to parenthood: studies of intersubjectivity in mothers and fathers," in *Early Parenting and Prevention of Disorder: Psychoanalytic Research at Interdisciplinary Frontiers*, eds R. Emde and M. L. Bohleber (London: Routledge: Karnac Books), 129–164.

Bartha-Doering, L., Alexopoulos, J., Giordano, V., Stelzer, L., Kainz, T., Benavides-Varela, S., et al. (2019). Absence of neural speech discrimination in preterm infants at term-equivalent age. *Dev. Cogn. Neurosi.* 39:100679. doi: 10.1016/j.dcn.2019.100679

Bettes, B. A. (1988). Maternal depression and motherese: temporal and intonational features. *Child Dev.* 59, 1089–1096. doi: 10.2307/1130275

Bourvis, N., Singer, M., Saint Georges, C., Bodeau, N., Chetouani, M., Cohen, D., et al. (2018). Pre-linguistic infants employ complex communicative loops to engage mothers in social exchanges and repair interaction ruptures. *R. Soc. Open Sci.* 5:e170274. doi: 10.1098/rsos.170274

Chetouani, M., Mahdhaoui, A., and Ringeval, F. (2014). Time-scale feature extractions for emotional speech characterization. *Cogn. Comput.* 1, 194–201. doi: 10.1007/s12559-009-9016-9

Cohen, D., Cassel, R., Saint-Georges, C., Mahdhaoui, A., Laznik, M. C., Apicella, F., et al. (2013). Do parentese prosody and fathers' commitment facilitate social interaction in infants who later develop autism? *PLoS ONE* 8:e61402. doi: 10.1371/journal.pone.0061402

Cooper, R. P. (1993). The effect of prosody on young infants' speech perception. *Adv. Infancy Res.* 8, 137–167.

Cooper, R. P., and Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Dev.* 61, 1584–1595. doi: 10.2307/1130766

Covi, L. (1986). New concepts and treatments for anxiety. *Md. Med. J.* 35:821.

Cuijlits, I., van de Wetering, A. P., Endendijk, J. J., van Baar, A. L., Potharst, E. S., and Pop, V. J. M. (2019). Risk and protective factors for pre- and postnatal bonding. *Infant Ment. Health J.* 40, 768–785. doi: 10.1002/imhj.21811

DeCasper, A. J., Lecanuet, J. P., Busnel, M. C., Granier-Deferre, C., and Maugeais, R. (1994). Fetal reactions to recurrent maternal speech. *Infant Behav. Dev.* 17:159164. doi: 10.1016/0163-6383(94)90051-5

Dupoux, E., and Mehler, J. (1990). *Naître Humain*. Paris: Odile Jacob.

Evers, S., Doran, L., and Schellenberg, K. (1998). Influences on breastfeeding rates in low income communities in Ontario. *Can. J. Public Health* 89, 203–207. doi: 10.1007/BF03404475

Feldman, R. (2016). The neurobiology of mammalian parenting and the biosocial context of human caregiving. *Horm. Behav.* 77, 3–17. doi: 10.1016/j.yhbeh.2015.10.001

Fernald, A., and Kuhl, P. K. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behav. Dev.* 10, 279–293. doi: 10.1016/0163-6383(87)90017-8

Fernald, A., and Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Dev. Psychol.* 20, 104–113. doi: 10.1037/0012-1649.20.1.104

Figueiredo, B., Dias, C., Brandão, S., Canário, C., and Nunes-Costa, R. (2013). Breastfeeding and postpartum depression: state of the art review. *J. Pediatr.* 89, 332–338. doi: 10.1016/j.jped.2012.12.002

Fisher, C., and Tokura, H. (1995). The given-new contract in speech to infants. *J. Mem. Lang.* 34, 287–310. doi: 10.1006/jmla.1995.1013

Hernandez-Reif, M., Kendrick, A., and Avery, D. M. (2018). Pregnant women with depressive and anxiety symptoms read, talk, and sing less to their fetuses. *J. Affect. Disord.* 15, 532–537. doi: 10.1016/j.jad.2017.12.108

Herrera, E., Reissland, N., and Shepherd, J. (2004). Maternal touch and maternal child-directed speech: effects of depressed mood in the postnatal period. *J. Affect. Disord.* 81, 29–39. doi: 10.1016/j.jad.2003.07.001

Kaasen, A., Helbig, A., Malt, U. F., Naes, T., Skari, H., and Haugen, G. (2010). Acute maternal social dysfunction, health perception, and psychological distress after ultrasonographic detection of a fetal structural anomaly. *BJOG* 117, 1127–1138. doi: 10.1111/j.1471-0528.2010.02622.x

Kaplan, P. S., Bachorowski, J.-A., Smoski, M. J., and Zinser, M. (2001). Role of clinical diagnosis and medication use in effects of maternal depression on infant-directed speech. *Infancy* 2, 537–548. doi: 10.1207/S15327078IN0204_08

Kaplan, P. S., Bachorowski, J.-A., and Zarlengo-Strouse, P. (1999). Child-directed speech produced by mothers with symptoms of depression fails to promote associative learning in 4-month-old infants. *Child Dev.* 70, 560–570. doi: 10.1111/1467-8624.00041

Kaplan, P. S., Bachorowski, J. A., Smoski, M. J., and Hudenko, W. J. (2002). Infants of depressed mothers, although competent learners, fail to learn in response to their own mothers' infant-directed speech. *Psychol. Sci.* 13, 268–271. doi: 10.1111/1467-9280.00449

Kersuzan, C., Gojard, S., Tichit, C., Thierry, X., Wagner, S., Nicklaus, S., et al. (2014). Prévalence de l'allaitement à la maternité selon les caractéristiques des parents et les conditions de l'accouchement. Résultats de l'Enquête Elfe maternité, France metropolitaine, 2011. *Bull. Epidémiol. Hebd.* 27, 440–449. Available online at: http://www.invs.sante.fr/neh/2014/27/2014_27_1.html

Mahdhaoui, A., Chetouani, M., Cassel, R. S., Saint-Georges, C., Parlato, E., Laznik, M. C., et al. (2011). Computerized home video detection for motherese may help to study impaired interaction between infants who become autistic and their parents. *Int. J. Methods Psychiatr. Res.* 20, e6–e18. doi: 10.1002/mpr.332

Malm, M. C., Hildingsson, I., Rubertsson, C., Rådestad, I., and Lindgren, H. (2016). Prenatal attachment and its association with foetal movement during pregnancy—a population based survey. *Women Birth* 29, 482–486. doi: 10.1016/j.wombi.2016.04.005

Mazzeschi, C., Pazzagli, C., Radi, G., Raspa, V., and Buratta, L. (2015). Antecedents of maternal parenting stress: the role of attachment style, prenatal attachment, and dyadic adjustment in first-time mothers. *Front. Psychol.* 6:1443. doi: 10.3389/fpsyg.2015.01443

Outters, V., Schreiner, M., Behne, T., and Mani, N. (2020). Maternal input and infants' response to infant-directed speech. *Infancy* 25, 478–499. doi: 10.31219/osf.io/mw9es

Parlato, E., Chetouani, M., Cadic, J. M., Viaux, S., Xavier, J., Ouss, L., et al. (2020). The emotional component of Infant directed-speech: a cross-cultural study using machine learning. *Neuropsychiatr. Enfance. Adolesc.* 68, 106–113. doi: 10.1016/j.neurenf.2019.10.004

Petersen, J., and Jahn, A. (2008). Suspicious findings in antenatal care and their implications from the mothers' perspective: a prospective study in Germany. *Birth* 35, 41–49. doi: 10.1111/j.1523-536X.2007.00210.x

Pisoni, C., Garofoli, F., Tzialla, C., Orcesi, S., Spinillo, A., Politi, P., et al. (2016). Complexity of parental prenatal attachment during pregnancy at risk for preterm delivery. *J. Matern. Fetal Neonatal Med.* 29, 771–776. doi: 10.3109/14767058.2015.1017813

Ramus, F. (1999). "La discrimination des langues par la prosodie: modélisation linguistique et études comportementales," in *De la caractérisation à l'identification des langues, Actes de la 1ère journé d'étude sur l'identification automatique des langues*, ed F. Pellegrino (Lyon: Editions de l'Institut des Sciences de l'Homme), 186–201.

Raskin, A., and Crook, T. (1976). Sensitivity of rating scales completed by psychiatrists, nurses and patients to antidepressant drug effects. *J. Psychiatr. Res.* 13, 31–41. doi: 10.1016/0022-3956(76)90007-8

Røhder, K., Væver, M. S., Aarestrup, A. K., Jacobsen, R. K., Smith-Nielsen, J., and Schiøtz, M. L. (2020). Maternal-fetal bonding among pregnant women at psychosocial risk: the roles of adult attachment style, prenatal parental reflective functioning, and depressive symptoms. *PLoS ONE* 15:e0239208. doi: 10.1371/journal.pone.0239208

Saint-Georges, C., Chetouani, M., Cassel, R., Apicella, F., Mahdhaoui, A., Muratori, F., et al. (2013). Motherese in interaction: at the cross-road of emotion and cognition? (a systematic review). *PLoS ONE* 8:e78103. doi: 10.1371/journal.pone.0078103

Saito, Y., Aoyama, S., Kondo, T., Fukumoto, R., Konishi, N., Nakamura, K., et al. (2007). Frontal cerebral blood flow change associated with infant-directed speech. *Arch. Dis. Child. Fetal Neonatal* 92, F113–F116. doi: 10.1136/adc.2006.097949

Sherer, K. R., (1986). Vocal affect expression: a review and a model for future research. *Psychol. Bull.* 99, 143–165. doi: 10.1037/0033-2909.99.2.143

Spinelli, M., Fasolo, M., and Mesman, J. (2017). Does prosody make the difference? a meta-analysis on relations between prosodic aspects of infant-directed speech and infant outcomes. *Dev. Rev.* 44, 1–18. doi: 10.1016/j.dr.2016.12.001

Tordjman, S., Zenasni, F., and Granier-Deferre, C. (2004). *Presentation and Validation of the Sensations During Pregnancy and Life Events Questionnaire*. Aix-en-Provence: International Society for Developmental Psychobiology.

Trainor, L. J., Austin, C. M., and Desjardins, R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychol. Sci.* 11, 188–195. doi: 10.1111/1467-9280.00240

Viaux-Savelon, S., Decherf, M., Bodeau, N., Ville, Y., Marey, Y., Cohen, C., et al. (2020). Foetal chromosomal microarray genetic screening for minor ultrasound anomalies affects maternal representations and emotional state: an exploratory study. *Devenir* 32, 105–177. doi: 10.3917/dev.202.0105

Viaux-Savelon, S., Dommergues, M., Rosenblum, O., Bodeau, N., Aidane, E., Philippon, O., et al. (2012). Soft markers detected at fetal scan induce perturbation of early mother-infant interactions. *PLoS ONE* 7:e30935. doi: 10.1371/journal.pone.0030935

Watson, M., Hall, S., Langford, K., and Marteau, T. (2002). Psychological impact of the detection of soft markers on routine ultrasound scanning: a pilot study investigating the modifying role of information. *Prenat. Diagn.* 22, 569–575. doi: 10.1002/pd.373

Weisman, O., Chetouani, M., Saint-Georges, C., Bourvis, N., Delaherche, E., Zagoory-Sharon, O., et al. (2016). Dynamics of non-verbal vocalizations and hormones during father-infant interaction. *IEEE Trans. Affect. Comput.* 7, 337–345. doi: 10.1109/TAFFC.2015.2478468

Williams, C. E., and Stevens, K. N. (1972). Emotions and speech: some acoustic correlates. *J. Acoust. Soc. Am.* 52, 1238–1250. doi: 10.1121/1.1913238

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 19 July 2021
doi: 10.3389/fpsyg.2021.688002

# Vocalic Intrusions in Consonant Clusters in Child-Directed vs. Adult-Directed Speech

*Nina Gram Garmann*[1,2*†], *Pernille Hansen*[2,3†], *Hanne Gram Simonsen*[2†], *Elisabeth Holm*[1], *Eirik Tengesdal*[4], *Brechtje Post*[5] *and Elinor Payne*[6]

*[1] Department of Early Childhood Education, OsloMet – Oslo Metropolitan University, Oslo, Norway, [2] MultiLing – Center for Multilingualism in Society Across the Lifespan, Department of Linguistics and Scandinavian Studies, University of Oslo, Oslo, Norway, [3] Department of Humanities, Inland Norway University of Applied Sciences, Hamar, Norway, [4] Department of Linguistics and Scandinavian Studies, University of Oslo, Oslo, Norway, [5] Faculty of Modern and Medieval Languages and Linguistics, University of Cambridge, Cambridge, United Kingdom, [6] Faculty of Linguistics, Philology and Phonetics, University of Oxford, Oxford, United Kingdom*

OPEN ACCESS

**Edited by:**
*Adrian Garcia-Sierra, University of Connecticut, United States*

**Reviewed by:**
*Yang Zhang, University of Minnesota Twin Cities, United States*
*Lama K. Farran, University of West Georgia, United States*

***Correspondence:**
*Nina Gram Garmann*
*nina.garmann@oslomet.no*

*† These authors have contributed equally to this work and share first authorship*

**Specialty section:**
*This article was submitted to Developmental Psychology, a section of the journal Frontiers in Psychology*

**Received:** *30 March 2021*
**Accepted:** *15 June 2021*
**Published:** *19 July 2021*

**Citation:**
*Garmann NG, Hansen P, Simonsen HG, Holm E, Tengesdal E, Post B and Payne E (2021) Vocalic Intrusions in Consonant Clusters in Child-Directed vs. Adult-Directed Speech. Front. Psychol. 12:688002. doi: 10.3389/fpsyg.2021.688002*

In this paper, we investigate a prosodic-phonetic feature in child-directed speech within a dynamic, complex, interactive theoretical framework. We focus on vocalic intrusions, commonly occurring in Norwegian word initial consonant clusters. We analysed child-directed speech from nine Norwegian-speaking mothers to their children, aged 2;6, 4, and 6 years, and compared the incidence and duration of vocalic intrusions in initial consonant clusters in these data with those in adult-directed speech and child speech. When viewed overall, vocalic intrusion was found to be similar in incidence in child- and adult-directed speech. However, closer examination revealed differential behaviour in child-directed speech for certain conditions. Firstly, a difference emerged for one particular phonetic context: While vocalic intrusions in /Cr/ clusters are *frequent* in adult-directed speech, their presence is *near-categorical* in child-directed speech. Secondly, we found that the duration of vocalic intrusions was longer in child- than in adult-directed speech, but only when directed to 2;6-year-olds. We argue that vocalic intrusions in child-directed speech may have both a bonding as well as a didactic function, and that these may vary according to the age of the child being addressed.

**Keywords: child-directed speech (CDS), consonant clusters, language acquisition, Norwegian, prosodic-phonetic biases, vocalic intrusions**

## INTRODUCTION

In infant- and child-directed speech (IDS and CDS), adults are known to adjust their speech in various ways. For example, IDS and CDS have shorter and less complex sentences (Snow, 1972) with fewer false starts and hesitations. Adults repeat their utterances more to 2-year-olds than to 10-year-olds (Snow, 1972), place key words at the end of an utterance, or sometimes in isolation, and produce them with more emphatic stress (Aslin et al., 1996). On a phonological level, IDS and CDS are reported to be generally more exaggerated in their intonation with a higher pitch and wider pitch range, and are slower in tempo (Cruttenden, 1994).

In a recent review article, Wang et al. (2018) discuss acoustic properties of IDS, that is, speech to children younger than 24 months. They report that a large body of literature shows that prosodic modifications such as higher pitch, larger pitch variability, slower tempo and longer vowel duration are attested in IDS when compared to ADS across a wide range of languages (Cristià, 2013). Concerning segmental properties, they report fewer studies and the findings are more mixed, possibly due to differences between languages. For instance, Kuhl et al. (1997) have reported a more expanded vowel space in IDS than in ADS (for American English, Russian and Swedish), indicating hyperarticulation, while others have found a reduced vowel space (Benders, 2013, for Dutch and Englund and Behne, 2006, for Norwegian), suggestive of hypoarticulation. For the Norwegian vowels /æ(ː), ø(ː), o(ː), y(ː), ʉ(ː), e(ː)/, Englund (2018) also found evidence for hypoarticulation in IDS compared to ADS, for example with more front articulation and less lip protrusion in IDS; the lack of rounding was possibly attributable to mothers smiling to their infants when talking to them. For consonants, VOT values have been found to increase (Englund, 2005, for Norwegian), to be maintained (Baran et al., 1977, for English) or to decrease (Sundberg and Lacerda, 1999, for Swedish) in IDS compared to ADS. Some studies have found consonants to be more clearly articulated in IDS than in ADS (Cristià, 2010; Dilley et al., 2014, for English), while the opposite has also been found (Martin et al., 2015, for Japanese).

Wang et al. (2018) lay out how IDS changes over time as the child develops and note that parents adjust both to the child's chronological age, and also, in the case of cochlear implantation in children with hearing loss, to their peers matched in hearing experience. They conclude therefore that parents may modify their speech to children to adapt to the latter's needs. When children are still young infants, prosodic exaggeration may be more important, while later on, it may be that other linguistic information, like segmental information, is of higher value to the child. Some characteristics of IDS disappear already during the child's first year, while other characteristics may persist over a longer time span. Rattanasone et al. (2013) showed that IDS-specific tonal characteristics of Cantonese-speaking mothers' speech to their infants had already disappeared at 12 months, whereas for example Stern et al. (1983) showed that the tonal characteristics of IDS in American English diminished over time, but that there were still differences between CDS and ADS at 24 months.

The findings of Wang et al. (2018) relate to IDS, that is, to speech addressed to infants in the earliest stages of their development. Although the literature is sparser on CDS than IDS, there is evidence that adults speak differently to children even when they are older than 24 months. In a study comparing mothers' speech to children aged 2, 4, and 6 in Catalan, English and Spanish to speech addressed to adults, Payne et al. (2010) found that speech was both proportionally more vocalic and containing more even-timed vocalic intervals in CDS than in ADS, characteristics that were shown to mirror those of the children's own speech. They also reported that these characteristics in the CDS did not change across the child age-span covered. Poulain and Brauer (2018) examined different aspects of child-directed communication to German children aged between 2 and 6 years: mean length of utterance, pointing, and variability of pitch. They found that the mothers adapted their behaviour to the advancing abilities of their children. As to prosody, variability of pitch decreased with age: there was a significant difference between ADS and CDS at 2, 3, and 5 years, but not at 6 years, so this phonetic adaptation disappeared sometime between 5 and 6 years. Comparing mothers' and fathers' speech to 2- and 5-year-olds with ADS, Warren-Leubecker and Bohannon (1984) found that mothers adopted a higher pitch when speaking to both 2-year-olds and 5-year-olds, but that they had a wider pitch range when speaking to the younger children. The fathers raised their pitch and increased their pitch range to the 2-year-olds, but did not adjust their pitch when speaking to 5-year-olds compared to when speaking to adults. This shows that mothers, at least, speak differently to their children even when the latter are older than 24 months, and also that there are changes in CDS as the child grows older.

In addition to investigating whether and how adults modify their speech when addressing infants and children, and whether and how this varies as a function of child age, research in this area has sought to identify the function(s) of CDS. Building on a large body of research, Wang et al. (2018, p. 19) conclude that IDS may have three possible functions: "to maintain infants' attention, to communicate affect, and to be didactic." According to Wang et al. (2018), there is evidence in favour of the hypothesis on adults using IDS because children are attracted to 'happy' speech that communicates affect. They claim that there is less evidence for a didactic function to the IDS register. However, there may be a difference in the function of IDS and CDS registers, as the characteristics of the dyadic relationship and communicative priorities shift. Fernald and Mazzie (1991) suggest that while IDS may have the function of drawing the child's attention, the function of CDS may be more of a didactic one where more distinct speech may help the child to segment individual words from the stream of fluent speech. Even though children start the process of segmenting individual words from the speech stream earlier than from 24 months, this process may also be relevant later, and the distinct speech may serve to enhance morpho-syntactic and phonological characteristics of the words. Identifying possible functions is not a straightforward task, however, and one might be able to attribute various possible functions to a single characteristic. For example, the repetition of words and utterances in CDS may have a didactic function (e.g., to facilitate the learning of lexical items or word shape) as well as being a way of keeping the conversation running and linking it to the child's interest, thus combining attention-seeking, the communication of affect and a didactic function.

Independently of the underlying function of these modifications, the evidence clearly shows that adults change the way they speak when addressing infants and young children as compared to when speaking to adults. As we have seen, these speech accommodations occur on different levels, that is, segmental phonetic, prosodic, or higher-level linguistic, and potentially with different degrees of speaker awareness. Certain properties of CDS may be characterised as relatively 'local,' for example the placement of greater emphasis on a given word, while others may be more pervasive throughout a stretch of speech, such as slower speech tempo. Some that may *appear* to be relatively local, for example a segmental difference in vowel quality, may actually arise from a more general effect of hyperarticulation, which results in an expanded vowel space and, in turn, in changed vowel qualities for individual vowels. Certain characteristics of CDS may be relatively language-independent, particularly those that are less didactic in purpose and more attributable to general, (quasi-)universal strategies of attracting and maintaining attention which we might plausibly assume to have adapted toward general properties of the (developing) human perceptual system. Other strategies may more deliberately draw attention to structural features that are specific to a given language, for example the exaggeration of geminate duration, thereby emphasising a lexical contrast, and we might plausibly attribute a more didactic motivation to these.

Any variation in the input, as characterised by modifications apparent in CDS, may potentially shape the child speech acquisition process, by shifting the distribution of patterns to which the child is exposed. Nevertheless, this is not taken to be a passive process. Vihman and Velleman argue that while patterns in the ambient language shape the acquisition process, the process is also influenced by the child's phonetic skills and her own emerging phonological patterns: 'the onset of phonological systematisation is superimposed upon ongoing phonetic learning' (2000: 265, see also Vihman, 2017). This process of interaction between salient properties in the speech the child hears on the one hand, and the child's own phonetic abilities and emergent phonological system on the other, may result in apparent discontinuities in development, with phonological structure arising from phonetic patterns in a way that is not only gradual. While a child's first words may result from her matching her own productions to what she hears, (the 'articulatory filter,' Vihman, 1993), this is not purely a mechanical process; instead, 'certain phonetic structures are exploited and generalised,' in the formation of nascent phonological systems, or 'templates,' which may lead to non-adult-like adaptations. Thus, while guided by the child's own phonetic development and the phonetic patterns in the input speech she is exposed to, the nascent phonological system exerts, in turn, its own pull on the child's speech productions. As a result, any influence of particular CDS characteristics will itself be mediated through this self-organising process (cf. Davis and Bedore, 2013). The child does not simply mirror what the adult does. There is, instead, a dynamic interaction between the adult's speech – with its own structures and patterns that may be modified in addressing the child – and the child's emerging structures and speech patterns. For Davis and Bedore (2013), the acquisition process also involves a dynamic interaction between the input (speech input from communication partners as well as their extrinsic critical guidance, cf. the didactic purpose discussed above) and the child's developing intrinsic biological and cognitive abilities, but their model places a greater emphasis on functional pressures arising from the child's need to connect and communicate, and hence, they identify the child's growing capacity for social interaction as a third key factor in child phonological development.

As well as being a dynamic process, which we imagine may shift according to discourse and wider context and as the child develops, this interaction is also complex, in that the structures in both the adult's speech and the child's speech are, inevitably, implemented phonetically, adding further scope for divergence. Languages, and language varieties, vary not only in their phonological structure, but also in how that structure is implemented. Thus, aside from features that are more clearly either phonological (e.g., pertaining to a lexical contrast or phonotactics) or general phonetic (e.g., pertaining to general articulatory skills), language-specific *linguistic-phonetic* features are also a body of knowledge to be acquired. These are phonetic (i.e., non-contrastive) aspects of a given language, or variety of language, that a child must master in order to be a native or near-native speaker, for example cross-linguistic variation in whether [s] is produced as laminal or apical, or in VOT for the cueing of voicing contrasts. Such linguistic-phonetic features may also pertain to temporal coordination of gestures and their association with prosody, and thus to the production of connected speech, in what can be characterised as prosodic-phonetic tendencies (or *biases*; see Payne, 2016, for a fuller discussion). Thus, prosodic-phonetic biases are pervasive, systematic and language-specific phonetic patterns in the implementation of phonological structure. Evident in adult-directed speech, they may be more or less salient to the listener, including the infant listener. As such, we can consider them as the language-specific phonetic 'packaging' which conveys phonological structure, and thus acquisition of that structure for a particular language is mediated via these patterns.

One example of a prosodic-phonetic bias in Norwegian are vocalic intrusions in the production of consonant clusters, the characteristics of which in ADS we briefly describe here. In contrast to most variants of English, Norwegian has a so-called open transition between consonants in a cluster, where the first consonant is released before the onset of the next one (Endresen, 1991, p.127, see also Bradley, 2007). This often leads to vocalic intrusions between a sequence of consonants in a cluster in adult speech (e.g., [ˈbᵊɭoː] for [ˈbɭoː] *blå*, 'blue'). In a recent study, we found vocalic intrusions in 30.6% of instances for (Urban East) Norwegian, while for (Southern British) English, vocalic intrusions were nearly non-existent (affecting only 0.6% of instances) (Garmann et al., 2021). In the same study, vocalic intrusions were found to be more common in ADS when C2 is a liquid, with the greatest incidence occurring when C2 is a rhotic tap or flap, and a particularly high incidence when this was combined with a voiced stop in C1 position. We observed that 'the incidence of vocalic intrusion in Norwegian is gradient and clearly shaped at least in part by articulatory considerations.' (2021, p. 22). Hence, while not obligatory in any phonetic context, vocalic intrusions are nevertheless common in Norwegian, and neither their incidence nor duration are dependent on speech rate in ADS.

It is important to reflect here on what this means for diverging inputs into the child speech acquisition process. In terms of phonological structures, the evidence available in the input may be very similar (and indeed the phonotactics of English and Norwegian are quite similar). Thus, if the English-ambient child and the Norwegian-ambient child are exposed to similar input – at least with respect to this particular variable – their nascent phonological systems, and thus their own productions should also be fairly similar (at least no more dissimilar than between two children of the same ambient language). However, if we consider not just the phonological structures but also *how they are implemented*, we can model actual divergence in inputs, and make different predictions about the children's early productions. Indeed, in the same study (Garmann et al., 2021), we found that, while vocalic intrusions were also evident in English child speech [as one possible strategy for tackling clusters, and one that is attested to some degree cross-linguistically (McLeod et al., 2001)], they were far less prevalent than in Norwegian child speech. This could indicate that in selecting a strategy for tackling clusters infants are influenced by distributional patterns in their ambient speech input. In other words, young children are sensitive to speech patterns of different degrees of granularity, and these include not just which segments can be juxtaposed in connected speech, and where, but crucially also *how* they are juxtaposed, that is, the fine detail of intersegmental coordination. Furthermore, these provide another type of evidence from which their own nascent systems are forged. Indeed, these vocalic intrusions were found to be even more frequent and longer in duration in Norwegian child speech than in Norwegian ADS, with some generalisation to phonetic contexts for which there was no incidence in ADS, for example /sC/ clusters. Together these pieces of evidence strongly suggest that vocalic intrusions, arising from a particular setting of temporal and articulatory coordination, are a pervasive *bias* (a strong but non-obligatory tendency) in the production of Norwegian consonant clusters, and one that shapes the acquisition pathway of Norwegian-ambient infants.

The fact that children generalise vocalic intrusion to other phonetic contexts suggests that *linguistic-phonetic knowledge at the implementation level* is part and parcel of the child's nascent phonological system. Note that we are not suggesting that the child is interpreting the vocalic intrusion as having the status of a phonological segment – something which is theoretically possible but for which there is insufficient evidence. Rather, we are proposing that knowledge about the *implementation* of phonological structure should be seen as part of *knowledge about phonological structure*. And thus, a child's nascent phonological system will also include knowledge of how that system is implemented, and just as the system itself may diverge from the adult system, so may (language-specific) properties relating to its implementation. In part this also depends on the distributional properties of the input, and indeed raises the question as to whether these biases may be subject to modification in CDS. If they are, we may also ask what form this modification might take, and to what extent such modification may be interpreted as *deliberate*, or *incidental*, as the unplanned consequence of slower speech tempo, for example. Here, we investigate the role that CDS may have in mediating the prosodic-phonetic bias of vocalic intrusion in Norwegian in the child acquisition process. In our earlier study, we found that child speech displays more vocalic intrusion in consonant clusters than does ADS (Garmann et al., 2021), which suggests vocalic intrusion may be more prevalent in CDS than in ADS. This could come about quite incidentally: since vocalic intrusions may be influenced by the slower tempo and exaggerated prosody that characterise CDS, we hypothesise that CDS has longer and more frequent vocalic intrusions than ADS. Even though we did not find any influence of speech rate on the incidence and duration of vocalic intrusions in ADS, slower speech rate is a known characteristic of CDS (Cruttenden, 1994) and could thus be a determining factor in the incidence of vocalic intrusions in CDS. In this scenario, a greater incidence of vocalic intrusions would simply emerge from other CDS behaviours. If based on these general properties of CDS, there would be no reason to expect any differences in the phonetic contexts in which the vocalic intrusions occur; on the contrary, we would expect the pattern of vocalic intrusions to be very similar to that in ADS, only with longer durations and potentially higher incidence. These would be the result of generalised CDS strategies which can be interpreted as having broader functions of increasing the infant's attention, increasing closeness in the dyadic relationship ('bonding') and a general facilitation of comprehension (i.e., not focussing on any structures in particular). This would constitute a kind of speech accommodation that acknowledges the child's different capacities and knowledge, and renders adult speech patterns more transparent (through slowing down and exaggeration).

Another scenario is that adults adjust the prevalence and/or distribution of vocalic intrusions when addressing children. Such changes would be difficult to attribute solely to broader, more generalised CDS strategies such as tempo, and would point to a more focussed and localised strategy (however, conscious or not) of speech accommodation. The potential reasons for doing so are multiple, and would likely affect to some degree the nature of the adjustments being made. One possibility is that adults accommodate toward the child's own speech patterns. CDS would under this scenario show patterns of vocalic intrusion that more closely mirror those evident in child speech, for example through a greater incidence of intrusions and/or incidence in phonetic contexts in which intrusions are detected in child speech but not in ADS. Similarly with the previous scenario, this would also constitute a kind of speech accommodation that acknowledges the child's different capacities and knowledge, but rather than rendering adult speech patterns more transparent, chooses to close the communicative gap by adjusting speech patterns *toward* those of the child. We would attribute this kind of accommodation to a desire to increase the closeness of the dyadic relationship ('bonding') and potentially to facilitate comprehension. A longer duration of intrusions could also fall under this scenario, since as well as an attention-calling device, it can also be seen as a mirroring of the child's own productions.

Another possibility is that adults hyperarticulate consonant clusters to clarify the elements of the cluster in what is potentially an instructive, or didactic, way. Evidence of this might include paying close, exaggerated attention to phonetically difficult articulations, for example the pronunciation of [ɾ] (which is challenging in many languages, Bernhardt and Stemberger, 2018). They may also make their ADS cluster patterns more categorical in CDS to show the children how a particular cluster is typically produced and reduce ambiguity through making segment boundaries clearer.

Thus, there are various kinds of adjustments that one might expect in CDS, motivated potentially by different functions. These might affect incidence within and across different

**TABLE 1 |** Plausible underlying motivations for vocalic intrusions in CDS (compared to ADS) and their potential phonetic realisation.

| Intention | Function | Potential evidence | Explanatory notes |
|---|---|---|---|
| Attracting attention | Bonding/Didactic (communicative) | Higher incidence, longer intrusions | Increase in acoustic salience |
| Expressing affect | Bonding | Higher incidence, longer intrusions | Increase in acoustic salience |
| Emphasising | Didactic (linguistic) | Higher incidence, longer intrusions | Selective increase in acoustic salience to facilitate word learning or comprehension of phrase structure |
| Instructing: mastering clusters | Didactic (phonetic) | Higher and/or more systematic incidence, longer intrusions | Exaggerating properties of ADS. Facilitates articulation of particular clusters (e.g., with rhotics) |
| Instructing: sounding 'Norwegian' | Didactic (linguistic-phonetic) | Same incidence, longer intrusions | Patterning with, or exaggerating, ADS to help acquire Norwegian-appropriate cluster transitions (prosodic-phonetic biases) |
| Mirroring child speech patterns | Bonding, facilitating comprehension | Higher incidence, longer intrusions, more phonetic contexts | Patterning more closely with CS, changes as children develop |

phonetic categories, and duration, differently, resulting in quite a complex picture. This is further complicated by the fact that mapping underlying motivations to observable behaviours may be ambiguous (i.e., a given behaviour, such as longer vocalic intrusions, may plausibly map onto more than one motivation) and the fact that underlying motivations are, of course, only speculative. Nevertheless, some plausible relationships can be posited. In **Table 1**, we set out a schematic overview of these for vocalic intrusions in Norwegian CDS, and suggest the potential phonetic evidence for these, as compared with ADS.

A strong prediction, across almost all kinds of possible motivating factors, is that vocalic intrusions will be of longer duration in CDS than in ADS. Both generalised properties of CDS, with universal communicative and bonding goals, and more didactically-orientated strategies are compatible with longer vocalic intrusions.

A slightly less strong prediction can be made about the incidence of vocalic intrusion in CDS when compared with ADS. Generalised CDS properties of slower tempo and exaggerated prosody would conspire to increase incidence, and this would be compatible with the general goals of attracting attention for increased affect and the conveyance of emotion, and for marking emphasis more transparently. It would also be compatible with the mirroring of child speech (for affect enhancement and facilitating comprehension), and for the phonetic didactic goal of instructing how to produce difficult clusters.

Finally, with respect to the distribution of vocalic intrusion in CDS, most potential motivating factors would predict little or no deviation from the distribution observed in ADS. There are two exceptions, however. Firstly, the goal of facilitating phonetic mastery of certain clusters, as well as highlighting the presence of the short [ɾ] as C2 in a cluster, that is, being didactic, could result in a higher incidence of intrusions in those specific phonetic contexts. Secondly, mirroring child speech patterns could result in a distribution pattern that deviates from ADS: Possible intrusions in CDS in phonetic contexts that are characteristic of child speech, but not of ADS, would be examples of adults mirroring the children in CDS, suggesting a bonding function, rather than a didactic one.

A further consideration is that the relationship between caregiver and child is dynamic, as are the communicative needs and abilities of the child. Thus tracing potential changes in CDS over a range of child ages can be enlightening. Knowing that prosodic characteristics of IDS and CDS appear to change over the age span of the infant/child being addressed, we expect vocalic intrusions to be less prominent in CDS directed toward older children than to younger ones.

With these considerations in mind, we analyse data from mothers speaking to 2;6-, 4- and 6-year-olds to investigate the following hypotheses:

(1) Vocalic intrusions have a higher incidence in CDS, at least when addressed to 2;6-year-olds, than in ADS.
(2) Vocalic intrusions have the same distribution with respect to phonetic contexts in CDS as in ADS.
(3) Vocalic intrusions are of longer duration in CDS, at least when addressed to 2;6-year-olds, than in ADS, but shorter than in child speech.
(4) Vocalic intrusions in CDS become less prevalent and of shorter duration as the children being addressed grow older.

## MATERIALS AND METHODS

### Data

Our data consist of speech recordings from nine mothers reading a story to their child (CDS), compared to four mothers reading sentences to a research assistant (ADS), and nine children playing a naming game with their mothers, with words from the story they had just heard (CS). The ADS and CS data have already been reported on by Garmann et al. (2021), and serve as a basis for comparison here. The CDS, analysed specifically for this paper, was elicited from three Norwegian mothers of 2;6-year-olds, three mothers of 4-year-olds and three mothers of 6-year-olds. All nine mothers as well as their children were native speakers of Urban East Norwegian (Kristoffersen, 2000) and lived in and around Oslo. There is an overlap between three of the subjects providing ADS and CDS data, and between seven of the mothers and their children.

Audio was recorded using a Zoom Handy Recorder H2 with built-in microphones. The mothers and their children took part in a larger study comparing phonetic microvariation in the production of consonants in both child and adult speech in English and Norwegian, reported on by Payne et al. (2015, 2017) and Garmann et al. (2021). The studies were reviewed and approved by NSD – Norwegian Centre for Research Data under the reference number 36466. All participants or their legal guardians/next of kin provided written informed consent based on written and oral information.

To investigate cluster production, the materials were constructed around a set of target words with a variety of word initial consonant clusters. For ADS, the target words were included in a list of sentences, which the mothers were asked to read aloud to a researcher. To elicit CDS in a comparable setting, the target words were also included in a story of text and pictures made as a PowerPoint presentation. The CDS data consist of the speech of the mothers reading the story to their children. On the basis of this story, a naming game was constructed with pictures of the target words. The CS data consist of the children's responses when playing the naming game with their mothers. For comparability, the CDS data for analysis were selected based on the CS cluster types analysed in Garmann et al. (2021). **Table 2** shows the number of analysed instances across phonetic contexts and data sets.

## Phonetic Analysis

Following Garmann et al. (2021), the productions were categorised into four groups, based on the occurrence of a (possible) vowel intrusion visible in spectrogram and waveform, as analysed by using the programme Praat (Boersma and Weenink, 2016). These four categories were defined in Garmann et al. (2021, p. 10) as:

(a) Clear vocalic intrusion: "a clearly definable period of high amplitude, voicing, formant structure, and of an easily measurable duration."
(b) Relatively clear vocalic intrusion: "evidence of a post-consonantal period with lower amplitude than for (a), which may be either not fully voiced or with weak formant structure, and shorter in duration (or more difficult to measure) than for (a)."
(c) Possible 'masked' vocalic intrusion: "segment boundaries are hard to ascertain, e.g., because of a period of post-release aspiration and/or devoicing in an approximant may overlay a vocalic interval."
(d) Definitely no vocalic intrusion: "no intervening acoustic material or discontinuity between C1 and C2."

To assess the validity of this categorisation, 21% of the CDS words were blind-coded by a rater not involved in the original scoring. The agreement between raters concerning the existence of a vocalic intrusion (a + b vs. c + d) was 83%. For the ADS and CS, Garmann et al. (2021) reported a corresponding agreement of 75%. The duration of the vocalic intrusions in category (a) and (b) were measured in Praat, again following the methodology of Garmann et al. (2021).

To check for speech rate as a possibly confounding factor, we made a comparison between CDS and ADS: Garmann et al. (2021) measured the number of syllables per second in four identical ADS sentences as they were produced by three of the mothers of 2;6-year-olds. We now measured the number of syllables per second in four identical CDS sentences produced by the same three mothers. According to these measurements, the speech rate was significantly slower in CDS than in ADS (median 261 vs. 205 ms per syllable, $W = 37.5$, $p = 0.0496$), but there was no connection between these three mothers' speech rate and the incidence or duration of vowel intrusions in their consonant clusters, as the mother with the slowest speech had just as many intrusions as the mother with the fastest speech, and only marginally longer intrusions (median 34 ms) than the two others (with a median of 31 and 32 ms, respectively).

## Statistical Analysis

Hypotheses regarding the incidence of vocalic intrusions [i.e., categories (a) and (b)] were tested with chi-squared ($\chi^2$) tests, or Fisher's exact test for expected values at 4 or below. The hypotheses concerning the duration of vocalic intrusion were investigated using Wilcoxon rank sum tests, preferred over $t$-tests due to deviations from a normal distribution. These non-parametric statistical tests were deemed suitable for the data set consisting of 517 observations in total, with 192 observations from CDS compared to 159 from ADS and 156 from CS.

The study focuses on differences between samples rather than individual variation, based on Vihman et al. (1994), who found little individual differences at the segmental level within groups of mothers speaking the same language. In our study as mentioned above, all mothers spoke the same dialect, and the speech samples were scripted. Each mother's speech showed considerable variation in the duration of vocalic intrusions, but there was no significant variation in neither incidence nor duration between the mothers: The incidence of vocalic intrusions did not differ significantly between the mothers with the most (MNR with 7 out of 10 measured clusters) and the fewest (HI with 6 out of 13) intrusions ($p = 0.06$), and there was no significant difference between the mothers with the longest (336 ms) and shortest (129 ms) median duration of intrusion ($W = 25$, $p = 0.2$).

The statistical analyses were carried out in R 4.0.2 (R Core Team., 2020) using RStudio 1.3 (RStudio Team., 2020). We used the stats package (R Core Team., 2020) to run tests and the package ggplot2 (Wickham, 2016) for visual inspection and preparation of figures.

# RESULTS

## Categorisation

The proportions of the four categories of intrusion in the three datasets are shown in **Figure 1**. It is of note that category (c) constituted only 15% of the cluster productions in CDS, a significantly smaller proportion than Garmann et al. (2021) reported for ADS [30%, $\chi^2(1) = 10.36$, $p = 0.001$] and CS [29%, $\chi^2(1) = 8.56$, $p = 0.003$]. This category consists of the cases where segment boundaries could not be determined from the acoustic signal, due to the ambiguous alignment of certain articulatory parameters. For example, in a voiceless stop cluster + lateral, a medial period of voiceless aperiodicity could

**TABLE 2 |** A breakdown of cluster types analysed in this paper, with incidences for each category of speech data.

| Phonetic context | | Clusters | CS data | CDS data | ADS data |
|---|---|---|---|---|---|
| Stop + liquid | Stop + /r/ | /pr, br, tr, dr, kr, gr/ | 40 | 43 | 56 |
| | Stop + /l/ | /pl, bl, kl, gl/ | 24 | 19 | 32 |
| Fricative (non-s) + liquid | | /fl, fr/ | 25 | 21 | 16 |
| S-clusters | | /sp, st, sk, sl, sn, sm, sv/ | 67 | 60 | 72 |
| Total | | | 156 | 143 | 176 |

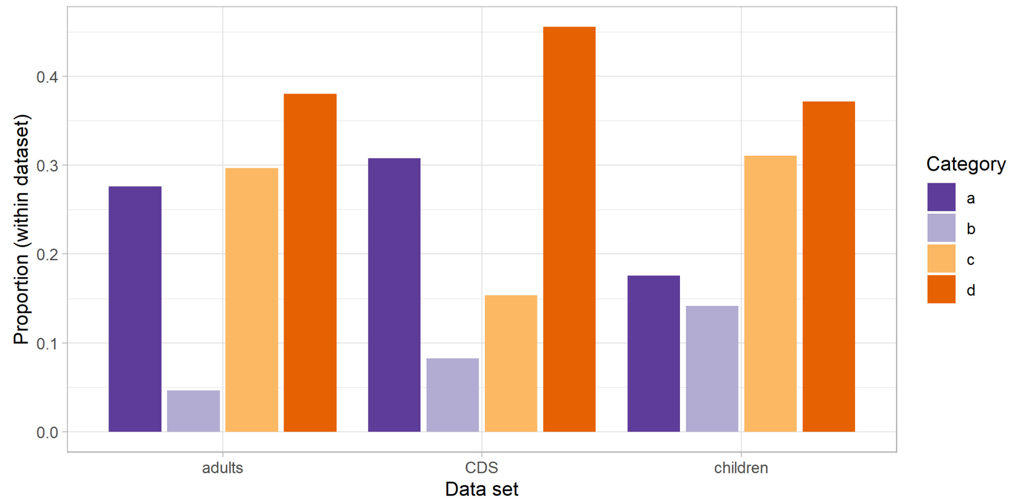



**FIGURE 1 |** Distributions of the four categories of intrusion across the three datasets.

be attributed to post-aspiration of the voiceless stop, which could coincide with a (devoiced) vocalic intrusion, and/or a period of devoicing in the lateral.

The fact that this kind of ambiguous articulation was significantly reduced in CDS would suggest that adults avoided it, opting instead either for a more clear-cut open transition (and a definite, voiced period of vocalic intrusion), hence adding to categories (a) or (b), or potentially the reverse, suppressing post-release aspiration and therefore devoicing in a following lateral. As can be observed in **Figure 1**, the data suggest a greater inclination toward the latter: While all other categories [namely (a), (b), and (d)] have more incidences in CDS than in ADS, the difference is more notable for (d), although not significant in either. The decision was made to exclude category (c) from any further calculations, thus making the estimate of intrusion a conservative one (Garmann et al., 2021). However, this observed difference in how mothers categorise their articulations suggests a desire to avoid segmentally 'ambiguous' sequences (even if they are completely natural in ADS).

## Incidence

With regard to the incidence of vocalic intrusion, there was no significant difference between CDS (46.1%) and ADS [46.9%, $\chi^2(1) = 0$, $p = 1$], when viewed overall. Following the assumption that traits of CDS are more likely to be present in speech addressed to younger children, we isolated the CDS directed toward the 2;6-year-olds and compared this to ADS. Still, there was no significant difference in incidence [$\chi^2(1) = 0.724$, $p = 0.395$]. This means that, viewed across phonetic contexts, and compared with ADS, CDS does not have more vocalic intrusion than ADS.

Next, we investigated potential differences in the incidence of vocalic intrusion in the CDS by child age. The adults addressing 2;6-year-olds had vocalic intrusion in 53% of their consonant clusters, while the adults speaking to the older children produced 40% of their consonant clusters with a vocalic intrusion. According to a chi-squared test, this difference was significant [$\chi^2(1) = 5.17$, $p = 0.023$]. There were no significant differences in the number of intrusions between adults addressing 4-year-olds (42%) compared to adults addressing 6-year-olds (38%). Thus, while the differences are not sufficient to make a significant difference when compared with ADS, there is evidence that mothers are behaving differently toward 2;6 year-olds, when compared with 4 and 6 year-olds.

Comparing children's speech and CDS within each age group (see **Table 3**), we found a significant difference in incidence between mothers and children at 2;6 [$\chi^2(1) = 5.14$, $p = 0.023$], but not at 4 [$\chi^2(1) = 1.35$, $p = 0.245$] or 6 [$\chi^2(1) = 0.20$, $p = 0.6543$]. The figures indicate a correspondence between CDS and CS, but with a delay: The proportions of vocalic intrusions in children's speech at 4 and 6 resemble the proportions in CDS at 2;6 and

**TABLE 3 |** Number and proportion of vocalic intrusions in produced clusters in CDS and CS by age of the child.

| Child age | CDS | CS |
|---|---|---|
| 2;6 | 23/36 (64%) | 9/28 (32%) |
| 4 | 24/57 (42%) | 23/41 (56%) |
| 6 | 19/50 (38%) | 15/33 (45%) |

4, respectively (For a discussion of the development of vocalic intrusion in CS, cf. Garmann et al., 2021).

## Distribution by Cluster Type

As shown in **Table 4**, the vocalic intrusions generally occurred in the same phonetic contexts in CDS and in ADS, but CDS also had a few intrusions in non-liquid contexts. There were no significant differences between the two datasets in the incidence of intrusions when the second consonant was a non-liquid ($p = 1$ according to Fisher's exact test), or a lateral [$\chi^2(1) = 0.04$, $p = 0.844$]. However, intrusions were significantly more common in CDS than in ADS when the second consonant was a tap or a flap [$\chi^2(1) = 5.67$, $p = 0.017$]. There was only one occurrence in CDS of a consonant cluster with a tap or flap as C2 produced without a vocalic intrusion, namely a production of *krakk* 'stool' addressed to a 6-year-old. In other words, in CDS there is no apparent increased incidence of vocalic intrusions when viewed across phonetic contexts, but there does appear to be a greater incidence within a specific context, namely pre-rhotic. It should be noted that vocalic intrusion is almost omnipresent in this context. Thus, adults are differentiating their speech toward children in terms of extent of vocalic intrusion for a given phonetic context, making something that occurs gradiently in ADS, categorically occurring in CDS. In other words, it is a more systematic feature of CDS than of ADS.

While we found no vocalic intrusions at all in non-liquid C2 clusters in ADS, in effect /sC/ clusters, there were 7 occurrences in this phonetic context in the CDS data, in the words *smokk* 'pacifier,' *snørr* 'snot,' *svane* 'swan,' and *sverd* 'sword.' These instances were evenly distributed across age groups. The difference between CDS and ADS is, as noted above, not statistically significant, but the mere occurrence is interesting, particularly because we also saw intrusions in this context in children's speech, as shown in **Table 4**. This indicates that mothers may be mirroring their children when it comes to phonetic contexts. There was no significant difference between incidence in CDS and children's speech when the C2 was a non-liquid [$\chi^2(1) = 0.27$, $p = 0.600$] or a tap/flap [$\chi^2(1) = 1.18$, $p = 0.277$], but vocalic intrusions were significantly less common in CDS than in CS when the second consonant was a lateral [$\chi^2(1) = 9.24$, $p = 0.002$].

**TABLE 4 |** The number of instances with a vowel intrusion and the number of clusters measured in the three data sets, by C2 category.

| | C2 non-liquid | C2 lateral | C2 tap/flap | Total |
|---|---|---|---|---|
| CS: intrusions | 10/60 (17%) | 9/11 (82%) | 28/31 (90%) | 36/102 (35%) |
| CDS: intrusions | 7/60 (12%) | 7/30 (23%) | 52/53 (98%) | 66/143 (46%) |
| ADS: intrusions | 0/43 (0%) | 6/20 (30%) | 54/65 (83%) | 60/128 (47%) |

## Duration

Turning to the duration of the vocalic intrusions, there was no significant difference between CDS (median = 29 ms), when viewed overall, and ADS (median = 22 ms, $W = 1390$, $p = 0.078$), as shown in **Figure 2**. However, when we looked specifically at the speech directed toward the youngest children, the difference between CDS and ADS *was* significant: The intrusions in CDS were significantly longer when adults addressed 2;6-year-olds (median = 32 ms) than when adults addressed adults ($W = 804$, $p = 0.018$), see **Figure 3**. This difference was not found when comparing ADS with CDS addressed to 4-year-olds (median = 28 ms, $W = 658$, $p = 0.708$) or 6-year-olds (median = 25 ms, $W = 580$, $p = 0.267$). Thus, in terms of *duration of intrusion*, adults are differentiating their CDS as a function of the age of the child being addressed. In other words, vocalic intrusion is arguably a more *salient* feature of CDS addressed to very young children.

Looking at the duration of vocalic intrusions between CDS overall (median = 29 ms) and CS overall (median = 43 ms), the former was significantly shorter than the latter ($W = 811$, $p < 0.001$). As can be seen from **Figure 2**, there may be a development corresponding to what occurs in children's speech, where Garmann et al. (2021) reported a significant reduction in duration with age. However, there is a lot of variation masking possible changes in CDS over time: Investigating the duration of vocalic intrusions in the CDS data divided by the children's age, we found no significant differences between the CDS addressed to 2;6-, 4-, and 6-year-olds. Children and mothers appear to approach each other over time: There was a significant difference in duration of the vocalic insertions between CDS to 2;6-year-olds (median = 32 ms) and the 2;6-year-olds themselves

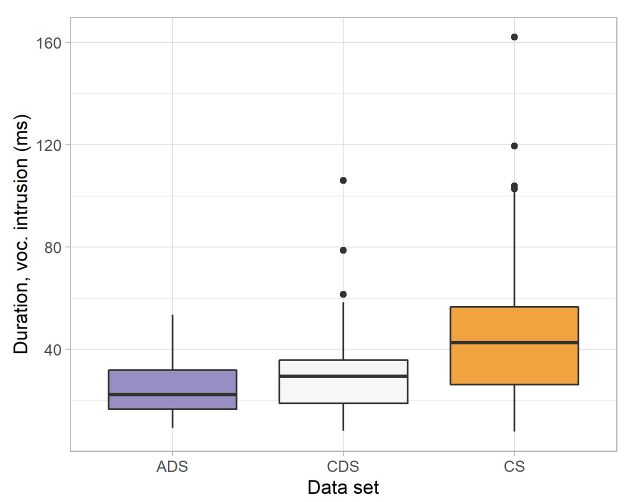



**FIGURE 2 |** The duration of vocalic intrusions in consonant clusters produced in ADS, CDS, and CS.

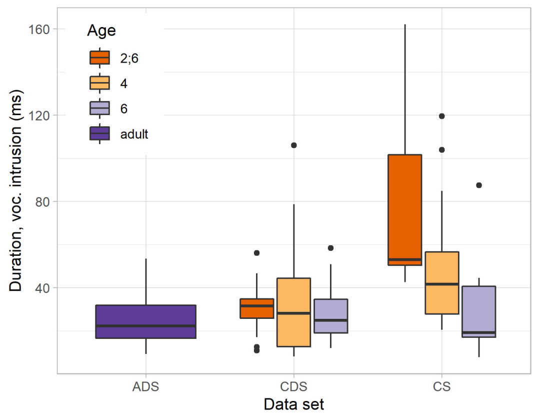



**FIGURE 3 |** Duration of vocalic intrusion in ADS, in CDS to different age groups (2;6, 4, and 6) and in CS in different age groups (2;6, 4, and 6).

(median = 53 ms, $W = 201$, $p < 0.001$), and a smaller, but still significant difference between mothers (median = 28 ms) and children (median = 41 ms) at 4 years ($W = 358$, $p = 0.015$), but no difference between mothers (median = 25 ms) and children (median = 19 ms) at age 6 ($W = 92$, $p = 0.611$).

## DISCUSSION

In this paper, we set out to study vocalic intrusions in the production of clusters in Norwegian CDS. This prosodic-phonetic bias has been observed in Norwegian ADS as well as in CS. We investigated whether vocalic intrusions are present also in CDS, and if so, whether they occur with the same incidence and patterns of distribution. We analysed CDS from mothers addressed to children aged 2;6, 4, and 6, and compared these data to ADS and CS data collected and analysed by Garmann et al. (2021). Furthermore, we divided the CDS data by age to look for changes as children grow older. We tested four hypotheses, namely that vocalic intrusions (1) have a higher incidence in CDS than in ADS; (2) have the same distribution with respect to phonetic contexts in CDS as in ADS; (3) are of longer duration in CDS than in ADS but shorter than in children's speech, and (4) their incidence and duration in CDS diminish as children grow older.

Regarding our first hypothesis, we found that when viewed overall, there is the same prevalence of vocalic intrusions in CDS and ADS. This suggests that increasing the incidence of vocalic intrusion is not used as a generalised strategy either to increase affect or bonding or for didactic purposes. However, when isolating the speech addressed to the youngest children, we did find intrusions to be significantly more common than in ADS. We interpret this as indicating that bonding strategies concerned with attracting attention, conveying affect and mirroring the child's speech behaviour, together with didactic goals of instructing how to convey emphasis and to produce difficult clusters, are more important in speech addressed to the younger children.

Moreover, we found a significantly smaller proportion of clusters analysed as category (c) clusters (i.e., clusters where the segment boundaries are hard to identify) in CDS when compared to ADS. This indicates that even if the incidence of intrusions generally is the same in CDS and ADS, mothers tend to produce their clusters more clearly either with or without intrusions when speaking to children. In other words, there is evidence that some adjustments may be made to clarify an ambiguous segmental boundary within a cluster, which may play some phonological didactic function in terms of reinforcing the segmental composition of a sequence.

Concerning our second hypothesis, the vocalic intrusions are generally found in the same phonetic contexts in CDS and ADS, with two exceptions: Firstly, albeit only marginally, CDS contained instances of vocalic intrusion in non-liquid /sC/ clusters, which to some extent mirrors the greater cross-context generalisation of this feature in child speech. In addition to possible mirroring of child speech behaviour, this shows an *expansion* in the range of incidence, even if the overall level of incidence is the same. Perhaps more important, however, is the finding that vocalic intrusions were significantly more common (and indeed almost entirely categorically present) in CDS than ADS in clusters that could be considered to be particularly challenging for the child, that is when the second consonant was a tap or a flap (e.g., in *brannmann* 'fireman,' ['bɾɑn.mɑn], and *glass* 'glass,' ['gɾɑs]). This suggests that while vocalic intrusion may not be more prevalent overall in CDS, it does appear to be *systematically* applied in those phonetic contexts in which it is frequently – though not categorically – applied in ADS. In view of the facts that the Norwegian tap [ɾ] is short and can be difficult to perceive without a preceding intrusion (Bradley, 2007) and being one of the latest speech sounds to be acquired by children speaking Urban East Norwegian (Fintoft et al., 1983), this would appear to be motivated by a phonetic didactic intent.

When it comes to the third hypothesis, a longer duration in vocalic intrusion was strongly predicted, being compatible with a number of possible motivations. Longer durations are associated with greater acoustic salience, which are compatible with general strategies for attracting more attention and conveying affect. The strongest version of this hypothesis was not confirmed, as there were no significant differences in the duration of intrusions between CDS in general and ADS, but the intrusions had longer durations in CDS addressed to 2;6-year-olds than in ADS. Our interpretation is that this particular modification is primarily underpinned by bonding strategies, since Wang et al. (2018) have suggested that the purpose of IDS directed toward younger infants is more geared toward increasing closeness. There may also be some didactic intent in drawing attention to specific words by emphasising them. If there were some form of didactic intent related to cluster production specifically, we might expect it to persist in CDS directed toward older children, but this appears not to be the case. We also note that while vocalic intrusions are longer in CS than in either CDS or ADS, they are particularly long in the speech of 2;6-year-olds, hence providing possible evidence for a (albeit somewhat subtle) mirroring effect in the CDS addressed to that age group. We interpret a mirroring behaviour as another bonding strategy, since it seeks to close the distance between the behaviour of the child and that of the speaker. In mirroring articulation strategies of the child, it also arguably facilitates comprehension. While longer intrusions are compatible with various didactic intents (e.g., showing higher linguistic structure such as emphasis, or showing the linguistic-phonetic detail of how to 'sound Norwegian,' or simply to aid phonetic mastery of a universally difficult articulation), we would expect these to be employed also, or even especially, with older children, which is not the case. We thus conclude that longer vocalic intrusions are principally a bonding strategy with possibly a didactic function of word learning, employed selectively toward the youngest children.

Thus far, it would appear that mothers do indeed modify their productions of consonant clusters, and in a number of quite specific ways. Firstly, they reduce the proportion of clusters with ambiguous segmental boundaries [category (c) clusters]. Secondly, they systematise a phonetically natural tendency within a specific category of clusters (those with a rhotic C2), and marginally extend this tendency to other phonetic contexts (i.e., with a non-liquid). Thirdly, they phonetically exaggerate the intrusion, through lengthening, to children under 4 years only. From the literature, we know that we adapt our speech to our interlocutors in myriad ways, and that different properties of CDS may fade at different stages, thus revealing complex strategic modification of speech in parent–child interactions. Recall that Payne et al. (2015) found that mothers modified their rhythm in CDS to children aged 2–6, with no change over time, while Warren-Leubecker and Bohannon (1984) found fathers to use a higher pitch and wider pitch range when speaking to 2-year-olds, but not to 5-year-olds. The mothers in the same study adapted in both pitch and pitch range to both 2- and 5-year-olds.

This marginal extension and exaggeration of intrusions in CDS appears to echo reported patterns in CS in which we found that intrusions were also generalised beyond the ADS phonetic contexts (Garmann et al., 2021). A critical question is whether children are simply replicating this pattern in the CDS input, that is, their behaviour is driven by phonetic evidence in the input, or if they are in fact going beyond this and imposing their own structural constraints on their output. If they extend and/or exaggerate significantly more than in CDS, that would indicate a degree of abstract mediation, along the lines of Vihman and Velleman (2000). Further research, comparing CS and CDS more directly and for a wider range of discourse contexts is needed to establish this.

As for the fourth hypothesis concerning a decline of both incidence and duration of intrusions as children grow older, we found a higher incidence in speech addressed to 2;6-year-olds than to the older children, but no significant difference in duration of the intrusions. Comparing mothers and children, we found significantly more vocalic intrusions among the children at age 2;6, but no significant difference at age 4 or 6. Correspondingly, the intrusions were significantly shorter in CDS than in CS at age 2;6 and 4 (although with a smaller difference), while the difference had disappeared completely at age 6. Hence, concerning vocalic intrusions in Norwegian, it appears that exaggerations in incidence and duration are a property of CDS addressed to young children only. Why does this property of CDS disappear so early?

The finding is in line with those of Warren-Leubecker and Bohannon (1984) for pitch qualities in fathers' speech. While the function of speech addressed to infants (IDS) may be to draw the child's attention, speech addressed to children above 2 years (CDS) may have more of a didactic function, where more distinct speech may support children in the further segmentation of fluent speech (Fernald and Mazzie, 1991). It is, however, difficult to say which function the CDS to the 2;6-year-olds serve, which is different from 4- to 6-year-olds. It could be that some of the explanation can be found in the setting studied in this paper, namely reading a story. It is likely that 2;6-year-olds have more difficulties in focussing on the reading task than the older children, which may suggest that the mothers tried to keep their children's attention. On the other hand, it could also be that the mothers subconsciously aim to help the children in recognising clusters or perceiving their constituents, and that the CDS in this situation therefore had a more didactic function. The mothers were telling a picture-based story presenting words that they perhaps were not certain that their children knew, preparing them for a later naming task (see Garmann et al., 2021). In support of the didactic function of CDS, the mothers of 2;6-year-olds addressed their children with a frequency and duration of vocalic intrusions similar to the productions of 4-year-olds, and correspondingly, the 4-year-olds received input similar to the productions of 6-year-olds. Hence, the mothers may be guiding their children to more ADS-like speech, being one step ahead of their children.

As we have noted in the introduction, there are few studies reporting on CDS to children older than 2, and even fewer reporting on the path that caregivers take, modifying the details of their CDS register to move away from child-directed adaptations. Thus, more is known about the transition from IDS to CDS than about modifications within later stages of CDS, or indeed about the transition from CDS to ADS. Our study has made a contribution by establishing that the incidence of vocalic intrusions is higher and the duration is longer in CDS addressed to 2;6-year-olds than to older children as well as adults, and that compared to ADS, Norwegian CDS is at the same time both more systematic (near-obligatory vocalic intrusions before taps and flaps) and closer to child speech (producing vocalic intrusions also in non-liquid /sC/ clusters). Finally, mothers appear to guide their children by mirroring them while at the same time staying one step ahead.

In the Section "Introduction," we outlined a range of different intentions that mothers may have in adapting their speech to children, and two possible underlying functions: bonding and didactic. Most – though not all – of the behaviours underpinned by these intentions pull in the same direction, namely to a higher incidence and longer duration of vocalic intrusions in CDS, contributing to an increase in acoustic salience. Hence, these two properties may fill multiple intentions and underlying functions at the same time. As they are only present in speech addressed to the 2;6-year-olds, it is possible that the specific intentions of attracting attention and instructing sounding Norwegian lose their importance some time before age 4.

There is evidence that other behaviours persist beyond this age. Two of the intentions discussed in the introduction appear to be at play throughout the age range in this study, namely instructing the mastery of Norwegian clusters, with a didactic function, and mirroring children, with a bonding function. These two pull in different directions. From a phonetic didactic point of view, we would expect CDS to be more systematic. However, for the purpose of bonding, mothers may subconsciously try to mirror their children's speech patterns, producing intrusions also where they do not occur in ADS. We see both these functions at play in our data. First, the CDS data we have reported here very clearly show that mothers are much more unambiguously employing vocalic intrusion in *specific* phonetic contexts, and overwhelmingly in the pre-tap/flap context. By increasing the systematicity of vocalic intrusion for this category of cluster, they are arguably driving this home even more categorically, especially if we consider that potentially ambiguously segmented clusters [category (c)] are resolved more definitively as category (d) productions. However, we also see seven intrusions in CDS in one phonetic context where it occurs in CS but not ADS, namely in non-liquid clusters. The mere existence, even if they are few, suggests mirroring, and by that a bonding function possibly accompanied by a desire to facilitate comprehension. Note that the fact that the systematicity does not change over the time period that we have studied would suggest that it has a different function from the increased phonetic salience of the longer duration of vocalic intrusion in speech addressed to 2;6-year-olds.

## CONCLUSION

We advocate a wider approach to the investigation of dynamic variations according to individual and interactional factors, to extend to pervasive characteristics of intersegmental timing and coordination. Further investigation is needed to consider how this approach could be integrated into existing dynamic interactive models of phonological acquisition (cf. Vihman and Velleman, 2000; Davis and Bedore, 2013). Vocalic intrusions in consonant clusters are a property, or prosodic-phonetic bias, of Norwegian (Garmann et al., 2021), as a language with an open transition between the consonants in a cluster (Endresen, 1991). It would be interesting to see whether this phenomenon is treated similarly by parents speaking for example Bulgarian or Portuguese to their children, since children speaking these languages produce vocalic intrusions in clusters (Ignatova et al., 2018; Ramalho and Freitas, 2018). This paper has shown how this particular prosodic-phonetic bias is subject to quite detailed and stratified modification in CDS, arguably with multiple functions. It thus highlights the potential importance of such biases to adaptive speech variation used in a variety of discourse contexts, and for different purposes. For example, we might conceive of increased vocalic intrusion as a strategy of hyperarticulation. To increase our understanding of the function that this property has in increasing comprehension, it would be interesting to investigate consonant clusters in speech directed to other groups for whom register adaptations have been observed, for example as L2 learners, elderly persons or individuals with receptive language difficulties, or for speech in poor listening conditions. More broadly, we advocate for an approach that incorporates three fundamental aspects of CDS. Firstly, CDS is *dynamic*, and we thus need to differentiate by child age, and to trace a longer trajectory. Secondly, CDS is *complex*, and thus we need to tease apart different aspects of speech (e.g., phonetic vs. phonological, segmental vs. longer domain aspects of connected speech). And finally, the relationship between CDS and CS is *interactive*, and to properly understand this requires close analysis of CDS in relation to child speech. This encompasses both the possible interaction of bonding and didactic functions in the dyadic relationship, and the role that phonetic-prosodic biases in the implementation of phonology in CDS, play in the construction of the child's phonology.

## STUDY LIMITATIONS

We have looked at CDS, and compared it with ADS and CS, within a specific and limited interaction context. Although we tried to make the elicitation situation as similar as possible between data sets, the different reading contexts for ADS and CDS are not identical. The sample in the study is relatively small, so the results should be interpreted with caution. A larger data set with more participants covering other situations might yield different patterns of vocalic intrusions in CDS and would also lead to better possibilities for generalisation. Ideally, we would also have included CDS and ADS data from all the same mothers, as well as CS data from their children to reinvestigate the conclusion in Vihman et al. (1994) that there is little individual variation between mothers as to segmental properties of CDS. Furthermore, we found variation in both speech rate as well as incidence and duration of vocalic intrusions between the mothers while speaking to their children, but no connections between speech rate and intrusions. However, as the data set was limited in size, we cannot rule out the possibilities of such connections. Moreover, adding another data point between age 2 and 3 could shed light on when a higher incidence and a longer duration of vowel intrusions fade in CDS; it can furthermore be interesting to study CDS to children who are older than 6 years of age.

We found that mothers adapted to their child's language level with age, but having more detailed knowledge about the children's language skills could have informed us on how parents adjust the proportion of and phonetic contexts in which they produce vocalic intrusions in consonant clusters. Whereas our data were cross-sectional, a longitudinal study of children in the age range 2–6 or older could also tell us more about the individual ways in which we adapt to our children, and whether there is a correspondence between the systematicity of vocalic intrusions before taps in individual parents' speech and their children's mastery of the tap. This would align our work with an increasing body of research which shows that young children are sensitive to variable acoustic information in the input that is non-phonemic, and that this sensitivity is reflected in their production (e.g., coda stop release in Sim and Post, under review, and vocalic intrusion in clusters, here), highlighting the importance of the precise quality of the input, in addition to the quantity of patterns of realisation.

As Warren-Leubecker and Bohannon (1984) found differences between mothers and fathers with respect to high pitch and pitch variation, it could also be interesting to investigate all potential characteristics of CDS in both mothers and fathers. Nevertheless, the lion's share of the research on CDS since has focussed on mothers', or female caregivers', speech. There is a clear need for also including fathers/male caregivers in future research and investigating the variable of gender of the child being addressed. Moreover, we do not know whether there are differences between parents depending on social factors like maternal education. A further variable of potential interest is sibling order and the degree to which infants/young children are also communicating with each other, and whether this 'child input' further affects the trajectory of children's articulatory fine-tuning.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are not readily available because of ethical and legal restrictions related to sound files. Requests to access the datasets should be directed to NG.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by NSD – Norwegian Centre for Research Data. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## AUTHOR CONTRIBUTIONS

EP and BP initiated the ACT-project within which this manuscript is situated. NG, PH, and HS contributed to the conception and design of this particular study and organised the data collection and data analyses in Praat. NG obtained financing for the Norwegian analyses. ET and EH analysed the data in Praat and participated in discussions of research questions, analyses and categorisation of data. PH performed the statistical analyses. With the help of PH, NG, HS, ET, and EH prepared and held a conference presentation on the study. Together with NG, PH, HS, EP, and BP prepared the data collection and EP trained research assistants, as well as EH and ET, in analysing the data. NG, PH, and HS wrote the first and second drafts of the manuscript and finalised it. EP and BP contributed with considerable portions of text to the first and second drafts. ET and NG edited the final text. All authors contributed to manuscript revision, read, and approved the submitted version.

## FUNDING

This research was funded by a British Academy Small Research Grant SG122210 'The acquisition of consonant timing: a study in cross-linguistic micro-variation' (EP), by funding from the University of Oslo's Center for Multilingualism in Society across the Lifespan (MultiLing, The Research Council of Norway through its Centres of Excellence scheme, project number 223265), and from the Faculty of Education and International Studies at OsloMet – Oslo Metropolitan University. OsloMet has funded the publication of this manuscript.

## ACKNOWLEDGMENTS

We thank the participants in the study, parents and children, for providing invaluable data for this piece of research. We would also like to thank Henrik Torgersen and Maria Evjen for their assistance in coding the data, Matias Jentoft for blind coding and help with the reference list. The two reviewers have provided insightful comments and suggestions. Last, but not least, we thank our late friend Inger Moen for invigorating discussions on clusters.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.688002/full#supplementary-material

## REFERENCES

Aslin, R. N., Woodward, J. Z., LaMendola, N. P., and Bever, T. G. (1996). "Models of word segmentation in fluent maternal speech to infants," in *Signal to Syntax: Bootstrapping From Speech to Grammar in Early Acquisition*, eds J. L. Morgan and K. Demuth (New York, NY: Psychology Press), 117–134.

Baran, J. A., Laufer, M. Z., and Daniloff, R. (1977). Phonological contrastivity in conversation: a comparative study of voice onset time. *J. Phon.* 5, 339–350. doi: 10.1016/S0095-4470(19)31204-5

Benders, T. (2013). Mommy is only happy! Dutch mothers' realisation of speech sounds in infant-directed speech expresses emotion, not didactic intent. *Infant Behav. Dev.* 36, 847–862. doi: 10.1016/j.infbeh.2013.09.001

Bernhardt, B. M., and Stemberger, J. P. (2018). Tap and trill clusters in typical and protracted phonological development: conclusion. *Clin. Linguist. Phon.* 32, 563–575. doi: 10.1080/02699206.2017.1370496

Boersma, P., and Weenink, D. (2016). *Praat: Doing Phonetics by Computer [Computer program]. Version 6.0.17*. Available online at: http://www.praat.org, (accessed May 9, 2016)

Bradley, T. G. (2007). Morphological derived-environment effects in gestural coordination: a case study of Norwegian clusters. *Lingua* 117, 950–985. doi: 10.1016/j.lingua.2006.05.004

Cristià, A. (2010). Phonetic enhancement of sibilants in infant-directed speech. *J. Acoust. Soc. Am.* 128, 424–434. doi: 10.1121/1.3436529

Cristià, A. (2013). Input to language: the phonetics and perception of infant-directed speech. *Lang. Linguist. Compass* 7, 157–170. doi: 10.1111/lnc3.12015

Cruttenden, A. (1994). "Phonetic and prosodic aspects of baby talk," in *Input and Interaction in Language Acquisition*, eds C. Gallaway and B. Richards (Cambridge: Cambridge University Press), 135–152. doi: 10.1017/cbo9780511620690.008

Davis, B. L., and Bedore, L. M. (2013). *An Emergence Approach to Speech Acquisition: Doing and Knowing*. New York, NY: Psychology Press.

Dilley, L. C., Millett, A. L., McAuley, J. D., and Bergeson, T. R. (2014). Phonetic variation in consonants in infant-directed and adult-directed speech: the case of regressive place assimilation in word-final alveolar stops. *J. Child Lang.* 41, 155–175. doi: 10.1017/S0305000912000670

Endresen, R. T. (1991). *Fonetikk og fonologi. Ei elementær innføring*, 2nd Edn. Oslo: Universitetsforlaget.

Englund, K., and Behne, D. M. (2006). Changes in infant directed speech in the first six months. *Infant Child Dev.* 15, 139–160. doi: 10.1002/icd.445

Englund, K. T. (2005). Voice onset time in infant directed speech over the first six months. *First Lang.* 25, 219–234. doi: 10.1177/0142723705050286

Englund, K. T. (2018). Hypoarticulation in infant-directed speech. *Appl. Psycholing.* 39, 67–87. doi: 10.1017/S0142716417000480

Fernald, A., and Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Dev. Psychol.* 27, 209–221. doi: 10.1037/0012-1649.27.2.209

Fintoft, K., Bollingmo, M., and Feilberg, J. (1983). *4 år: en undersøkelse av normalspråket hos norske 4-åringer*. Trondheim: Universitetet i Trondheim, Norges Lærerhøgskole.

Garmann, N. G., Simonsen, H. G., Hansen, P., Holm, E., Post, B., and Payne, E. (2021). Cross-linguistic variation in word-initial cluster production in adult and child language: evidence from English and Norwegian. *J. Child Lang.* 48, 1–30. doi: 10.1017/S0305000920000069

Ignatova, D., Bernhardt, B. M., Marinova-Todd, S., and Stemberger, J. P. (2018). Word-initial trill clusters in children with typical versus protracted phonological development: Bulgarian. *Clin. Linguist. Phon.* 32, 506–522. doi: 10.1080/02699206.2017.1359853

Kristoffersen, G. (2000). *The Phonology of Norwegian*. Oxford: Oxford University Press.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., et al. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science* 277, 684–686. doi: 10.1126/science.277.5326.684

Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., et al. (2015). Mothers speak less clearly to infants than to adults: a comprehensive test of the hyperarticulation hypothesis. *Psychol.Sci.* 26, 341–347. doi: 10.1177/0956797614562453

McLeod, S., van Doorn, J., and Reed, V. A. (2001). Normal acquisition of consonant clusters. *Am. J. Speech Lang. Pathol.* 10, 99–110. doi: 10.1044/1058-0360(2001/011)

Payne, E. (2016). "Prosodic-phonetic biases and the construction of phonological markedness," in *Helt fabelaktig! Festskrift til Hanne Gram Simonsen på 70-årsdagen*, eds H.-O. Enger, M. I. N. Knoph, K. E. Kristoffersen, and M. Lind (Oslo: Novus), 181–198.

Payne, E., Post, B., Astruc, L., Prieto, P., and Vanrell, M. (2010). "Rhythmic modification in child directed speech," in *Prosodic Universals. Comparative Studies in Rhythmic Modeling and Rhythm Typology*, ed. M. Russo (Roma: Aracne Editrice), 147–184.

Payne, E., Post, B., Garmann, N. G., and Simonsen, H. G. (2015). "VC timing acquisition: integrating phonetics and phonology," in *Proceedings of the International Congress of Phonetic Sciences*, ed. The Scottish Consortium for ICPhS (Oxford: University of Oxford), 0764.

Payne, E., Post, B., Garmann, N. G., and Simonsen, H. G. (2017). "The acquisition of long consonants in Norwegian," in *The Phonetics and Phonology of Geminate Consonants*, Vol. 2, ed. H. Kubozono (Oxford: Oxford University Press), 130–162.

Poulain, T., and Brauer, J. (2018). The changing role of mothers' verbal and nonverbal behavior in children's language acquisition. *First Lang.* 38, 129–146. doi: 10.1177/0142723717735427

R Core Team. (2020). *R: A Language and Environment for Statistical Computing [Computer Software]*. Vienna: R Foundation for Statistical Computing.

Ramalho, A. M., and Freitas, M. J. (2018). Word-initial rhotic clusters in typically developing children: European Portuguese. *Clin. Linguist. Phon.* 32, 459–480. doi: 10.1080/02699206.2017.1359857

Rattanasone, N. X., Burnham, D., and Reilly, R. G. (2013). Tone and vowel enhancement in Cantonese infant-directed speech at 3, 6, 9, and 12 months of age. *J. Phon.* 41, 332–343. doi: 10.1016/j.wocn.2013.06.001

RStudio Team. (2020). *RStudio: Integrated Development Environment for R [Computer Software]*. Boston, MA: RStudio PBC.

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

Stern, D. N., Spieker, S., Barnett, R. K., and MacKain, K. (1983). The prosody of maternal speech: infant age and context related changes. *J. Child Lang.* 10, 1–15. doi: 10.1017/S0305000900005092

Sundberg, U., and Lacerda, F. (1999). Voice onset time in speech to infants and adults. *Phonetica* 56, 186–199. doi: 10.1159/000028450

Vihman, M. M. (1993). Variable paths to early word production. *J. Phon.* 21, 61–82. doi: 10.1016/s0095-4470(19)31321-x

Vihman, M. M. (2017). Learning words and learning sounds: advances in language development. *Br. J. Psychol.* 108, 1–27. doi: 10.1111/bjop.12207

Vihman, M. M., Kay, E., de Boysson-Bardies, B., Durand, C., and Sundberg, U. (1994). External sources of individual differences? A cross-linguistic analysis of the phonetics of mothers' speech to 1-yr-old children. *Dev. Psychol.* 30, 651–662. doi: 10.1037/0012-1649.30.5.651

Vihman, M. M., and Velleman, S. L. (2000). The construction of a first phonology. *Phonetica* 57, 255–266. doi: 10.1159/000028478

Wang, Y., Houston, D. M., and Seidl, A. (2018). "Acoustic properties of infant-directed speech," in *The Oxford Handbook of Voice Perception*, eds S. Frühholz and P. Belin (New York, NY: Oxford University Press), 93–116.

Warren-Leubecker, A., and Bohannon, J. (1984). Intonation patterns in child-directed speech: mother-father differences. *Child Dev.* 55, 1379–1385. doi: 10.2307/1130007

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag.


**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 04 August 2021
doi: 10.3389/fpsyg.2021.688242

# Relationship Between Parent Vowel Hyperarticulation in Infant-Directed Speech and Infant Phonetic Complexity on the Level of Conversational Turns

*Ulrika Marklund*[1,2,3]*, *Ellen Marklund*[4] and *Lisa Gustavsson*[3,4]

[1]*Division of Sensory Organs and Communication, Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden,* [2]*Department of Neurology, Speech and Language Clinic, Danderyd Hospital, Stockholm, Sweden,* [3]*Division of Speech and Language Pathology, Department of Clinical Science, Intervention and Technology, Karolinska Institutet, Stockholm, Sweden,* [4]*Phonetics Laboratory, Stockholm Babylab, Department of Linguistics, Stockholm University, Stockholm, Sweden*

OPEN ACCESS

***Edited by:***
*Adrian Garcia-Sierra,*
*University of Connecticut,*
*United States*

***Reviewed by:***
*Naomi Yamaguchi,*
*Université de la Sorbonne Nouvelle Paris III, France*
*Carl Dunst,*
*Orelena Hawks Puckett Institute, United States*
*Kyle Danielson,*
*University of Toronto, Canada*

****Correspondence:***
*Ulrika Marklund*
*ulrika.marklund@liu.se*

***Specialty section:***
*This article was submitted to Developmental Psychology, a section of the journal Frontiers in Psychology*

***Received:*** *30 March 2021*
***Accepted:*** *11 June 2021*
***Published:*** *04 August 2021*

***Citation:***
*Marklund U, Marklund E and Gustavsson L (2021) Relationship Between Parent Vowel Hyperarticulation in Infant-Directed Speech and Infant Phonetic Complexity on the Level of Conversational Turns. Front. Psychol. 12:688242. doi: 10.3389/fpsyg.2021.688242*

When speaking to infants, parents typically use infant-directed speech, a speech register that in several aspects differs from that directed to adults. Vowel hyperarticulation, that is, extreme articulation of vowels, is one characteristic sometimes found in infant-directed speech, and it has been suggested that there exists a relationship between how much vowel hyperarticulation parents use when speaking to their infant and infant language development. In this study, the relationship between parent vowel hyperarticulation and phonetic complexity of infant vocalizations is investigated. Previous research has shown that on the level of subject means, a positive correlational relationship exists. However, the previous findings do not provide information about the directionality of that relationship. In this study the relationship is investigated on a conversational turn level, which makes it possible to draw conclusions on whether the behavior of the infant is impacting the parent, the behavior of the parent is impacting the infant, or both. Parent vowel hyperarticulation was quantified using the vhh-index, a measure that allows vowel hyperarticulation to be estimated for individual vowel tokens. Phonetic complexity of infant vocalizations was calculated using the Word Complexity Measure for Swedish. Findings were unexpected in that a negative relationship was found between parent vowel hyperarticulation and phonetic complexity of the immediately following infant vocalization. Directionality was suggested by the fact that no such relationship was found between infant phonetic complexity and vowel hyperarticulation of the immediately following parent utterance. A potential explanation for these results is that high degrees of vowel hyperarticulation either provide, or co-occur with, large amounts of phonetic and/or linguistic information, which may occupy processing resources to an extent that affects production of the next vocalization.

**Keywords: turn-taking, infant-directed speech, phonetic complexity, vowel hyperarticulation, conversational turns, vhh-index, Word Complexity Measure for Swedish, WCM-SE**

## INTRODUCTION

This study investigates the relationship between parents' infant-directed speech (IDS) and the developing speech production of the infant. In terms of IDS, the focus lies on the specific characteristic of vowel hyperarticulation (VH) often—but not always—found in IDS (e.g., Kuhl et al., 1997; Englund and Behne, 2006; Benders, 2013; Kalashnikova and Burnham, 2018). VH in IDS has been linked to both various language outcomes (Liu et al., 2003; Hartman et al., 2017; Kalashnikova and Burnham, 2018; García-Sierra et al., 2021; Marklund et al., accepted) and immediate facilitation of word recognition (Song et al., 2010), suggesting that it may have an impact on infant language learning and/or processing. On the other hand, from a phonetic point of view, it is to be expected that the degree of VH in parents' speech varies with the perceptual requirements of the child (Lindblom, 1990; Buz et al., 2016; Marklund and Gustavsson, 2020), whether that be different stages of language development or dynamic in-the-moment fluctuations in focus and attention. This means that VH is a characteristic of IDS which is likely highly susceptible to contextual influence, and that it may well be the case that the varying degrees of VH in parent IDS is an adaptive response to perceived cues from the infant. Therefore, the present study focuses on the directionality of any potential relationship between VH and infant speech production. The specific aspect of infant speech production under investigation is phonetic complexity (PC) of infant vocalizations. The reason for this focus is that a positive correlational relationship between VH of parent IDS and PC of infant vocalizations has previously been established on a subject level (Marklund et al., accepted). This study aims to determine whether such a relationship can be found on the level of individual conversational turns, and if so, if any directionality can be established.

### Infant Behavior Influences Parent IDS

It has been demonstrated that at least some aspects of IDS are part of a feedback loop between parent and infant, in which parents respond to infants' in-the-moment reactions to their speech by amplifying or attenuating certain IDS characteristics. The pitch of mothers' IDS to their four-month-old infants can be manipulated by interrupting this feedback loop (Smith and Trainor, 2008). In the study, mothers interacted with their infants *via* monitors. Mothers could both see and hear their infants, while the infants could neither see nor hear their mothers. Instead, research assistants interacted with the infants (out of sight of the mothers) and modulated their interaction based on the momentary pitch characteristics of the mothers' speech. Providing positive interaction to the infant contingent upon mothers' high-pitched utterances resulted in more high-pitched utterances from the mother than providing positive interaction to the infant contingent upon mothers' low-pitched utterances. This suggests that mothers' pitch modulations in IDS are at least partially a response to the infants' behavior in response to them (Smith and Trainor, 2008). Similarly, several characteristics of IDS were shown to be attenuated in mothers' speech when the immediate feedback loop with their two- to four-month-old infants was interrupted by playing previously recorded video instead of live video (Braarud and Stormark, 2008).

It has also been reported that mothers respond differently to different types of infant vocalizations; for example, more mature infant vocalizations elicit a vocal response from the mother more frequently than less mature infant vocalizations (Albert et al., 2018).

When it comes to VH, mothers' articulation of vowels was impacted when the feedback loop was interrupted as they interacted with their six- to seven-month-old infants (Lam and Kitamura, 2012). Interacting with their infants *via* a video link, mothers showed less VH when their infants were able to see them but not hear them, compared to when the infants could both see and hear them (Lam and Kitamura, 2012).

To summarize, infant behavior—including vocalizations—influences the specific realization of parent IDS in the moment. This has been shown for a number of IDS characteristics, including VH. VH is a result of spontaneous communicative adaptation to the perceptual and linguistic demands of the interlocutor (Lindblom, 1990), in this case the infant. One source of information to the level of infants' linguistic proficiency is the maturity of their vocalization. Therefore, it is reasonable to posit that PC of infant vocalizations may impact VH in parents' responses.

### Parent Behavior Influences Infant Language

The linguistic, prosodic, and articulatory modifications that parents use when speaking IDS to their infants are thought to impact both infant language development in the long term and infant language production and perception in the short term. For example, overall amount of IDS in everyday speech input at seven to eleven months is positively correlated with language outcomes at five years of age (Weisleder and Fernald, 2013), and amount of IDS in a one-on-one setting at 11 and 14 months of age is correlated with productive vocabulary at 24 months (Ramírez-Esparza et al., 2014, 2017a) as well as word production at 33 months (Ramírez-Esparza et al., 2017b). IDS also facilitates in-the-moment aspects of language development such as word learning (Ma et al., 2011; Graf Estes and Hurley, 2013), statistical learning (Bosseler et al., 2016) and word recognition (Singh et al., 2009).

Parent social and vocal behavior has also been shown to influence infant vocal behavior of the child. For example, amount of IDS in a one-on-one setting correlates with amount of infant speech output (Ramírez-Esparza et al., 2014), and the prosodic variations in parent IDS are associated with high levels of infant vocalizations (Dunst et al., 2012; Spinelli et al., 2017). Contingent vocal feedback from parents leads to more mature vocalizations, syllabic rather than vocalic in 3-month-olds (Bloom et al., 1987), and syllabic rather than vocalic and more canonical syllables in 8-month-olds (Goldstein et al., 2003). At 9.5 months of age, infants whose mothers responded to vocalizations with words produced more consonant–vowel syllables, while infants whose mothers responded with a long vowel sound produced more fully voiced vocalizations (Goldstein and Schwade, 2008). This demonstrates that the phonetic content of parent utterances can have an impact on the phonetic realization of infant vocalizations.

When it comes to the specific characteristic of VH in parents' IDS, it has been shown to predict later vocabulary size (Hartman et al., 2017; Kalashnikova and Burnham, 2018), facilitate word recognition (Song et al., 2010), and correlate with concurrent perceptive phonological development (Liu et al., 2003; García-Sierra et al., 2021). However, only one study has so far demonstrated a correlation between VH and infant vocal production, specifically PC of infant vocalizations, at 12 months of age (Marklund et al., accepted).

To summarize, parent behavior—both in terms of IDS realization and temporally contingent social feedback—influences infant language, either long term and/or in the moment. When it comes to VH in parent IDS and PC of infant vocalizations, a positive correlational relationship between them has been shown (Marklund et al., accepted), but any potential momentary impact is yet to be established.

## This Study

This study focuses on the relationship between parent VH and PC of infant vocalizations. A positive relationship between the two has previously been established on a subject level (Marklund et al., accepted), leaving unanswered, and highlighting, the question of directionality. Does the phonetic maturity of infant vocalization influence the articulatory behavior of the parent, and/or does the clarity of parents' articulation influence the vocal behavior of the infant? Based on previous findings reviewed above, both explanations are plausible. Attempting to shed light on this issue, the present study focuses on the relationship between parent VH and infant PC on a turn level. The VH of parent utterances immediately preceding and following infant vocalizations is calculated and related to the PC of the vocalization.

This study uses vhh-index, a measure of VH that normalizes across vowel type and speaker, and thus makes it possible to estimate and compare VH of individual vowel tokens. This measure has been used in a previous study on VH in Swedish IDS to 12-month-olds, where it was motivated from phonetic theory and compared to traditional measures of VH for validation purposes (Marklund and Gustavsson, 2020). The rationale for using the vhh-index in the current study is that, contrary to traditional measures, it is calculated on the level of individual vowel tokens, permitting analysis on a turn level. Previous studies on the relationship between VH and infant language have used vowel space area (Liu et al., 2003; Hartman et al., 2017; Kalashnikova and Burnham, 2018; García-Sierra et al., 2021; Marklund et al., accepted). Vowel space area is calculated on a subject level and can thus not be analyzed on the level of individual turns. On a subject level, both vhh-index calculated on all vowel types and vhh-index calculated only on point vowels have previously shown comparable results to vowel space area measures of VH (Marklund and Gustavsson, 2020).

The measure of infant vocalization maturity used in the present study is the Word Complexity Measure for Swedish (WCM-SE; Marklund et al., 2018). The WCM-SE can be used as a measure of phonological maturity, that is, it may be used to measure the stability of a developing phonological system. However, since the infants taking part in the present study are only 12 months of age and as such are not expected to have much of a phonological system in place yet, the WCM-SE is in this case used to estimate PC of infant vocalizations. This is possible since the phonological complexity parameters included in the WCM-SE are also based on PC as detailed in a previous paper (Marklund et al., 2018).

# MATERIALS AND METHODS

## Participants

Nineteen infants and their parents participated in this study (9 girls, 10 boys; 12 mothers, 7 fathers). At the time of recording the material, the infants were approximately 12 months old (mean = 12.0, range = 11.5–12.3, SD = 0.2). All infants were born full term (within three weeks of due date) and monolingual (defined as both parents speaking only Swedish with the infant). The majority of the parents ($n$ = 15) had university education, and all had completed high school (which entails three non-obligatory years of education after the mandatory nine to ten years of basic education in Sweden). The participants constitute a subset of a larger group of subjects ($n$ = 72), taking part in a longitudinal study in which parent–child dyads were recorded during free play every three to six months, when the child was between three months and three years.[1]

Participants were selected for inclusion in the present study if (a) there was a recording from the 12-month visit, (b) the infant was monolingual, and (c) there was sufficient ADS material (recorded at the 27-month visit, from the same parent as in the 12-month visit) to include in the VH analysis. The study has been approved by the Regional Ethics Review Board (2015/63-31). For the original longitudinal study, recruitment was conducted *via* mail. Addresses of infants in the appropriate age range and living in the greater Stockholm area were obtained *via* the Swedish Tax Agency, and their parents were invited to participate in the study. Parents received memory sticks with all their audio and video recordings as thanks for their participation in the longitudinal study.

## Recordings

Audio and video recordings of parent–infant interaction were made at Stockholm Babylab, the Phonetics Laboratory, Stockholm University. Parent–infant dyads (one parent and the infant) were recorded in a comfortable carpeted studio equipped with age-appropriate furniture and toys. Video and audio recordings were made with three cameras (Canon XA10) mounted on the walls of the studio to capture all

[1]The longitudinal study was part of the MINT-project (MAW 2011.0070, PI Gerholm).

angles of the parent interacting with the infant. A fourth camera (GoPro Hero3), attached to the parent's chest, enabled video uptake of the infant facing the parent. To capture high-quality audio, an additional three microphones were used. Omnidirectional wireless lavalier microphones (Sennheiser EW 100 G2) were mounted on parent and infant, and one room microphone (AKG SE 300 B) was mounted on a high shelf. In the present study, audio from the two lavalier microphones was used, since this enables high-quality close-up recordings of the parent's speech and the infant's vocalizations with minimal interference from the other speaker.

Each infant was recorded together with the parent for approximately 10 min, providing the infant vocalizations and the parent IDS material for the current study. The experimenter instructed parents to interact, play, and talk with their infant as they typically would at home. After instructions and equipment arrangements, the experimenters left the studio, closed the door, and monitored the session from the adjacent control room.

## Vowel Hypo- and Hyperarticulation Estimations in Parent Speech Material

Estimation of VH in parent's IDS was performed as a part of a previous study (Marklund and Gustavsson, 2020), and detailed information about the procedure can be found there. In brief, parent speech was quasi-orthographically transcribed by a team of researchers and research assistants, then automatically segmented, converted to IPA, and aligned with their audio files using the web service WebMAUS General 5.33 of the Bavarian Archive for Speech Signals at the University of Munich (Schiel, 1999; Kisler et al., 2017). Formants were estimated for the audio recordings, using default settings in Praat (Boersma and Weenink, 2018), except for formant ceiling and number of expected formants which were varied as part of a procedure for more robust formant estimation *via* formant ceiling optimization (Escudero et al., 2009). Since reliability of formant estimations decreases considerably with higher fundamental frequency ($f_o$), vowel tokens with a median $f_o$ exceeding 350 Hz were excluded (Monsen and Engebretson, 1983).

VH was quantified using a novel measure, the vhh-index, which entails speaker and vowel normalization, so that VH can be estimated for each individual vowel token (Marklund and Gustavsson, 2020). A midpoint of the acoustic vowel space defined by $F_1$ and $F_2$ was calculated for each individual participant based on all available tokens (in this case vowels found in both ADS and IDS). This point in space served as the absolute zero point on a scale of VH, representing extreme hypoarticulation. For each vowel type, the mean formant values were then calculated, and that point in acoustic space served as the midpoint of an individual VH scale for each vowel type. The zero point and the midpoint were used to calculate a theoretical maximum of the individual VH scale, representing hyperarticulation. Individual vowel tokens were then placed along this scale and given a vhh-index based on where they appear in the acoustic space.

## Phonetic Complexity Estimations in Infant Speech Material

Infant vocalizations were transcribed in ELAN 5.8-5.9 (Sloetjes and Wittenburg, 2008) using IPA. The transcriptions were performed by two experienced phoneticians (authors UM and LG) according to a protocol developed for compatibility with WCM-SE (Marklund et al., 2018).

The protocol entailed transcribing all sounds present in the Swedish phoneme inventory as described in Engstrand (1999), with the addition of a number of other common allophones (**Table 1** and **Figure 1**). The Swedish vowel inventory consists of seventeen vowel qualities, some of which are considered pairs of phonemically contrasted long and short vowels. However, this phonematic quantity distinction is not only based on the duration of the vowel relative to adjacent consonants, but also its spectral quality (see Elert, 1964; Hadding-Koch and Abramson, 1964). Only the spectral quality was transcribed, and no length signs were used, since vowel quality is sufficient to distinguish the vowels that are awarded points in WCM-SE (**Figure 1**). Segments that were not possible to interpret as Swedish phonemes were annotated with "C" if consonant-like and "V" if vowel-like. If it was not possible to determine whether the sound was a consonant or a vowel, it was denoted with a square. Syllable boundaries and primary stress were also marked up in each vocalization.

All infant vocalizations were transcribed. They could consist of words, syllables, babbling, or isolated speech sounds. Laughter, crying, fuzzing, coughing, effort sounds, and vegetative sounds such as breathing, sneezes, and hiccups were not transcribed. Overlapping speech and distorted sounds were excluded. Boundaries between vocalizations were based on silence (pause or breath) and thus not dependent on interpretation of lexical content or on other linguistic information such as intonation.

Two recordings were annotated by both annotators independently to check inter-transcriber agreement. Percentage of matching characters for each transcribed vocalization was compared. Characters were IPA consonants, IPA vowels (treated as a single category in the inter-rater comparison, unless they were long, front, and rounded, i.e., relevant to the WCM-SE measure, in which case their vowel quality was taken into account), syllable markers, stress markers, and vocalization boundary markers. Inter-transcriber agreement of which vocalizations were transcribed was 70%, and out of those the average transcription inter-transcriber agreement was 78%.

To operationalize complexity in infant vocalizations, the WCM-SE was used (Marklund et al., 2018). Based on a number of phonological/phonetic complexity parameters in three domains, a complexity score is calculated for each vocalization based on how many of the parameters are present in it (**Table 2**). For example, the Swedish word *elefant* ("elephant") produced as ele'fant results in 6 WCM-SE points, since it has more than two syllables, non-initial stress, a word final consonant, 1 consonant cluster, 1 liquid, and 1 fricative. *Sko* ("shoe") produced as sku: results in 3 points and *màma* ("mommy") produced as màma results in 0 points. WCM-SE points were calculated for each vocalization based on the transcriptions using a script written in R 3.5.0-4.0.2 (R Core Team, 2018).

**TABLE 1 |** The Swedish consonants used in the transcription. Consonants not recognizable as any of those phonemes were marked as "C." Adapted from IPA Chart from International Phonetic Association.

| | Bilabial | Labiodental | Dental | Retroflex | Alveolar | Palatal† | Velar | Uvular | Glottal |
|---|---|---|---|---|---|---|---|---|---|
| Plosive | p b | | t d | ʈ ɖ | | | k g | | |
| Nasal | m | | n | ɳ | | | ŋ | | |
| Trill | | | r | | | | | ʀ | |
| Tap/flap | | | ɾ | | | | | | |
| Fricative | | f v | s | ʂ ʐ | | j | | ʁ | h |
| Approximant | | | | | ɹ | | | | |
| Lat. approximant | | | l | ɭ | | | | | |

*†also /ɧ/ (voiceless dorso-palatal/velar fricative) and /ɕ/ (voiceless alveolo-palatal fricative).*

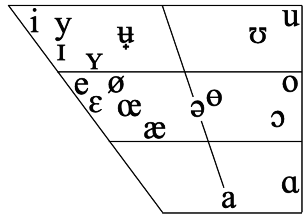



**FIGURE 1 |** The Swedish vowels used in the transcription. Vowels not recognizable as any of those phonemes were marked as "V." Adapted from IPA Chart from International Phonetic Association.

**TABLE 2 |** The WCM-SE measure as implemented in the present study, based on Marklund et al. (2018). Each transcribed vocalization was given a WCM-SE score, after which a subject mean score was calculated.

| Domains | Complexity parameter | N points |
|---|---|---|
| Word patterns | >2 syllables | 1 per vocalization |
| | Non-initial stress | 1 per vocalization |
| Syllable structures | Word-final consonant | 1 per vocalization |
| | Consonant cluster† | 1 per occurrence |
| Sound classes | Velar consonant [k], [g], [ŋ], [ɧ] | 1 per occurrence |
| | Liquid [l], [ɭ], [ɹ] | 1 per occurrence |
| | Fricative‡ [f], [v], [s], [ʐ], [ʁ], [ʂ], [j], [h], [ɧ], [ɕ] | 1 per occurrence |
| | Voiced fricative [v], [ʐ], [ʁ], [j] | 1 per occurrence |
| | Trill [r], [ʀ] | 3 per occurrence |
| | Long, front, rounded vowel [y], [ø], [ʉ̟] | 1 per occurrence |

*†Only consonant clusters within syllables were counted. ‡Vocalization-final [h] excluded, since that is likely to be release of breath.*

**TABLE 3 |** Examples of transcriptions of infant vocalizations, and WCM-SE calculations for the vocalizations. Syllable onsets are denoted by "." and stress by "ˈ".

| Example transcription | Complexity parameters | Points |
|---|---|---|
| .ə.ˈbm | Non-initial stress, word-final consonant, consonant cluster | 3 |
| .ˈC.C | Word-final consonant | 1 |
| .ˈtɛ.kɛ.tæ | >2 syllables, velar consonant [k] | 2 |
| .ˈhɪ.Cɪ.V | >2 syllables, fricative [h] | 2 |
| .hɔŋ.gɛ.ˈjɛ | >2 syllables, non-initial stress, velar consonants [ŋ] and [g], fricative [h] | 5 |
| ˈV.V | – | 0 |

Examples of WCM-SE calculations of the material in this study can be seen in **Table 3**.

## Data and vhh-index Measures

Data consist of infant vocalization WCM-SE score and vhh-index measures of the preceding and following parent utterances. Cases where an infant vocalization was preceded or followed by another infant vocalization were excluded. Since the vhh-index is novel, and token-based measures of VH have not been used previously, a number of vhh-index measures were included for exploratory purposes. All VH measures were calculated on the level of utterances, that is, they are based on all vowels for which vhh-index could be calculated within a single parent utterance. The measures were mean vhh-index, max vhh-index, vhh-index range, hyperarticulation ratio (number of vowels with vhh-index > 50 over total number of vowels), weighted mean vhh-index, and weighted max vhh-index. The weighted mean and max vhh-index entails multiplying the vhh-index with the duration of the vowel, to give more weight to longer vowels and less weight to shorter vowels. The purpose of weighting the vowel tokens like this is to reflect their relative salience in the speech signal; a vowel with long duration entails longer exposure to its particular spectral properties than a vowel with shorter duration.

## Analyses

The analyses were performed using linear mixed models. Linear mixed models are conceptually similar to regular linear regression models, except that they also account for within-subject variation, essentially allowing the model to disregard between-subject variation in favor of variation related to the independent variable.

Two linear mixed effects regressions were calculated for each of the measures of vhh-index on utterance level (mean, max, range, ratio, weighted mean, and weighted max), one on data points in which the parent utterance preceded the infant vocalization (parent–infant turns), and one on data points in which the parent utterance followed the infant vocalization (infant–parent turns). In the case of parent–infant turns, the predicted variable was infant vocalization WCM-SE score, and the fixed effects variable was the parent utterance vhh-index measure. In the case of infant–parent turns, the predicted variable was the parent utterance vhh-index measure, and the fixed effects variable was infant vocalization WCM-SE score. In both cases, random variable was participant, that is, parent–infant dyad (intercept only).

# RESULTS

Data points with infant vocalizations that were outliers (thresholds: Q ± 3*IQR) in terms of WCM-SE score were removed ($n = 4$), leaving 580 unique vocalizations to be included in the analysis, with a mean WCM-SE score of 1.6. Outliers in terms of formant values were removed on a vowel token level prior to calculating the vhh-index, as were tokens with $f_o$ exceeding 350 Hz (Marklund and Gustavsson, 2020). The number of vowel token outliers removed was 886, and the number of high $f_o$ tokens removed was 580, leaving a total of 7,688 vowel tokens. The average vhh-index for parent utterances was thus calculated on the remaining tokens within each utterance, and as such no utterances were considered outliers. The number of unique parent utterances included in the analysis was 855, and the average vhh-index was 73.7. In the analysis, 379 parent–infant turns and 476 infant–parent turns were included.

**TABLE 4 |** Summary of the fixed effects of the analysis of the measure mean vhh-index in parent–infant turns (A) and infant–parent turns (B). No significant effects were found.

| | Est. | SE | *t* |
|---|---|---|---|
| **A. Fixed effects: Parent–infant turns** | | | |
| Intercept | 1.36 | 0.25 | 5.55 |
| Parent utterance mean vhh-index | <–0.01 | < 0.01 | –0.27 |
| **B. Fixed effects: Infant–parent turns** | | | |
| Intercept | 72.8 | 3.37 | 21.6 |
| Infant vocalization WCM-SE score | 0.41 | 1.18 | 0.34 |

**TABLE 5 |** Summary of the fixed effects of the analysis of the measure max vhh-index in parent–infant turns (A) and infant–parent turns (B). No significant effects were found.

| | Est. | SE | *t* |
|---|---|---|---|
| **A. Fixed effects: Parent–infant turns** | | | |
| Intercept | 1.29 | 0.24 | 5.47 |
| Parent utterance max vhh-index | <0.01 | <0.01 | 0.38 |
| **B. Fixed effects: Infant–parent turns** | | | |
| Intercept | 106.8 | 6.75 | 15.8 |
| Infant vocalization WCM-SE score | 1.66 | 2.07 | 0.80 |

**TABLE 6 |** Summary of the fixed effects of the analysis of the measure vhh-index range in parent–infant turns (A) and infant–parent turns (B). No significant effects were found.

| | Est. | SE | *t* |
|---|---|---|---|
| **A. Fixed effects: Parent–infant turns** | | | |
| Intercept | 1.28 | 0.22 | 5.82 |
| Parent utterance vhh-index range | <0.01 | <0.01 | 0.93 |
| **B. Fixed effects: Infant–parent turns** | | | |
| Intercept | 61.4 | 6.15 | 9.98 |
| Infant vocalization WCM-SE score | 1.35 | 2.13 | 0.64 |

For mean vhh-index, max vhh-index, vhh-index range, and vhh-index ratio, no significant results were found (**Tables 4–7**). The weighted mean vhh-index of the parent utterance significantly predicted the infant vocalization WCM-SE score in parent–infant turns (**Table 8**). Surprisingly, the relationship was negative, with a change of −0.022 (95% CI -0.041| -0.003) in WCM-SE score for every increase of 1 in weighted mean vhh-index. This means that a difference between neutral articulation (vhh-index = 50) and hyperarticulation (vhh-index = 100) is associated with a difference of 1.1 WCM-SE points in the infant vocalization, reflecting the addition of one complex element. No significant

**TABLE 7 |** Summary of the fixed effects of the analysis of the measure vhh-index ratio in parent–infant turns (A) and infant–parent turns (B). No significant effects were found.

| | Est. | SE | *t* |
|---|---|---|---|
| **A. Fixed effects: Parent–infant turns** | | | |
| Intercept | 1.35 | 0.26 | 5.28 |
| Parent utterance vhh-index ratio | −0.03 | 0.21 | −0.16 |
| **B. Fixed effects: Infant–parent turns** | | | |
| Intercept | 0.68 | 0.02 | 31.0 |
| Infant vocalization WCM-SE score | <−0.01 | 0.01 | −0.86 |

**TABLE 8 |** Summary of the fixed effects of the analysis of the measure weighted mean vhh-index in parent–infant turns (A) and infant–parent turns (B). According to the t-as-z approach to estimate statistical significance (thresholds ±1.96; Luke, 2017), the effect of weighted mean vhh-index in the parent utterance can be considered significant in infant–parent turns (marked by an asterisk).

| | Est. | SE | *t* |
|---|---|---|---|
| **A. Fixed effects: Parent–infant turns** | | | |
| Intercept | 1.49 | 0.22 | 6.64 |
| Parent utterance weighted mean vhh-index | −0.02 | 0.01 | −2.24* |
| **B. Fixed effects: Infant–parent turns** | | | |
| Intercept | 9.13 | 0.65 | 14.0 |
| Infant vocalization WCM-SE score | −0.31 | 0.30 | −1.04 |

**TABLE 9 |** Summary of the fixed effects of the analysis of the measure weighted max vhh-index in parent–infant turns (A) and infant–parent turns (B). According to the t-as-z approach to estimate statistical significance (thresholds ±1.96; Luke, 2017), the effect of weighted max vhh-index in the parent utterance can be considered significant in infant–parent turns (marked by an asterisk).

| | Est. | SE | *t* |
|---|---|---|---|
| **A. Fixed effects: Parent–infant turns** | | | |
| Intercept | 1.49 | 0.23 | 6.49 |
| Parent utterance weighed max vhh-index | −0.01 | 0.01 | −2.06* |
| **B. Fixed effects: Infant–parent turns** | | | |
| Intercept | 15.3 | 1.22 | 12.6 |
| Infant vocalization WCM-SE score | −0.13 | 0.52 | −0.26 |

relationship was found between parent utterance weighted mean vhh-index and infant vocalization WCM-SE score in infant–parent turns (**Table 9**). The same patterns were found for weighted max vhh-index (**Table 5**), that is, a change of −0.012 (95% CI -0.024| -0.001) in WCM-SE score for every increase of 1 in weighted max vhh-index in parent–infant turns, but no significant relationship was found in infant–parent turns.

# DISCUSSION

The results show a negative relationship between parent VH in IDS to their 12-month-old infants and the PC of infant vocalizations on a turn level; specifically, the more hyperarticulated a parent utterance is, in terms of mean and max vhh-index weighted for vowel duration, the less phonetically complex the following infant vocalization is, in terms of WCM-SE score.

This is a somewhat surprising finding since previous findings on the same data show a positive relationship between parent VH and PC of infant vocalizations on the level of individual dyads (Marklund et al., accepted). Based on previous findings, it was expected that if any relationship was found, it would be a positive one, that is, a high degree of VH in the parent utterance would be followed by high PC in the infant vocalization, or high PC in the infant vocalization would be followed by a high degree of VH in the parent utterance.

In the previous study, the positive correlation between infants' WCM-SE scores and parents' VH (measured in vowel space area) could indicate that parents' articulation impact infants' production and/or that infants' production impact parents' articulation, or that a third, underlying variable mediates the relationship. For example, it is possible that articulatory adaptiveness is a specific realization of a general communicative adaptiveness, and that other components of this general adaptiveness may be the driving factors for any potential benefit for language development, rather than VH in itself.

In the present study, both the direct impact of VH on a turn level and the directionality of any potential effect were investigated. The negative relationship that was found between infant WCM-SE score and parent vhh-index suggests that there is a direct, in-the-moment causality between the two, and directionality of the effect was indicated by the fact that the effect was only significant in parent–infant turns.

Had the effect been significant in both directions, one potential interpretation could have been that parents are responsive and use a high degree of VH to support the linguistic needs of infants with less mature vocalizations overall. However, previous studies have shown that parent VH is typically attenuated rather than increased in interaction with atypically developing infants or infants at risk for developmental delays (Lam and Kitamura, 2010; Kalashnikova et al., 2018). It is therefore not necessarily the case that increased VH would be expected in response to immature vocalizations either. Regardless, the effect was found only in parent–infant turns, suggesting that it is the parents' articulation that has an impact on the infant vocalization.

There is no reason to believe that an infant would try less hard in their production as a direct response to high degrees of VH in the preceding parent utterance. However, high levels of hyperarticulation in the input might mean more or novel phonetic information to process for the infant. This could potentially leave less energy or focus for the infant in regard to the next task, that is, production of the next vocalization. This is in line with the *resource limitation hypothesis* which states that task demands (in addition to the developmental level of the infant) may impact attention to critical details of the speech signal (Curtin et al., 2011). Another possibility is that there might be other characteristics of parent utterances with high VH that elicit other types of responses from the infant. As an example, also in line with the resource limitation hypothesis, VH typically occurs when introducing words that are rare or have dense phonological neighborhoods (Munson and Solomon, 2004), and so utterances with incidental high VH may claim additional processing resources because they introduce new or complex linguistic information.

There are limitations to this study that should be acknowledged. The study has a relatively small sample size, although in line with previous similar studies (e.g., Liu et al., 2003; Hartman et al., 2017; Kalashnikova and Burnham, 2018; García-Sierra et al., 2021). However, by using the vhh-index as the measure of VH and linear mixed models for the analysis, multiple data points per participant could be used. Nevertheless, the small sample may be a contributing factor to the null findings in most of the vhh-measures.

There are a few things to take into consideration with regards to the complexity measure of the infant vocalizations, the WCM-SE score (Marklund et al., 2018). The choice to use this measure, even if the infants that took part in the study were too young to have a phonological system in place, was motivated because the measure is based on phonetic principles. Nevertheless, it makes assumptions of an underlying phonological system in the making and this may affect the complexity score. An additional issue is that transcription of young infants' vocalizations is notoriously difficult. The choice here was to use quite broad phonetic transcriptions compatible with the WCM-SE grading, but phonetic segments are quite detailed and perhaps not optimal representations of early vocalizations. On a related note, the interrater reliability in this study can be considered low, especially if compared to transcriptions of adult speech. It can, however, be considered a reasonable level of agreement when transcribing young infants' vocalizations on this level of detail. Previous studies with less detailed annotations (e.g., canonical babbling vs. non-canonical babbling or syllable counts) report between 70 and 84% interrater reliability (Warlaumont and Ramsdell-Hudock, 2016; Lieberman et al., 2019), and the agreement in this study lies within that range (70 and 78%).

In addition, the study uses a method to quantify VH in the parent speech, the vhh-index, which has only recently been developed (Marklund and Gustavsson, 2020) and not yet evaluated thoroughly. There are ways in which this measure could be improved on theoretical grounds, for example, the current algorithm for placing tokens in the hyperarticulation direction of the hypo-/hyper-scale does not punish deviations in one of the two formant frequencies, and such deviations should in theory impact the calculated vhh-index. Applying this novel measure in multiple ways increases the risk for spurious significant findings, but since no previous research exists as basis for methodological choices, this was deemed justified in this study.

Furthermore, high $f_o$ is one of the most consistently reported characteristics of IDS (e.g., Fernald et al., 1989); however, tokens with $f_o$ of more than 350 Hz were excluded since they are highly unreliable in terms of formant measures (Monsen and Engebretson, 1983). These exclusions introduce a potential validity issue. However, keeping high $f_o$ tokens in the analysis is not a viable option, since the fact that their formant estimation likely are inaccurate would be a major reliability issue.

There is also the possibility that the fact that recordings were made in a laboratory impacted the way that parents and infants interacted. However, previous research has shown both that young children speak similarly in different contexts such as laboratory setting or at home (Stevenson et al., 1986; Bornstein et al., 2000), and that mothers' speech is similar in laboratory setting and home, in particular after the first few minutes (Stevenson et al., 1986; although see Belsky, 1980 for differences in amount of parent speech in laboratory vs. home settings). In this study, parents were familiar with the laboratory setting since the families visited regularly to make recordings as part of the longitudinal study. The recording used in this study was made during parents' and infants' third or fourth visit to the laboratory. The literature also confirms that the presence of an observer does not necessarily negatively affect the nature of interaction (Gardner, 2000). Parents in this study were aware that various aspects of parent-child interaction were to be studied but they were not informed specifically about any analyses of articulation, which minimizes the risk that they would articulate in a way that was less natural to them.

The unexpected findings are difficult to interpret and explain in the light of existing knowledge, and one reason is that this study is the first of its kind. Given this, as well as the limitations listed above, it is premature to talk about new insights into the relationship between VH in parent IDS and infant speech production based on the findings this study. They have, however, contributed to new thoughts about how perceptual processing demands potentially impacts infant production, which need to be addressed in future studies, together with further evaluation of the VH and PC measures used in this study.

In conclusion, the present study reports a negative relationship between VH in parent utterances and PC in immediately following infant vocalizations. No relationship was found between PC in infant vocalization and VH of the immediately following parent utterance. That is, a negative relationship between parent VH and infant PC was found on the level of conversational turns, and the directionality suggested was that parent utterances influence infant vocalizations rather than the opposite.

## DATA AVAILABILITY STATEMENT

Tabular data generated for this study are available at the Open Science Framework at: https://osf.io/jw32d/.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Regional Ethics Committee in Stockholm, Sweden (2015/63-31). Written informed consent to participate in this study was provided by the adult participants and the infant participants' legal guardian/next of kin.

## AUTHOR CONTRIBUTIONS

All authors contributed to the article and approved the submitted version. UM, EM, and LG: study design, drafting the manuscript, and critical revisions of the manuscript. UM and LG: data collection (part) and transcriptions. EM: data processing and analyses. All authors approved the submission.

## FUNDING

The research presented in this manuscript was funded by Riksbankens Jubileumsfond (RJ P17-0175, PI LG). Speech material was recorded as part of a project funded by Marcus and Amalia Wallenberg Foundation (MAW 2011.0070, PI Tove Gerholm).

## ACKNOWLEDGMENTS

The authors would like to thank all families participating in the MINT project as well as Tove Gerholm for permission to use the data and also thank Tove Gerholm and David Pagmar for data collection and Freya Eriksson, Alice Gustavsson, Mika Matthis, Linnea Rask, Johanna Schelhaas, and Sofia Tahbaz for transcriptions of parent speech used for VH estimation.

## REFERENCES

Albert, R. R., Schwade, J. A., and Goldstein, M. H. (2018). The social functions of babbling: acoustic and contextual characteristics that facilitate maternal responsiveness. *Dev. Sci.* 21:e12641. doi: 10.1111/desc.12641

Belsky, J. (1980). Mother-infant interaction at home and in the laboratory: A comparative study. *J. Genet. Psychol.* 137, 37–47. doi: 10.1080/00221325.1980.10532800

Benders, T. (2013). Mommy is only happy! Dutch mothers' realisation of speech sounds in infant-directed speech expresses emotion, not didactic intent. *Infant Behav. Dev.* 36, 847–862. doi: 10.1016/j.infbeh.2013.09.001

Bloom, K., Russell, A., and Wassenberg, K. (1987). Turn taking affects the quality of infant vocalizations. *J. Child Lang.* 14, 211–227. doi: 10.1017/S0305000900012897

Boersma, P., and Weenink, D. (2018). Praat: Doing phonetics by computer [Computer program]. Version 6.0 37. Available at: http://www.praat.org/ (Accessed February 3-September 8, 2018).

Bornstein, M. H., Haynes, O. M., Painter, K. M., and Genevro, J. L. (2000). Child language with mother and with stranger at home and in the laboratory: A methodological study. *J. Child Lang.* 27, 407–420. doi: 10.1017/S0305000900004165

Bosseler, A. N., Teinonen, T., Tervaniemi, M., and Huotilainen, M. (2016). Infant directed speech enhances statistical learning in newborn infants: An ERP study. *PLoS One* 11:e0162177. doi: 10.1371/journal.pone.0162177

Braarud, H. C., and Stormark, K. M. (2008). Prosodic modification and vocal adjustments in mothers' speech during face-to-face interaction with their two-to four-month-old infants: A double video study. *Soc. Dev.* 17, 1074–1084. doi: 10.1111/j.1467-9507.2007.00455.x

Buz, E., Tanenhaus, M. K., and Jaeger, T. F. (2016). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. *J. Mem. Lang.* 89, 68–86. doi: 10.1016/j.jml.2015.12.009

Curtin, S., Byers-Heinlein, K., and Werker, J. F. (2011). Bilingual beginnings as a lens for theory development: PRIMIR in focus. *J. Phon.* 39, 492–504. doi: 10.1016/j.wocn.2010.12.002

Dunst, C., Gorman, E., and Hamby, D. (2012). Preference for infant-directed speech in preverbal young children. *Center Early Lit. Learn.* 5, 1–13.

Elert, C.-C. (1964). *Phonologic Studies of Quantity in Swedish*. Uppsala: Almqvist and Wiksell.

Englund, K., and Behne, D. (2006). Changes in infant directed speech in the first six months. *Infant. Child. Dev.: Int. J. Res. Pract.* 15, 139–160. doi: 10.1002/icd.445

Engstrand, O. (1999). "Swedish," in *Handbook of the International Phonetic Association: A Guide to the usage of the International Phonetic Alphabet* (Cambridge: Cambridge University Press), 140–142.

Escudero, P., Boersma, P., Rauber, A. S., and Bion, R. A. (2009). A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *J. Acoust. Soc. Am.* 126, 1379–1393. doi: 10.1121/1.3180321

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., and Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *J. Child Lang.* 16, 477–501. doi: 10.1017/S0305000900010679

García-Sierra, A., Ramírez-Esparza, N., Wig, N., and Robertson, D. (2021). Language learning as a function of infant directed speech (IDS) in Spanish: Testing neural commitment using the positive-MMR. *Brain Lang.* 212:104890. doi: 10.1016/j.bandl.2020.104890

Gardner, F. (2000). Methodological issues in the direct observation of parent–child interaction: do observational findings reflect the natural behavior of participants? *Clin. Child. Fam. Psychol. Rev.* 3, 185–198. doi: 10.1023/A:1009503409699

Goldstein, M. H., King, A. P., and West, M. J. (2003). Social interaction shapes babbling: testing parallels between birdsong and speech. *Proc. Natl. Acad. Sci.* 100, 8030–8035. doi: 10.1073/pnas.1332441100

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

Graf Estes, K., and Hurley, K. (2013). Infant-directed prosody helps infants map sounds to meanings. *Infancy* 18, 797–824. doi: 10.1111/infa.12006

Hadding-Koch, K., and Abramson, A. (1964). *Duration Versus Spectrum in Swedish Vowels: Some Perceptual Experiments*. New York: Haskins Laboratories.

Hartman, K. M., Ratner, N. B., and Newman, R. S. (2017). Infant-directed speech (IDS) vowel clarity and child language outcomes. *J. Child Lang.* 44, 1140–1162. doi: 10.1017/S0305000916000520

Kalashnikova, M., and Burnham, D. (2018). Infant-directed speech from seven to nineteen months has similar acoustic properties but different functions. *J. Child Lang.* 45, 1035–1053. doi: 10.1017/S0305000917000629

Kalashnikova, M., Goswami, U., and Burnham, D. (2018). Mothers speak differently to infants at-risk for dyslexia. *Dev. Sci.* 21:e12487. doi: 10.1111/desc.12487

Kisler, T., Reichel, U., and Schiel, F. (2017). Multilingual processing of speech via web services. *Comput. Speech Lang.* 45, 326–347. doi: 10.1016/j.csl.2017.01.005

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., et al. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science* 277, 684–686. doi: 10.1126/science.277.5326.684

Lam, C., and Kitamura, C. (2010). Maternal interactions with a hearing and hearing-impaired twin: similarities and differences in speech input, interaction quality, and word production. *J. Speech Lang. Hear. Res.* 53, 543–555. doi: 10.1044/1092-4388(2010/09-0126)

Lam, C., and Kitamura, C. (2012). Mommy, speak clearly: induced hearing loss shapes vowel hyperarticulation. *Dev. Sci.* 15, 212–221. doi: 10.1111/j.1467-7687.2011.01118.x

Lieberman, M., Lohmander, A., and Gustavsson, L. (2019). Parents' contingent responses in communication with 10-month-old children in a clinical group with typical or late babbling. *Clin. Linguist. Phon.* 33, 1050–1062. doi: 10.1080/02699206.2019.1602848

Lindblom, B. (1990). "Explaining phonetic variation: a sketch of the HandH theory," in *Speech Production and Speech Modelling*. eds. W. J. Hardcastle and A. Marhcal (Netherlands: Springer), 403–439.

Liu, H. M., Kuhl, P. K., and Tsao, F. M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Dev. Sci.* 6, F1–F10. doi: 10.1111/1467-7687.00275

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behav. Res. Methods* 49, 1494–1502. doi: 10.3758/s13428-016-0809-y

Ma, W., Golinkoff, R. M., Houston, D. M., and Hirsh-Pasek, K. (2011). Word learning in infant-and adult-directed speech. *Lan. Learn. Dev.* 7, 185–201. doi: 10.1080/15475441.2011.579839

Marklund, E., and Gustavsson, L. (2020). The dynamics of vowel hypo-and hyperarticulation in Swedish infant-directed speech to 12-month-olds. *Front. Commun.* 5:523768. doi: 10.3389/fcomm.2020.523768

Marklund, U., Marklund, E., Schwarz, I. C., and Lacerda, F. (2018). Introducing WCM-SE: the word complexity measure phonetically justified and adapted to Swedish. *Clin. Linguist. Phon.* 32, 1042–1053. doi: 10.1080/02699206.2018.1493620

Monsen, R. B., and Engebretson, A. M. (1983). The accuracy of formant frequency measurements: a comparison of spectrographic analysis and linear prediction. *J. Speech Lang. Hear. Res.* 26, 89–97. doi: 10.1044/jshr.2601.89

Munson, B., and Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *J. Speech Lang. Hear. Res.* 47, 1048–1058. doi: 10.1044/1092-4388(2004/078)

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ramírez-Esparza, N., García-Sierra, A., and Kuhl, P. K. (2014). Look who's talking: speech style and social context in language input to infants are linked to concurrent and future speech development. *Dev. Sci.* 17, 880–891. doi: 10.1111/desc.12172

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

Ramírez-Esparza, N., García-Sierra, A., and Kuhl, P. K. (2017b). Look who's talking NOW! Parentese speech, social context, and language development across time. *Front. Psychol.* 8:1008. doi: 10.3389/fpsyg.2017.01008

Schiel, F. (1999). "Automatic phonetic transcription of non-prompted speech." in *Proceedings of the 14th International Congress on Phonetic Sciences*, August 1–7, 1999; 106–110.

Singh, L., Nestor, S., Parikh, C., and Yull, A. (2009). Influences of infant-directed speech on early word recognition. *Infancy* 14, 654–666. doi: 10.1080/15250000903263973

Sloetjes, H., and Wittenburg, P. (2008). "Annotation by category – ELAN and ISO DCR." in *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*, May 28–30, 2008.

Smith, N. A., and Trainor, L. J. (2008). Infant-directed speech is modulated by infant feedback. *Infancy* 13, 410–420. doi: 10.1080/15250000802188719

Song, J. Y., Demuth, K., and Morgan, J. (2010). Effects of the acoustic properties of infant-directed speech on infant word recognition. *J. Acoust. Soc. Am.* 128, 389–400. doi: 10.1121/1.3419786

Spinelli, M., Fasolo, M., and Mesman, J. (2017). Does prosody make the difference? A meta-analysis on relations between prosodic aspects of infant-directed speech and infant outcomes. *Dev. Rev.* 44, 1–18. doi: 10.1016/j.dr.2016.12.001

Stevenson, M. B., Leavitt, L. A., Roach, M. A., Chapman, R. S., and Miller, J. F. (1986). Mothers' speech to their 1-year-old infants in home and laboratory settings. *J. Psycholinguist. Res.* 15, 451–461.

Warlaumont, A. S., and Ramsdell-Hudock, H. L. (2016). "Detection of total syllables and canonical syllables in infant vocalizations." in *INTERSPEECH 2016*, Sepetember 8–12, 2016; 2676–2680.

Weisleder, A., and Fernald, A. (2013). Talking to children matters: early language experience strengthens processing and builds vocabulary. *Psychol. Sci.* 24, 2143–2152. doi: 10.1177/0956797613488145

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

ORIGINAL RESEARCH
published: 01 September 2021
doi: 10.3389/fpsyg.2021.708887

# English-Speaking Adults' Labeling of Child- and Adult-Directed Speech Across Languages and Its Relationship to Perception of Affect

*Melanie Soderstrom*[1*], *Marisa Casillas*[2], *Megan Gornik*[1], *Alexis Bouchard*[3], *Sarah MacEwan*[4], *Anahita Shokrkon*[5] *and John Bunce*[6]

[1] Department of Psychology, University of Manitoba, Winnipeg, MB, Canada, [2] Department of Comparative Human Development, University of Chicago, Chicago, IL, United States, [3] Département d'Éducation, Université de Saint-Boniface, Winnipeg, MB, Canada, [4] Department of Audiology, University of British Columbia, Vancouver, BC, Canada, [5] Department of Psychological Science, University of Alberta, Edmonton, AB, Canada, [6] Department of Human Development and Women's Studies, California State University East Bay, East Bay, CA, United States

OPEN ACCESS

**Edited by:**
Maria Spinelli,
University of Studies G. d'Annunzio Chieti and Pescara, Italy

**Reviewed by:**
Gloria Gagliardi,
University of Bologna, Italy
Erika Parlato-Oliveira,
Federal University of Minas Gerais, Brazil

***Correspondence:**
Melanie Soderstrom
m_soderstrom@umanitoba.ca

**Specialty section:**
This article was submitted to Developmental Psychology, a section of the journal Frontiers in Psychology

**Received:** 12 May 2021
**Accepted:** 04 August 2021
**Published:** 01 September 2021

**Citation:**
Soderstrom M, Casillas M, Gornik M, Bouchard A, MacEwan S, Shokrkon A and Bunce J (2021) English-Speaking Adults' Labeling of Child- and Adult-Directed Speech Across Languages and Its Relationship to Perception of Affect. Front. Psychol. 12:708887. doi: 10.3389/fpsyg.2021.708887

Child-directed speech, as a specialized form of speech directed toward young children, has been found across numerous languages around the world and has been suggested as a universal feature of human experience. However, variation in its implementation and the extent to which it is culturally supported has called its universality into question. Child-directed speech has also been posited to be associated with expression of positive affect or "happy talk." Here, we examined Canadian English-speaking adults' ability to discriminate child-directed from adult-directed speech samples from two dissimilar language/cultural communities; an urban Farsi-speaking population, and a rural, horticulturalist Tseltal Mayan speaking community. We also examined the relationship between participants' addressee classification and ratings of positive affect. Naive raters could successfully classify CDS in Farsi, but only trained raters were successful with the Tseltal Mayan sample. Associations with some affective ratings were found for the Farsi samples, but not reliably for happy speech. These findings point to a complex relationship between perception of affect and CDS, and context-specific effects on the ability to classify CDS across languages.

**Keywords: child-directed speech, infant-directed speech, positive affect, universality, cross-language perception**

## INTRODUCTION

Many decades of research have supported the idea that adults speak in a specialized register, with particular acoustic and linguistic features, when speaking to infants and young children than when speaking to adults (Soderstrom, 2007; Golinkoff et al., 2015). This special form of speech, known as child-directed speech (CDS), encompasses a wide variety of particular characteristics, including higher and more variable pitch, simplified vocabulary, shortened utterances, and changes to articulatory/phonetic properties, and it has been found to support early language development. Indeed, studies suggest that exposure to more and higher quality CDS is associated with faster vocabulary development, while overheard speech generally is not (Weisleder and Fernald, 2013; Ramírez-Esparza et al., 2014). However, recently a longstanding debate about the role of cultural specification in CDS has again bubbled to the surface (e.g., Ferguson, 1978; Ochs, 1982; Pye, 1986; and more recently, Golinkoff et al., 2019; Sperry et al., 2019)–i.e. to what extent is CDS (and its role in language development) universal? Although this debate is complex and multifaceted, one important component is the finding that in cultures in which CDS is not prominent, children appear to meet language development milestones on a roughly similar timeframe (Brown, 2011, 2014; Casillas et al., 2020a,b).

There are two components to the question of CDS across languages and cultures. One is that of quantity–to what extent do different cultures use the CDS register (or talk to their infants in any register)[1] at similar rates? Although theoretically important, in the current study we set this question aside and ask a different question: How universal are the features of CDS? In other words, when adults do talk to their infants/young children (henceforth simply "infants"), however rare or common this may be, is their speech recognizable as CDS by speakers of another language?

A distinctive CDS register has been documented in a wide variety of languages, ranging from a multitude of Western and Asian languages (e.g., Fernald et al., 1989; Fernald and Morikawa, 1993; Soderstrom, 2007 and references therein), and geographically diverse places such as the Middle East (e.g., Zeidner, 1983; Farran et al., 2016) Kenya and Fiji (Broesch and Bryant, 2015) and Vanuatu (Broesch and Bryant, 2018), leading some to suggest that this register is indeed a universal feature of human interaction between caregivers and infants. More questionable, perhaps, is the idea that CDS has similar interactive functions and takes similar forms cross-culturally (e.g., Ferguson, 1978; Fernald et al., 1989). While documented similarities exist across a broad spectrum of languages and cultures, cross-linguistic and cross-cultural work suggests that there are few—if any— truly universal features of language directed to children. The lack of universals is due in large part to variation in caregivers' views about how children should be socialized as recipients and producers of language (e.g., Stross, 1972; Heath, 1983; Bernstein Ratner and Pye, 1984; Ochs and Schieffelin, 1984, 2011; Pye, 1986; Schieffelin and Ochs, 1986; Rogoff et al., 1993; Ingram, 1995; Gaskins, 2006) but has also been linked to typological variation across languages (e.g., an absence of consonant cluster simplification because the language has few clusters to begin with) or use of some CDS-related cues for other social means (e.g., high pitch use when talking to high-status adult speakers; Pye, 1986).

Across a range of unrelated cultural communities, the idea of special "babytalk" words, linguistic simplifications of any kind, or adult interpretations of infant communicative intent is seen as detrimental to children's language development or even inappropriate given children's lower social status or lack of potential as an addressee (e.g., Heath, 1983; Ochs and Schieffelin, 1984; Pye, 1986; LeVine et al., 1994; see Gaskins, 2006 for a review). With infants and toddlers alike, patterns of caregiver responsiveness to children's bids for attention also varies given cultural norms. For example, caregivers in some contexts more consistently respond to negative than positive infant vocalizations and do so more often through non-verbal than verbal means (e.g., Yucatecy Maya and Gusii vs. Euro-American caregivers; Gaskins, 1990; LeVine et al., 1994), meanwhile older children's verbalized needs are met in some contexts with responses that do not invite further contributions from the child (e.g., Quiché Maya, Kaluli, and Tseltal Maya caregivers; Stross, 1972; Ochs and Schieffelin, 1984; Pye, 1986; Brown, 2011, 2014) or are implemented via a caregiver of more proximal social status to the child (e.g., Samoa; Ochs and Schieffelin, 1984). A common thread through most of these non-urban, traditional contexts, is that the child is encouraged to meet the demands of their interactional milieu and not the other way around; in her review, Gaskins (2006) lays out at least three dimensions of child socialization that may affect CDS content and format, including caregiver ideas about: (a) the acceptability of infants broadcasting their positive/negative inner experiences, (b) in what circumstances infants are allowed to influence the actions of others, and (c) how infants can and should go about exploring the physical world. Cross-cultural variation along these and other dimensions renders affective or communicatively functional universals of CDS highly unlikely. Indeed, even similar apparent patterns of behavior may derive from different cultural motives; e.g., the lack of simplification in speech to children is done by Kaluli caregivers to support robust language development and by Samoan caregivers to maintain the status relations that permeate all other aspects of daily life (Ochs and Schieffelin, 1984).

Even among the language communities where CDS is reported to be distinct from ADS in ways that partly overlap with the distinction in English and other urban, Western linguistic contexts, there is variation in the strength and character of its implementation (e.g., Bernstein Ratner and Pye, 1984; Kitamura et al., 2001; Broesch and Bryant, 2018). Indeed, it has been suggested that North American English is a particularly extreme example of the phenomenon (Fernald et al., 1989; Shute and Wheldall, 1989), which may introduce bias in our understanding, since the majority of studies in child language come from North America. This variation calls into question how robustly universal CDS may be.

One characteristic of CDS that is highly relevant to the question of its universality is its reported relationship to positive affect. The primary prosodic characteristics of CDS, i.e., higher and more variable pitch, are also associated with the communication of positive affect or friendliness (e.g., Fernald, 1992; Kalashnikova et al., 2017), and some studies have even suggested that the well-established infant preference for listening to CDS (The ManyBabies Consortium, 2020) may primarily stem from a preference for positive affect (Singh et al., 2002). For this reason, questions about the universality of the expression of affect and the universality of CDS may be intertwined. While there is evidence that the vocal expression of emotion may be recognized above chance performance across disparate cultures (e.g., Sauter et al., 2010; Chronaki et al., 2018), there also appears to be substantial variation in its expression and perception, with advantages in perception of native-language expression

[1] In the literature, this term is often used interchangeably to refer to speech directed toward children regardless of linguistic or acoustic characteristics, and speech that contains the expected register features regardless of addressee.

(Chronaki el al.). The extent to which affect may have universal components remains controversial (see e.g., Gendron et al., 2015; Sauter et al., 2015).

In the domain of CDS, recent work by Bryant and colleagues provides evidence in favor of universality by showing that adults can identify both CDS and speaker communicative intentions across very different cultures and languages. In one study (Bryant and Barrett, 2007), adults from a Shuar (hunter-horticulturalist) village in Ecuador were able to discriminate, at about 70% accuracy, CDS from ADS samples spoken by English adults. They were also able to discriminate four categories of communicative intention (attention, prohibition, approval, comfort), with somewhat better performance in the CDS than the ADS samples. This latter result is consistent with similar findings within an English-to-English study (Fernald, 1989). Similar results to the Shuar findings were found in a later study involving adults from a Turkana (pastoralist) community in Northwestern Kenya, although there was less evidence for a role of CDS in facilitating the recognition of intention (Bryant et al., 2012).

In the current study, we add to this small body of research on cross-cultural perception of CDS. We build on the existing studies by Bryant and colleagues in several ways. First, we extend the analysis to two new languages/cultural contexts in order to add depth to the question of universality and cross-cultural similarities and differences: one an Iranian urban Farsi-speaking community (Experiment 1), and the other a small-scale subsistence Tseltal Mayan community in Southern Mexico (Experiment 2). In our study, we also "turn the tables" by asking English-speaking participants to discriminate samples from these other languages. Additionally, we add an element of ecological validity to the analysis by using audio samples that were recorded in a semi-naturalistic elicitation task (Farsi sample) and fully naturalistic realworld recordings (Mayan sample). In Experiment 2, we present data collected from trained research assistants (Experiment 2a) and from naive listeners (Experiment 2b). Finally, we explicitly examine the relationship between identification of CDS and perception of positive affective emotions. Specifically, we ask the following questions:

1) Can English speakers accurately discriminate CDS from ADS in two unfamiliar languages, Tseltal and Farsi?
2) Are speakers more likely to label speech as CDS if they perceive it to contain high positive affect (regardless of its actual status of CDS/ADS)?

In addressing these questions, we note that the questions of a potential relationship between CDS and affectively positive speech, the potential universality of CDS, and the potential universality of affective features are not straightforward to disentangle experimentally. See **Table 1** for a summary of predicted outcomes given various hypotheses. In the current study, since we have no access to a "ground truth" identification of the speakers' intended affective communication, we start from the assumption (possibly unwarranted, but at least partially supported by the above-cited literature) that some degree of universal characteristics of affect are perceivable across languages–in other words that English speakers will hear happy speech produced by Farsi and Tseltal Mayan speakers as happy. We will return to this assumption in the Discussion section in order to more fully address the implications of this limitation.

**TABLE 1 |** Overview of predicted outcomes for the present analysis, given various (simplified) ground truth conditions.

| Ground truth | | Predictions | | |
|---|---|---|---|---|
| **CDS is universally tied to positive affect** | **CDS has universal prosodic feature(s)** | **Non-native labeling accuracy** | **CDS label relates to affect rating** | **Implications for CDS universality** |
| True | True | **High** | **High** | CDS is universally identifiable via both prosody and affect |
| False | False | **Low** | **High**[a] | CDS is *not* universally identifiable on the basis of prosody or affect |
| True | False | **High** | **High** | CDS is universally identifiable via affect but not prosody |
| False | True | **High** | **Low** | CDS is universally identifiable via prosody but not affect |

[a] *Here we assume that our English-speaking participants, in the absence of other options, will rely on affect in their judgments.*

# EXPERIMENT 1: FARSI

## Materials and Methods

### Adult Participant Raters

English-speaking participant raters were recruited through the Introductory Psychology subject pool at the University of Manitoba in Winnipeg, Canada, and received course credit for their participation. The research was approved through the Psychology-Sociology Research Ethics Board of the University of Manitoba, and informed consent was obtained from each participant. Inclusion criteria included English as a primary language spoken and normal hearing. A total of 41 participants were included in the final sample. One participant was excluded after partially completing the study as they self-identified as being familiar with Farsi.

### Stimuli

Farsi recordings were collected from a sample of mother-infant dyads in Tehran, Iran during a playgroup for mothers and babies. The research was approved through the Psychology-Sociology Research Ethics Board of the University of Manitoba, and informed consent was obtained in Farsi from each participant from an Iranian native (the 6th author) who collected the samples. Dyads were recorded in an adjacent room to the playgroup, however there is some ambient noise from the playgroup in the recordings. In total, recordings from $N = 9$ mothers were used in the rating study. Infants were aged 2–9 months.

CDS and ADS samples were collected using a semi-naturalistic task developed for the elicitation of CDS samples (The ManyBabies Consortium, 2020). For the CDS samples, mothers took a series of ten objects out of a bag and talked about the objects with their infant one at a time. The same procedure was implemented for the ADS using the Farsi-speaking researcher as the interlocutor. Recordings were later segmented and transcribed in ELAN (Wittenburg et al., 2006) by a different native Farsi speaker at the University of Manitoba. Each utterance was tagged as child-directed or adult-directed.

### Rating Procedure

This experiment was run in the context of an honors thesis on the part of the 4th author examining the relationship between affective/emotion labels and CDS. Audio clips were segmented from the recordings using custom processing software written by our lab based on the transcripts. Utterances that had been tagged as being produced by an adult were pulled from the larger recording and turned into short wav files. Each clip was identified as to whether it was directed to an adult or the infant based on the annotator tags. These wav files were then randomly shuffled so that the clips were no longer in chronological or recording order. The recordings were then split into four relatively even stimulus groups with between 150 and 174 clips in each group.

Each participant was assigned randomly to one of the four stimulus groups. The study took 1–1.5 h to complete, and participants were offered a 5-min break halfway through to avoid effects of fatigue. Using a custom Python script developed in our lab, the utterances were presented as individual audio clips to each participant, who was asked to identify them as child-directed or adult-directed using the appropriate button, and their confidence in this rating on a scale from 0 to 4. For each clip, they were also asked to rate the extent to which the speaker in each clip exhibited the following characteristics: Happy, angry, sad, soothing, loving, exaggerated. The scale ranged from 1 to 5 with 3 as "neutral." They were also asked to rate the extent to which the background noise was distracting (on the same scale), as a check on the quality of the audio. Background noise ratings were low indicating this was not a significant concern. An exit survey assessed their knowledge of infant-directed speech and the characteristics they used in deciding how to classify each clip (these data were not analyzed further for this study).

## Analysis

The two primary hypotheses were pre-registered on OSF prior to conducting the analyses: https://osf.io/wq5af However, because the Farsi data originate from an honors thesis project, some of the authors were aware of the findings of a similar analysis conducted on these data prior to the pre-registration.

### Confirmatory Analyses

All statistical analyses were conducted in R with the lme4 packages (Bates et al., 2015; R Core Team, 2020) and all figures were generated with ggplot2 (Wickham, 2016).

Analysis scripts and raw anonymized data are available at https://github.com/BLLManitoba/LabellingPaperData2020. Our dependent measures Accuracy (main model) and Addressee (exploratory model) are both binary measures–correct/incorrect and cds/ads, respectively.

**TABLE 2 |** The Experiment 1 counts, means, and standard deviations for Accuracy broken out by cds, ads, and overall performance.

| | n | mean | sd |
|---|---|---|---|
| **Experiment 1** | | | |
| cds | 2,323 | 0.84 | 0.37 |
| ads | 3,827 | 0.79 | 0.40 |
| overall | 6,150 | 0.82 | 0.38 |
| **Experiment 2a** | | | |
| cds | 2,938 | 0.78 | 0.41 |
| ads | 2,351 | 0.84 | 0.37 |
| overall | 5,289 | 0.81 | 0.39 |
| **Experiment 2b** | | | |
| cds | 2,039 | 0.51 | 0.50 |
| ads | 592 | 0.52 | 0.50 |
| overall | 2,631 | 0.52 | 0.50 |

We used a binomial mixed-effects logistic regression of accuracy to both determine whether there are differences in English speakers' ability to identify ADS and CDS in an unfamiliar language (hypothesis 1) and whether positive affect plays a role in determining intended addressee (hypothesis 2).

The simple effects included in the main model were Addressee (cds/ads), Confidence, sounds Happy, sounds Loving, sounds Soothing, and sounds Exaggerated[2]. We also included the interaction terms between the positive affect measures and Addressee. Note that these models only take one reference level as the default for comparison for each factorial predictor (e.g., it will compare one level of the affect measure to each of the others but doesn't do full pairwise comparisons between the different levels of affect). We set the neutral rating as our reference group (affect measures). Therefore our models give us pairwise difference information between neutral rating and each of the other affect rating levels, but not for the pairwise differences between other levels (e.g., somewhat happy vs. extremely happy).

### Exploratory Analysis

To further explore the role positive affect plays in identifying CDS compared to ADS, we fit a binomial mixed-effects logistic regression with Addressee (cds = 0, ads = 1) as our dependent measure. The simple effects in this model were positive affect measures of sounded happy, loving, soothing, and exaggerated. We again used the "*Neutral*" rating as our reference group.

## Results

Descriptive statistics for Accuracy are shown in **Table 2**. **Figure 1** provides a breakdown of the ratings for each affect measure by addressee. Note that the affective rating for all four measures was higher for child-directed than adult-directed speech.

[2]The preregistered model accidentally included one negative affect measure (sounds sad)–here we present the model without this effect for the sake of clarity, but inclusion does not change the results.

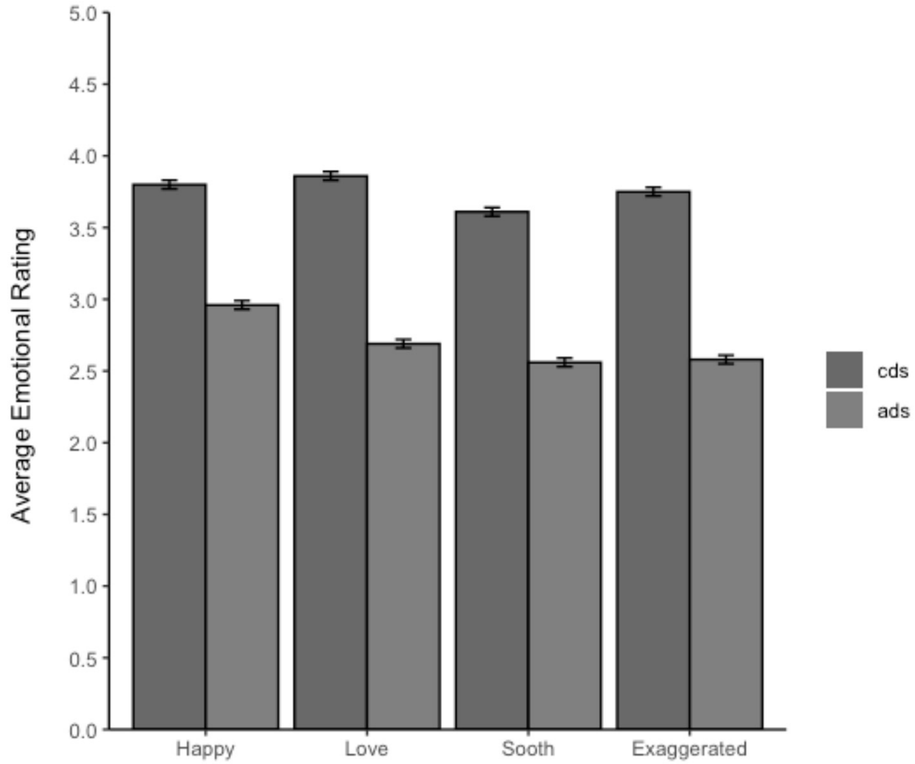



**FIGURE 1 |** Average rating of each of the positive affect measures by addressee. The error bars represent the 95% CIs. The rating scale was 1–5 with 1 being "*extremely not sounding*" like that emotion, 3 being neutral, and 5 being "*sounded extremely*" like that emotion. Thus, average ratings >3 suggest a tendency to rate the speech as containing this emotion and values <3 suggest the speech was rated as not sounding like that emotion.

## Confirmatory Analyses

As predicted, across all the clips participants' average accuracy identifying a speaker's intended addressee in an unfamiliar language was well above chance (82.1%, sd = 0.38). Our model of Accuracy included Addressee, Confidence, sounded Happy, sounded Sad, sounded Loving, sounded Soothing, sounded Exaggerated, the interaction terms for addressee with each of the 4 positive affect measures, and random intercepts by recording [$N = 6150$, log-likelihood = −2,179.1, overdispersion estimate = 2.61; formula: Accuracy ∼ Addressee + Confidence + happy + love + sooth + exaggerate + Addressee*happy + Addressee*love + Addressee*sooth + Addressee*exaggerate + (recording)].

The participants' Accuracy significantly differed by Addressee, Confidence, and the interactions between Addressee and sounded Soothing, sounded Loving and sounded Exaggerated. Accuracy was significantly lower for ads compared to cds ($B = -0.57$, SE = 0.15, $z = -3.81$, $p < 0.001$). Participants' accuracy significantly improved with higher ratings of confidence ($B = 0.29$, SE = 0.04, $z = 7.06$, $p < 0.001$). The model also revealed significant interactions between addressee and speech that sounded soothing, loving, or exaggerated.

To fully interpret the interactions, we used planned comparisons with a Bonferonni correction (alpha adjusted to <0.01) to determine how the specific affect ratings impacted accuracy for cds relative to ads. Across the three significant affect measures a similar pattern of results emerged. We found a significant decrease in accuracy for cds tagged as "*Somewhat not loving*" compared to ads ($B = -0.87$, SE = 0.27, $z = -3.21$, $p < 0.01$), a significant relative increase in the accuracy of cds tagged as "*Somewhat Loving*" ($B = 1.41$, $SE = 0.23$, $z = 5.98$, $p < 0.001$) and "*Extremely Loving*" ($B = 2.07$, $SE = 0.64$, $z = 3.22$, $p < 0.01$), and no differences in cds and ads that was rated as neutral or Extremely Not Loving ($p > 0.24$). For Soothing, we found a significant relative increase in the accuracy of cds tagged as "*Somewhat Soothing*" ($B = 1.56$, $SE = 0.24$, $z = 6.57$, $p < 0.001$) and "*Extremely Soothing*" ($B = 2.99$, $SE = 0.71$, $z = 4.19$, $p < 0.001$), and a marginal difference in cds and ads that was rated as neutral ($B = -0.51$, $SE = 0.29$, $z = -1.74$, $p = 0.08$). Finally, we found that accuracy significantly increased relative to ads for cds clips labeled "*Somewhat Exaggerated*" ($B = 0.98$, $SE = 0.21$, $z = 4.78$, $p < 0.001$) and "*Extremely Exaggerated*" ($B = 2.48$, $SE = 0.57$, $z = 4.38$, $p < 0.001$), and significantly decreased in accuracy for cds clips labeled "*Somewhat Not Exaggerated*" ($B = -1.68$, $SE = 0.22$, $z = -7.71$, $p < 0.001$) and "*Extremely Not Exaggerated*" ($B = -2.24$, $SE = 0.43$, $z = -5.23$, $p < 0.001$). The full results of the best-fit model are reported in **Supplementary Table 1**. Overall, for each of loving, soothing and exaggerated (but not happy), in general higher ratings led to higher likelihood of accurately identifying CDS compared with ADS. In other words, being rated as loving, soothing

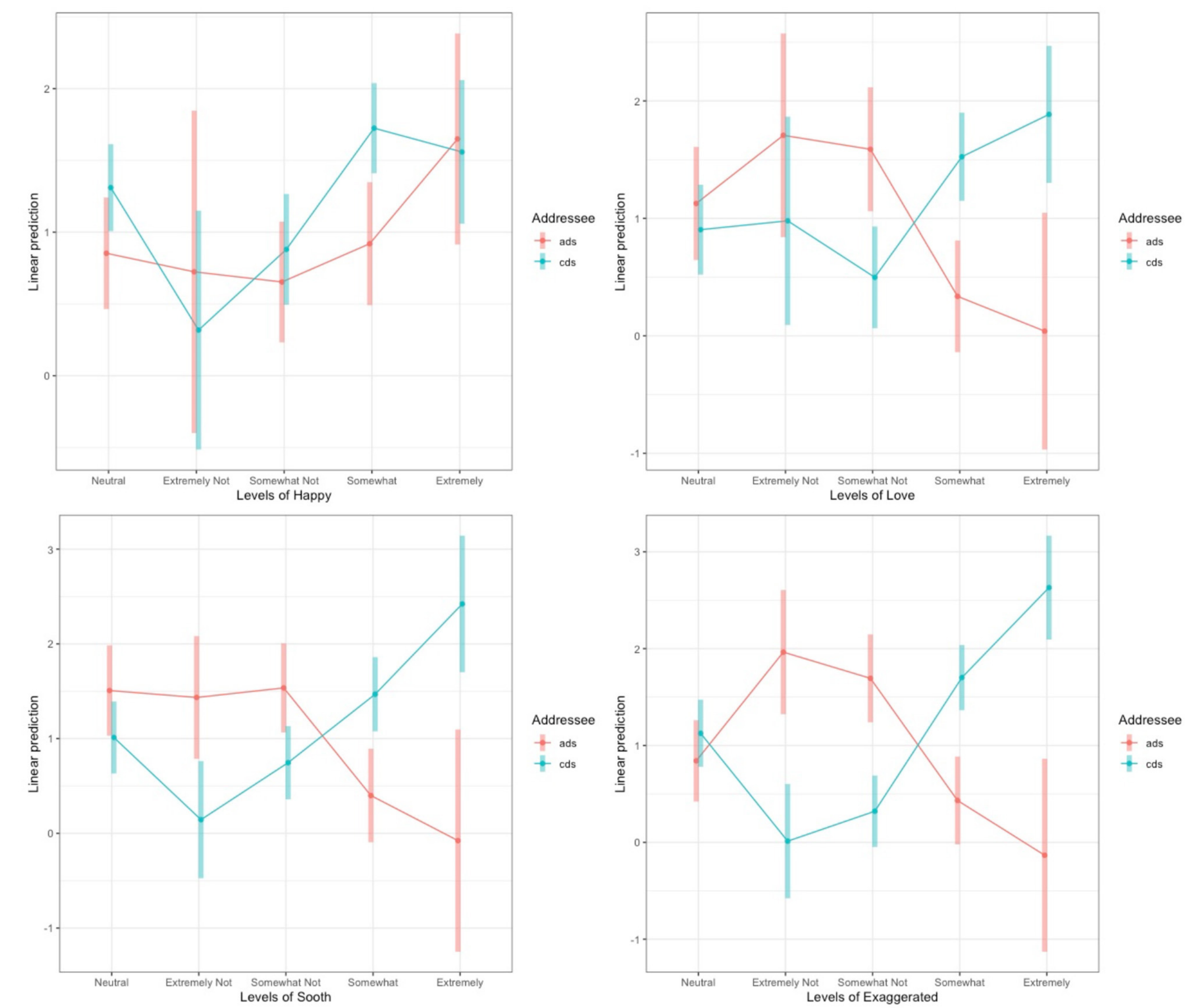



**FIGURE 2 |** Levels of sounded Happy (top left), sounded Loving (top right), sounds Soothing (bottom left), and sounded Exaggerated (bottom right) for the interactions with Addressee for Experiment 1. The log odds of each affect level are plotted on the y-axis and the shaded areas represent the 95% CIs. The interactions between Addressee and Loving, Soothing and Exaggerated were significant. Neutral is graphed on the left as it was the reference level.

and exaggerated increased the likelihood of being labeled as CDS.

## Exploratory Analysis

To probe the effect of emotion rating on cds labeling more directly, our exploratory model used addressee (as identified by the raters, not ground truth) as the dependent measure and included the predictors sounded happy, sounded loving, sounded soothing, sounded exaggerated and the random intercepts by recording [$N = 6{,}150$, log-likelihood $= -2{,}660.8$, overdispersion estimate $= 23.05$; formula = Addressee $\sim$ happy + love + sooth + exaggerate + (recording)].

The model revealed that speech tagged as Extremely Not Happy was marginally more likely to be identified as cds compared to ads ($B = -0.52$, $SE = 0.31$, $z = -1.66$, $p = 0.098$), speech tagged as "*somewhat not happy*" was significantly more likely to be ads ($B = 0.55$, $SE = 0.10$, $z = 5.49$, $p < 0.001$), clips labeled "*Somewhat Happy*" were significantly more often labeled cds ($B = -0.69$, $SE = 0.09$, $z = -8.07$, $p < 0.001$) and "*Extremely Happy*" clips were marginally more often cds ($B = -0.34$, $SE = 0.18$, $z = -1.91$, $p = 0.056$). We found that cds clips were significantly more frequently labeled "*Somewhat Loving*" ($B = -0.71$, $SE = 0.10$, $z = -6.86$, $p < 0.001$) and "*Extremely Loving*" ($B = -1.68$, $SE = 0.28$, $z = -6.10$, $p < 0.001$). For Soothing, we found that cds clips were significantly more frequently labeled "*Somewhat soothing*" ($B = -0.58$, $SE = 0.11$, $z = -5.55$, $p < 0.001$) and "*Extremely soothing*" ($B = -1.51$, $SE = 0.29$, $z = -5.24$, $p < 0.001$), and clips labeled "*Somewhat Not Soothing*" were significantly more frequently ads ($B = 0.25$, $SE = 0.11$, $z = 2.28$, $p < 0.05$). Finally, we found that cds clips were significantly more frequently labeled "*Somewhat Exaggerated*" ($B = -1.14$, $SE = 0.09$, $z = -12.53$, $p < 0.001$) and "*Extremely Exaggerated*" ($B = -2.54$, $SE = 0.24$, $z = -10.68$, $p < 0.001$), and significantly more frequently ads was labeled "*Somewhat Not Exaggerated*" ($B = 0.39$, $SE = 0.09$, $z = 4.40$, $p <$ 0.001) and "*Extremely Not Exaggerated*" ($B = 0.94$, $SE = 0.17$, $z = 4.88$, $p < 0.001$). **Figure 2** shows the interaction between the four affect measures and Addressee and the full results of the best-fit model are reported in **Supplementary Table 2**. Overall, these results are consistent with the confirmatory analysis that being rated as loving, soothing and exaggerated increased the likelihood of being labeled as CDS, while the results for happy were more mixed.

## EXPERIMENT 2: TSELTAL MAYAN

Experiment 1 provided evidence in favor of our first hypothesis, that English-speaking adults could discriminate CDS from ADS in an unfamiliar language (Farsi). It also showed that there was a relationship between those judgments and raters' perception that the clips were "loving," "soothing" and/or "exaggerated" (but less clearly for "happy").

In Experiment 2, we conducted a parallel analysis on speech from a Tseltal Mayan sample. These clips differed from the Farsi clips on a number of characteristics in addition to the language: They were highly naturalistic samples from everyday life (vs. semi-naturalistic speech from an elicitation task), and came from a rural, horticulturalist community in Mexico vs. a urban industrialized community in Iran. In Experiment 2, for the CDS/ADS distinction, we also present rating data collected from a small number of trained research assistants (Experiment 2a) in addition to the naive participant raters (Experiment 2b).

### Materials and Methods

#### Adult Participant Raters

For Experiment 2a, raters were five trained research assistant transcriber-annotators in the Baby Language Lab at the University of Manitoba. All spoke Canadian English as their primary language and had normal hearing. The ratings were collected as part of their normal duties processing the audio files as transcriber-annotators.

For Experiment 2b, participant raters were recruited as in Experiment 1. Inclusion criteria included English as a primary language spoken and normal hearing. Thirty-two participants rated the samples for CDS/ADS and a separate 32 participants rated the samples for affect. Two additional participants' data were excluded: One started the experiment but decided to stop labeling after a handful of clips. Another participant in the affect group completed a unique set of clips designed to fill in missing data that the other participants in the affect group had not labeled during the experiment. However, it was ultimately decided not to include these data as it would have added problematic complexity to the model.

#### Stimuli

The Tseltal recordings were collected in 2015 from children growing up in a rural, horticulturalist Tseltal Mayan village in the Chiapas highlands of Southern Mexico. The research was approved through the Radboud University Social Sciences Ethics Committee and informed consent was collected verbally and interactively in Tseltal from the members present in each recorded household. Children and their families were visited on the morning of the recording day by the 2nd author and a local research assistant who would conduct informed consent and give instructions before fitting the target child with an elastic vest containing a lightweight stereo audio recorder (Olympus WS-832) across the chest and a miniature photo camera (Narrative Clip 1) on the shoulder strap. The researchers then left the child and family to go about their ordinary business for the day, returning 7–11 h later to collect the recording equipment. The original corpus contains recordings from 55 children under 4;0. A subset of $N = 10$ children under 3;0 were selected for manual annotation of language activity (see Brown, 2011, 2014 and Casillas et al., 2020b for more information regarding the Tseltal cultural context). These recordings are available via the Casillas HomeBank repository (Casillas et al., 2017).

Nine 5-min clips were randomly selected from throughout each of the 10 children's recording days and were then fully annotated for all hearable speech from the target child and others in their environment by the 2nd author and a local research assistant. Each stretch of non-target-child speech was assigned to a speaker identity (e.g., the child's aunt/brother/etc.), annotated for addressee (e.g., target-child-directed, other-child-directed, adult-directed, etc.), transcribed, and loosely translated into Spanish (Casillas et al., 2020b) in ELAN (Wittenburg et al., 2006). In the present study we used the speaker identity and intended addressee annotations to extract relevant clips to present to English-speaking participants (see below).

#### Rating Procedure: Trained Raters

These data were collected as part of the general processing and classification of the Tseltal Mayan samples described above in preparation for future studies (such as Experiment 2b). Stimuli were generated by running custom software similar to the software used to process the Farsi language recordings. The software took the previously labeled and tagged longform Tseltal language recordings and generated short wav files of utterances directed at adults and children that were made by male and female adult speakers. In the current analysis only the data from female speakers was included for greater consistency with the analyses in Experiments 1 and 2b. In total, there were 5,291 clips that met these criteria. However, 2 clips were missed during the analysis, so a total of 5,289 clips were included in the sample. The order of the clips was not randomized. Trained research assistants labeled the clips according to their perception of whether the utterances were adult or child directed, or whether the clip was too noisy and should be classified as junk ($N = 210$; due to the naturalistic nature of these recordings in multi-speaker, and variable indoor and outdoor rural environments, there was a high degree of ambient noise). These latter were excluded from further analysis. The research assistant also gave a rating of their confidence on a scale of one to five, with one being not at all confident, and five being fully confident in their assessment of the clip's directive target. Affective ratings were not collected in this analysis.

### Rating Procedure: Naive Raters

This experiment was run as part of an undergraduate research project examining the relationship between maternal and infant vocal affect in naturalistic interactions. A total of 2,631 clips were randomly sub-selected from those in 2a (excluding noisy clips). Clips were generated from solely female adult speakers' utterances that were tagged as being directed at an adult or a child. We then randomly grouped the clips into through 3 roughly equal groups. Participants were randomly assigned to one of the sets. Half of the participants were instructed to label addressee (ads/cds group) and the other half were assigned to label the speakers' affect (affect group) to ultimately form 32 dyads. Infant vocal affect was also rated by a separate group of participants but those data are not reported here.

The study took 45–60 min to complete. Each participant would rate clips until they were finished or their hour time slot was up. Using a custom Python script based on that of Experiment 1, the utterances were presented as individual audio clips to each participant. Participants in the ADS/CDS group were asked to identify them as child-directed or adult-directed using the appropriate button, and their confidence in this rating on a scale from 0 to 4. Participants in the affect group rated the clips according to the categories of loving, soothing, happy, and excited, with a scale from 1 "neutral" to 5 "extremely [category]." Note that these categories are similar, but not identical, to those used in Experiment 1.

## Analysis

As noted above, the two primary hypotheses were pre-registered on OSF prior to conducting the analyses: https://osf.io/wq5af.

### Confirmatory Analyses (Trained Raters)

Similar to Experiment 1, we used a binomial mixed-effects logistic regression of accuracy to determine whether there are differences in English speakers' ability to identify ADS and CDS in an unfamiliar language (hypothesis 1). The simple effects included in the main model were Addressee (cds/ads) and Confidence (from the CDS/ADS group).

### Confirmatory Analyses (Naive Raters)

Confirmatory analyses were conducted parallel to those described for the Farsi data. The simple effects included in the main model were Addressee (cds/ads), Confidence, sounds Happy, sounds Loving, sounds Soothing, and sounds Excited and the interaction terms of each affective measure with Addressee. We again set the neutral rating as our reference group (affect measures), though note that neutral in this model was the lowest rating (1) rather than the middle rating (3). As a reminder, the data from Addressee and from affect ratings were collected from different (paired) participants in this model.

### Exploratory Analysis (Naive Raters)

As with Experiment 1, we fit a binomial mixed-effects logistic regression with Addressee (cds = 0, ads = 1) as our dependent measure. The simple effects in this model were positive affect measures of sounded happy, loving, soothing, and excited. We again used the "*Neutral*" rating as our reference group.

## Results

Descriptive statistics for Accuracy are shown in **Table 2**, **Figure 3** provides a breakdown of the ratings in Experiment 2b for each affect measure by addressee. Note that the affective rating for all four measures did not differ between child-directed and adult-directed speech.

### Confirmatory Analyses (Trained Raters)

Similar to Experiment 1, across all the clips raters' average accuracy identifying a speaker's intended addressee in an unfamiliar language was well above chance at 80.1% (sd = 0.39, see Figure). However, unlike the participants in Experiment 1 the labellers in Experiment 2 had higher accuracy tagging ads clips ($M = 0.84$, SD = 0.37) compared to cds clips ($M = 0.78$, SD = 0.41). Our binomial mixed-effects logistic regression of Accuracy included Addressee and Confidence as fixed, and random intercepts by recording [$N = 5{,}289$, log-likelihood = −2,332.4, overdispersion estimate = 1.32; formula = Accuracy ∼ Addressee + Confidence + (recording)]. Accuracy was significantly lower for cds compared to ads ($B = -0.83$, SE = 0.09, z = −9.44, $p < 0.001$). Higher confidence predicted improved accuracy ($B = 0.89$, SE = 0.05, $z = 17.73$, $p < 0.001$).

### Confirmatory Analyses (Naive Raters)

Counter to our prediction and unlike the participants in the prior two analyses, the naive labellers' average accuracy identifying a speaker's intended addressee in an unfamiliar language was approximately at chance ($M = 0.52$, SD = 0.50) and this was true for both cds ($M = 0.51$, SD = 0.50) and ads ($M = 0.52$, SD = 0.50). To explore the participants' performance statistically, we used a similar model structure from Experiment 1. Our model of Accuracy included fixed effects for Addressee, Confidence, sounded Happy, sounded Loving, sounded Soothing, sounded Excited, the interaction terms for addressee and the 4 positive affect measures, and random intercepts by recording [$N = 2{,}631$, log-likelihood = −1757.3, overdispersion estimate = 2.87; formula = Accuracy ∼ Addressee + Confidence + happy + love + sooth + excited + Addressee*happy + Addressee*love + Addressee*sooth + Addressee*excited + (recording)].

The participants' Accuracy significantly differed by Addressee, Confidence, and the interaction between Addressee and sounded Excited. Accuracy was significantly lower for cds compared to ads ($B = -0.79$, SE = 0.21, $z = -3.83$, $p < 0.001$). Surprisingly, there was a decrease in participants' accuracy the higher they rated their confidence ($B = -0.14$, SE = 0.21, $z = -3.88$, $p < 0.001$). The model also revealed a significant interaction between addressee and speech that sounded excited.

To fully interpret the interaction, we again used planned comparisons with a Bonferonni correction (alpha adjusted to <0.01) to determine how the specific affect ratings impacted accuracy for cds and ads. We found that accuracy significantly increased for cds clips labeled "*Somewhat Excited*" ($B = 0.81$, $SE = 0.27$, $z = 2.96$, $p < 0.01$), "*More Excited*" ($B = 1.91$, $SE = 0.38$, $z = 5.05$, $p < 0.001$), and clips labeled "*Extremely*

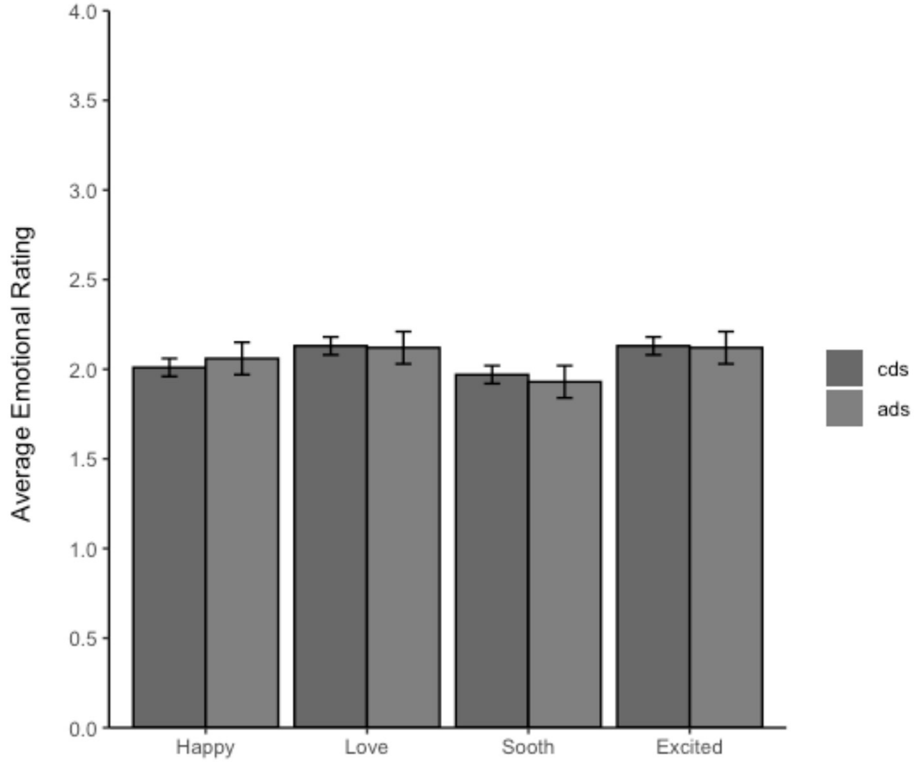



**FIGURE 3 |** Average ratings of the positive affect measures by addressee from Experiment 2b. The error bars represent the 95% CIs. The ratings on this scale were from 1 to 5 with 1 being neutral and 5 being "*sounded extremely*" like that emotion.

*Excited*" ($B = 1.38$, $SE = 0.52$, $z = 2.63$, $p < 0.01$). There were no differences at the "*Neutral*" and "*little excited*" levels (ps > 0.21). **Figure 4** shows the interaction between the four affect measures and Addressee and the full results of the best-fit model are reported in **Supplementary Table 3**.

### Exploratory Analysis (Naive Raters)

As in Experiment 1, to further explore the role that positive affect may play in identifying child-directed speech compared to adult-directed speech, we fit a mixed-effects with Addressee (cds = 0, ads = 1) as our dependent measure. We included the predictors sounded happy, sounded loving, sounded soothing, sounded excited and random intercepts by recording [$N = 2{,}631$, log-likelihood $= -1141.5$, overdispersion estimate $= 5.15$; formula = Addressee $\sim$ happy + love + sooth + excited + (recording)]. The model revealed that clips labeled "*Extremely soothing*" were significantly more likely to be labeled as ads compared to cds ($B = 1.00$, $SE = 0.35$, $z = 2.84$, $p < 0.01$). The full results of the best-fit model are reported in **Supplementary Table 4**.

# DISCUSSION

## Summary of Findings

Across 3 experiments, we examined the ability of Canadian English-speaking adults to identify CDS in two unfamiliar languages/communities, and the relationship between their judgments and measures of emotion/positive affect. Results of the first Experiment, with a semi-naturalistic sample of speech from Farsi-speaking mothers, found high accuracy for naive raters in identifying both ADS and CDS, with somewhat higher accuracy for CDS. Identification as CDS was correlated with higher rated levels of "loving," "soothing," and "exaggerated" speech characteristics, but not clearly with "happy" speech. In a second analysis rating speech from a Tseltal Mayan community as ADS or CDS, trained research assistants also showed high rates of accuracy, although accuracy was higher for ADS than CDS. However, a third experiment with naive raters assessing these same Tseltal Mayan samples found very poor accuracy, and an inverse relationship between confidence and accuracy (i.e., more confident ratings were less likely to be accurate). In this last study, identification as CDS was related to higher "excited" scores, but was not reliably associated with the other characteristics.

## Implications for Universality of CDS Characteristics

If we first consider the first two experiments (i.e., Experiment 1 and 2a) on their own, our findings support the idea that CDS is identifiable across different languages/cultures (at least, these specific ones). They also provide support for a relationship between affect and CDS (at least for Farsi, as affect was not rated in Experiment 2a). These findings are largely consistent with the

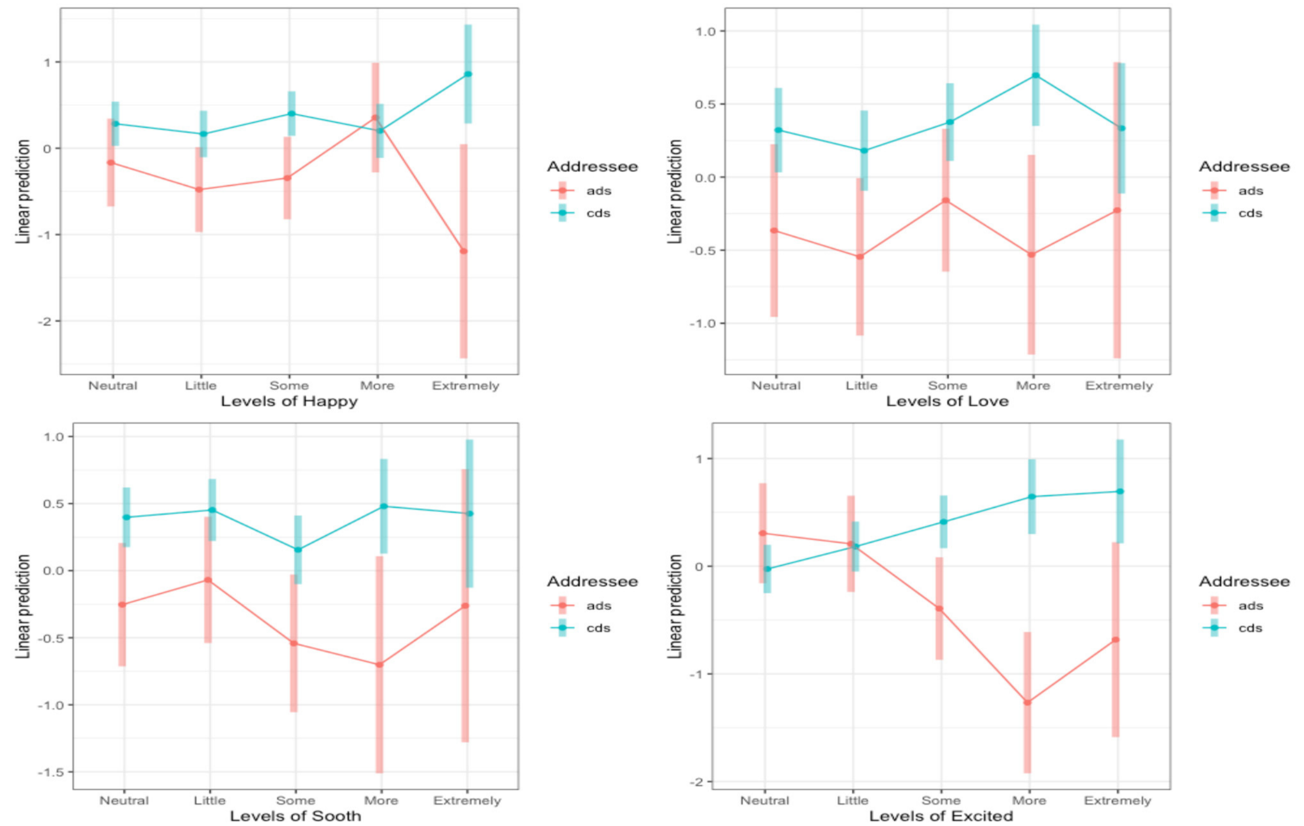



**FIGURE 4 |** Levels of sounded Happy (top left), sounded Loving (top right), sounds Soothing (bottom left), and sounded Excited (bottom right) for the interactions with Addressee in Experiment 2b. The log odds of each affect level are plotted on the y-axis and the shaded areas represent the 95% CIs. Only the interaction between Addressee and Excited was significant.

prior work by Bryant and colleagues (Bryant and Barrett, 2007; Bryant et al., 2012). However, before delving into the details of Experiment 2b, a number of nuances to these basic findings must be articulated.

First, based on these findings alone, we cannot differentiate between two possible interpretations of these findings (see **Table 1**). In one interpretation, CDS is both tied to positive affect and contains universal expressions unique to the CDS register (i.e., separate from a general expression of positive affect). In this interpretation, raters used both affective characteristics and prosodic characteristics unique to CDS in their judgments. However, it is equally possible that raters made their judgments solely on the basis of affect and not on any prosodic characteristics of CDS unrelated to affect.

Second, the findings with respect to the happiness rating are unexpected and intriguing. Recall that prior work on the relationship between affect and CDS has suggested that infants' attraction to CDS is mediated by "happy talk" (Singh et al., 2002). It is therefore of interest that "happy" talk was not reliably associated with raters' judgments of CDS, while characteristics of loving, soothing and exaggerated were. This is consistent with prior research suggesting that various affective states are communicated with CDS (Fernald, 1989), and suggests that "happy" talk *per se* may not be a driving factor in the perception of CDS.

We next turn to the findings of Experiment 2b. These findings do not fit neatly with any of the patterns predicted in **Table 1** we found both poor performance in discriminating CDS from ADS, and little relationship between the affective characteristics and rating of CDS, with the exception of excited speech (and less reliably inversely with soothing speech in the exploratory analysis). Of note, level of confidence was inversely related to performance, suggesting that the raters were relying on a somewhat systematic, but inaccurate, metric for CDS. As a first pass, these findings can be interpreted as a failure to identify CDS in the Tseltal Mayan sample. Harder to ascertain is why, particularly relative to the success with naive raters on Farsi in Experiment 1 and with trained research assistants on Tseltal Mayan in Experiment 2a. One salient possible interpretation is that the cultural context of Tseltal Mayan is more distinct from Canadian English language/culture than that of Farsi, making identification of CDS more challenging. However, the success of the research assistants suggests that this identification is not impossible, just hard. Having spent time working with naturalistic language files of this type may have given the research assistants an "edge" in detecting subtle cues based on their greater experience with these kinds of samples despite their lack of access to the conversational context or meaning of the speech. It is important to note that there are other possible reasons for the apparent greater difficulty in the Tseltal Mayan samples, however. For example, the samples were taken in a fully naturalistic, everyday life, context, whereas the elicitation task of the Farsi samples may have served to exaggerate some characteristics of CDS. Second, there was an age difference between the Farsi and the Tseltal Mayan infants, which may influence the type and degree of CDS used. Moreover, the Tseltal Mayan samples included both speech to the target child and to other nearby children, whereas the Farsi samples were restricted to the infants under study, which could also have affected the nature of the speech samples.

The reason that the pattern of results for Experiment 2b did not appear in our **Table 1** is because of two assumptions inherent in the predictions. First, that emotions are detectable across cultures, and second that in the absence of salient direct cues to CDS, raters would rely on their judgments of the emotional characteristics. These findings suggest that the first of these assumptions, and to an extent the second, is incorrect. Unfortunately, we do not have ground truth measures of the intended affective communication in the Tseltal Mayan samples (or from a third party Tseltal Mayan listener), so it is not possible to determine whether the raters were perceiving happiness, lovingness, soothingness or excitedness in the same way as a Tseltal Mayan speaker, nor whether such cues existed in the samples at all. Moreover, the affective judgments made in Experiment 2b were made by a separate group of participants than those making the CDS/ADS discrimination, so we cannot ask this question at the level of individual participants, but only at the level of group based judgments regarding affect and CDS. Nonetheless, our findings suggest that lovingness, soothingness and happiness were not used by the raters in making their judgment regarding whether an utterance was CDS or ADS. Instead, the raters relied at least in part on the degree of excitedness they perceived in the speech–but reliance on this characteristic did not lead them to accurate judgments.

## Limitations, Conclusions, and Future Directions

In interpreting these findings, it is important to point out that we did not conduct a systematic or fulsome exploration of how classification of CDS/ADS (and its relationship to positive affect) might vary across languages, language typologies, cultural contexts or communities. Our analyses were conducted purely over samples of convenience regarding two non-English languages to which we had access. These languages and communities differ in important ways from North American culture and language and from each other, but are far from representing the vast diversity of infant linguistic experience. Moreover, differences both in the context over which the speech was sampled and the methodological details in the ratings collection limit our ability to directly compare the findings across the two analyses and identify with certainty the reason for the differential findings. In particular, scripted speech (e.g., Singh et al., 2002) or semi-structured, often object-focused, activities such as those used in the elicitation task with the Farsi sample (e.g., Fernald, 1989; The ManyBabies Consortium, 2020) underly much of the work investigating CDS in Western contexts. However, the register is recognizably present in other data types, including brief free-play sessions (e.g., Kitamura et al., 2001) and daylong recordings similar to those used in our Tseltal Mayan sample (e.g., Bergelson et al., 2019), which cross a range of at-home naturalistic circumstances.

Nonetheless, these three analyses, together with the prior work by Bryant and colleagues, are an important first step in teasing apart these thorny questions of the universality of CDS and the relationship between CDS and the perception of affect. With the above limitations in mind, our findings suggest that the answer to these questions is not straightforward. Our findings are consistent with the decades-long literature on the acoustic and linguistic characterizations of CDS itself (e.g. Fernald et al., 1989; Soderstrom, 2007)—we see evidence both for shared properties and variation across languages in the crosslanguage perception of CDS. Gaining a window into the extent to which true "universals" may be established will require a much broader and systematic examination across different language and cultural typologies. Our findings also suggest that perception of CDS (and affect) outside of one's own language may be highly sensitive to the context in which the speech was collected. This may be particularly true due to the laboratory-based, decontextualized conditions in which our raters made their judgments. Our findings further suggest that affect does affect raters' perception of CDS, but not in a simple way–contrary to our expectations, we did not find a consistent relationship across either study between ratings of happiness and raters' perception of CDS.

One additional question left unanswered by the research so far is the specific characteristics adult raters use to make their judgments. Our starting assumption is that pitch characteristics play a primary role in these judgments, although the specific acoustic features underlying this relationship remain unidentified. Both the current study and the prior work by Bryant and colleagues, in providing evidence for cross-language identification, rule out the possibility that an understanding of the substantive content of the speech (e.g., topic) are necessary. However, it is possible that other characteristics such as articulatory/phonetic features may play a role.

In sum, elements of CDS appear discriminable across vastly different languages and cultural contexts, and this discrimination is tied to affective characteristics of the speech. However, the relationship between affective and other properties of speech and the characterization of CDS is complex and highly context-sensitive.

## DATA AVAILABILITY STATEMENT

The datasets and analytic scripts used in this study can be found https://github.com/BLLManitoba/LabellingPaperData2020.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Psychology-Sociology Research Ethics Board, University of Manitoba. The participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

AS and MC collected and contributed the speech samples. AB and JB conducted Experiment 1. MG and SM conducted Experiment 2. JB conducted the analyses. MS, JB, and MC were primarily responsible for writing the manuscript. All authors approved the final draft.

## FUNDING

MS was supported by funding from the Social Sciences and Humanities Research Council of Canada (Insight Grant: 435-2015-062; Digging Into Data award: 869-2016-0003). MC was supported by an NWO Veni Innovational Scheme Grant (275-89-033).

## ACKNOWLEDGMENTS

We were grateful for the contributions of the participant families who allowed us to record their interactions with their infants and the support of the many research assistants who contributed to this project.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.708887/full#supplementary-material

## REFERENCES

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Bergelson, E., Casillas, M., Soderstrom, M., Seidl, A., Warlaumont, A. S., and Amatuni, A. (2019). What do North American babies hear? a large-scale cross-corpus analysis. *Dev. Sci.* 22:e12724. doi: 10.1111/desc.12724

Bernstein Ratner, N. B., and Pye, C. (1984). Higher pitch in BT is not universal: acoustic evidence from Quiche Mayan. *J. Child Lang.* 11, 515–522. doi: 10.1017/S0305000900005924

Broesch, T., and Bryant, G. A. (2018). Fathers' infant-directed speech in a small-scale society. *Child Dev.* 89, e29–e41. doi: 10.1111/cdev.12768

Broesch, T. L., and Bryant, G. A. (2015). Prosody in infant-directed speech is similar across western and traditional cultures. *J. Cogn. Dev.* 16, 31–43. doi: 10.1080/15248372.2013.833923

Brown, P. (2011). "The cultural organization of attention," in *Handbook of Language Socialization*, eds A. Duranti, E. Ochs, and and B. B. Schieffelin (Malden, MA: Wiley-Blackwell), 29–55.

Brown, P. (2014). "The interactional context of language learning in Tzeltal," in *Language in Interaction: Studies in Honor of Eve V. Clark*, eds I. Arnon, M. Casillas, C. Kurumada, and B. Estigarribia (Amsterdam, NL: John Benjamins), 51–82.

Bryant, G. A., and Barrett, H. C. (2007). Recognizing intentions in infant-directed speech: evidence for universals. *Psychol. Sci.* 18, 746–751. doi: 10.1111/j.1467-9280.2007.01970.x

Bryant, G. A., Liénard, P., and Clark Barrett, H. (2012). Recognizing infant-directed speech across distant cultures: evidence from Africa. *J. Evol. Psychol.* 10, 47–59. doi: 10.1556/JEP.10.2012.2.1

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

Casillas, M., Brown, P., and Levinson, S. C. (2020a). Early language experience in a Papuan community. *J. Child Lang.* 1–23.

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

Chronaki, G., Wigelsworth, M., Pell, M. D., and Kotz, S. A. (2018). The development of cross-cultural recognition of vocal emotion during childhood and adolescence. *Sci. Rep.* 8, 1–17. doi: 10.1038/s41598-018-26889-1

Farran, L. K., Lee, C. C., Yoo, H., and Oller, D. K. (2016). Cross-cultural register differences in infant-directed speech: an initial study. *PLoS ONE* 11:e0151518. doi: 10.1371/journal.pone.0151518

Ferguson, C. A. (1978). "Talking to children: a search for universals," in *Universals of Human Language: Vol. 1 Method and Theory*, ed J. H. Greenberg (Stanford, CA: Stanford University Press), 203–224.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: is the melody the message? *Child Dev.* 60, 1497–1510. doi: 10.2307/1130938

Fernald, A. (1992). "Human vocalizations to infants as biologically relevant signals: An evolutionary perspective," in *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, eds J. Barkow and L. Cosmides (New York, NY: Oxford University Press), 391–428.

Fernald, A., and Morikawa, H. (1993). Common themes and cultural variations in Japanese and American mothers' speech to infants. *Child Dev.* 64, 637–656. doi: 10.2307/1131208

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., and Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *J. Child Lang.* 16, 477–501. doi: 10.1017/S0305000900010679

Gaskins, S. (1990). *Mayan exploratory play and development* (Ph.D. dissertation). Department of Education (Educational Psychology), University of Chicago, Chicago, IL, United States.

Gaskins, S. (2006). "Cultural perspectives on infant-caregiver interaction," in *Roots of Human Sociality: Culture, Cognition, and Interaction*, eds N. J. Enfield and S. C. Levinson (Oxford: Berg), 279–298.

Gendron, M., Roberson, D., and Barrett, L. F. (2015). Cultural variation in emotion perception is real: a response to Sauter, Eisner, Ekman, and Scott (2015). *Psychol. Sci.* 26, 357–359. doi: 10.1177/0956797614566659

Golinkoff, R. M., Can, D. D., Soderstrom, M., and Hirsh-Pasek, K. (2015). (Baby) talk to me: the social context of infant-directed speech and its effects on early language acquisition. *Curr. Dir. Psychol. Sci.* 24, 339–344. doi: 10.1177/0963721415595345

Golinkoff, R. M., Hoff, E., Rowe, M. L., Tamis-LeMonda, C. S., and Hirsh-Pasek, K. (2019). Language matters: denying the existence of the 30-million-word gap has serious consequences. *Child Dev.* 90, 985–992. doi: 10.1111/cdev.13128

Heath, S. B. (1983). "Learning how to talk in Trackton (Ch. 3)," in *Ways With Words* (Cambridge: Cambridge University Press), 73–112.

Ingram, D. (1995). The cultural basis of prosodic modifications to infants and children: a response to Fernald's universalist theory. *J. Child Lang.* 22, 223–233. doi: 10.1017/S0305000900009715

Kalashnikova, M., Carignan, C., and Burnham, D. (2017). The origins of babytalk: smiling, teaching or social convergence? *R. Soc. Open Sci.* 4:170306. doi: 10.1098/rsos.170306

Kitamura, C., Thanavishuth, C., Burnham, D., and Luksaneeyanawin, S. (2001). Universality and specificity in infant-directed speech: pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant Behav. Dev.* 24, 372–392. doi: 10.1016/S0163-6383(02)00086-3

LeVine, R. A., Dixon, S., LeVine, S., Richman, A., Leiderman, P. H., Keefer, C. H., et al. (1994). *Child Care and Culture: Lessons From Africa*. Cambridge: Cambridge University Press.

Ochs, E. (1982). Talking to children in Western Samoa. *Lang. Soc.* 11, 77–104. doi: 10.1017/S0047404500009040

Ochs, E., and Schieffelin, B. B. (1984). "Language acquisition and socialization: three developmental stories and their implications," in *Culture Theory: Essays on Mind, Self, and Emotion*, eds R. A. Schweder and R. A. LeVine (Cambridge: Cambridge University Press), 276–320.

Ochs, E., and Schieffelin, B. B. (2011). "The theory of language socialization," in *The Handbook of Language Socialization*, eds A. Duranti, E. Ochs, and B. B. Schieffelin (Chichester: A John Wiley & Sons, Ltd),1–21.

Pye, C. (1986). Quiché Mayan speech to children. *J. Child Lang.* 13, 85–100. doi: 10.1017/S.0305000900000313

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Ramírez-Esparza, N., García-Sierra, A., and Kuhl, P. K. (2014). Look who's talking: speech style and social context in language input to infants are linked to concurrent and future speech development. *Dev. Sci.* 17, 880–891. doi: 10.1111/desc.12172

Rogoff, B., Mistry, J., Gönc,ü, A., Mosier, C., Chavajay, P., and Heath, S. B. (1993). Guided participation in cultural activity by toddlers and caregivers. *Monogr. Soc. Res. Child Dev.* 58, 1–179. doi: 10.2307/1166109

Sauter, D. A., Eisner, F., Ekman, P., and Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proc. Nat. Acad. Sci. U.S.A.* 107, 2408–2412. doi: 10.1073/pnas.0908239106

Sauter, D. A., Eisner, F., Ekman, P., and Scott, S. K. (2015). Emotional vocalizations are recognized across cultures regardless of the valence of distractors. *Psychol. Sci.* 26, 354–356. doi: 10.1177/0956797614560771

Schieffelin, B. B., and Ochs, E. (1986). Language socialization. *Annu. Rev. Anthropol.* 15, 163–191. doi: 10.1146/annurev.an.15.100186.001115

Shute, B., and Wheldall, K. (1989). Pitch alterations in British motherese: some preliminary acoustic data. *J. Child Lang.* 16, 503–512. doi: 10.1017/S0305000900010680

Singh, L., Morgan, J. L., and Best, C. T. (2002). Infants' listening preferences: baby talk or happy talk? *Infancy* 3, 365–394. doi: 10.1207/S15327078IN0303_5

Soderstrom, M. (2007). Beyond babytalk: re-evaluating the nature and content of speech input to preverbal infants. *Dev. Rev.* 27, 501–532. doi: 10.1016/j.dr.2007.06.002

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

Stross, B. (1972). Verbal processes in Tzeltal speech socialization. *Anthropological Linguist.* 14, 1–13.

The ManyBabies Consortium (2020). Quantifying sources of variability in infancy research using the infant-directed-speech preference. *Advan. Methods Pract. Psychol. Sci.* 3, 24–52. doi: 10.1177/2515245919900809

Weisleder, A., and Fernald, A. (2013). Talking to children matters: early language experience strengthens processing and builds vocabulary. *Psychol. Sci.* 24, 2143–2152. doi: 10.1177/0956797613488145

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). "ELAN: a professional framework for multimodality research," in *Proceedings of the Fifth International Conference on Language Resources and Evaluation* (Genoa), 1556–1559.

Zeidner, M. (1983). 'Kitchie-koo' in modern Hebrew: the sociology of Hebrew baby talk. *Int. J. Soc. Lang.* 41, 93–113. doi: 10.1515/ijsl.1983.41.93

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 09 September 2021
doi: 10.3389/fpsyg.2021.723366

# The Effects of a Parent-Implemented Language Intervention on Late-Talkers' Expressive Skills: The Mediational Role of Parental Speech Contingency and Dialogic Reading Abilities

*Chiara Suttora[1]*, Mariagrazia Zuccarini[1], Arianna Aceti[2,3], Luigi Corvaglia[2,3], Annalisa Guarini[1†] and Alessandra Sansavini[1*†]*

[1] *Department of Psychology "Renzo Canestrari", University of Bologna, Bologna, Italy,* [2] *Department of Medical and Surgical Sciences, University of Bologna, Bologna, Italy,* [3] *Neonatology and Neonatal Intensive Care Unit, S. Orsola-Malpighi Hospital, Bologna, Italy*

OPEN ACCESS

**Edited by:**
*Ilaria Grazzani, University of Milano-Bicocca, Italy*

**Reviewed by:**
*Veronica Ornaghi, University of Milano-Bicocca, Italy*
*Maria Carmen Usai, University of Genoa, Italy*

***Correspondence:**
*Chiara Suttora*
*chiara.suttora@unibo.it*
*Alessandra Sansavini*
*alessandra.sansavini@unibo.it*

*† These authors share last authorship*

**Specialty section:**
*This article was submitted to Developmental Psychology, a section of the journal Frontiers in Psychology*

**Received:** *10 June 2021*
**Accepted:** *30 July 2021*
**Published:** *09 September 2021*

**Citation:**
*Suttora C, Zuccarini M, Aceti A, Corvaglia L, Guarini A and Sansavini A (2021) The Effects of a Parent-Implemented Language Intervention on Late-Talkers' Expressive Skills: The Mediational Role of Parental Speech Contingency and Dialogic Reading Abilities. Front. Psychol. 12:723366. doi: 10.3389/fpsyg.2021.723366*

Several qualitative and quantitative features of parental speech input support children's language development and may play a critical role in improving such process in late talkers. Parent-implemented interventions targeting late-talkers have been developed to promote children's language outcomes by enhancing their linguistic environment, i.e., parental speech input. This study investigated the effect of a parent-implemented intervention in increasing late talkers' expressive skills through modifications in structural and functional features of parental speech input. Forty-six thirty-one-month-old late talkers differing in their birth condition (either low-risk preterm or full-term) participated in the study with a parent; 24 parent-child dyads received a parent-implemented intervention centered on dialogic reading and focused stimulation techniques, whereas the other 22 dyads constituted the control group. At pre- and post-intervention, dyads took part in a parent-child shared book-reading session and both parental and child's speech measures were collected and examined. Results showed that the intervention positively affected parents' use of responses and expansions of children's verbal initiatives, as well as the parental amount of talking over reading, whereas no structural features of parental input resulted modified. Mediation analyses pointed out that the intervention indirectly enhanced late-talkers' use of verbal types and tokens through changes in parental use of expansions and amount of talking over reading. As birth status was entered as a covariate in the analysis, these findings can be extended to children with different gestational age. We conclude that the parent-implemented intervention was effective in supporting late-talkers' gains in language development as a cascade result of the improvements in parental contingency and dialogic reading abilities. These promising findings suggest to examine not only children and parental outcomes but also the intervention mechanisms promoting changes in late-talkers' language development as a clearer view on such process can inform the development of feasible, ecological and effective programs.

**Keywords: parent-implemented intervention, child-directed speech, expressive language delay, late-talkers, preterm birth**

# INTRODUCTION

## Relationship Between Parental Speech and Child Language Development

The first 1,000 days of life are considered a fundamental time window in which children's developmental trajectories and future outcomes are shaped. Within this period, providing children with nurturing experiences such as responsive caregiving and adequate learning opportunities is vital (Britto et al., 2017). Language stimulation by parents and caregivers is one of these essential nurturing experiences (Golinkoff et al., 2019). Parents usually talk to their infants and children using a particular speech register also known as infant- or child-directed speech (IDS, CDS). Such input has specific prosodic (i.e., pitch, length of sounds, intensity), structural (i.e., quantitative aspects of speech, lexical and syntactic complexity), and functional features (i.e., directiveness, contingency, and tutorial function of parental utterances directed at the child) that make it an optimal input for toddlers developing language; child-directed parental utterances are typically high pitched and modulated, short in their length, built with a simple and redundant lexicon, and contingent to child's communicative bids (Tamis-LeMonda et al., 2001; Hoff and Naigles, 2002; Huttenlocher et al., 2002; Rowe, 2012).

Differences in structural and functional features of parental speech relate to variability in child language learning (Anderson et al., 2021). Concerning structural features, much research on parental input emphasizes the role of input quantity in predicting both children's rate of vocabulary growth and vocabulary skills (Huttenlocher et al., 1991, 2010; Hart and Risley, 1995; Weisleder and Fernald, 2013). Of no lesser importance, input quality, often calculated as word types and input complexity, expressed through the mean length of utterance index (MLU), also accounts for variability in children's lexical outcomes (Hoff and Naigles, 2002; Pan et al., 2005; Huttenlocher et al., 2010; Rowe, 2012; Anderson et al., 2021). Moreover, a very recent study (Silvey et al., 2021) indicates that not the absolute complexity of syntactic input captured in a specific time of development but the extent to which input complexity increases over time predicts children's grammatical outcomes.

Functional features of parental input are also thought to contribute to child language development. Parental ability to respond contingently to children's attentive focus and communicative initiatives is one of these features, with literature findings showing that differences in maternal contingent responding predict children's vocabulary growth (Tamis-LeMonda et al., 2001, 2014). A relevant role in determining children's linguistic outcomes is also played by parental recasts or reformulations of children's linguistic attempts which encompass all those responses in which parents imitate, expand, or reduce children's original verbal utterances (Taumoepeau, 2016). These reformulations are not only inherently contingent to children's initiative but provide them with relevant lexical and syntactic data. Parental repetitions draw a child's attention to his/her own verbal production, allowing a phonological comparison with the adult form; moreover, expansions provide the child with further relevant lexical and syntactic data linked to the original verbal production, exposing him/her to new learning opportunities. The role of structural and functional features of parental input addressed to children with delayed language development has also been investigated (Girolametto et al., 1999, 2002; Vigil et al., 2005; D'Odorico and Jacob, 2006; Levickis et al., 2018; Suttora et al., 2020). The term "late talkers" refers to those children who lag behind in several aspects of language, showing a slower rate of growth in language learning and limited expressive vocabulary (i.e., below the 10th percentile with respect to normative data), in absence of sensory, cognitive or socio-emotional difficulties (Hawa and Spanoudis, 2014). As late talkers represent a significant proportion of 2–3-year-old children—with prevalence ranging from 9 to 21% (Reilly et al., 2007; Korpilahti et al., 2016; Sansavini et al., 2021)—it is relevant from a clinical stance to describe the peculiarities of their linguistic milieu to capture which aspects of parental input could be enhanced and/or modified. With respect to the structural features of parental input, literature addressing late talkers' samples are coherent in finding no significant differences in terms of input quantity (i.e., lexical rate), quality (i.e., lexical diversity), and complexity (i.e., MLU) when this input is compared with that addressed to typically developing children (Paul and Elwood, 1991; Vigil et al., 2005; D'Odorico and Jacob, 2006; Suttora et al., 2020). However, an input characterized by high levels of grammatical complexity, lexical rate, and diversity has been linked to lower abilities in late talkers' spontaneous and reported lexical production (Girolametto et al., 1999; Suttora et al., 2020). According to Girolametto et al. (1999), this latter pattern of associations might be representative of an "idiosyncratic feedback loop," a circle in which children's linguistic impairment negatively affects their parents' input, which in turn constitutes a further complication for children's language improvement. In light of this, regardless of differences in input quality and/or quantity, late talkers could benefit from a less complex input characterized by shorter, simpler, and clearer utterances. As for the functional features of parental input, literature suggests that parents of late talkers are less contingent to their children displaying fewer responses to their children's initiations and fewer expansions than parents talking to their children with typical language skills (Vigil et al., 2005). Again, lower use of expansions in the input directed to late talkers predicts smaller vocabulary and expressive skills (Girolametto et al., 1999, 2002), also when assessed at 2-year distance (Levickis et al., 2018). According to these findings, interventions aimed at enhancing late talkers' linguistics environment, by improving parental responsiveness and expansions while keeping its complexity adjusted to children's communicative skills, might constitute a privileged route to support children with language delays.

## Parent-Implemented Interventions Supporting Late-Talkers' Language Development

Treatment options for late talkers include both direct and indirect interventions. The formers consist of individual treatment delivered by a speech-language therapist (SLT) in a clinical setting and may also involve parents who can be asked to do at-home activities with their child to support the treatment. The efficacy of direct interventions with SLTs is demonstrated by a Cochrane review considering studies involving children with phonological and lexical difficulties (Law et al., 2017).

Indirect interventions are programs in which parents—trained, guided, or supported by SLTs and/or psychologists with speech-language expertise—are the main providers of the treatment. Parent training can be individual or group based. As previously accounted, given the critical role of caregivers in supporting and enhancing their children's language development and the differences highlighted in parental speech directed at children with delayed language acquisition, programs designed to train caregivers how to best support language development are relevant components of effective intervention practices. In this direction, studies aimed at comparing directed versus parent-implemented interventions for late talkers revealed a lack of differences in their efficacy in enhancing children language skills (Roberts and Kaiser, 2011; DeVeney et al., 2017; Tosh et al., 2017) making parent-implemented interventions a valid option for early intervention. In these programs, parents are trained to use specific language and conversational strategies aimed at supporting their children's language learning by enhancing their linguistic environment. Specifically, parents are taught: (a) to follow children's attention and lead during conversation trying to get them focused on the exchange; (b) to increase their responsiveness to children's communicative and verbal initiatives, by recasting, imitating, and expanding their verbal productions; and (c) to limit an excessive use of questioning and/or directiveness in the input they address to them. These interventions can also include parent training on focused stimulation and dialogic book reading. In the first technique, parents are trained to repetitively use few selected target words during play or routine contexts (Girolametto et al., 1996a). In the second, parents are taught how to elicit conversation and turn-taking during a book sharing activity (Buschmann et al., 2009). Among parent-intervention programs, the Hanen Parent Programs (HPP; Manolson, 1992) is one of the most common, directed not only at children with primary language difficulties, but also at children with secondary linguistic issues, such as children with motor disorders, cerebral palsy, or autism spectrum disorder (Pennington and Thomson, 2007; Weitzman, 2013). In HPP caregivers are instructed on how to follow their children's attentional states and how to use specific responsive interaction strategies aimed at supporting children's language learning throughout daily routines. Summarizing, parent-implemented interventions aim at affecting late talkers' language skills through a cascading effect, i.e., because of modifications in their parents' input and conversational strategies (Roberts and Kaiser, 2012).

The efficacy of parent-implemented programs on late talkers' language outcomes is consistent and well-documented. Roberts and Kaiser's meta-analysis (Roberts and Kaiser, 2011) of 18 studies indicated that children participating to parent-implemented interventions scored better than controls in almost all measures of language development—observed and parent-reported—with greater effect sizes for measures of expressive morphosyntax and receptive vocabulary. These findings were confirmed even when entering intellectual disability as a moderator, as seven out of eighteen studies included in the meta-analysis involved children with cognitive disabilities, genetic syndromes, or autism. Narrowing their analysis to studies addressing children with language delay, Tosh et al.'s (2017) review reported very similar conclusions confirming that children enrolled in parent-implemented programs showed more favorable language outcomes than children in the control conditions. Finally, Heidlage et al.'s (2020) meta-analysis on 25 RCT studies indicated that, on average, parent-implemented language interventions have significant effects on children's expressive vocabulary, both when interventions focus on caregiver-child play routines and on book sharing activities.

As regards the effects of parent-implemented programs on caregivers' input and use of conversational strategies, literature findings are fewer and less clear. Roberts and Kaiser's (2011) meta-analysis concluded that parent-implemented programs positively impacted caregivers' outcomes, with particular regard to their responsiveness to children's communicative initiatives. However, among the studies examined, only four studies addressed interventions directed at parents of children with language delay (Girolametto et al., 1996a,b; Law et al., 1999; Baxendale and Hesketh, 2003). In Girolametto et al. (1996a), mothers in the intervention group significantly produced fewer words per minute and shorter utterances than mothers in the control group at post-intervention assessment, demonstrating an adjustment to children's communicative level. These mothers also showed greater use of focused stimulation on target words, which was one of the techniques modeled by the intervention. Baxendale and Hesketh (2003), by contrast, found no differences between parents in a HHP group and parents enrolled in conventional clinic therapy group in the use of expansions and imitation strategies, as for all participants there was a significant increase in the use of strategies such as imitation and expansion from pre- to post-intervention. Law et al. (1999) also failed to find significant effects of the intervention on parental outcomes.

More recent findings shed some light on the effects of parent-implemented interventions on caregivers' input and strategies. Heidlage et al.'s (2020) meta-analysis confirmed a significant effect of parent-implemented interventions on parents' responsiveness but considered this finding as preliminary as it was based on just five studies. Comparing the parent-training Enhanced Milieu Teaching (EMT) to usual care in a sample of 97 parent-child dyads, Roberts and Kaiser (2015) found that caregivers in the EMT group improved in all language facilitation strategies targeted in the intervention, namely the use of turn-taking, responsiveness, expansion and prompting as assessed during a 20 min play-based caregiver-child interaction. Similar results were underscored by Kruythoff-Broekman et al. (2019) comparing the Target Word program (part of the HHP procedure) to a usual-care control group. At 6-month post-intervention, parental language strategies, as measured with a rating scale during a 5-min parent-child interaction, resulted significantly improved, with an increase in the use of interactive strategies and a decrease of parental utterances aimed at putting pressure on the child. Additional results of Kruythoff-Broekman et al.'s (2019) study revealed that children whose parents reduced pressing behaviors significantly improved their expressive vocabulary and expressive syntax, suggesting a cascading effect of the modifications observed in parental input on children's gains in language development.

In short, although effects of parent-implemented interventions have been documented both on children linguistic outcomes and, to a lesser extent, on parental input and strategies, studies expressly addressing the effects of such interventions on children's gains in language skills through modifications in caregivers' use of input and conversational strategies have not been performed yet. This is our study's main intent.

## Aims of the Study

The present study aimed at investigating the effect of a parent-implemented language intervention in enhancing structural and functional features of parental communicative input to their own late talking children in the third year of life and eventually triggering positive cascading effects on children's lexical and grammatical skills.

Firstly, we investigated whether the intervention based on dialogic book reading impacted: (a) structural features of parent speech, such as lexical diversity, rate, and grammatical complexity; (b) functional features of parental input, such as the ability to respond contingently to their own child's verbal initiatives—by reformulating child's speech productions—and to engage the child in a conversation during the book sharing activity. As the intervention was mainly focused on promoting functional features of parental conversation, we expected to find a more significant impact of the intervention on these features rather than on structural ones (i.e., lexical diversity, rate, and grammatical complexity).

Secondly, we investigated the effects of the intervention on children's advances in language development. A significant increase of expressive lexicon in children's spontaneous speech, as regards lexical and grammatical measures, was expected as suggested by previous works that documented the efficacy of the same intervention on measures of children's lexical and grammatical skills collected through parental reports (Bello et al., 2019; Zuccarini et al., 2020). We hypothesized that this effect was triggered by parental input improvement determined by the intervention. As the intervention mainly addressed functional features of parental input, we expected that significant changes in these features would, in turn, positively impact on child's language development. As the intervention was provided to a group of children differing for birth condition (i.e., low-risk preterm and full-term) this variable was controlled in our analyses.

# MATERIALS AND METHODS

## Participants

Sixty-two parents with their late-talking children were invited to participate in the study. Fifty-nine out of them accepted to be enrolled in the study. Criteria of inclusion in the study consisted in children being monolingual or mainly exposed to the Italian language from birth onward, being either full-term (i.e., with a gestational age $\geq$ 37 weeks) or low-risk preterm (i.e., with a gestational age $<$ 37 weeks) and not having any severe neurological impairment and/or congenital malformations, visual, hearing, or motor impairments, or severe neonatal complications, or severe cognitive deficits (Bayley-III cognitive score $<$ 70).

With a convenience sample methodology parents were asked whether they would participate in the intervention condition. Thirty-one parents accepted to participate and 28 declined the invitation and were assigned to the control condition. Parental speech to the child and child spontaneous vocal productions were assessed during two assessments conducted at the Developmental Psychology Lab at the University of Bologna when children were around 31-month-old ($M$ = 31.13, $SD$ = 1.20)—pre-intervention assessment—and 37-month-old ($M$ = 37.06, $SD$ = 1.47)—post-intervention assessment. The parent-implemented language intervention lasted approximately 2 months, between the pre- and post-assessment. Eleven participants, 6 in the intervention and 5 in the control group, did not attend or complete the post-intervention assessment and were thus excluded from the data analysis. Other 2 dyads, one from the intervention and one from the control group, were also excluded as the parent who attended the pre and the post-intervention was different. Thus, the final sample consisted of 46 parents and their children with 24 parents participating in the intervention and 22 included in the control group. A flow diagram provides an overview of parents participating in the study (see **Supplementary Figure 1**).

The final sample included 17 parents of low-risk preterm children, born before 37 weeks of gestation, at the Sant'Orsola-Malpighi Hospital of the University of Bologna. Perinatal characteristics of the subgroup of low-risk preterm children are reported in **Supplementary Table 1**. The remaining participants ($n$ = 29) were parents of healthy full-term children that were born in the same hospital. Parents of low-risk preterm and full-term children were not equally distributed in the intervention and control groups with proportionally more parents of low-risk preterm children participating in the intervention (intervention $n$ = 12; control $n$ = 5) compared to parents of full-term children (intervention $n$ = 12; control $n$ = 17), $\chi^2(1, N = 46) = 3.66$, $p$ = 0.05.

**TABLE 1 |** Sociodemographic characteristics of participants in the entire sample.

| Participants' characteristics | Intervention group ($n$ = 24) | Control group ($n$ = 22) | $\chi^2/t$ (df) | $p$ |
|---|---|---|---|---|
| Gestational age (weeks), mean (SD) | 37.38 (3.12) | 38.03 (2.76) | 0.74 (44) | 0.464 |
| Birthweight (grams), mean (SD) | 2782.58 (942.69) | 2939.82 (648.12) | 0.64 (44) | 0.524 |
| Length of stay in hospital (days), mean (SD) | 13.58 (35.62) | 4.36 (5.63) | −1.19 (44) | 0.237 |
| Gender (Female), *n* (%) | 10 (41.7) | 7 (31.8) | 0.48 (1, 46) | 0.489 |
| Firstborn, *n* (%) | 14 (58.3) | 8 (36.4) | 3.08 (1, 46) | 0.214 |
| Twins, *n* (%) | 4 (16.7) | 4 (18.2) | 0.02 (1, 46) | 0.892 |
| Otitis media, *n* (%) | 1 (4.2) | 2 (9.1) | 0.46 (1, 46) | 0.499 |
| Family history of language and/or learning disorders (LLD), *n* (%) | 6 (25.0) | 4 (18.2) | 0.31 (1, 46) | 0.575 |
| Child-care center attendance, *n* (%) | 21 (87.5) | 13 (59.1) | 4.80 (1, 46) | **0.028** |
| Other parent input besides Italian, *n* (%) | 2 (8.3) | 1 (4.5) | 0.23 (1, 46) | 0.632 |
| Mother's age (years), mean (SD) | 40.04 (5.20) | 35.98 (4.69) | −2.79 (44) | **0.008** |
| Father's age (years), mean (SD) | 42.00 (5.09) | 38.70 (5.59) | −1.98 (41) | 0.055 |
| Mothers with high educational level (>13 years), *n* (%) | 17 (70.8) | 13 (59.1) | 0.70 (1, 46) | 0.404 |
| Fathers with high educational level (>13 years), *n* (%) | 11 (46.8) | 9 (40.9) | 0.11 (1, 46) | 0.736 |
| Mother's nationality (Italian), *n* (%) | 23 (95.5) | 21 (95.8) | 0.01 (1, 46) | 0.950 |
| Father's nationality (Italian), *n* (%) | 23 (95.5) | 21 (95.8) | 0.01 (1, 46) | 0.950 |
| Age at pre-intervention (months), mean (SD) | 30.86 (1.44) | 31.30 (1.06) | 1.65 (44) | 0.250 |
| Age at post-intervention (months), mean (SD) | 37.02 (1.44) | 37.13 (1.15) | 0.29 (44) | 0.777 |
| Pre and post-intervention interval (days), mean (SD) | 187.29 (52.21) | 177.36 (40.75) | −0.71 (44) | 0.479 |

*Significant results are displayed in bold.*

Biological and sociodemographic characteristics of children and parents in the intervention and control groups are described and compared in **Table 1**. The same table displays information about the age of children at the pre- and post-intervention assessment, as well as a measure of the time interval between pre- and post-intervention assessment. For children born preterm, age was corrected for weeks of prematurity to consider their level of neurobiological maturation as done in previous studies (Sansavini et al., 2011). Children in the intervention and control groups were similar in mostly all sociodemographic variables, with the only exception of their attendance to child-care centers that was higher for children in the intervention group. With regard to parental variables, mothers in the intervention group were significantly older than mothers in the control group.

## Procedure and Study Design

Children identified as late talkers—having an expressive vocabulary size at or below the 10th percentile for their age—through the use and normative values of the Italian version of the MacArthur Bates Communicative Development Inventories (MB-CDI), Words and Sentences Complete Form (Caselli et al., 2015; Rinaldi et al., 2019) were invited, around 31 months of age, with their parents at the Developmental Psychology Lab at the University of Bologna for an assessment of their communicative exchanges. The MB-CDI served as a tool to identify children as late-talkers. Dyads were observed and videotaped during a parent-child shared book-reading session during which both partners' speech was collected. One parent, more often the mother (except for two children whose father participated in the study), was asked to interact with his/her child by sharing two age-appropriate picture books seated at a child-table. Parents included in the intervention condition attended six 2-h intervention sessions with a trained psychologist. To test the effectiveness of the intervention, a pre-post-intervention assessment was used. Thus, parent-child dyads were invited, when children were around 37 months of age, to participate in a second videotaped book reading session. The pre-intervention session lasted on average 10 min (*SD* = 84 s), the post-intervention session approximately 9 min and 54 s (*SD* = 146 s).

## Parent-Administered Intervention Program

A 2-month-lenght parent-administered manualized intervention, named "Oltre il libro" (Girolametto et al., 2017), was used in the study. This is a dialogic book reading program consisting of 6 training sessions, of about 2 h each, directed at small groups of parents, normally 4–6 people per group. The intervention program is theoretically based on the interactive model of language intervention and it aims at fostering children language development by coaching parents in the use of different conversational strategies during book reading. The intervention aims at coaching parents for: (a) fostering turn-taking skills and promoting the use of extra-verbal cues as intonation, rhythm, and gestures; (b) adjusting their speech to their child's linguistic skills using simple sentences and redundant lexicon; (c) using, besides close-ended questions, open-ended wh-questions (e.g., "where is Anna hidden?" "why is the elephant sad?"); (d) implementing focused stimulation on target words that are already understood but not produced by the child yet; (e) expanding their child's verbal production (e.g., the child says "elephant" and the parent replies "yes, the elephant is sad as it cannot find a place to draw").

## Tools

Child's expressive vocabulary was assessed with the Words and Sentences Complete Form of the Italian MB-CDI (Caselli et al., 2015), that is a valid and reliable tool to investigate child lexical production and grammatical skills, as indicated by its widespread use in clinical contexts and empirical studies (Sansavini et al., 2019; Majorano et al., 2020; Zuccarini et al., 2020).

## Transcription and Coding

Parental speech directed to the child and child spontaneous speech productions observed during the video-recorded sessions were transcribed into CHAT format of the Child Language Data Exchange System (CHILDES, MacWhinney, 2000) by an experienced speech therapist blind to study hypotheses and child's age. The transcription unit was the utterance that was defined as any speech production, a vocal sound, a single word, or a sequence of words, delimited by a pause, a conversational turn, or a change in the intonation pattern (Craig, 1982). With respect to the child's speech, a vocal production was considered verbal and transcribed as a word when at least three of the following criteria were met: (a) occurred at least two times; (b) was phonetically similar to the target word; (c) had a specific referent; and (d) was recognized as a word by the parent during the exchange (Vihman and McCune, 1994). All the vocal productions that did not meet these criteria were transcribed in IPA and classified as unintelligible in transcriptions.

## Structural Features of Parental and Child's Speech

Once transcribed, child-directed parental utterances and child's speech production were analyzed with the CLAN software and different measures were obtained. Onomatopoeic productions as well as interjections and unintelligible speech were excluded from these analyses. CLAN automated analysis of the transcriptions generated the following indexes of quantity and complexity of parental and child's speech input: (a) the frequency of word Types as an index of lexical diversity; (b) the frequency of word Tokens as a measure of lexical rate; (c) the mean length of utterances (MLU), i.e., the ratio of words over utterances, as a measure of speech grammatical complexity. Raw frequencies were converted in rate per 10 min to control for session's length. Finally, to obtain measures of the change between pre- and post-intervention, deltas were computed for the abovementioned indexes of parental and child's speech by subtracting from values observed at post-intervention those computed at pre-intervention.

## Functional Features of Parental Input

A further analysis of the transcripts was conducted using CHIP, a CLAN software for the automatic coding and analysis of parent-child conversational interactions. CHIP automatically compares pairs of utterances in which the first is considered the source and the following the response. Through this comparison the software creates a series of additional tiers in the transcript in which responses or self-repetitions are examined. In this study, we focused on the adult tiers (*%adu)* in which parent's responses to child's utterances are evaluated. As parents can reply to his/her child in more than one turn following child's speech production, CHIP command searches parental responses within a six utterances window. In the present data the average distance between child's source and parental responses was low ($M = 1.08$, $SD = 0.54$ at pre-intervention; $M = 1.36$, $SD = 0.71$ at post-intervention). According to the study's main aims the following indexes were considered: (a) *Total Responses*, i.e., the total number of parental responses to child's utterances; (b) *Exacts*, i.e., the number of exactly matching responses (e.g., the child says "hat" and the parent replies "hat"); (c) *Reductions*, i.e., the number of responses in which there was an overlap of at least one word in the source and response utterances with deletions but no additions (e.g., the child says "big hat" and the parent replies "hat"; (d) *Expansions*, i.e., the number of responses containing only exact matches and additions (e.g., the child says "hat" and the parent replies "right, the hat!").

Moreover, a measure of the amount of parental talking over reading was computed by dividing the amount of talking tokens for the sum of parental talking and reading word tokens. A high value in Talking over Reading measure indicated that a parent spent most of the session engaging the child in a conversational exchange instead of reading the available books. To measures changes in functional indexes of parental speech from pre- to post-intervention delta measures were calculated.

## Reliability

The first author of this manuscript, who was blind to the child's age and birth status, transcribed 27% (12 at pre-intervention and 13 at post-intervention) of the parent-child sessions to establish transcription reliability. Reliability between the two transcribers was high, with a percent interrater agreement equal to 88% on the segmentation of parents' utterances and of 87% on the transcription of child's vocal utterances.

Interrater reliability for parental measures was assessed using the Intraclass Correlation Coefficient (*ICC*) with high levels of agreement resulting for all parent's measures ($ICCs > 0.85$). Interrater agreement on child's speech coding into intelligible, unintelligible, or mixed utterances was tested computing Cohen's Kappa which resulted equal to 0.83. Interrater reliability was more than substantial. Regarding child's linguistic outcomes (i.e., word types, word tokens, MLU), interrater agreement was achieved by calculating the *ICC*, resulting in optimal values with $ICC = 0.96$.

## Statistical Analyses

Analyses were computed using IBM SPSS Statistics 25 and the macro Process for SPSS (Hayes, 2018). Tests were bilateral with a statistical significance set at 0.05. Preliminary analyses of data distribution revealed that most of the study's variables were not normally distributed (Kolmogorov–Smirnov and Shapiro–Wilk tests, $ps < 0.01$). Therefore, rank transformation was applied to both parental and child measures.

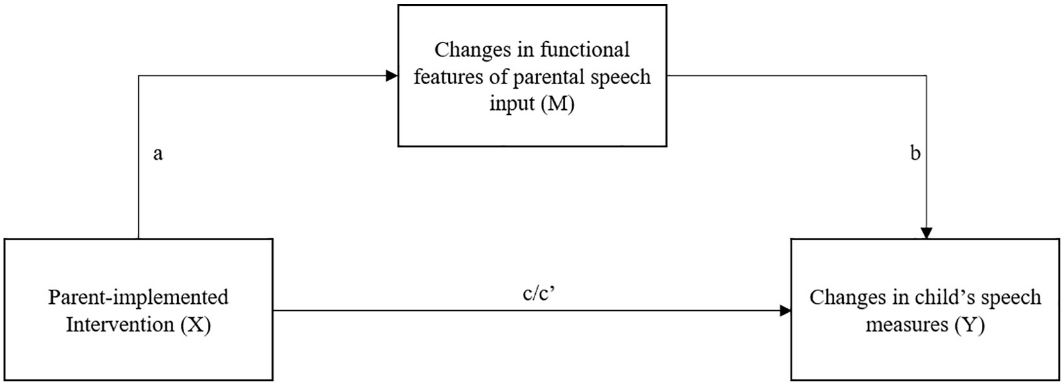



**FIGURE 1 |** Changes in functional features of parental speech input (M) operate mediationally between Parent-implemented Intervention (X) and Changes in child's speech measures (Y).

A set of multivariate MANCOVAs was preliminary carried out to verify that at the pre-intervention assessment parental speech and child's lexical and grammatical measures were comparable between the intervention and the control group. As the presence of low-risk preterm and full-term children differed between groups, birth condition was included as a covariate in the analyses.

To explore the effect of the intervention (intervention vs. control group) on parental structural and functional speech measures and child's linguistic measures over time (pre-intervention vs. post-intervention), several repeated measure MANCOVAs were conducted, controlling for birth condition.

Preliminary Pearson's correlational analyses were carried out to explore the associations between parental and child's speech delta measures. Subsequently, indirect effects of the intervention condition on child's delta speech outcomes through changes in parental speech were tested with mediation analyses using the macro PROCESS, model 4 (Hayes, 2018, p. 585). Unstandardized indirect effects were computed for each of 5,000 bootstrapped samples and the 95% confidence intervals were obtained. In **Figure 1** parental changes in speech input is a mediator (M) of the relationship between the parent-implemented language intervention (X) and child's gain in speech measures from pre- to post-intervention assessment (Y). These latter analyses only included parental measures that resulted significantly affected by the parent-implemented intervention. Again, birth condition was entered in the analyses as a covariate.

## Ethics Statement

The study met ethical guidelines for human subject protections, including adherence to the legal requirements of Italy, and it received formal approval from the Bologna Health Authority's Independent Ethics Committee (numbers of formal approval documents: EM 194/2017/U_ and EM 193–2018_76/2013/U/Sper/AOUBo). All parents gave informed written consent for study participation, data analysis, and data publication. No incentives or benefits were provided to participants.

# RESULTS

## Pre-intervention Assessment

In **Table 2**, the descriptive statistics describing parental and child's speech measures at pre-intervention were reported. No significant differences between parents and children in the control and intervention groups at the pre-intervention assessment were found (for details see **Supplementary Table 2**).

## Effects of the Parent-Implemented Intervention on Parental Speech Outcomes

**Table 2** summarized the descriptive statistics of parental speech outcomes at pre- and post-intervention assessment. Regarding parent's speech structural features (i.e., word types, tokens and MLU) no significant effects of intervention were found with the multivariate test indicating a lack of significant effect [$F_{(3, 41)} = 1.65$, $p = 0.192$, partial $\eta^2 = 0.108$]. Regarding parental speech functional features, the multivariate analysis yielded a significant effect [$F_{(5, 39)} = 2.47$, $p = 0.048$, partial $\eta^2 = 0.241$], with univariate results showing that the intervention significantly influenced *Total responses, Expansions, and Talking over Reading* measures (see **Table 2**). Parents that participated in the intervention showed a significant increase from pre- to post-assessment in the total responses to the child's verbal initiatives and in the use of utterances aimed to expand the child's productions when compared to the control group. Furthermore, in the intervention group, a significant increase in the amount of talking over reading was observed from pre- to post-intervention assessment.

## Effects of the Parent-Implemented Intervention on Child's Spontaneous Speech

The impact of the parent-implemented intervention on child's spontaneous speech outcomes resulted close to statistical significance with the multivariate test. $F_{(3, 41)} = 2.61$, $p = 0.064$, partial $\eta^2 = 0.160$. Considering the univariate results reported

**TABLE 2** | Means and standard deviations M (SD) for parental and child's speech measures at pre- and post-intervention assessment, Δ pre-post assessment, and results of repeated measures MANCOVAs performed on pre- and post-intervention data.

| | Pre-intervention assessment | | Post-intervention assessment | | Δ Pre-post assessment | | Univariate test Time X Intervention | | |
|---|---|---|---|---|---|---|---|---|---|
| | Intervention Group (*n* = 24) | Control Group (*n* = 22) | Intervention Group (*n* = 24) | Control Group (*n* = 22) | Intervention Group (*n* = 24) | Control Group (*n* = 22) | $F(1, 43)$ | $p$ | partial $\eta^2$ |
| **Parental speech—structural features** | | | | | | | | | |
| Types | 202.92 (38.44) | 182.32 (47.91) | 202.55 (45.31) | 186.99 (85.91) | −0.37 (38.86) | 4.67 (89.10) | 0.03 | 0.869 | 0.001 |
| Tokens | 603.13 (151.59) | 563.55 (201.46) | 618.59 (201.59) | 521.59 (243.62) | 15.46 (152.44) | −41.96 (213.25) | 1.17 | 0.285 | 0.026 |
| MLU | 3.25 (0.42) | 3.14 (0.51) | 3.47 (0.61) | 3.11 (0.38) | 0.22 (0.59) | −0.02 (0.37) | 2.47 | 0.124 | 0.054 |
| **Parental speech—functional features** | | | | | | | | | |
| Total Responses | 96.92 (55.57) | 108.05 (57.69) | 137.42 (85.50) | 94.77 (79.33) | 40.50 (97.30) | −13.27 (85.62) | 4.57 | **0.038** | 0.096 |
| Exacts | 3.63 (3.69) | 2.95 (3.40) | 5.00 (3.99) | 3.95 (4.20) | 1.38 (4.19) | 1.00 (3.52) | 0.87 | 0.356 | 0.020 |
| Reductions | 0.29 (0.69) | 0.60 (1.25) | 1.25 (1.80) | 0.59 (1.09) | 0.96 (2.01) | 0.00 (1.77) | 0.01 | 0.975 | 0.001 |
| Expansions | 4.29 (4.92) | 2.64 (3.26) | 9.63 (6.78) | 4.68 (5.35) | 5.33 (8.19) | 2.05 (5.24) | 4.06 | **0.050** | 0.086 |
| Talking over reading | 0.69 (0.14) | 0.71 (0.17) | 0.75 (0.20) | 0.62 (0.26) | 0.05 (0.20) | −0.09 (0.23) | 6.14 | **0.017** | 0.125 |
| **Child's speech** | | | | | | | | | |
| Types | 9.28 (9.09) | 6.04 (4.95) | 26.29 (21.71) | 15.79 (19.69) | 17.02 (23.81) | 9.75 (16.55) | 3.34 | 0.075 | 0.072 |
| Tokens | 6.11 (8.22) | 10.17 (10.84) | 56.60 (40.13) | 50.48 (78.97) | 50.49 (41.92) | 40.31 (76.44) | 7.41 | **0.009** | 0.147 |
| MLU | 1.05 (0.26) | 0.98 (0.33) | 1.36 (0.38) | 1.03 (0.50) | 0.31 (0.42) | 0.05 (0.50) | 3.25 | 0.079 | 0.070 |

*Significant results are highlighted in bold. Parental and child types and tokens are computed as rate per 10 min.*

in **Table 2**, the intervention significantly improved children's production of word tokens from the pre- to the post-intervention assessment. Also, child's production of word types and MLU were observed to increase due to the intervention but with *p*-values implying trends to statistical significance.

## Direct and Indirect Effects of the Parent-Implemented Language Intervention on Child's Speech Outcomes Through Changes of Parental Input

In **Table 2**, descriptive statistics of measure of change over time in parental and child's measures for the intervention and the control group are summarized. The results of Pearson's correlation analyses testing the associations between measures of change of child's and parental speech from pre- to post-intervention assessment are reported in the **Supplementary Materials** (see **Supplementary Table 3**). Almost every measure of change in parental speech—except for MLU—resulted positively and significantly associated to changes in child's word types, tokens, and MLU.

Subsequent analyses focused on the indirect effects of parent-implemented language intervention on child's speech changes in types and tokens production and MLU through changes in parental input, namely parents' use of total responses, expansions, and amount of talking over reading during sessions (see **Table 3** for the results of mediation analyses). Models, in which the indirect effect of the intervention via parental speech modification were significant, are also reported in the **Supplementary Material** (see **Supplementary Figure 2**).

The first set of models (I, II, III) assessed indirect effects of the parent-implemented intervention on child's changes in word types (I), tokens (II) and MLU (III) through parental changes in the use of total responses (see **Table 3**). All models yielded no significant direct and indirect effects, indicating the absence of significant mediation effects of changes in the parental use of total responses on the association between intervention and child's progresses in word types, tokens and MLU.

The second set of models considered parental use of expansions as a mediator between intervention and child's changes in word types (IV), tokens (V), and MLU (VI) (see **Table 3** and **Supplementary Figure 2**). In model IV the indirect effect of the intervention on child's changes in the production of word types resulted significant, whereas the direct effect was not. This implies that changes in the use of expansion by parents totally mediate the effect of the intervention on child's gain in word types. Model V yielded similar results, with the intervention significantly influencing changes in the production of word tokens via changes in parental use of expansions. No indirect effects of the intervention *via* parental expansions were found in model VI including child's MLU as the dependent variable.

The third set of analyses took in exam the indirect effect of intervention on child's changes in word types (VII), tokens (VIII), and MLU (IX) through changes in the amount of parental talking over reading (see **Table 3** and **Supplementary Figure 2**). In model VII and VIII the indirect effects of intervention on

**TABLE 3** | Estimated coefficients (unstandardized B and standardized coefficients a, b, c, c') for mediation model of changes in parental use of total responses, expansions, and amount of talking over reading.

| Model | | Total effect | | | | Direct effect | | | | Indirect effect | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | B | c | SE | 95% CI | B | c' | SE | 95% CI | B | a | b | ab | Boot SE | Boot 95% CI |
| **Δ Total responses** | | | | | | | | | | | | | | | |
| **Δ child's speech** | | | | | | | | | | | | | | | |
| I | Types | 7.61 | 0.27 | 4.06 | [−0.58, 15.79] | 3.64 | 0.53 | 3.62 | [−3.65, 10.94] | 3.96 | 0.55 | 0.53 | 0.29 | 2.56 | [−0.68, 9.48] |
| II | Tokens | 8.31 | 0.62 | 4.02 | **[0.20, 16.41]** | 4.24 | 0.32 | 3.53 | [−2.88, 11.37] | 4.06 | 0.55 | 0.55 | 0.30 | 2.48 | [−0.61, 9.27] |
| III | MLU | 8.87 | 0.66 | 3.90 | **[0.99, 16.74]** | 7.04 | 0.52 | 3.97 | [−0.97, 15.04] | 1.54 | 0.55 | 0.24 | 0.13 | 1.71 | [−0.68, 5.82] |
| **Δ Expansions** | | | | | | | | | | | | | | | |
| **Δ Child's speech** | | | | | | | | | | | | | | | |
| IV | Types | 7.61 | 0.57 | 4.06 | [−0.58, 15.79] | 2.21 | 0.16 | 3.65 | [−5.17, 9.58] | 5.40 | 0.70 | 0.58 | 0.41 | 2.91 | **[0.84, 12.15]** |
| V | Tokens | 8.30 | 0.62 | 4.02 | **[0.20, 16.41]** | 3.88 | 0.29 | 3.86 | [−3.91, 11.66] | 4.43 | 0.70 | 0.47 | 0.33 | 2.61 | **[0.47, 10.46]** |
| VI | MLU | 8.87 | 0.66 | 3.90 | **[0.99, 16.74]** | 8.30 | 0.62 | 4.20 | [−0.18, 16.79] | 0.57 | 0.70 | 0.06 | 0.04 | 1.73 | [−2.51, 4.75] |
| **Δ Talking over reading** | | | | | | | | | | | | | | | |
| **Δ Child's speech** | | | | | | | | | | | | | | | |
| VII | Types | 7.61 | 0.57 | 4.06 | [−0,58, 15.79] | 3.21 | 0.24 | 3.83 | [−4.52, 10.93] | 4.40 | 0.68 | 0.48 | 0.32 | 2.49 | **[0.36, 10.12]** |
| VIII | Tokens | 8.31 | 0.62 | 4.02 | **[0.20, 16.41]** | 3.42 | 0.25 | 3.65 | [−3.94, 10.77] | 4.89 | 0.68 | 0.53 | 0.36 | 2.62 | **[0.54, 11.01]** |
| IX | MLU | 8.87 | 0.66 | 3.90 | **[0.99, 16.74]** | 6.21 | 0.46 | 3.99 | [−1.85, 14.27] | 2.67 | 0.68 | 0.29 | 0.20 | 1.82 | [−0.13, 6.83] |

*Significant results are highlighted in bold.*

child's changes in word types and tokens through changes in the amount of parental talking over reading were statistically significant. Again, no significant direct and indirect effects of the intervention were observed in the model predicting child's MLU through amount of talking over reading. As a proof of concept, all models were further run using child's measures as mediators and parental outcomes as independent measures. No significant indirect effects resulted from these analyses.

# DISCUSSION

The present study aimed to investigate whether a parent-implemented language intervention targeted at parents of late-talkers might enhance the linguistic environment these children are exposed to through parental speech. We were particularly interested in evaluating which structural and functional features of parental speech could best benefit from an intervention based mainly on dialogic reading and focused stimulation techniques. This study further aimed to evaluate whether the intervention was effective in fostering children's advances in language development and, more importantly, whether these effects were direct or mediated by modifications in parental speech features, consistently with the cascading effects of parent training model (Roberts and Kaiser, 2012; Heidlage et al., 2020).

## Intervention Effects Through Changes in Parental Speech

The study results contributed with novel findings to the literature, showing that the parent-implemented language intervention was effective in supporting children's growth in expressive lexicon—with a significant increase in children's lexical diversity (word types) and rate (word tokens)—indirectly, i.e., via the enhancement of functional features of parental speech. Children's speech during parent-child interaction increased in terms of amount of words produced whereas parents demonstrated a growth in their responses to their child's verbal attempts particularly by expanding them, adding extra verbal material able to provide the child with new attributes with respect to the expressed original meaning (e.g., the child says "hat" and the parent replies "that's a very nice hat!"). Parents in the intervention group also showed better dialogic reading skills, spending less time reading to their children and more time conversating with them. The assessment of the direct and indirect effects of the intervention on children's expressive skills indicated that the intervention positively affected parents' use of expansions of children's verbal initiatives and parents' dialogic reading skills which, in turn, positively influenced children's increase in the spontaneous use of word tokens and types. Although parent-implemented interventions assume that children's difficulties in language development can be sustained by improving the quality of their linguistic environment, through changes in the parental input, studies documenting this cascading process are scant as most of the empirical work in this area only investigated a part of this process, namely the effects of intervention on children's gains in language development. Studies addressing whether parent-implemented language interventions increased parents' use of language and conversational strategies supporting language development are rare, as documented by the recent meta-analysis by Heidlage et al. (2020) that reported only five studies investigating this aspect. A recent work contributing to this debate (Kruythoff-Broekman et al., 2019) found that the parent-implemented Target Word program was effective in increasing parents' communicative interaction with their children and in decreasing those behaviors aimed at putting pressure on their children and that this latter reduction, in turn, resulted associated to children's progresses in expressive vocabulary and syntax. Unlike the present study, Kruythoff-Broekman et al.'s (2019) findings were not based on mediation analyses but on correlational models and no direction was tested, i.e., whether changes in parental intrusive behaviors might be a result of children's gains in language development or vice versa.

In our study, parents participating to the "Oltre il libro" intervention exhibited relevant changes in the way they verbally interact with their children, compared to the parents in the control group. Parents receiving the intervention increased their total responses to their children's verbal initiatives and, among total responses, used a greater amount of utterances intended at expanding their children's utterances. A positive impact of parent-implemented interventions on parental responsiveness and use of expansions was suggested by Roberts and Kaiser's (2011) meta-analysis and the more recent Heidlage et al. (2020), although both commented on the lack of strength of their findings due to the paucity of data supporting this conclusion. Our study contributed to reinforcing this finding, emphasizing the role of parent-implemented intervention in stimulating the parental use of total responses and expansions.

An increase of responses contiguous to children's verbal attempts might be determinant for language learning as children—given the temporal connection between their initiatives and parental replies—can more easily make connections between labels and referents available in the context (Tamis-LeMonda et al., 2014). However, in our data, the increase in the use of total responses to children's verbal attempts failed to mediate the effect of the intervention on children's gains in their lexical and grammatical skills. This finding may be explained by considering that responses' contiguity, if not accompanied by semantic contingency, can expose children to contents unrelated to their verbal initiatives, thus not immediately useful for their word learning (Tamis-LeMonda et al., 2014).

Different findings were observed regarding the increase of parental expansions which resulted as a significant mediator of the effects of the intervention on children's advances in lexical diversity (word types) and rate (word tokens). Through expansions, children are provided not only with the repetition of their own verbal production—a feedback mechanism that confirms children their intended meaning and provides them with a phonologically correct version of the production—but they are further exposed to new data as syntactic information is added to children's original verbal production. Moreover, this added material is likely to be semantically contingent to children's verbal attempt, helping them to refine and expand their knowledge about the word and its meaning (Taumoepeau, 2016). Studies addressing both children with typical language development and late talkers showed that parental use of expansions contributed to children's improvement in language development measures (Girolametto et al., 1999, 2002; Levickis et al., 2014, 2018; Taumoepeau, 2016). Positive associations between the use of expansions by parents and advances in language development of their late-talking children were observed when children's language outcomes were assessed either with standardized tools, as in Levickis et al. (2014), or with direct observation of children's spontaneous speech, as in the study of Girolametto et al. (2002).

Besides the significant improvements in the use of total responses and expansions, parents participating in the intervention also showed a significant increase in the talking over reading measure when compared to parents in the control condition. With this measure we intended to capture a parental dialogic reading style, as spending more time talking to the child—using prompts and connections to the child's experiences and wh-questions to elicit a communicative exchange—rather than reading aloud without including the child, represented one of the aspects modeled by the intervention. In this sense, the intervention positively affected parents' dialogic reading that used less verbatim reading of the text engaging their children in more verbal interaction over the shared books, and this change, in turn, favored children's increased use of word types and tokens. Our conclusion is in line with literature findings assessing the effects of dialogic parent-child book reading interventions in promoting children's language and literacy outcomes (Mol et al., 2008; Flack et al., 2018). A meta-analysis by Mol et al. (2008) highlighted that dialogic reading interventions are successful in fostering children's expressive vocabulary with younger children—preschoolers vs. kindergarteners—gaining the best out of these programs. Book reading interventions also resulted beneficial for children with limited expressive vocabularies, as in Hargrave and Sénéchal (2000) study that compared children receiving dialogic reading vs. regular reading interventions, and as in Tsybina and Eriks-Brophy (2010) addressing bilingual preschoolers with slow expressive vocabulary development.

Concerning the structural features of parental speech input our study indicated the absence of significant changes due to the intervention, even if adjustments of the input to match children's language abilities were modeled. Parents taking part in the intervention did not exhibit modifications in the lexical diversity (word types) and rate (word tokens) of the utterances directed to their children during the book sharing interaction nor in their speech grammatical complexity (MLU). Previous works evaluating modifications in parental speech due to parent-implemented interventions mostly took into exam changes in parental responsivity and use of conversational strategies. With regard to structural features of the input results are mixed. Girolametto et al. (1996a) found that the HPP was effective in decreasing maternal input complexity: mothers enrolled in the intervention used a slower rate of words per minute and shorter utterances when assessed at post-intervention. Differently, Roberts and Kaiser's (2011) meta-analysis, reported a lack of significant effects of parent-implemented interventions on parents' rate of speech. Overall, results regarding structural modifications of parental input are scant and inconsistent, and further studies are needed to shed light on this issue.

Summing up, our study brings new evidence of the effectiveness of parent-implemented interventions in affecting late talkers' growth in expressive lexicon—diversity (word types) and rate (word tokens)—highlighting how these effects are mediated by significant modifications in parental use of expansions and dialogic reading skill. Some limits and strengths of the study are discussed below.

## Limitations and Strengths of the Study

Some limitations must be considered. The first regards the lack of randomization in the assignment of the participants to the study's conditions. Participants' inclusion in the intervention or control conditions was performed with a convenience method which might lead to a selection bias regarding differences in motivation and readiness to endorse the intervention and to baseline differences in the target measures. While this latter issue was resolved by controlling for pre-assessment differences, the former cannot be really ruled out. Furthermore, the study sample was limited and, for this reason, we should be cautious in generalizing our findings to the late-talkers' population. Moreover, participants in the intervention and control group differed for their child-care center attendance, as more children in the intervention were enrolled in a child-care program. The opportunity of social engagement with peers and educators in such context could be beneficial for late-talkers' language development, as suggested by Chen et al. (2020) with regard to a peer effect for language development in preschoolers, although no data are available on this topic concerning the role of child-care center attendance on younger children. A second limit concerns the lack of a long-term follow-up assessing the maintenance of the effects of the intervention. Studies addressing long-term effects of parent-implemented interventions are mixed in their findings documenting both long-lasting effects of interventions on children's language scores and abilities and a lack of long-term effects in other cases (Buschmann et al., 2015; Kruythoff-Broekman et al., 2019). Ongoing monitoring of late talkers' language development receiving parent-implemented interventions is relevant to determine the timing and dose of such programs. A final limitation- which was mainly due to our sample size—regards the lack of examination of the role of birth condition in moderating the effects of the intervention on parental and child's speech outcomes. As preterm birth is considered a risk factor both for parent-child interaction (Bilgin and Wolke, 2015; Cambonie et al., 2017) and child's language development (Sansavini et al., 2010), it would be interesting to examine in future studies whether or not it might play a moderator role in this intervention. In this study, this variable was included as a covariate so that our results can apply to both low-risk preterm and full-term child-parent dyads.

The study also presents some relevant strengths. The first regards the nature of speech data collected as both parental and children speech measures were computed through a fine-grained analysis of the sessions' transcripts using a set of software specifically developed for such a complex material. Differently from other works in the field parents' total responses, reformulations and reading style were computed directly from the transcript and not assessed by using rating scales or global evaluations (Kruythoff-Broekman et al., 2019), and the same occurred for the analysis of child's spontaneous speech (Zuccarini et al., 2020). Another relevant strength regards the selected data analysis strategy and, in particular, the inclusion of models testing the direct and indirect effects of the parent-implemented intervention on children's advances in lexical and grammatical skills. Although this kind of analysis is neither new nor scarcely used, this is the first time that such analysis is used to test parent-implemented intervention effectiveness.

## Implications for Practice

The main implication concerns the effectiveness of parent-implemented language intervention in supporting the language development of children showing a language delay. Although effects of such interventions are largely documented, our study shed light on the mechanisms of such effectiveness showing that modifications in parent use of expansions and dialogic reading abilities have cascading effects on children's vocabulary growth and rate of speech. These results point to emphasize those features of parental training directed at enhancing functional features of parental speech input rather than at modifying structural features of such register. These findings can also inform other kinds of intervention, such as those delivered by child-care programs, and lead to the identification of language activities that can support language development of children attending child-care centers and preschool with poor language skills or at risk of delays.

# CONCLUSION

In this study, a parent-implemented intervention based on two main techniques, i.e., dialogic book reading and focused stimulation, was administered through six sessions during a two-month-period. This low-dosage intervention resulted in significant modifications of parental speech and children's expressive outcomes at post-intervention assessment; our results showed that modifications in parents' use of expansions and dialogic reading abilities mediated the effects of the intervention on children's increase in lexical diversity and rate of words during a parent-child interaction. Our findings call for greater attention not only to children's and parental outcomes but also to the intervention mechanisms promoting late-talkers' linguistic advances and stress that the parental ability to engage in a conversation over a shared focus—the book's storyline in our case—and respond contingently to child's verbal attempts should be central in designing ecological and effective parent-implemented language interventions.

# DATA AVAILABILITY STATEMENT

The dataset presented in this article is not readily available because it includes sensitive information about minors with developmental vulnerabilities. Requests to access the dataset should be directed to corresponding authors.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Bologna Health Authority's Independent Ethics Committee (numbers of formal approval documents: EM 194/2017/U_ and EM 193–2018_76/2013/U/Sper/AOUBo). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## AUTHOR CONTRIBUTIONS

CS, AS, AG, and MZ: conceptualization and methodology. MZ: data collection. AA and LC: medical aspects of methodology and medical data collection and supervision. CS: data transcription, data analysis. CS and MZ: data coding and data curation. CS, AS, and AG: writing—original draft and review and editing. AS and AG: funding acquisition. AS: project administration and supervision. All authors have read and approved the manuscript.

## FUNDING

This work was supported by grants from (a) AlmaIdea 2017 Grant Senior, University of Bologna, "Ritardo di linguaggio nei bambini nati pretermine: screening, valutazione e intervento" ("Language delay in preterm children: Screening, assessment and intervention") awarded to AS as main PI, and (b) Italian Ministry of Education, University and Research MIUR PRIN 2017 (2017HRCPE4_004) "Early markers of Language-Learning Impairment," awarded to AS, as PI of the Unit of Bologna. In addition, it was supported by grants from (c) Con i Bambini Impresa Sociale, "Trame Educative," awarded to AG as PI of the Unit Department of Psychology, University of Bologna.

## ACKNOWLEDGMENTS

We are grateful to the families and infants who participated in this study. We are also grateful to the health professionals of the Neonatology and Neonatal Intensive Care Unit—S. Orsola- Malpighi Hospital, University of Bologna—and, in particular, to Giacomo Faldella, for his support in funding acquisition and supervision of medical aspects of methodology, Rosina Alessandroni, for medical care of the preterm sample, and Silvia Savini for her support in contacting parents of preterm children. We would also like to thank the speech-language therapist Cecilia Gorini, for the transcription of parent and child speech into CHAT of CHILDES. Furthermore, we are grateful to Luigi Girolametto, University of Toronto, Canada, Arianna Bello, University of Roma Tre, Italy, and Maria Cristina Caselli, ISTC-CNR Rome, Italy, for discussing together about the structure and implementation of the "Oltre il libro" intervention program for supporting language development in late talkers. We also thank Lorena Remi for sharing with us her experience in implementing the "Oltre il libro" intervention program.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.723366/full#supplementary-material

**Supplementary Figure 1 |** Flow diagram of late talkers whose parents participated in the study.

**Supplementary Figure 2 |** Parental changes in expansions (models IV and V) and talking over reading (models VII and VIII) as mediators of the parent-implemented intervention effect on child's speech types and tokens.

**Supplementary Table 1 |** Clinical and Perinatal Characteristics of the Low-Risk Preterm Children.

**Supplementary Table 2 |** Descriptive statistics of parental and child's speech measure at the pre-intervention assessment for the intervention and the control group. The table also summarizes the results of multivariate MANCOVAs addressing groups differences at pre-intervention.

**Supplementary Table 3 |** Pearson's correlation coefficients among study's variables of change in parental and child's speech.

## REFERENCES

Anderson, N. J., Graham, S. A., Prime, H., Jenkins, J. M., and Madigan, S. (2021). Linking quality and quantity of parental linguistic input to child language skills: a meta-analysis. *Child Dev.* 92, 484–501. doi: 10.1111/cdev.13508

Baxendale, J., and Hesketh, A. (2003). Comparison of the effectiveness of the Hanen Parent Programme and traditional clinic therapy. *Int. J. Lang. Commun. Disord.* 38, 397–415. doi: 10.1080/1368282031000121651

Bello, A., Onofrio, D., Remi, L., Caselli, M. C., and Girolametto, L. (2019). La lettura dialogica per genitori di bambini con ritardo di linguaggio di 2-3 anni. *Psicol. Clin. Dello Sviluppo* 23, 159–168. doi: 10.1449/93139

Bilgin, A., and Wolke, D. (2015). Maternal sensitivity in parenting preterm children: a meta-analysis. *Pediatrics* 136, e177–e193. doi: 10.1542/peds.2014-3570

Britto, P. R., Lye, S. J., Proulx, K., Yousafzai, A. K., Matthews, S. G., Vaivada, T., et al. (2017). Nurturing care: promoting early childhood development. *Lancet* 389, 91–102. doi: 10.1016/S0140-6736(16)31390-3

Buschmann, A., Jooss, B., Rupp, A., Feldhusen, F., Pietz, J., and Philippi, H. (2009). Parent based language intervention for 2-year-old children with specific expressive language delay: a randomised controlled trial. *Arch. Dis. Child.* 94, 110–116. doi: 10.1136/adc.2008.141572

Buschmann, A., Multhauf, B., Hasselhorn, M., and Pietz, J. (2015). Long-term effects of a parent-based language intervention on language outcomes and working memory for late-talking toddlers. *J. Early Interv.* 37, 175–189. doi: 10.1177/1053815115609384

Cambonie, G., Muller, J. B., Ehlinger, V., Roy, J., Guédeney, A., Lebeaux, C., et al. (2017). Mother-infant interaction assessment at discharge and at 6 months in a French cohort of infants born very preterm: the OLIMPE study. *PLoS One* 12:e0188942. doi: 10.1371/journal.pone.0188942

Caselli, M. C., Bello, A., Rinaldi, P., Stefanini, S., and Pasqualetti, P. (2015). *Il Primo Vocabolario del Bambino: Gesti, Parole e Frasi. Forme Lunghe e Forme Brevi del Questionario e Valori di Riferimento per la Fascia 8–36 Mesi*. Milan: Franco Angeli.

Chen, J., Justice, L. M., Tambyraja, S. R., and Sawyer, B. (2020). Exploring the mechanism through which peer effects operate in preschool classrooms to influence language growth. *Early Child. Res. Q.* 53, 1–10. doi: 10.1016/j.ecresq.2020.02.002

Craig, H. K. (1982). Assessing language production in children: experimental procedures. Jon F. Miller. Baltimore: University Park Press, 1981. Pp. viii + 186. *Appl. Psycholinguist.* 3, 83–86. doi: 10.1017/s0142716400004185

D'Odorico, L., and Jacob, V. (2006). Prosodic and lexical aspects of maternal linguistic input to late-talking toddlers. *Int. J. Lang. Commun. Disord.* 41, 293–311. doi: 10.1080/13682820500342976

DeVeney, S. L., Hagaman, J. L., and Bjornsen, A. L. (2017). Parent-implemented versus clinician-directed interventions for late-talking toddlers: a systematic review of the literature. *Commun. Disord. Q.* 39, 293–302. doi: 10.1177/1525740117705116

Flack, Z. M., Field, A. P., and Horst, J. S. (2018). The effects of shared storybook reading on word learning: a meta-analysis. *Dev. Psychol.* 54, 1334–1346. doi: 10.1037/dev0000512

Girolametto, L., Bello, A., Onofrio, D., Remi, L., and Caselli, M. C. (2017). *Parent-Coaching Per L'intervento Precoce sul Linguaggio. Percorsi di Lettura Dialogica nel Programma" Oltre il Libro*. Madison, WI: Erikson.

Girolametto, L., Bonifacio, S., Visini, C., Weitzman, E., Zocconi, E., and Pearce, P. S. (2002). Mother-child interactions in Canada and Italy: linguistic responsiveness to late-talking toddlers. *Int. J. Lang. Commun. Disord.* 37, 153–171. doi: 10.1080/13682820110116794

Girolametto, L., Pearce, P. S., and Weitzman, E. (1996a). Interactive focused stimulation for toddlers with expressive vocabulary delays. *J. Speech Lang. Hear. Res.* 39, 1274–1283. doi: 10.1044/jshr.3906.1274

Girolametto, L., Pearce, P. S., and Weitzman, E. (1996b). The effects of focused stimulation for promoting vocabulary. *J. Child. Commun. Dev.* 17, 39–49. doi: 10.1177/152574019501700205

Girolametto, L., Weitzman, E., Wiigs, M., and Pearce, P. S. (1999). The relationship between maternal language measures and language development in toddlers with expressive vocabulary delays. *Am. J. Speech Lang. Pathol.* 8, 364–374. doi: 10.1044/1058-0360.0804.364

Golinkoff, R. M., Hoff, E., Rowe, M. L., Tamis-LeMonda, C. S., and Hirsh-Pasek, K. (2019). Language matters: denying the existence of the 30-million-word gap has serious consequences. *Child Dev.* 90, 985–992. doi: 10.1111/cdev.13128

Hargrave, A. C., and Sénéchal, M. (2000). A book reading intervention with preschool children who have limited vocabularies: the benefits of regular reading and dialogic reading. *Early Child. Res. Q.* 15, 75–90. doi: 10.1016/S0885-2006(99)00038-1

Hart, B., and Risley, T. R. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*. Baltimore, MD: Paul H Brookes Publishing. doi: 10.1177/074193259801900609

Hawa, V. V., and Spanoudis, G. (2014). Toddlers with delayed expressive language: an overview of the characteristics, risk factors and language outcomes. *Res. Dev. Disabil.* 35, 400–407. doi: 10.1016/j.ridd.2013.10.027

Hayes, A. F. (2018). *Introduction to Mediation, Moderation, and Conditional Process Analysis : A Regression-Based Approach*. New York, NY: Guilford Press.

Heidlage, J. K., Cunningham, J. E., Kaiser, A. P., Trivette, C. M., Barton, E. E., Frey, J. R., et al. (2020). The effects of parent-implemented language interventions on child linguistic outcomes: a meta-analysis. *Early Child. Res. Q.* 50, 6–23. doi: 10.1016/j.ecresq.2018.12.006

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

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., and Lyons, T. (1991). Early vocabulary growth: relation to language input and gender. *Dev. Psychol.* 27, 236–248. doi: 10.1037/0012-1649.27.2.236

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

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

Korpilahti, P., Kaljonen, A., and Jansson-Verkasalo, E. (2016). Identification of biological and environmental risk factors for language delay: the Let's talk STEPS study. *Infant Behav. Dev.* 42, 27–35. doi: 10.1016/j.infbeh.2015.08.008

Kruythoff-Broekman, A., Wiefferink, C., Rieffe, C., and Uilenburg, N. (2019). Parent-implemented early language intervention programme for late talkers: parental communicative behaviour change and child language outcomes at 3 and 4 years of age. *Int. J. Lang. Commun. Disord.* 54, 451–464. doi: 10.1111/1460-6984.12451

Law, J., Dennis, J. A., and Charlton, J. J. (2017). Speech and language therapy interventions for children with primary speech and/or language disorders. *Cochrane Database Syst. Rev.* 2017:CD012490. doi: 10.1002/14651858.CD012490

Law, J., Kot, A., and Barnett, G. (1999). *A Comparison of Two Methods for Providing Intervention to Three Year Old Children with Expressive/Receptive Language Impairment*. London: City University London.

Levickis, P., Reilly, S., Girolametto, L., Ukoumunne, O. C., and Wake, M. (2014). Maternal behaviors promoting language acquisition in slow-to-talk toddlers. *J. Dev. Behav. Pediatr.* 35, 274–281. doi: 10.1097/dbp.0000000000000056

Levickis, P., Reilly, S., Girolametto, L., Ukoumunne, O. C., and Wake, M. (2018). Associations between maternal responsive linguistic input and child language performance at age 4 in a community-based sample of slow-to-talk toddlers. *Child Care Health Dev.* 44, 776–783. doi: 10.1111/cch.12600

MacWhinney, B. (2000). The CHILDES project: tools for analyzing talk (third edition): volume I: transcription format and programs, volume II: the database. *Comput. Linguist*. 26:657. doi: 10.1162/coli.2000.26.4.657

Majorano, M., Guerzoni, L., Cuda, D., and Morelli, M. (2020). Mothers' emotional experiences related to their child's diagnosis of deafness and cochlear implant surgery: parenting stress and child's language development. *Int. J. Pediatr. Otorhinolaryngol.* 130:109812. doi: 10.1016/j.ijporl.2019.109812

Manolson, A. (1992). *It Takes Two to Talk: A Parent's Guide to Helping Children Communicate*, 3rd Edn. Toronto, ON: The Hanen.

Mol, S. E., Bus, A. G., De Jong, M. T., and Smeets, D. J. H. (2008). Added value of dialogic parent-child book readings: a meta-analysis. *Early Educ. Dev.* 19, 7–26. doi: 10.1080/10409280701838603

Pan, B. A., Rowe, M. L., Singer, J. D., and Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Dev.* 76, 763–782. doi: 10.1111/j.1467-8624.2005.00876.x

Paul, R., and Elwood, T. J. (1991). Maternal linguistic input to toddlers with slow expressive language development. *J. Speech Hear. Res.* 34, 982–988. doi: 10.1044/jshr.3405.982

Pennington, L., and Thomson, K. (2007). It takes two to talk – the Hanen Program® and families of children with motor disorders: a UK perspective. *Child Care Health Dev.* 33, 691–702. doi: 10.1111/j.1365-2214.2007.00800.x

Reilly, S., Wake, M., Bavin, E. L., Prior, M., Williams, J., Bretherton, L., et al. (2007). Predicting language at 2 years of age: a prospective community study. *Pediatrics* 120, e1441–e1449. doi: 10.1542/peds.2007-0045

Rinaldi, P., Pasqualetti, P., Stefanini, S., Bello, A., and Caselli, M. C. (2019). The Italian words and Sentences MB-CDI: normative data and concordance between complete and short forms. *J. Child Lang.* 46, 546–566. doi: 10.1017/S0305000919000011

Roberts, M. Y., and Kaiser, A. P. (2011). The effectiveness of parent-implemented language interventions: a meta-analysis. *Am. J. Speech Lang. Pathol.* 20, 180–199. doi: 10.1044/1058-0360(2011/10-0055)

Roberts, M. Y., and Kaiser, A. P. (2012). Assessing the effects of a parent-implemented language intervention for children with language impairments using empirical benchmarks: a pilot study. *J. Speech Lang. Hear. Res.* 55, 1655–1670. doi: 10.1044/1092-4388(2012/11-0236)

Roberts, M. Y., and Kaiser, A. P. (2015). Early intervention for toddlers with language delays: a randomized controlled trial. *Pediatrics* 135, 686–693. doi: 10.1542/peds.2014-2134

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech vocabulary development. *Child Dev.* 83, 1762–1774. doi: 10.1111/j.1467-8624.2012.01805.x

Sansavini, A., Guarini, A., and Caselli, M. C. (2011). Preterm birth: neuropsychological profiles and atypical developmental pathways. *Dev. Disabil. Res. Rev.* 17, 102–113. doi: 10.1002/ddrr.1105

Sansavini, A., Guarini, A., Justice, L. M., Savini, S., Broccoli, S., Alessandroni, R., et al. (2010). Does preterm birth increase a child's risk for language impairment? *Early Hum. Dev.* 86, 765–772. doi: 10.1016/j.earlhumdev.2010.08.014

Sansavini, A., Guarini, A., Zuccarini, M., Lee, J. Z., Faldella, G., and Iverson, J. M. (2019). Low rates of pointing in 18-month-olds at risk for autism spectrum disorder and extremely preterm infants: a common index of language delay? *Front. Psychol.* 10:2131. doi: 10.3389/fpsyg.2019.02131

Sansavini, A., Zuccarini, M., Gibertoni, D., Bello, A., Caselli, M. C., Corvaglia, L., et al. (2021). Language profiles and their relation to cognitive and motor skills at 30 months of age: an online investigation of low-risk preterm and full-term children. *J. Speech Lang. Hear. Res.* 64, 2715–2733. doi: 10.1044/2021_JSLHR-20-00636

Silvey, C., Demir-Lira, ÖE., Goldin-Meadow, S., and Raudenbush, S. W. (2021). Effects of time-varying parent input on children's language outcomes differ for vocabulary and syntax. *Psychol. Sci.* 32, 536–548. doi: 10.1177/0956797620970559

Suttora, C., Guarini, A., Zuccarini, M., Aceti, A., Corvaglia, L., and Sansavini, A. (2020). Speech and language skills of low-risk preterm and full-term late talkers: the role of child factors and parent input. *Int. J. Environ. Res. Public Health* 17, 1–22. doi: 10.3390/ijerph17207684

Tamis-LeMonda, C. S., Bornstein, M. H., and Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Dev.* 72, 748–767. doi: 10.1111/1467-8624.00313

Tamis-LeMonda, C. S., Kuchirko, Y., and Song, L. (2014). Why is infant language learning facilitated by parental responsiveness? *Curr. Dir. Psychol. Sci.* 23, 121–126. doi: 10.1177/0963721414522813

Taumoepeau, M. (2016). Maternal expansions of child language relate to growth in children's vocabulary. *Lang. Learn. Dev.* 12, 429–446. doi: 10.1080/15475441.2016.1158112

Tosh, R., Arnott, W., and Scarinci, N. (2017). Parent-implemented home therapy programmes for speech and language: a systematic review. *Int. J. Lang. Commun. Disord.* 52, 253–269. doi: 10.1111/1460-6984.12280

Tsybina, I., and Eriks-Brophy, A. (2010). Bilingual dialogic book-reading intervention for preschoolers with slow expressive vocabulary development. *J. Commun. Disord.* 43, 538–556. doi: 10.1016/j.jcomdis.2010.05.006

Vigil, D. C., Hodges, J., and Klee, T. (2005). Quantity and quality of parental language input to late-talking toddlers during play. *Child Lang. Teach. Ther.* 21, 107–122. doi: 10.1191/0265659005ct284oa

Vihman, M. M., and McCune, L. (1994). When is a word a word? *J. Child Lang.* 21, 517–542. doi: 10.1017/S0305000900009442

Weisleder, A., and Fernald, A. (2013). Talking to children matters: early language experience strengthens processing and builds vocabulary. *Psychol. Sci.* 24, 2143–2152. doi: 10.1177/0956797613488145

Weitzman, E. (2013). More than words—the Hanen Program for parents of children with autism spectrum disorder: a teaching model for parent-implemented language intervention. *Perspect. Lang. Learn. Educ.* 20, 96–111. doi: 10.1044/lle20.3.86

Zuccarini, M., Suttora, C., Bello, A., Aceti, A., Corvaglia, L., Caselli, M. C., et al. (2020). A parent-implemented language intervention for late talkers: an exploratory study on low-risk preterm and full-term children. *Int. J. Environ. Res. Public Health* 17, 1–20. doi: 10.3390/ijerph17239123

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

ORIGINAL RESEARCH
published: 24 September 2021
doi: 10.3389/fpsyg.2021.712647

# Acoustic-Lexical Characteristics of Child-Directed Speech Between 7 and 24 Months and Their Impact on Toddlers' Phonological Processing

*Margaret Cychosz*[1,2]*, *Jan R. Edwards*[1], *Nan Bernstein Ratner*[1], *Catherine Torrington Eaton*[3] *and Rochelle S. Newman*[1]

[1] *Department of Hearing and Speech Sciences, University of Maryland, College Park, MD, United States,* [2] *Center for Comparative and Evolutionary Biology of Hearing, College Park, MD, United States,* [3] *Department of Communication Sciences and Disorders, University of Texas Health Science Center at San Antonio, San Antonio, TX, United States*

OPEN ACCESS

**Edited by:**
*Chiara Suttora, University of Bologna, Italy*

**Reviewed by:**
*Tonya Bergeson, Butler University, United States*
*Chiara Cantiani, Eugenio Medea (IRCCS), Italy*

***Correspondence:**
*Margaret Cychosz*
*mcychosz@umd.edu*

**Specialty section:**
*This article was submitted to Developmental Psychology, a section of the journal Frontiers in Psychology*

**Received:** *20 May 2021*
**Accepted:** *05 July 2021*
**Published:** *24 September 2021*

**Citation:**
*Cychosz M, Edwards JR, Bernstein Ratner N, Torrington Eaton C and Newman RS (2021) Acoustic-Lexical Characteristics of Child-Directed Speech Between 7 and 24 Months and Their Impact on Toddlers' Phonological Processing. Front. Psychol. 12:712647. doi: 10.3389/fpsyg.2021.712647*

Speech-language input from adult caregivers is a strong predictor of children's developmental outcomes. But the properties of this child-directed speech are not static over the first months or years of a child's life. This study assesses a large cohort of children and caregivers ($n = 84$) at 7, 10, 18, and 24 months to document (1) how a battery of phonetic, phonological, and lexical characteristics of child-directed speech changes in the first 2 years of life and (2) how input at these different stages predicts toddlers' phonological processing and vocabulary size at 2 years. Results show that most measures of child-directed speech do change as children age, and certain characteristics, like hyperarticulation, actually peak at 24 months. For language outcomes, children's phonological processing benefited from exposure to longer (in phonemes) words, more diverse word types, and enhanced coarticulation in their input. It is proposed that longer words in the input may stimulate children's phonological working memory development, while heightened coarticulation simultaneously introduces important sublexical cues and exposes them to challenging, naturalistic speech, leading to overall stronger phonological processing outcomes.

**Keywords: acoustics, lexicon, child-directed speech, phonological neighborhood density, speech clarity, phonological development, nonword repetition**

## 1. INTRODUCTION

The speech and language that children hear early in life is a strong predictor of their linguistic outcomes (Hart and Risley, 1995; Rowe, 2008; Huttenlocher et al., 2010). Children who are exposed to more child-directed speech (CDS) from caregivers eventually produce more complex babbling shapes, process speech faster, and grow larger vocabularies (Hoff, 2003; Weisleder and Fernald, 2013; Ferjan Ramírez et al., 2019). The slow speaking rate, dynamic pitch modulations, and shortened utterance lengths that characterize the unique CDS register are hypothesized to draw infants' attention and scaffold developmentally-appropriate linguistic models, facilitating speech-language development (Ferguson, 1977; Furrow et al., 1979; Fernald, 1985; Cooper and Aslin, 1990; Soderstrom, 2007; Wang et al., 2021). While the acoustic properties of CDS, and their changes over development, have been documented (Liu et al., 2009; Ko, 2012; Cristia, 2013; Hartman et al., 2017; Kalashnikova and Burnham, 2018), less is known about the phonological-lexical characteristics of CDS. This is a clear blindspot. For one thing, many of the acoustic measures previously studied, like vowel dispersion or speaking rate, covary with measures like word frequency and phonological neighborhood density. Consequently, previous observations about the effects of acoustic properties of CDS on children's speech-language outcomes could instead be attributable to lexical properties of the input. Furthermore, despite evidence that the quantity and quality of CDS predict a number of early phonological outcomes including babbling, infant speech perception, and toddler lexical processing—and suggestions in the literature that CDS quantity *should* impact children's phonological processing—no work to date has explicitly studied the relationship between CDS and children's phonological processing. This, again, leaves a critical gap in our understanding of development because different kinds of linguistic input may matter more, or less, to different developmental outcomes (e.g., lexical processing, speech discrimination) at different developmental stages. It is thus essential that we understand not simply *if* input matters for children's phonological processing, but also *how* and *when*.

To that end, this study has two goals. In a large cohort of English-learning children ($n = 84$), observed from 7 to 24 months, we first document longitudinal changes in a comprehensive battery of (North American) CDS characteristics. Then, in the same cohort, we model static CDS measures at 7, 10–11, and 18 months to predict toddlers' phonological processing and vocabulary sizes at 24 months. In doing so, we document how an array of co-varying lexical, phonological, and acoustic characteristics of CDS change over development, allowing us to disentangle and model how they contribute to toddlers' speech-language outcomes at 24 months.

## 2. PREVIOUS LITERATURE

### 2.1. Changes in Child-Directed Speech With Age

Child-directed speech (CDS) is not a static construct over the first years of a child's life. From infancy through early toddlerhood, the mean length of CDS utterances increases (Stern et al., 1983), the vowel space and individual vowel categories expand and then contract (Bernstein Ratner 1984; Liu et al. 2009; Hartman et al. 2017, but see Burnham et al. 2015; Kalashnikova and Burnham 2018), and speaking rate increases (Ko, 2012; Sjons et al., 2017; Raneri et al., 2020). Fundamental frequency baselines and ranges also change non-linearly throughout this time (Stern et al., 1983; Kitamura et al., 2001; Kitamura and Burnham, 2003; Vosoughi and Roy, 2012; Han et al., 2018). While these studies demonstrate a fairly comprehensive understanding of age-related changes in the acoustics of CDS, our descriptions of phonological-lexical characteristics of CDS, as well as how they change over time, are more superficial. For example, it is well known that the number of word types, and frequency of rare words, in CDS increases as children age (Rowe, 2012; Glas et al., 2018). But there are far more sophisticated methods available to model the phonological-lexical properties of linguistic input. For example, sublexical organization of the lexicon can be modeled using PHONOLOGICAL NEIGHBORHOOD DENSITY, or the number of words differing from a target word by one phoneme (e.g., the neighbors *sat* and *cat*). Words with many neighbors are said to reside in dense neighborhoods while those with few neighbors reside in sparse neighborhoods. Another related sub-lexical covariate is PHONOTACTIC PROBABILITY, or the likelihood of a sound sequence in a language (e.g., *blick* is higher probability than *bnick* in English). Word frequency and word length (in phonemes) are likewise rarely included in models of CDS.

There are three reasons why it is important to model these phonological-lexical characteristics of CDS and their impact on children's speech-language outcomes. First, there is strong evidence that phonological-lexical properties of the input should impact children's phonological processing. Equivalent research has established such a relationship for children's word learning; for example, children learn words from dense neighborhoods first, followed by sparser neighborhoods (Storkel, 2004a; Storkel and Hoover, 2011; Kern and dos Santos, 2018; Zamuner and Thiessen, 2018). Second, the organization of the lexicon is reflected in *adult* speech production. Adults phonetically reduce (shorten segment durations, contract the vowel space, coarticulate more) in high-frequency relative to low-frequency words (Gahl, 2008; Bell et al., 2009), and in words from dense phonological neighborhoods relative to words from sparse neighborhoods (Scarborough, 2005; Gahl et al., 2012)[1]. Adults, and children, also reflect the structure of their lexicon in their speech as they produce high phonotactically probable sequences more smoothly (shorter durations) than low probability sequences (Edwards et al. 2004; see Vitevitch and Luce 2016 for overview). Adults are also known to reflect word structure in their speech. For example, adults compensate for prosodic structure via COMPENSATORY SHORTENING, the phenomenon where word duration decreases as word length increases (Munhall et al., 1992; Harrington et al., 2015).

Finally, modeling characteristics of the lexicon, like phonological neighborhood density, in CDS is important because many of these measures are known to impact children's lexical access and processing. Dense phonological neighborhoods have been shown to inhibit lexical retrieval in children (Newman and German, 2002; Arnold et al., 2005), while high-frequency words, especially from sparse neighborhoods, are more rapidly recognized (Metsala, 1997)[2]. Evidence from the nonword repetition paradigm has also demonstrated that children process phonemically-shorter and phonotactically-probable words better than longer and/or less probable words (Gathercole et al., 1991; Edwards et al., 2004).

[1]Previous reports of greater dispersion in words from dense neighborhoods were generally due to confounds with lexical frequency which Gahl et al. (2012) were able to statistically control.

[2]Dense neighborhoods inhibit adults' lexical access and recognition as well (Luce and Pisoni, 1998; Vitevitch and Luce, 1998), but this is less relevant for the study at hand.

The conclusion from the above findings is that the acoustic properties of children's input, and the speed and accuracy of children's ensuing word recognition, vary systematically by word structure and the organization of the lexicon. Consequently, in development, it is not sufficient to only model CDS parameters like word types or tokens. Furthermore, and most critically, much work suggests that hyperarticulated CDS, like an expanded vowel space, may facilitate certain linguistic outcomes (Liu et al., 2003; Hartman et al., 2017). But as this body of research demonstrates, hyperarticulation (i.e., speech clarity) varies according to word structure and statistics: speakers hyperarticulate low-frequency words, phonetically reduce long words relative to short, etc.. So the effects of these co-varying parameters, acoustic and lexical, need to be disentangled for children's development.

## 2.2. Nonword Repetition: An Important Indicator of Speech-Language Development

Phonological working memory, or the ability to recall sequences of phones, is a critical prerequisite to process—and thus learn—speech and language (Gathercole, 2006; Pierce et al., 2017). Children with stronger phonological working memories grow larger vocabularies (Adams and Gathercole, 1995; Baddeley et al., 1998), construct more complex syntactic constituents (Adams and Gathercole, 1995, 1996), and develop stronger phonological awareness skills (Michalczyk et al., 2013; Erskine et al., 2020).

Phonological working memory is often assessed using nonword repetition (NWR) tasks where participants process, temporarily store, and repeat novel, phonotactically-permissible sequences of phones (e.g., *blick*). NWR closely mimics novel word learning processes. In the task, children must not only activate phonological representations in response to an auditory stimulus, just as they do in, for example, looking-while-listening tasks, but they must also organize their speech motor-schemata to articulate a novel sequence of sounds. As such, NWR ability is unsurprisingly one of the strongest and most consistent predictors of children's future speech, language, and literacy development (Gathercole and Baddeley, 1989; Gathercole, 2006).

A variety of linguistic and experiential factors impact NWR accuracy, including stimulus length (Gathercole et al., 1991), phonological complexity (Szewczyk et al., 2018), and phonotactic probability/wordlikeness (Gathercole et al., 1991; Edwards et al., 2004; Szewczyk et al., 2018). Participants' vocabulary size (Gathercole and Baddeley, 1989; Munson et al., 2005; Hoff et al., 2008) and real-word repetition accuracy (Torrington Eaton et al., 2015) also predict NWR performance. Thus, although NWR was originally assumed to be a language-neutral diagnostic measure of phonological working memory, relevant experiential predictors, such as phonotactic probability, demonstrate that biologically-endowed working memory *and* experience with language, together, predict performance on the task (Gathercole and Baddeley, 1989; MacDonald and Christiansen, 2002).

One experiential predictor that developmental researchers have long assumed should predict children's NWR ability is linguistic input. There could be an indirect effect of input on NWR. Children who hear more CDS in their environments grow larger vocabularies (Hoff, 2003) and this vocabulary knowledge could result in a reorganization and early maturation of the lexicon, increasing NWR accuracy (Gathercole and Baddeley, 1989; Munson et al., 2005; Hoff et al., 2008).

Evidence for possible *direct* effects of language input on NWR, however, comes from a few different sources. First, studies show that children who receive more language input process speech faster during lexical identification tasks (Hurtado et al., 2008; Weisleder and Fernald, 2013), suggesting that there could be a similar relationship for NWR tasks. Elsewhere, in bilingual children, dual language exposure explains 20-25% of the variance in their NWR abilities at 22 months (Parra et al., 2011), meaning that bilingual children who are exposed to more of one of their languages have stronger NWR abilities in that language (though these exposure effects have not been found in older bilinguals; Core et al. 2017; Farabolini et al. 2021). Furthermore, children in an indigenous society with low reported rates of CDS were reported to have lower NWR scores than age-matched peers elsewhere in the literature (Cristia et al., 2020). And finally, a series of computational modeling studies that manipulated parameters of the input were able to replicate known developmental NWR patterns in 2- 6-year-olds, suggesting direct effects of input on NWR outcomes (Jones, 2016).

The assumption underlying all of this work is that children who are spoken to more, and engage in more conversations with caregivers, should have more practice processing incoming speech, encoding new words, and articulating novel sequences of phones—all skills that are implicated during NWR. However, despite these assumptions, and the strong evidence suggesting that such a link between input and phonological processing exists, to date no study has explicitly evaluated what parameters of the input predict children's NWR performance, a critical gap that the current study fills.

## 2.3. Research Questions

In sum, there are clear lexical (word types and tokens) and acoustic (vowel space, speaking rate) effects on children's speech-language development. However, many of these parameters vary systematically by the structure of the lexicon. The first goal of this paper is to document age-related changes in a battery of acoustic and lexical parameters of North American CDS. We ask:

1. How do organizational characteristics of the lexicon—phonological neighborhood density, word frequency, phonotactic frequency, word length—that are so predictive of adult speech production and children's lexical processing, change in CDS over development? Relatedly, how do frequently studied CDS measures, such as number of word types, change over development in this sample?

Then, to disentangle the effects of these co-varying acoustic and lexical parameters of children's input, we evaluate how each parameter predicts children's phonological processing and vocabulary sizes at 24 months.

2. What is the unique contribution of each acoustic and lexical input parameter, at 7, 10–11, and 18 months, for children's NWR and expressive vocabulary size at 24 months?

To answer these questions, we measure a host of acoustic and lexical parameters of CDS in semi-naturalistic caregiver-child interactions over the first 2 years of life as well as children's outcomes at 24 months.

# 3. METHODS

## 3.1. Participants

Eighty-six mother-child pairs participated in the study. All children were born full-term, had normal hearing and vision, and heard primarily American English (approximately ≥85%) in the home at the time of initial recruitment (one child was also beginning to be exposed to Spanish at 7 months). Four children were in bilingual childcare settings at 24 months. All pairs were followed longitudinally from when the child was 7 to 24 months, and participated in a number of speech-language tasks including speech segmentation, phonological processing, and receptive and expressive vocabulary assessments, as well as free play sessions between the mother and child to elicit CDS samples.

Prior to analysis, two caregiver-child dyads were removed completely: one where the child did not complete the vocabulary assessments at 18 or 24 months and another where all of the transcripts of the mother-child interactions were unavailable for analysis. The gender distribution for the final sample of 84 participants was $n = 49$ female and $n = 35$ male children (see **Table 1** for age information). Family socioeconomic status was quantified as mother's education level: 79 mothers (94%) had at least a college degree (1 did not respond). Caregivers identified their children's race/ethnicity as follows: 7 African American children (8.33%), 2 African American and white (2.38%), 3 Asian American and white (3.57%), 66 white (78.57%), 2 white Hispanic (2.38%), 3 non-white Hispanic (3.57%), and 1 child of mixed race and ethnicity. Forty-six (54.76%) of children were first-born, 32 (38.10 %) second-born, 4 (4.76%) third-born, and 1 each was the fourth- and fifth-born in the families.

An additional $n = 39$ children participated in the research program but either could not complete the NWR task ($n = 30$) or scored below the 10th percentile on the MacArthur-Bates Communicative Development Inventory (MB-CDI) (Fenson et al., 2007) at 24 months ($n = 9$); the data from these caregiver-child dyads are not analyzed here. We elected to remove the children who scored below the 10th percentile because that can be considered the cut-off for clinical diagnosis and it was not possible to ascertain diagnoses of language-related disorders (e.g., developmental language disorder) via other means given the children's young ages. See Torrington Eaton et al. (2015) for further details on participant exclusion.

## 3.2. Procedures

### 3.2.1. Adult Language Samples

To elicit CDS samples, each caregiver-child pair participated in a free-play session in the lab at 7, 10, 11, and 24 months. Approximately half of the participants ($n = 40$, 47.62% of the sample) also completed a play session at 18 months. For the purposes of this study, the 10 and 11 month timepoint data were combined: $n = 56$ dyads (67.47% of the sample) contributed the CDS sample at 10 months and $n = 27$ (32.53%) contributed at 11 months.

During the play session, caregivers were instructed to interact and speak with their child naturally, as if they were at home. Participants were provided with a number of standardized toys and board books to ensure that a sufficient number of target vowels and segments were elicited over the course of the interaction. Caregivers were recorded with an Audio-Technica AT 8531 lavalier microphone connected to a Marantz PMD 660 solid-state recorder. Each session lasted between 15 and 20 minutes.

Two recordings, one at 11 months and another at 24 months, were removed because they were only approximately 5 minutes in length or shorter. An additional two 7-month recordings were removed due to poor audio quality/unavailability. Transcriptions from the caregiver-child play sessions in the lab can be found in the NewmanRatner corpus, available on CHILDES (https://childes.talkbank.org/access/Eng-NA/NewmanRatner.html) (MacWhinney, 2000; Newman et al., 2016).

### 3.2.2. Word Repetition Tasks

At their 24-month visit, children completed a real word and corresponding nonword repetition task. Our goal in the NWR task was to evaluate phonological errors attributable to breakdowns in speech processing. However, 2-year-olds regularly make phonemic substitutions, due to ongoing articulatory maturation, that do not reflect their phonological processing skills. Consequently, we administered a real word repetition task, in addition to the nonword task, to control for children's articulatory skill during NWR (see section 3.3.4). To further ensure that we were evaluating children's phonological processing skills, and not their articulatory maturity, we also excluded late-emerging consonants such as /ɹ/ from the stimuli. See Torrington Eaton et al. (2015) for extensive modeling of these children's nonword and real word repetition results.

Stimuli for the real word and nonword repetition tasks consisted of $n = 11$ nonwords and $n = 11$ corresponding real words ($n = 4$ one-syllable, $n = 4$ two-syllable, and $n = 3$ three-syllable in each condition), matched for target consonants and consonant-vowel transitions by word condition (see **Table A1** for stimuli list). Stimuli were adapted from Hoff et al. (2008). For the real word repetition task, children were handed small toys representing the target word and were prompted to repeat the word after the experimenter. For nonword task administration, the child was handed a brightly colored stuffed animal and the experimenter asked the child to repeat the "funny name." The experimenter produced each item no more than two times before continuing to the next item. The real word repetition task was always administered before the nonword task to familiarize children with the task.

### 3.2.3. Vocabulary Measurement

Children's vocabulary size was assessed at each timepoint in the longitudinal investigation. The MB-CDI (Fenson et al., 2007) was administered at 7, 10, 11 and 24 months for all children, and at 18 months for the $n = 40$ children tested at that timepoint (receptive vocabulary was assessed at ages 7, 10, and 11 months and expressive at 18 and 24 months). Additionally, the Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4) (Dunn and Dunn, 2007) and Expressive One-Word Picture

**TABLE 1** | Child age and caregiver-child play session statistics.

| Timepoint | 7mos | 10–11mos | 18mos | 24mos |
|---|---|---|---|---|
| Child age: mean (SD) range | 7.5 (0.33) 6.93–8.27 | 10.5 (0.64) 8.77–12.13* | 18.26 (0.57) 17.2–19.23 | 24.56 (0.57) 23.27–26.33 |
| # of transcribed play sessions | 82 | 83 | 40 | 83 |
| # and % analyzed for coarticulation | 81 (98.78) | 82 (98.8) | 40 (100) | 83 (100) |
| # and % analyzed for vowels | 75 (91.46) | 78 (93.98) | 40 (100) | 82 (98.8) |

**Reflects child age (10 or 11 months) during collection of the CDS sample used in the analysis.*

Vocabulary Test, 3rd edition (EOWPVT-3) (Brownell, 2000) were administered at 24 months. For the PPVT-4 we report raw scores because standard scores are only available for children older than 30 months.

## 3.3. Data Processing

### 3.3.1. Cleaning Caregiver-Child Transcripts

From the caregiver-child transcripts, we excluded all onomatopoeia, exclamations (e.g., "ick!"), and proper names (except places likely to be common to all children in the sample such as "Maryland"), resulting in $n = 3{,}463$ word types across all timepoints and speakers. From these word types, contractions were excluded from the calculation of phono-lexical measures, as they are not included in the lexical statistics dictionary we used [$n = 252$ (7.28%) word types removed]; contractions were not excluded from the acoustic analysis.

### 3.3.2. Measures of Phono-Lexical Diversity

A number of phono-lexical characteristics were calculated over transcripts of the caregiver-child interactions:

- Phonotactic Probability: probability of each word type in the transcript based on its average biphone positional probability in American English
- Phonological Neighborhood Density: number of phonological neighbors of each word type in the transcript
- Word Length: length, in phonemes, of each word type in the transcript
- Word Frequency: each word type's frequency in American English

We additionally computed the Type:Token Ratio (TTR) of each transcript, as well as the Mean Average Type:Token Ratio (MATTR) which is less sensitive to speech sample length than TTR (Fergadiotis et al., 2013). MATTR was computed over a 10-word token moving window (i.e., for a window of x tokens, MATTR is computed over tokens 1-x, 2-x, etc.). Finally, we computed the Type and Token count of each transcript. Many of these phono-lexical statistics are highly correlated, so they are evaluated separately for the statistical modeling.

The TTR, MATTR, Type count, and Token count were computed using Computerized Language ANalysis (CLAN) software program (MacWhinney, 2000). The lexical statistics were calculated using the Irvine Phonotactic Online Dictionary (IPhOD) (Vaden et al., 2009). The IPhOD computes phonotactic probability from biphone co-occurrence in English. Phonological neighborhood density statistics in the IPhOD were made according to Vitevitch and Luce (1999) and word frequency estimates in the dictionary were derived from the American English SUBTLEX database (Brysbaert and New, 2009). Word frequency and phonotactic frequency were log transformed prior to analysis. Following Storkel (2004b), phonotactic probability was additionally z-score normalized to control for word length confounds. For words with multiple pronunciation variants in the IPhOD, we selected the variant with the highest phonotactic probability. These lexical statistics are reported over word types, not tokens, within each speech sample, which is consistent with previous research.

Although estimates of neighborhood density and phonotactic probability based on children's lexicons are available (Storkel, 2004a), we elected to compute these measures over adult lexicons because our interest was in what components of *adult* speech best predicted children's phonological outcomes. Lexical statistics calculated over adult and child speech corpora are also strongly correlated (Guo et al., 2015).

### 3.3.3. Acoustic Analysis

Given the large amount of acoustic data generated from 84 children, at multiple timepoints, we conducted the acoustic analysis over a subset—the second 5-min chunk—of each caregiver-child play session at 7, 10–11, and 24 months (excluding the 18-month sample since this was only collected from a subset of the dyads). The 5-min subsets of each play session were segmented into Praat TextGrids (Boersma and Weenink, 2020) and force-aligned to the phone level (McAuliffe et al., 2017). One of two trained phoneticians then hand-checked each TextGrid and adjusted the alignment as necessary. Because acoustic measures can be sensitive to segmentation, alignment was standardized in several ways. Periodicity in the waveform and formant structure in the spectrogram marked vowels. Vowels were distinguished from glides by the presence of a steady-state formant. In the absence of a steady-state formant, 50% of the segment was devoted to the glide and 50% to the vowel. Utterance-initial plosives were segmented at the start of their release. Nasals were identified by anti-formants and depressed intensities in the spectrogram and fricatives by high-frequency energy in the spectrogram and aperiodicity in the waveform. Speech that was whispered or yelled was removed from acoustic analysis as were all words whose spectral shape could not be deduced in the spectrogram due to phonetic reduction. Overlapping speech (i.e., with target child) was also marked to be excluded from analysis.

We computed four measures of hyperarticulation in the CDS samples: vowel dispersion, vowel space area, segment duration, and adjacent consonant-vowel/vowel-consonant coarticulation. To compute vowel dispersion and vowel space area, the first and second formant (F1 and F2) frequencies at the midpoint of the three peripheral /i, a, u/ vowels were measured using a triple formant tracker running Inverse Filter Control (Watanabe, 2001), Entropic Signal Processing System's "autocorrelation", and Entropic Signal Processing System's "covariance" formant trackers (https://github.com/megseekosh/vocal_tract_vowel). Then, the median F1 and F2 from the three trackers were computed. Formant measurements were Lobanov-normalized to account for speaker-specific anatomical differences (Lobanov, 1971); all vowel results were replicated with unnormalized data as well, except where noted. The caregiver's average vowel space area was measured, at each timepoint, using the `phonR` package in R (see McCloy 2016 for detail on measurement technique). Finally, the *dispersion* of each vowel token (calculated from word types to avoid data skew due to high-frequency words within the transcript) was calculated as the distance of each vowel token along F1 and F2 from each speaker's median F1 and F2 values.

We implemented coarticulation as the acoustic distance between adjacent phones, using a custom Python script running Librosa packages (McFee et al. 2015; see Gerosa et al. 2006; Cychosz et al. 2019 for further details). Specifically, Mel-frequency log-magnitude spectra were averaged over the entirety of each target phone; coarticulation was then the Euclidean distance between the averaged spectra of neighboring phones. We did not compute coarticulation within (1) stop-vowel sequences because it was not possible to delimit the closure portion of utterance-initial stops or (2) voiceless glottal fricative-vowel sequences due to the weak spectral signature of those fricatives. Coarticulation was computed for all remaining consonant manners.

Because unstressed vowels are highly reduced in American English, the hyperarticulation measures involving vowels were only made over stressed vowels/sequences containing a stressed vowel (including if the vowel-consonant transcended a syllable boundary). We additionally only computed the hyperarticulation measures in content words, which is in keeping with previous work on the interaction of vowel space, coarticulation, and the lexicon (Gahl et al., 2012; Zellou and Scarborough, 2015).

We assessed changes in speaking rate by modeling segment duration, and not explicitly calculating maternal speaking rate as number of syllables/minute, for example, because speaking rate is highly correlated with segment duration and there have been recent reports on age-varying changes in maternal speaking rate in this corpus (Raneri et al., 2020).

#### 3.3.4. Nonword Repetition Scoring

The nonword stimuli contained $n = 33$ phoneme targets to be scored. To ensure that NWR errors were attributable to children's phonological processing, and not articulatory limitations, each phoneme produced in the NWR condition was compared to the equivalent phoneme in the real word condition. If the phoneme was produced incorrectly in both conditions, it was assumed to be attributable to articulatory limitations, and was not marked incorrect in the nonword condition. If the phoneme was produced incorrectly in only the nonword but not real-word condition, it was marked incorrect in the nonword condition. Nonwords that children failed to repeat after two experimenter prompts were also marked as inaccurate.

## 4. RESULTS

Data were analyzed in the RStudio computing environment (version: 1.4.1103; RStudio Team 2020). Data visualizations were created with `ggplot2` (Wickham, 2016). Modeling was conducted using the `lme4` (Bates et al., 2015) and `lmerTest` (Kuznetsova et al., 2017) packages. Pairwise comparisons and model summaries were presented with `emmeans` (Lenth, 2021) and `Stargazer` (Hlavac, 2018). Model parameter significance was determined via a combination of log-likelihood comparisons between models, AIC estimations, and $p$-values from model summaries. Relevant variables were mean-centered prior to model fitting. All modeling and analysis scripts are included in the affiliated GitHub repository (https://github.com/megseekosh/cds-processing).

### 4.1. Age-Related Changes in Child-Directed Speech

Descriptive statistics for the acoustic-lexical CDS measures at 7, 10–11, 18, and 24 months are included in **Table 2** and outlined in **Figures 1**–**3**). To evaluate these age-related changes in CDS, we fit a series of linear mixed effects models to predict each CDS measure. Each model included a random intercept of child-caregiver dyad and a fixed effect of timepoint.

There were significant effects of timepoint for all CDS measures except phonotactic probability, indicating that the input measures changed as children aged. Word type and token count increased significantly between each timepoint sampled, replicating previous work (Rowe, 2012), except 7 to 10–11 months (see **Table S1** in the **Supplementary Materials** for pairwise comparisons of timepoints). The modeling also demonstrated how phono-lexical statistics of the input changed with child age. There was a significant, negative effect of the 24-month timepoint on the measures, indicating that children heard less frequent, longer words, from sparser neighborhoods, at 24 months than the other points (pairwise comparisons in **Table S2** in **Supplementary Materials**).

Finally, there were significant changes in the acoustics between 7 and 24 months and 10–11 and 24 months where the speech became significantly faster (7–24 month changes: $\beta = -7.32$ $t = -7.21$ $p < 0.001$), but less coarticulated (7–24 month: $\beta = 0.45$, $t = 5.94$ $p < 0.001$) and produced with a more expanded vowel space (7–24 month: $\beta = 0.99$, $t = 3.46$, $p = 0.002$) (see **Table S3** in **Supplementary Materials** for all pairwise comparisons). Given that vowels tend to reduce, and coarticulation increases, in faster speech, this pattern in the acoustics was somewhat surprising. However, as discussed in the introduction, many lexical statistics covary with acoustic properties so even as parents were speaking faster to their older children, the fact that they were using more diverse, lower-frequency words could explain the relative hyperarticulation in

**TABLE 2 |** Descriptive statistics of child-directed speech characteristics at 7, 10–11, 18, and 24 months.

| | 7mos Mean (SD) range | 10–11mos Mean (SD) range | 18mos Mean (SD) range | 24mos Mean (SD) range |
|---|---|---|---|---|
| Types | 246.63 (67.22) 49–392 | 237.76 (53.56) 30–339 | 261.28 (59.54) 120–360 | 292.68 (54.73) 59–419 |
| Tokens | 873.38 (313.16) 79–1,566 | 864.88 (277.02) 38–1,508 | 1,012.91 (302.73) 309–1,530 | 1,247.22 (346.29) 105–1,973 |
| TTR | 0.3 (0.06) 0.18–0.62 | 0.29 (0.06) 0.2–0.79 | 0.27 (0.05) 0.19–0.5 | 0.24 (0.04) 0.17–0.56 |
| MATTR | 0.88 (0.04) 0.78–0.95 | 0.87 (0.03) 0.74–0.95 | 0.89 (0.02) 0.83–0.93 | 0.91 (0.02) 0.82–0.95 |
| Biphone probability | 0.37 (1.73) −1.66–11.66 | 0.37 (1.71) −1.6–11.66 | 0.36 (1.7) −1.71–11.66 | 0.36 (1.71) −1.71–11.66 |
| Word frequency | 5.69 (2.56) −2.81–10.64 | 5.72 (2.53) −3.91–10.64 | 5.63 (2.53) −2.53–10.64 | 5.48 (2.57) −3.91–10.64 |
| Phon. neighborhood density | 20.55 (13.63) 0–50 | 20.65 (13.66) 0–50 | 20.56 (13.63) 0–50 | 19.94 (13.6) 0–50 |
| Word length | 3.48 (1.37) 1–12 | 3.47 (1.35) 1–12 | 3.49 (1.35) 1–14 | 3.59 (1.4) 1–13 |
| Coarticulation (spectral distance) | 6.89 (4.23) 0.93–34.75 | 6.91 (4.15) 1.04–36.91 | NA | 7.37 (4.23) 0.97–38.9 |
| Phone duration (ms) | 90.21 (70.78) 20.38–1,180 | 90.02 (72.73) 20.15–1,050 | NA | 83.45 (59.42) 20.13–880 |
| Vowel space area | 6.74 (1.43) 4.15–10.12 | 6.78 (1.52) 1.69–10.06 | NA | 7.7 (2.04) 1.5–14.65 |

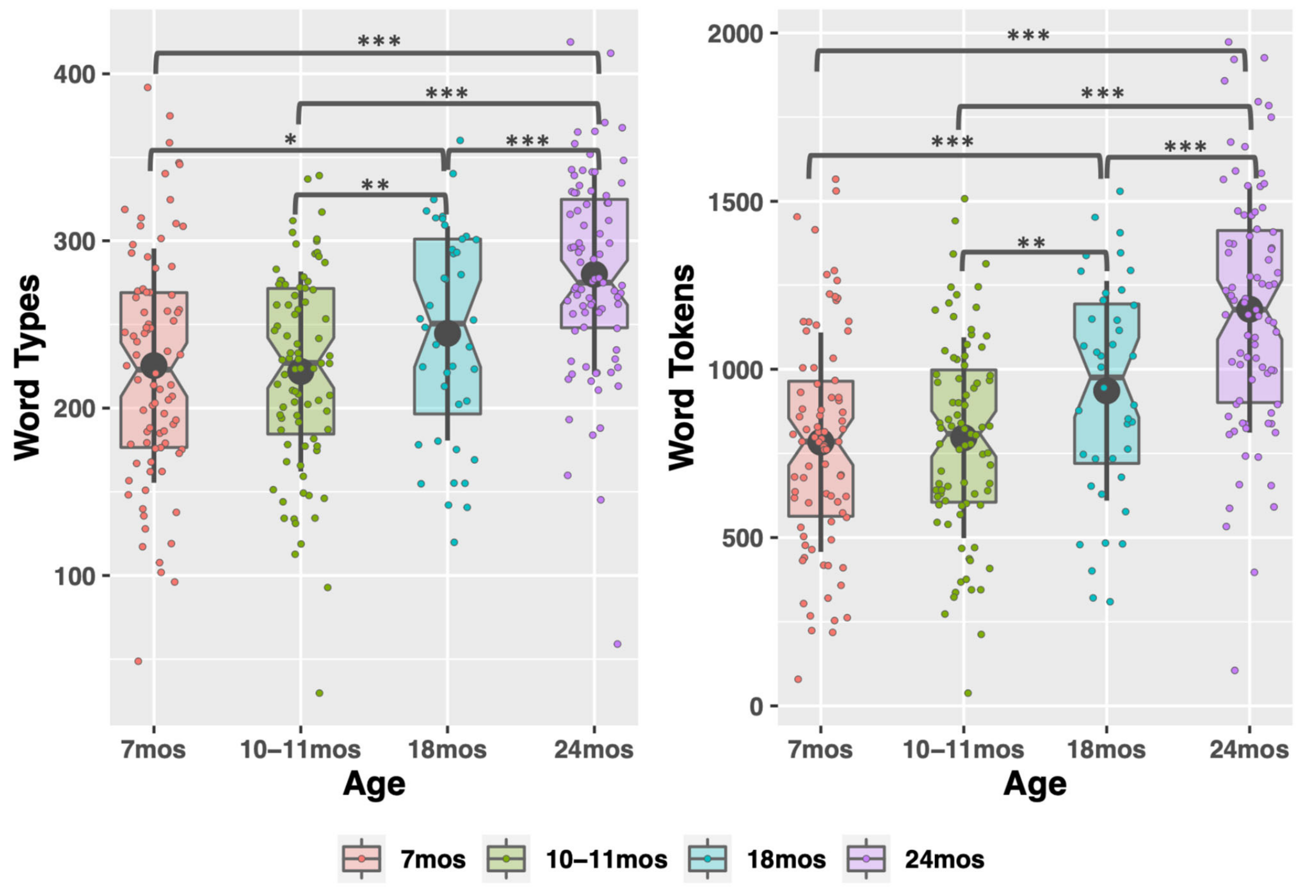



**FIGURE 1 |** Word type and token count in child-directed speech between 7 and 24 months. Large, gray points indicate mean; whiskers indicate 1 SD from mean. Notches indicate median. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

their speech at 24 months. (Again, acoustics were not measured at 18 months.)

In the descriptive statistics, one additional pattern emerged. Overall, the data trend is for CDS properties to resemble adult-directed speech more as children age. The exception to this is at 10–11 months, where many of the measures exhibit a hyper CDS register. There are, on average, fewer word types and tokens at 10 months than 7 months (the upper range of word count is also lower at 10 months). Words at 10 months tend to come from denser neighborhoods and be more coarticulated. We emphasize that the differences between 7 and 10–11 months are simply trends—no significant differences between the timepoints

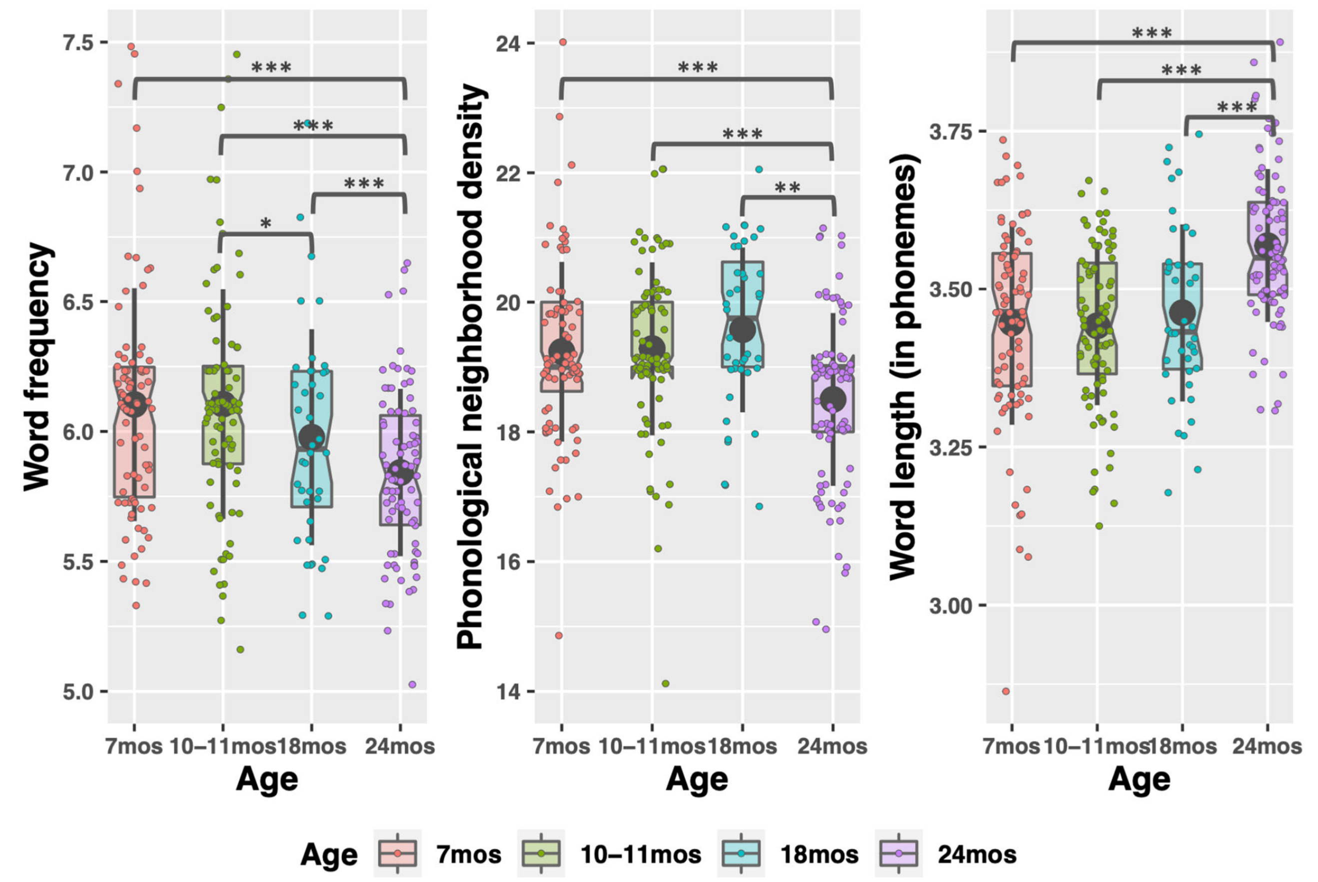



**FIGURE 2 |** Phono-lexical characteristics of child-directed speech between 7 and 24 months. Large, gray points indicate mean; whiskers indicate 1 SD from mean. Notches indicate median. Individual datapoints indicate median word frequency and phonological neighborhood density (left and center figures), or mean word length (right figure). *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

emerged in the modeling and there are no reliable differences between them. However, the trend suggests that parents may use a slightly more exaggerated CDS register at 10–11 months, as compared to just 3 months prior. We return to this point in the Discussion.

## 4.2. Modeling Relationships Between CDS, Phonological Processing, and Vocabulary Size

Having established that the quantity and quality of CDS speech differs by child age, we next evaluated how individual CDS differences explained the children's outcomes (NWR accuracy (i.e., phonological processing) and vocabulary size) at 24 months. Descriptive statistics of the children's outcomes are listed in **Table 3**, including the vocabulary measures at 7, 10, 11, and 18 months. Children varied greatly in performance on the NWR task (28–100% accuracy), and there was a similarly large range of vocabulary sizes at each timepoint sampled (i.e., 62–664 at 24 months). Expressive vocabulary size (MB-CDI) at 24 months is positively correlated with NWR accuracy at the same age ($r = 0.26$, $p = 0.02$), corroborating previous work on the relationship between the measures (Munson et al., 2005; Hoff et al., 2008)[3].

To model how the acoustic-lexical features of CDS predicted the children's outcomes, we fit a series of linear regression models outlining the relationship between input at the earlier stages—7, 10, and 18 months of age—on the children's outcomes at 24 months. Because there were different effects of acoustic and lexical CDS parameters by child age on NWR accuracy and vocabulary, we model acoustic and lexical parameters separately in the following sections.

### 4.2.1. Modeling Predictors of Phonological Processing

Linear models were fit to predict each child's accuracy on the NWR task. To ensure that any effect of the CDS measures on NWR performance was attributable to the *input*, we needed to control for well-known baseline covariates of input (Maternal Education) and NWR (vocabulary size). Consequently, all NWR

[3]We report on the relationship between NWR and concurrent vocabulary (24 months) measure because only approximately half of the children completed measures at 18 months.

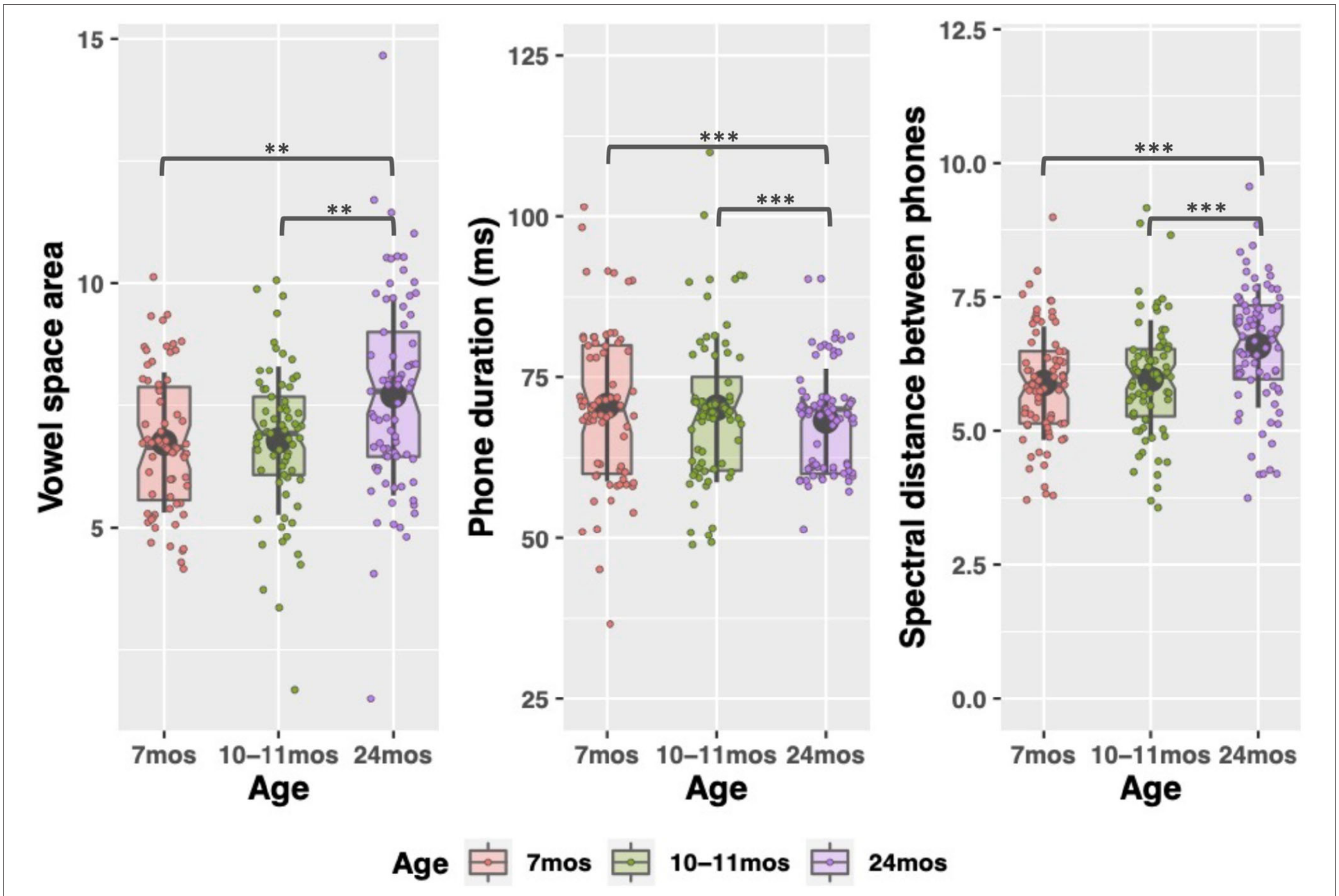



**FIGURE 3 |** Acoustic characteristics of child-directed speech between 7 and 24 months. Large, gray points indicate mean; whiskers indicate 1 SD from mean. Notches indicate median. Individual datapoints indicate each caregivers' vowel space size (left figure), or median phone duration and coarticulation (center and right figures). *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

**TABLE 3 |** Child outcome measures at 7, 10, 11, 18, and 24 months.

| | 7mos<br>Mean (SD) range | 10mos<br>Mean (SD) range | 11mos<br>Mean (SD) range | 18mos<br>Mean (SD) range | 24mos<br>Mean (SD) range |
|---|---|---|---|---|---|
| MB-CDI (receptive)* | 9.55 (13.7) 0–83 | 41.81 (51.13) 0–359 | 62.49 (59.89) 1–331 | NA | NA |
| MB-CDI (expressive) | NA | NA | NA | 112.03 (108.6) 2–472 | 355.94 (150.32) 62–664 |
| PPVT-4 (raw) | NA | NA | NA | NA | 32.85 (12.36) 12–60 |
| EOWVT (stan.) | NA | NA | NA | NA | 97.94 (12.38) 55–118 |
| Nonword rep. accuracy | NA | NA | NA | NA | 0.65 (0.16) 0.28–1 |

**MB-CDI measures receptive vocabulary from 7 to 11 months and expressive from 18 to 24 months.*

modeling includes these variables. In all cases, we modeled vocabulary concurrently with input since we wanted to control for the *predictive* nature of vocabulary for NWR and not simply its correlation with NWR at 24 months.

We next evaluated the role of each potential phono-lexical parameter: Word Frequency, Word Length (in phonemes), Phonological Neighborhood Density, Phonotactic Probability, Number of Word Types, Number of Word Tokens, MATTR, and TTR. Because the latter four variables are unnested, meaning they only provide one observation per transcript, while others are nested (i.e., each word type present in the transcript contributes an observation of phonotactic probability), we could not directly compare all of the variables in a straightforward manner. So we first present models of the nested variables, like Word Frequency, then the unnested variables, like Number of Word Types, and finally we propose a solution to model all phono-lexical parameters simultaneously.

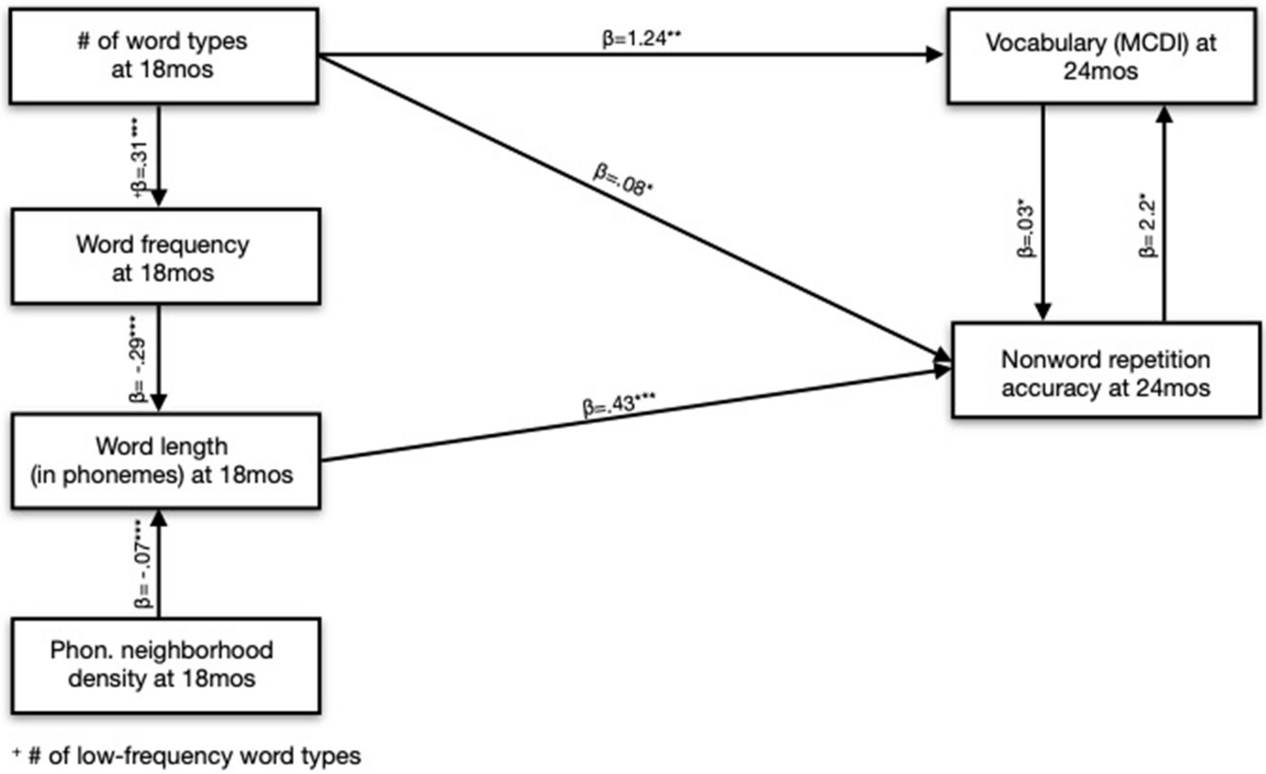



**FIGURE 4 |** Lexical predictors at 18 months of nonword repetition accuracy and vocabulary size at 24 months.

In our modeling of nested lexical CDS parameters, we found effects of Word Length, Word Frequency, and Phonological Neighborhood Density in the CDS sample at 18 months on the children's NWR accuracy at 24 months. Children who heard longer words, less frequent words, and words from sparser neighborhoods, tended to have higher NWR accuracy. We were interested in the distinct contribution of each of these variables, but they are highly correlated (i.e., high-frequency words tend to be shorter) (Correlation matrices included in **Supplementary Materials**). So, to determine which correlated parameter(s) resulted in the best model fit, we regressed out each parameter's contribution to the model (Gahl et al., 2012). Specifically, our model-fitting procedure consisted of the following steps:

1. We fit a series of simple linear models predicting the role of each correlated parameter on the other. For example, we fit a model predicting the role of Word Frequency on Word Length. The resulting residuals from that model represented the contribution of Word Length not attributable to Word Frequency.
2. We included the calculated residuals and the ambiguous parameter (representing Word Frequency or Word Length) in linear models predicting outcomes like NWR or vocabulary.
3. We evaluated if the calculated residuals predicted the developmental outcome, above and beyond the ambiguous parameter.

In a model with Word Length residuals and Word Frequency (where Word Frequency could indicate either the role of Word Frequency *or* Word Length), we found that Word Length residuals predicted NWR accuracy. However, in a model with Word Frequency residuals and Word Length (where Word Length could indicate either the role of Word Frequency *or* Word Length), Word Frequency residuals did *not* predict NWR accuracy. From these results, we concluded that there was a direct effect of Word Length on NWR accuracy: children who heard longer words had stronger NWR skills. We also concluded that the observed effect of Word Frequency on NWR accuracy was indirect and explained by Word Length: children who heard less-frequent words demonstrated better phonological processing skills, but only because less-frequent words tend to be longer (in phonemes) (**Figure 4** and **Table 4**).

We carried out a similar procedure to evaluate the relationship between Phonological Neighborhood Density and NWR accuracy. In a model with Word Length residuals and Phonological Neighborhood Density (where Phonological Neighborhood Density could indicate Phonological Neighborhood Density or Word Length), Word Length residuals predicted NWR accuracy. However, in a model with Phonological Neighborhood Density residuals and Word Length (where Word Length could indicate either parameter), Phonological Neighborhood Density residuals did *not* predict NWR accuracy. On the basis of this modeling, we also concluded that the effects of Phonological Neighborhood Density at 18 months were explained by Word Length. Again, children's phonological processing appears to benefit from hearing words from sparser neighborhoods, but this relationship is entirely explained by the fact that sparser words tend to be longer in length.

Lastly, we found effects of Word Length and Word Frequency at 7 months on NWR accuracy at 24 months, controlling for the children's receptive vocabulary size at 7 months. Following the same technique just outlined to regress out the contribution of the two correlated variables (Word Length and Word Frequency), we found that *both* Word Length and Word Frequency at 7 months predicted NWR at 24 months (see **Supplementary Materials** for visual and model summary). Altogether, however, the model of phono-lexical input at 18 months was a better fit to the NWR outcome.

We next assessed how the unnested lexical variables, like Word Token Count, predicted NWR outcomes. Only the parameter Number of Word Types at 18 months improved upon a model controlling for Maternal Education and the children's expressive vocabulary at 18 months: children who heard more word types at 18 months had higher NWR accuracy at 24 months. None of the remaining parameters (Word Tokens, TTR, MATTR) at any timepoint improved upon the baseline model.

To conclude the lexical modeling, we wanted to evaluate the contributions of Word Length and Number of Word Types at 18 months on NWR accuracy. When comparing nested and unnested independent variables such as these, researchers typically condense the nested variable to avoid overinflating the effect of the unnested variable (Foster-Johnson and Kromrey, 2018). One could, for example, model the average length of all word types and compare it to the number of word types. However, we assumed that caregivers did not necessarily differ in the average length of words in their speech. All speakers must use a large number of short, function words to communicate, resulting in little between-caregiver variability in a hypothetical parameter such as "average word frequency." Instead, we hypothesized that caregivers would vary in the number of outlier observations—in this case, long words—in their speech.

With this idea in mind, we calculated the median word length of all word types uttered by all caregivers at 18 months. The median word length was four phonemes. Then, for each caregiver, we counted how many words they produced that were equal to or longer than (in phonemes) this median word length. The result of this calculation was a new unnested parameter that we created called Number of Long Words. Crucially, because Number of Long Words was unnested, we could directly compare it to Number of Word Types in a model predicting NWR outcomes.

Number of Long Words and Number of Word Types are necessarily correlated (the more distinct words you use, the longer your average word length). So, we regressed out the effect of each of these variables on the other to calculate residuals following the same method previously outlined. In a model predicting NWR outcomes, neither residuals for the parameter Number of Word Types nor Number of Long Words was significant. This result indicated that it was not possible to disentangle the effect of Number of Long Words from Number of Word Types on NWR accuracy: the modeling suggests that both variables, together, predict NWR accuracy (**Figure 5**).

**TABLE 4 |** Modeling the effect of lexical CDS parameters at 18 months on nonword repetition at 24 months.

| | **Word length Estimate** ***p-value*** **(95% CI)** | **Word types Estimate** ***p-value*** **(95% CI)** |
|---|---|---|
| Intercept | $\beta = 43.36^{***}$ (42.15, 44.57) $t = 70.31$ $p < 0.001$ | $\beta = 22.14$ (−3.30, 47.57) $t = 1.71$ $p = 0.10$ |
| Word length | $\beta = 0.43^{***}$ (0.19, 0.66) $t = 3.55$ $p < 0.001$ | |
| Word types | | $\beta = 0.08^{*}$ (0.004, 0.16) $t = 2.06$ $p = 0.05$ |
| Exp. vocab. (18 months)† | $\beta = -0.03$ (−0.03, −0.03) | $\beta = -0.03$ (−0.07, 0.01) $t = -1.30$ $p = 0.21$ |
| Mat. Ed. | $\beta = 7.38$ (7.05, 7.71) | $\beta = 7.27^{**}$ (2.44, 12.10) $t = 2.95$ $p = 0.01$ |
| Observations | 8,317 | 38 |
| Residual Std. Error | 14.79 ($df = 8{,}313$) | 14.84 ($df = 34$) |
| F Statistic | $704.19^{***}$ ($df = 3$; 8,313) ($p < 0.001$) | $4.38^{*}$ ($df = 3$; 34) ($p = 0.02$) |

*$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.*

†*In models containing unnested variables (Maternal Education and Vocabulary Size) and nested variables that are actually of interest (i.e., Word Length), alpha values and standard errors are artificially inflated so these statistics are not reported.*

Our final models predicting NWR evaluated the influence of acoustic CDS features: Vowel Space Size, Vowel Token Dispersion, Coarticulation, and Phone Duration. The effect of acoustic features was only apparent in the CDS sample from 10 to 11 months, not 7 months (acoustics were not measured at 18 months).

We found no effect of Vowel Space Size or Vowel Token Dispersion, at any time point, on the children's NWR outcomes, after controlling for Receptive Vocabulary at 11 months and Maternal Education[4]. However, both Coarticulation and Phone Duration at 10 months negatively predicted the children's NWR: children who heard slower, less coarticulated speech performed worse on the NWR task (**Table 5**).

We were interested in teasing apart the roles of Phone Duration and Coarticulation for NWR accuracy. However, the

[4]For the 10–11 month timepoint models, we modeled the children's vocabulary at 11 months, even though we had samples at 10 months as well, because 2 children were reported to not recognize any words at 10 months.

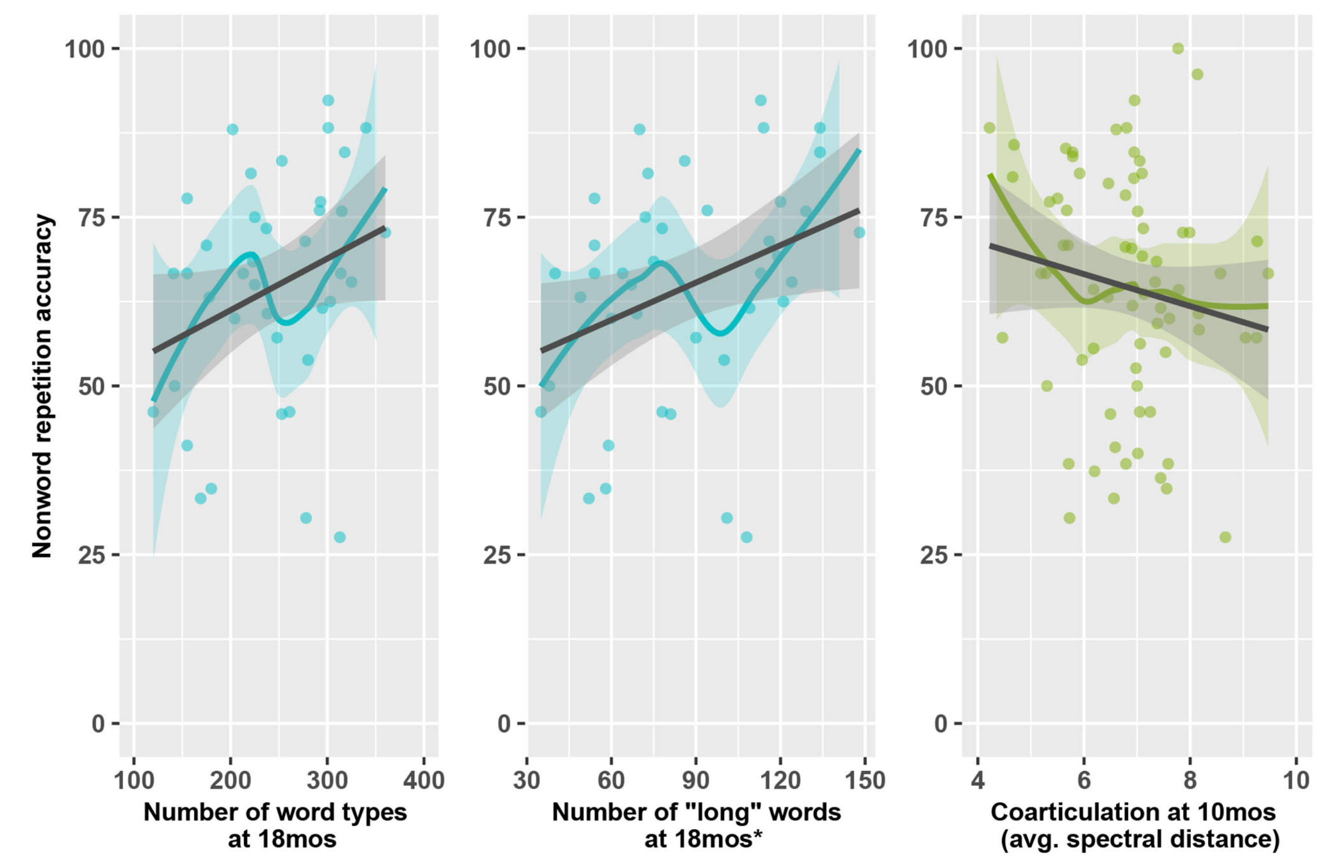



**FIGURE 5 |** Acoustic-lexical predictors of nonword repetition at 24 months. On the left graph, increased spectral distance indicates less coarticulation. Gray regression represents model predictions. Colored points and local regressions represent original data. Ribbons represent 95% confidence intervals. *Word types containing 4+ phonemes.

parameters are related because speakers tend to coarticulate more in faster speech (Gay, 1981). In our modeling of NWR accuracy, the best fit only included the variable Coarticulation (at 10 months), not Phone Duration. Adding Phone Duration to this model resulted in a slightly worse fit, and Phone Duration was not significant in the model summary. Nevertheless, we elected to include Phone Duration in the final model to control for the effect of speaking rate. Consequently, our Coarticulation parameter in the final model more accurately reflects the *unique* contribution of Coarticulation on NWR outcomes, controlling for speaking rate (**Figure 6**). Overall, however, we conclude that Coarticulation completely mediates the effect of Phone Duration on the children's NWR.

Additionally, although phonetic reduction, like coarticulation, is positively correlated with lexical statistics such as word frequency and word length (after controlling for frequency), we did not find effects of Word Length or Word Frequency at 10 months on the children's NWR outcomes. Thus, the effect of Coarticulation is not merely masking lexical effects such as Word Length: children who heard more coarticulated speech at 10 months performed better on the NWR task, irrespective of word length.

#### 4.2.2. Modeling Predictors of Expressive Vocabulary Size

In the final section, we modeled the effects of acoustic-lexical CDS parameters at 7, 10, and 18 months on expressive vocabulary size at 24 months. Linear models were fit to predict each child's reported expressive vocabulary size at 24 months. We again included baseline covariates known to predict children's vocabulary outcomes (Maternal Education and Child Gender): children of mothers with more years of education had larger vocabularies and girls had larger vocabularies than boys. All subsequent modeling includes these variables. As before, we first evaluate the nested lexical variables, then the unnested variables, and finally the acoustic parameters.

In a model predicting the children's expressive vocabulary size at 24 months, we found significant effects of Word Frequency and Word Length at 18 months. (We additionally found effects for these variables at 7 and 10 months, but the best model fit was again found for the measures at 18 months.) As before, we attempted to disentangle the effects of Word Frequency and Word Length by regressing the variables out. In doing so, we found effects of Word Frequency residuals in a model containing Word Frequency residuals and Word Length (where Word Length could indicate Word Frequency or Word Length),

**TABLE 5 |** Modeling the effect of acoustic CDS parameters at 10–11 months on nonword repetition at 24 months.

| | Estimate $^{p\text{-value}}$ (95% CI) |
|---|---|
| Intercept | $\beta = 50.24^{***}$ |
| | (48.05, 52.44) |
| | $t = 44.88$ |
| | $p < 0.001$ |
| Spectral distance | $\beta = -0.17^{**}$ |
| | (−0.28, −0.06) |
| | $t = -2.97$ |
| | $p = 0.004$ |
| Phone duration | $\beta = -4.99$ |
| | (−13.81, 3.83) |
| | $t = -1.11$ |
| | $p = 0.27$ |
| Recep. vocab. (11 months)$^{\dagger}$ | $\beta = 0.02$ |
| | (0.01, 0.03) |
| Mat. Ed. | $\beta = 3.69$ |
| | (3.13, 4.24) |
| Observations | 5,241 |
| Residual Std. Error | 17.09 ($df = 5{,}236$) |
| F Statistic | 50.72*** ($df = 4$; 5,236) ($p < 0.001$) |

*$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.*

$^{\dagger}$*In models containing unnested variables (Maternal Education and Vocabulary Size) and nested variables that are actually of interest (i.e., Word Length), alpha values and standard errors are artificially inflated so these statistics are not reported.*

but no reliable effect of Word Length residuals in a model containing Word Length residuals and Word Frequency. This result indicates a direct effect of Word Frequency on vocabulary outcomes. Additionally, it shows that the effect of Word Length at 18 months on children's vocabulary outcomes at 24 months is entirely explained by Word Frequency: unsurprisingly, children who heard less frequent words tended to have larger vocabularies.

We next evaluated the contribution of unnested parameters like Number of Word Types. As anticipated from previous work, Number of Word Types, at 7, 10, and 18 months, significantly predicted the children's vocabulary sizes at 24 months: children who heard more diverse words went on to have larger vocabularies. Number of Word Tokens was likewise significant, but Number of Word Types provided the best model fit, indicating, as previous work has established (Rowe, 2012), that word diversity was of greater importance than raw word quantity for children's vocabulary development.

We next wanted to evaluate the contributions of Word Frequency and Number of Word Types at 18 months on children's vocabulary sizes at 24 months. To compare nested and unnested variables, we followed the same steps previously outlined: first, we calculated the median frequency of all word types produced by all caregivers at 18 months. Then, we calculated the number of words below the word frequency median that the caregiver produced. The result was an unnested parameter, Number of Low Frequency Words, that we compared to the unnested parameter Number of Word Types by regressing out the effect of each variable on the other and modeling the ensuing residuals.

We found a significant effect of Number of Word Types residuals in a model with those residuals and Number of Low Frequency Words. However, we did not find an effect of Number of Low Frequency Words residuals in a model with those residuals and Number of Word Types. This result led us to conclude that the Number of Word Types is the most relevant predictor of children's vocabulary outcomes: word frequency is only predictive in that if caregivers use more diverse words, they will, necessarily, eventually use words with lower statistical frequency in English (**Figure 4** and **Table 6**).

As a final step in our modeling of vocabulary outcomes, we wanted to evaluate the contribution of the acoustic CDS parameters on the children's vocabulary outcomes. Unlike the NWR outcome, we found a significant, positive effect of increased, unnormalized Vowel Space Size at 10 months on expressive vocabulary size at 24 months (there was no effect of Vowel Token Dispersion at 7 or 10 months—the timepoints where acoustics were measured)[5]. However, an unnested model including only the parameter Word Type Count, as well as baseline covariates of Maternal Education and Gender, provided a better fit to the data so we conclude that Vowel Space Size is not a reliable predictor of vocabulary size in this dataset once lexical diversity of the input is considered.

We again found effects of Coarticulation and Phone Duration at 10 months on the expressive vocabulary outcomes. Specifically, children who heard slower, less coarticulated speech at 10 months had larger vocabularies at 24 months, controlling for Gender and Maternal Education, meaning that the direction of the effect of Coarticulation and Phone Duration differed by outcome (vocabulary vs. NWR). Again, speaking rate and coarticulation are correlated since speakers tend to coarticulate more in faster speech. Coarticulation improved upon a model with just Phone Duration, and both parameters were significant in the final model summary, leading us to conclude that both Phone Duration and Coarticulation, together, explained vocabulary outcomes. However, Coarticulation does, in part, mediate the effect of Phone Duration since (1) both parameters were significant in the model and (2) increased speaking rates *cause* increased coarticulation, but increased coarticulation does not cause speaking rates to increase (**Table 7**).

## 5. DISCUSSION

Child-directed speech changes over the first years of a child's life, with ramifications for speech and language development (Stern et al., 1983; Huttenlocher et al., 2010; Ko, 2012; Hartman et al., 2017; Kalashnikova et al., 2018; Silvey et al.,

[5]There was no effect of normalized Vowel Space Size on vocabulary outcomes, leading us to hypothesize that previous reports on the developmental benefits of expanded vowel spaces could be attributable to an expanded f0 range, something that is controlled for in normalized vowel data.

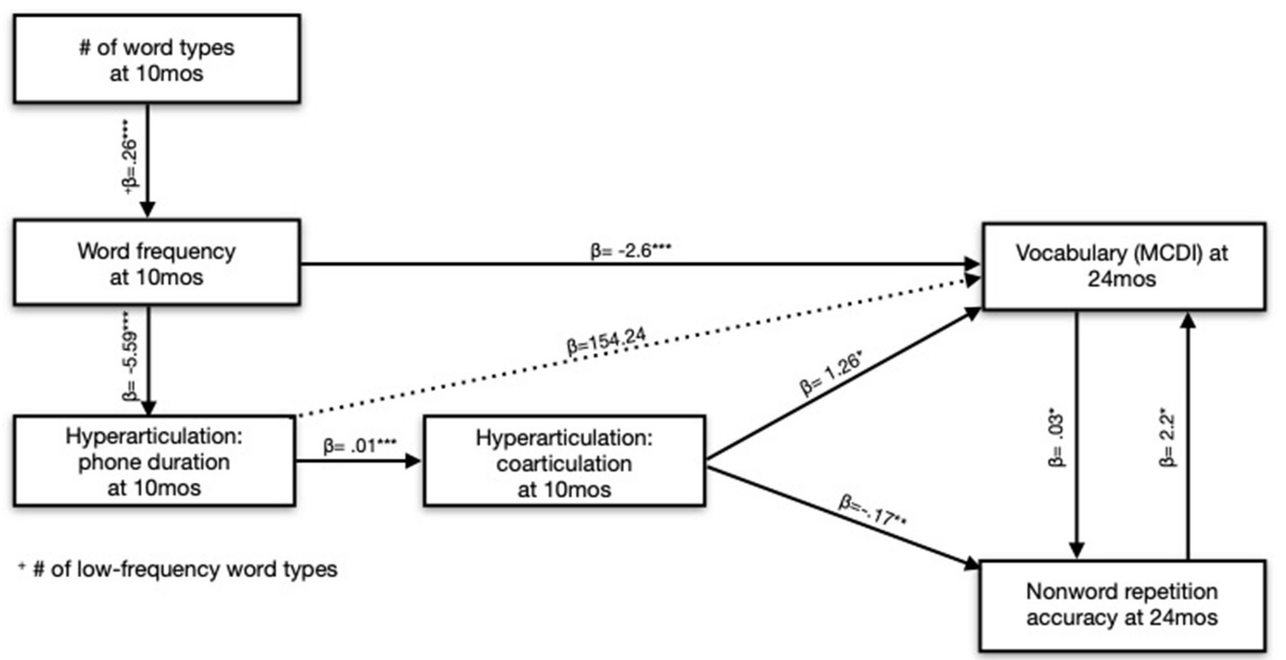



**FIGURE 6 |** Acoustic predictors of nonword repetition accuracy and vocabulary size at 24 months.

2021). While age-related acoustic changes in CDS are well-documented, lexical statistics such as phonotactic probability and word frequency—which, crucially, are reflected in the acoustics of adult speech—are not. Consequently, the first goal of this paper was to document longitudinal changes in an exhaustive set of acoustic, phonological, and lexical characteristics in North American CDS between 7 and 24 months of age. Unsurprisingly, most of the CDS characteristics we measured did change as children aged. However, the measures did not necessarily progress over development in directions anticipated from previous work. Instead, we found that the most hyperarticulated speech occurred at 24 months, even as other characteristics of CDS became more adult-like, and we also observed a tendency for somewhat simplified CDS at 10–11 months.

## 5.1. Age-Related Changes in Child-Directed Speech

CDS is frequently described as a hyperarticulated speech register (Fernald, 2000), with classic theories arguing that the expanded vowel space enhances and clarifies acoustic categories (Kuhl et al., 1997). Hyperarticulation in CDS, along with other classic CDS characteristics such as a dynamic fundamental frequency, slower speaking rate, and shortened utterance length, is thought to reduce into an adult-directed speech register as children age. Yet caregivers here tended to hyperarticulate the most at 24 months, the oldest developmental stage observed, at a time when their speech might otherwise be expected to at least *start* resembling a more adult-directed register. There was also a trend—that did not emerge as significant in the modeling—for CDS characteristics to increase at 10–11 months relative to 7.

To a certain extent, North American caregivers are thought to modify parameters of their input, including the phonetics and phonology, to accommodate children's developing linguistic capacities (Snow, 1972; Gros-Louis et al., 2006; Leung et al., 2020). For example, caregivers' vowel spaces tend to expand as children start learning words (Dilley et al., 2014). So it is possible that the hyperarticulation we observed at 24 months stems from caregivers' implicit attempts to highlight phonological contrasts and elucidate individual segments in the input as their children are learning more words. However, we believe that the hyperarticulation at 24 months could have an additional source: the relationship between phono-lexical statistics and speech production. A coarse summary of the relationship between the structure of the lexicon and speech production is that phonetic reduction accompanies language use: short, probabilistic, and high-frequency words, from dense neighborhoods, tend to be phonetically reduced. And one defining characteristic that we observed of the CDS at 24 months was the overall use of longer, lower-frequency words, from sparser neighborhoods. It could thus be that the hyperarticulation observed at 24 months is not necessarily attributable to more extreme CDS at this timepoint or caregivers' implicit attempts to elucidate phonetic categories; rather, this hyperarticulation could reflect the statistically predictable properties of words that caregivers used when speaking to their children at that age.

An alternative explanation for the hyperarticulation at 24 months, and the trend for increased CDS at 10–11 months relative to 7, is that parents may only fine-tune aspects of their input after a certain developmental stage. A caregiver may assume that accommodation is unnecessary before their child has achieved certain levels of linguistic and conceptual maturity. And children's linguistic capacities, especially lexical and phonological, do change rapidly and noticeably over the time period sampled. At 7 months, typically-developing infants have just begun producing consonant-vowel transitions and reduplicated syllables (e.g., "bababa") (Fagan, 2009). But by 10–11 months, a sizeable proportion of infants' vocalizations contain these transitions and reduplications, which are produced at increasingly faster speeds and with more fully-resonant vowels (Oller, 2000). Then, at 18 months, most infants have begun producing single, recognizable words and by 24 months their vocabularies are expanding rapidly during fast-mapping.

As caregivers are more likely to respond to infants' speech-like than non-speech-like vocalizations (Warlaumont et al., 2014), and to differentiate their feedback by the quality of infant vocalizations (Gros-Louis et al., 2006), we might expect input to differ between many of these timepoints. Specifically, we may observe hyperarticulation at 24 months, and a trend towards hyper CDS at 10–11 months, because caregivers could be engaging in cooperative communication (Renzi et al., 2017). They may recognize a need for linguistic accommodation to their infants at these ages thanks to, ironically, the infants' more advanced phonological and lexical capabilities and propensity to engage in contingent interaction compared to earlier timepoints.

Consequently, Goldilocks zones of infant phonological and lexical development—infants who are increasingly responsive and phonologically mature but not as linguistically advanced as young toddlers—may explain the hyperarticulation at 24 months and the trend toward hyper CDS (reduction in word type and token count, as well as neighborhood density and word frequency) at 10–11 months.

## 5.2. Language Input Drives Phonological Processing

The effects of language input on children's early lexical and morphosyntactic development have long been observed (Hoff, 2003; Huttenlocher et al., 2010; Bernstein Ratner, 2013; Weisleder and Fernald, 2013). Results concerning the role of input on phonological development, especially phonological processing and NWR, have been less conclusive. On the one hand, computational modeling and behavioral research on populations naturally-differing in input experience (bilingual children, cultures with low reported CDS rates) suggest that input *could* play a substantial role in some areas of speech development (Parra et al., 2011; Jones, 2016; Cristia et al., 2020). However, unlike other areas of language development, speech production interacts directly with the child's developing articulatory capabilities, potentially rendering production more immune to external factors such as adult input. Consequently, the second goal of this paper was to examine how the acoustic-lexical characteristics of CDS predicted children's NWR at 24 months. Given the strong, bi-directional relationships between children's NWR abilities and vocabulary sizes, we additionally modeled predictors of vocabulary growth. Overall, we found strong evidence for multi-faceted effects of input on NWR, suggesting that past null results could stem from the input measures assessed. The distinct effects of lexical diversity, word length, and hypoarticulation (coarticulation) on children's speech-language outcomes are addressed in the following sections.

**TABLE 6 |** Modeling the effect of lexical CDS parameters at 18 months on expressive vocabulary at 24 months.

| | Estimate $^{p\text{-value}}$ (95% CI) |
|---|---|
| Intercept | $\beta = -67.15$ |
| | (−320.88, 186.58) |
| | $t = -0.52$ |
| | $p = 0.61$ |
| Word types | $\beta = 1.24^{**}$ |
| | (0.47, 2.02) |
| | $t = 3.13$ |
| | $p = 0.004$ |
| Gender: male | $\beta = -75.80$ |
| | (−176.75, 25.15) |
| | $t = -1.47$ |
| | $p = 0.15$ |
| Mat. Ed. | $\beta = 39.41$ |
| | (−8.89, 87.72) |
| | $t = 1.60$ |
| | $p = 0.12$ |
| Observations | 40 |
| Residual Std. Error | 150.37 ($df = 36$) |
| F Statistic | $4.41^{**}$ ($df = 3$; 36) ($p = 0.01$) |

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

### 5.2.1. Lexical Diversity and Word Length Predict Phonological Processing

Lexical diversity in children's input, above and beyond quantity, results in stronger outcomes for just about every area of language development (Huttenlocher et al., 2010; Rowe, 2012). Here the relationship between lexical diversity and phonological processing/NWR could be explained as children who hear more word types from caregivers have more practice encountering, and potentially repeating, new words varying in phonological structure, length, and semantic content—skills relied upon during NWR. There are potentially additional, more indirect effects of lexical diversity on NWR as well. For example, children who are exposed to more diverse words in their input may also restructure their lexicons, including phonological neighborhoods, at a younger age relative to children who are repeatedly exposed to the same words (Charles-Luce and Luce, 1990; Storkel, 2004a). Among other effects, this lexical restructuring results in greater phonological awareness and phonological abstraction, allowing the children to repeat novel sequences of phones during the NWR task.

NWR ability is a key metric of phonological working memory (Gathercole, 2006; Pierce et al., 2017). Children who perform better on the task are better able to encode, remember, and

**TABLE 7 |** Modeling the effect of acoustic-lexical CDS parameters at 10 months on expressive vocabulary at 24 months.

| | Word frequency Estimate $^{p\text{-value}}$ (95% CI) | Coarticulation Estimate $^{p\text{-value}}$ (95% CI) |
|---|---|---|
| Intercept | $\beta = 323.56^{***}$ | $\beta = 310.80^{***}$ |
| | (313.01, 334.12) | (292.21, 329.39) |
| | $t = 60.09$ | $t = 32.77$ |
| | $p < 0.001$ | $p < 0.001$ |
| Word frequency | $\beta = -2.60^{***}$ | |
| | (−3.48, −1.72) | |
| | $t = -5.81$ | |
| | $p < 0.001$ | |
| Spectral distance | | $\beta = 1.26^{*}$ |
| | | (0.27, 2.26) |
| | | $t = 2.49$ |
| | | $p < 0.02$ |
| Phone duration | | $\beta = 154.24^{***}$ |
| | | (77.52, 230.96) |
| | | $t = 3.94$ |
| | | $p < 0.001$ |
| Gender: male | $\beta = -54.57^{***}$ | $\beta = -60.16^{***}$ |
| | (−59.08, −50.06) | (−68.32, −52.01) |
| | $t = -23.74$ | $t = -14.46$ |
| | $p < 0.001$ | $p < 0.001$ |
| Mat. Ed. | $\beta = 17.89^{***}$ | $\beta = 22.50^{***}$ |
| | (15.13, 20.64) | (17.67, 27.34) |
| | $t = 12.72$ | $t = 9.12$ |
| | $p < 0.001$ | $p < 0.001$ |
| Observations | 16,214 | 5,255 |
| Residual Std. Error | 144.25 ($df = 16{,}210$) | 148.55 ($df = 5{,}250$) |
| F Statistic | $261.84^{***}$ ($df = 3$; 16,210) ($p < 0.001$) | $78.91^{***}$ ($df = 4$; 5,250) ($p < 0.001$) |

*$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.*

*†In models containing unnested variables (Maternal Education and Gender) and nested variables that are actually of interest (i.e. Word Frequency), alpha values and standard errors are artificially inflated, so these statistics are not reported.*

articulate speech sounds. Modeling in this paper demonstrated that children who were exposed to longer words (in phonemes) performed better on the task, even after controlling for numerous variables, such as word frequency, that covary with word length. The effect of word length upon children's NWR accuracy could operate in the following manner: over time, children who are exposed to a higher modal word length in the input may hone their ability to remember sequences of phones—that is, their input provides them increased opportunity to develop their phonological working memories, a central component evoked during NWR.

Together, lexical diversity and word length dually contribute to NWR abilities at 24 months. To develop the skills required for NWR, children must be exposed to *diverse* words that allow them to practice novel word repetition and potentially restructure their lexicons. And children must also be exposed to somewhat *longer* words that exercise their phonological working memories. Lexical diversity and word length are still relatively coarse measures of the input. Both measures encompass a variety of constructs. For example, it could be not just the *length* of words in the input that promotes phonological processing but also the syllabic complexity of those words. We did not find effects of phonotactic probability, which may reflect phonological complexity to a certain extent, upon the children's outcomes. But our samples also showed little variability in phonotactic probability between or within speakers, so it could be that larger and/or more naturalistic samples would show effects of more detailed input measures such as phonological complexity or phonotactic probability upon children's phonological processing.

### 5.2.2. Hypoarticulation Drives Phonological Processing; Hyperarticulation Drives Vocabulary Growth

The final predictor of children's speech-language ability at 24 months was the degree of coarticulation in the caregiver's speech at 10 months. An oft-repeated tenet in studies of CDS is that clearer speech leads to better linguistic outcomes, perhaps because CDS helps elucidate phonological categories and demarcates word boundaries, permitting syntactic bootstrapping (Gleitman, 1990). We did indeed find beneficial effects of clear speech at 10 months on children's vocabulary sizes: children of caregivers with more expanded vowel spaces, who spoke more slowly, and coarticulated less, grew larger vocabularies (the effect of vowel space size did not remain relevant after factoring in word type count, however). So, it was initially somewhat surprising to find a beneficial effect of *hypo*articulation, instantiated as increased coarticulation, upon children's phonological processing. We were, once again, able to control for a number of (though certainly not all) confounding variables, such as word frequency and speaking rate, that could otherwise explain the relationship between hypoarticulation and NWR. So the question remains: how does children's phonological processing benefit from hearing speech that is *more* coarticulated?

There are two mechanisms that potentially explain the beneficial effects of hypoarticulation upon phonological processing. First, it is important to clarify that coarticulation is more than random noise and variation in the speech signal. Rather, it provides important, contextual cues about word and segmental identity (Mann and Repp, 1980; Soli, 1981; Mattys et al., 2005; Gow and McMurray, 2007), facilitating word recognition in children as young as 18 months (Mahr et al., 2015). One principle of coarticulation is that it is largely planned (Whalen, 1990), meaning that speakers may subconciously manipulate variability in their speech to enhance communication, including to young children (Zellou and Scarborough, 2015). Thus, the first way that hypoarticulation drives phonological development is via the enhanced sublexical cues that maximally, naturalistically coarticulated input provides.

The other, complementary way that hypoarticulation could facilitate phonological processing outcomes may require reframing our assumptions about the developmental benefits of clear speech. It is often assumed that speech variability introduces noise, overlap, and confusion for infant and child learners, rather than an opportunity for children to scaffold into the adult speech stream. Relative to traditional CDS registers, adult-directed speech is phonetically reduced: it is spoken faster, resulting in more coarticulation and compromised phonological contrasts[6]. Consequently, children who receive more coarticulated speech in their input are exposed to highly confusable, overlapping speech categories, but they are also exposed to highly naturalistic speech exemplars that reflect a typical adult-directed speech register. Rather than a hindrance to development then, highly reduced, naturalistic speech—that nevertheless stems from a predictable source like the child's central caregiver—may prepare children to parse phonological units from a variety of speech registers, not just simplified CDS.

It is important to consider the developmental stages where we observed effects of hypoarticulation vs. word diversity and length upon the children's NWR accuracy. We did not measure the acoustics of CDS at 18 months, but we did not find a concurrent effect of coarticulation in caregiver speech at 24 months upon NWR at the same age. Nor did we find relationships between coarticulation in caregiver speech at 7 months and later NWR. While these null results cannot entirely rule out a role of hypoarticulation at 7 or 24 months—it could be that we didn't have sufficient word types to determine an effect at 7 months, for example—they do suggest that effects of hypoarticulation upon phonological processing outcomes may be limited to a certain developmental period. Why do we observe effects of coarticulation at 10 months, but not the other time periods? And why do we observe effects of lexical statistics, such as word type counts, at 7 and 18 months but *not* 10 months?

We believe these results demonstrate that, for phonological processing, it matters more *how* caregivers speak to 10–11 month-olds than the words they use. Parents who coarticulate more in the speech directed to their children are also speaking faster, thereby reducing their phonological contrasts, all factors that may be preparing their children to process and parse naturalistic speech. This more naturalistic input may even be preparing infants to benefit from overheard, adult-directed speech. It is obvious that a simplified CDS register, with its shortened utterances, isolated words, and longer pauses between utterances, helps infants break into the speech stream at, for example, 6.5–7.5 months (Nelson et al., 1989; Thiessen et al., 2005). Seven- to 8-month-old infants also have stronger lexical recognition and recall for words presented in an infant-directed register than an adult-directed register (Singh et al., 2009). Furthermore, in this study, we still found a clear speech benefit for the children's vocabulary outcomes. But conversely, after a certain point in development, children who are only exposed to easily-segmentable phonemes, syllables, and words may not develop the strongest phonological parsing abilities, making them less prepared to take advantage of more naturalistic, overheard and/or adult-directed speech in their environments.

[6]A faster speaking rate doesn't have to imply reduced speech intelligibility: with training, talkers can produce a clear speech register at a conversational speaking rate (Krause and Braida, 2004).

Taken together, these three predictors of phonological processing—coarticulation, lexical diversity, and word length—suggest a complex, time-varying effect of input upon children's phonological processing outcomes. As such, it is not entirely surprising that previous work on this topic has proven inconclusive. For one thing, some effects of the input, such as word length, may be specific to certain phonological outcomes like NWR. As discussed above, the type of effect, acoustic vs. lexical, also appears to depend upon the timepoint studied.

Since we only sampled the children and their caregivers at discrete, non-random timepoints, we cannot definitively say that certain features (i.e., hypoarticulation) will always best stimulate phonological processing at certain developmental stages (i.e., 24 months). But these results may instead have some broader implications. Caregivers and early educators could consider modifying their speech-language patterns (speed, acoustic reduction, lexical diversity) in accordance with a child's developing linguistic capabilities, gradually increasing the prevalence of adult-directed speech characteristics as children age. Furthermore, there are many benefits of CDS beyond its slower speed and repetitiveness. Infants and children are also attracted to CDS registers because, relative to adult-directed speech, CDS is typified by greater pitch modulations (e.g., Kitamura et al. 2001), more eye-to-eye contact and positive affect between caregivers and children (Singh et al., 2002), and caregivers' exaggerated facial and bodily movements (Brand et al., 2002; Green et al., 2010). So adults could consider combining some aspects of adult-directed speech (e.g., faster speech rates, hypoarticulation) that scaffold the development of phonological processing skills with some aspects of CDS (e.g., positive affect, exaggerated facial expressions) that draw and maintain infants' and children's attention to the speech signal and conversational exchange.

Previous work on input in language development has been somewhat biased to certain outcome measures (vocabulary tests) and input measures (quantity and semantic quality of lexical items) because these are relatively straightforward measures to collect and compute. But a complete model of the role of input in development, one that predicts individual variability in speech production outcomes as well as more traditional measures such as vocabulary size or speech perception, clearly needs to incorporate a diverse set of acoustic and lexical parameters of the input, as this study has demonstrated.

## 5.3. Future Work

This work assessed children's input at 7, 10–11, 18, and 24 months and found time-varying CDS patterns with different effects on children's speech-language outcomes. Going forward, it will be important to sample input at additional, more regular time periods, particularly between 10–11 and 18 months. We cannot say, for example, if these age-related changes in CDS are linear or undergo additional changes at periods that were not observed.

Additionally, although our in-lab CDS samples allowed us to collect the high-quality audio required for the acoustic analysis, these play sessions likely do not entirely reflect typical caregiver-child interactions in the home. They are also of limited length (15–20 min). Some measures may be more biased than others by this sampling method. For example, while we believe that a 15–20 min interaction in the lab may reflect the diversity of word types typical of the caregiver's speech, this sampling strategy may not reflect word token count (and thus measures based on word token counts such as TTR and MATTR). Lexical, and especially phonetic, transcription is a lengthy, painstaking process, but going forward we should strive to collect high-quality acoustic samples of maximally-naturalistic CDS in the home to corroborate the results that we derived from the semi-naturalistic caregiver-child interactions in this paper.

## 6. CONCLUSIONS

The characteristics of child-directed speech (CDS) change over the first years of a child's life. Understanding how these changes unfold, and the consequences they have for children's speech-language development, is a key part of understanding the role of input for language development. We measured lexical, phonological, and acoustic properties of CDS at 7, 10–11, 18, and 24 months and found that the most significant changes in CDS occur in the second year of life. However, the developmental trend of CDS does not always progress to a more adult-directed speech register as children age. Rather, caregivers use a greater number and diversity of words at 24 months, increasing their use of low-frequency words, from sparser phonological neighborhoods, and driving hyperarticulation in their speech. Consequently, another source of hyperarticulation in CDS, beyond caregivers' implicit attempts to highlight phonological contrasts, may be lexical statistics.

We additionally measured how these properties of CDS predicted children's phonological processing and vocabulary at 24 months. Children's phonological processing benefited most from *hypo*articulation at 10 months, and longer, more diverse word types at 18 months, while vocabulary benefited from hyperarticulation and lexical diversity. Thus, novel measures of CDS, beyond lexical quantity and quality, demonstrated how language input could drive phonological development. Taken together, these results demonstrate how different characteristics of CDS vary by children's age, and how those characteristics promote speech-language development at distinct developmental stages.

## DATA AVAILABILITY STATEMENT

The NewmanRatner corpus analyzed for this study can be found in the CHILDES component of Talkbank (https://childes.talkbank.org/access/Eng-NA/NewmanRatner.html).

## ETHICS STATEMENT

This study was approved by the University of Maryland, College Park Institutional Review Board. Participants provided written consent to participate.

## AUTHOR CONTRIBUTIONS

RN and NB collected the data. MC, RN, and JE designed the research. CT scored the children's productions. MC analyzed the child-directed speech samples, conducting the modeling, and wrote the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was funded by National Institute on Deafness and Other Communication Disorders grant T32DC000046 to the University of Maryland (MC) and National Science Foundation grant BCS 0745412 (RN and NB).

## ACKNOWLEDGMENTS

We thank the following individuals for their part in recruiting families, testing children, and transcribing parent-child interactions: Katrina Ablorh, Candace Ali, Saher Ali, Alison Arnold, Megan Askew, Amelie Bail, Catherine Bender, Taryn Bipat, Devon Brunson, Michelle Cass, Danielle Chazen, Alyssa Cook, Jennifer Coon, Sara Davis, Justine Dombroski, Sara Dougherty, Sara Edelberg, Daniel Eisenberg, Meaghan Ervin, Lauren Evans, Andrea Farina, Josefina Fernandez, Lauren Fischer, Andrea Fisher, Arielle Gandee, Richard Garcia, Whitney Goodrich-Smith, Eliana Groskin, Kelly Hartman, Natalie Hein, Sean Hendricks, Laura Horowitz, Tim Howell, Megan Janssen Crenshaw, Mina Javid, Amanda Jensen, Jamie Karen, Caroline Kettl, Michelle Keenan, Esther Kim, Stephanie Lee, Perri Lieberman, Rachel Lieberman, Danielle Lindenger, Rachel Lipinski, Katie Lippitt, Debbie Martinez, Jenn McCabe, Kerry McColgan, Eileen McLaughlin, Kelly McPherson, Debra Mirazhi, Giovanna Morini, Vidda Moussavi, Molly Nasuta, Ashley Nimmo, Courtenay O'Connor, Sabrina Panza, Amanda Pasquarella, Catie Penny, Elise Perkins, Jenna Poland, Lauren Polovoy, Rachel Ports, Rachel Rhodes, Allie Rodriguez, Maria Rodriguez, Christina Royster, Julia Sampson, Judith Segal, Katie Shniderman, Veronica Son, Mara Steinberg, Sarah Steele, Anna Synnesvedt, Justine Taweel, Allison Temple, Dena Tran, Lisa Tuit, Hillary Tyler, Eugene Vassilas, Susan Veppumthara, Krista Voelmle, Chelsea Vogel, Amanda Wildman, Cavena Williams, Kimmie Wilson, Catherine Wu, Donna Zack-Williams, and Michelle Zobel. Additional thanks to Michele Liquori, Olivia Accomando, Farheen Ahmed, Aeshah Tawfik, Max Scribner, Samara Yalisove, Kavya Ganesan, Kaliyah Gowing, and Jillian Weinman for their assistance processing the adult language samples.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.712647/full#supplementary-material

## REFERENCES


Adams, A.-M., and Gathercole, S. E. (1995). Phonological working memory and speech production in preschool children. *J. Speech Lang. Hear. Res.* 38, 403–414. doi: 10.1044/jshr.3802.403

Adams, A.-M., and Gathercole, S. E. (1996). Phonological working memory and spoken language development in young children. *Q. J. Exper. Psychol. Sect. A* 49, 216–233. doi: 10.1080/713755610

Arnold, H., Conture, E., and Ohde, R. (2005). Phonological neighborhood density in the picture naming of young children who stutter: preliminary study. *J. Fluency Disord*. 30, 125–148. doi: 10.1016/j.jfludis.2005.01.001

Baddeley, A., Gathercole, S., and Papagno, C. (1998). The phonological loop as a language learning device. *Psychol. Rev*. 105, 158–173. doi: 10.1037/0033-295X.105.1.158

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw*. 67, 1–48. doi: 10.18637/jss.v067.i01

Bell, A., Brenier, J. M., Gregory, M., Girand, C., and Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *J. Mem. Lang*. 60, 92–111. doi: 10.1016/j.jml.2008.06.003

Bernstein Ratner, N. (1984). Patterns of vowel modification in mother–child speech. *J Child Lang*., 11, 557–578. doi: 10.1017/S030500090000595X

Bernstein Ratner, N. (2013). Why talk with children matters: clinical implications of infant- and child-directed speech research. *Semin. Speech Lang*. 34, 203–214. doi: 10.1055/s-0033-1353449

Boersma, P., and Weenink, D. (2020). *Praat: Doing Phonetics by Computer [Computer Program]. Version 6.1.36*. Available online at: http://www.praat.org/

Brand, R. J., Baldwin, D. A., and Ashburn, L. A. (2002). Evidence for 'motionese': modifications in mothers' infant-directed action. *Dev. Sci*. 5, 72–83. doi: 10.1111/1467-7687.00211

Brownell, R. (2000). *Expressive One-Word Picture Vocabulary Test, 3rd Edn*. Novato, CA: Academic Therapy Publications.

Brysbaert, M., and New, B. (2009). Moving beyond Kučera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav. Res. Methods* 41, 977–990. doi: 10.3758/BRM.41.4.977

Burnham, E. B., Wieland, E. A., Kondaurova, M. V., McAuley, J. D., Bergeson, T. R., and Dilley, L. C. (2015). Phonetic modification of vowel space in storybook speech to infants up to 2 years of age. *J. Speech Lang. Hear. Res*. 58, 241–253. doi: 10.1044/2015_JSLHR-S-13-0205

Charles-Luce, J., and Luce, P. A. (1990). Similarity neighbourhoods of words in young children's lexicons. *J. Child Lang*. 17, 205–215. doi: 10.1017/S0305000900013180

Cooper, R., and Aslin, R. (1990). Preference for infant-directed speech in the first month after birth. *Child Dev*. 61, 1584–1595. doi: 10.2307/1130766

Core, C., Chaturvedi, S., and Martinez-Nadramia, D. (2017). "The Role of language experience in nonword repetition tasks in young bilingual Spanish-English speaking children," in *Proceedings of the 41st Annual Boston University Conference on Language Development* (Somerville, MA: Cascadilla Press), 179–185.

Cristia, A. (2013). Input to language: the phonetics and perception of infant-directed speech. *Lang. Linguist. Compass*. 7, 157–170. doi: 10.1111/lnc3.12015

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

Cychosz, M., Edwards, J. R., Munson, B., and Johnson, K. (2019). Spectral and temporal measures of coarticulation in child speech. *J. Acoust. Soc. Am. Express Lett*. 146, EL516–EL522. doi: 10.1121/1.5139201

Dilley, L. C., Millett, A. L., Mcauley, J. D., and Bergeson, T. R. (2014). Phonetic variation in consonants in infant-directed and adult-directed speech: The case of regressive place assimilation in word-final alveolar stops. *J. Child Lang*. 41, 155–175. doi: 10.1017/S0305000912000670

Dunn, L. M., and Dunn, D. M. (2007). *Peabody Picture Vocabulary Test-Fourth Edition (PPVT-4)*. Circle Pines, MN: AGS.

Edwards, J., Beckman, M. E., and Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in novel word repetition. *J. Speech Lang. Hear. Res*. 57, 421–436. doi: 10.1044/1092-4388(2004/034)

Erskine, M. E., Munson, B., and Edwards, J. R. (2020). Relationship between early phonological processing and later phonological awareness: evidence from nonword repetition. *Appl. Psycholinguist*. 41, 319–346. doi: 10.1017/S0142716419000547

Fagan, M. K. (2009). Mean length of utterance before words and grammar: Longitudinal trends and developmental implications of infant vocalizations. *J. Child Lang*. 36, 495–527. doi: 10.1017/S0305000908009070

Farabolini, G., Rinaldi, P., Caselli, M. C., and Cristia, A. (2021). Nonword repetition in bilingual children: the role of language exposure, vocabulary scores and environmental factors. *Speech Lang. Hear*. 1–16. doi: 10.1080/2050571X.2021.1879609

Fenson, L., Marchman, V., Thal, D. J., Dale, P., Reznick, J., and Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories User's Guide and Technical Manual, 2nd Edn*. San Diego, CA: Singular.

Fergadiotis, G., Wright, H. H., and West, T. M. (2013). Measuring lexical diversity in narrative discourse of people with aphasia. *Am. J. Speech Lang. Pathol*. 22, S397–S408. doi: 10.1044/1058-0360(2013/12-0083)

Ferguson, C. (1977). "Baby talk as a simplified register," in *Talking to Children: Language Input and Acquisition*, eds C. Snow and C. Ferguson (Cambridge, UK: Cambridge University Press), 209–235.

Ferjan Ramírez, N., Lytle, S. R., Fish, M., and Kuhl, P. K. (2019). Parent coaching at 6 and 10 months improves language outcomes at 14 months: a randomized controlled trial. *Dev. Sci*. 22:e12762. doi: 10.1111/desc.12762

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant. Behav. Dev*. 8, 181–195. doi: 10.1016/S0163-6383(85)80005-9

Fernald, A. (2000). Speech to infants as hyperspeech: knowledge-driven processes in early word recognition. *Phonetica* 57, 242–254. doi: 10.1159/000028477

Foster-Johnson, L., and Kromrey, J. D. (2018). Predicting group-level outcome variables: an empirical comparison of analysis strategies. *Behav. Res. Methods* 50, 2461–2479. doi: 10.3758/s13428-018-1025-8

Furrow, D., Nelson, K., and Benedict, H. (1979). Mothers' speech to children and syntactic development: some simple relationships. *J. Child Lang*. 6, 423–442. doi: 10.1017/S0305000900002464

Gahl, S. (2008). Time and thyme are not homophones: the effect of lemma frequency on word durations in spontaneous speech. *Language* 84, 474–496. doi: 10.1353/lan.0.0035

Gahl, S., Yao, Y., and Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *J. Mem. Lang*. 66, 789–806. doi: 10.1016/j.jml.2011.11.006

Gathercole, S. E. (2006). Nonword repetition and word learning: the nature of the relationship. *Appl. Psycholinguist*. 27, 513–543. doi: 10.1017/S0142716406060383

Gathercole, S. E., and Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: a longitudinal study. *J. Mem. Lang*. 28, 200–213. doi: 10.1016/0749-596X(89)90044-2

Gathercole, S. E., Willis, C., Emslie, H., and Baddeley, A. D. (1991). The influences of number of syllables and wordlikeness on children's repetition of nonwords. *Appl. Psycholinguist*. 12, 349–367. doi: 10.1017/S0142716400009267

Gay, T. (1981). Mechanisms in the control of speech rate. *Phonetica* 38:148–158. doi: 10.1159/000260020

Gerosa, M., Lee, S., Giuliani, D., and Narayanan, S. (2006). "Analyzing children's speech: an acoustic study of consonants and consonant-vowel transition," in *2006 IEEE International Conference on Acoustics Speed and Signal Processing Proceedings, Vol. 1* (Toulouse: IEEE), 393–396.

Glas, L., Rossi, C., Hamdi-Sultan, R., Batailler, C., and Bellemmouche, H. (2018). Activity types and child-directed speech: a comparison between French, Tunisian Arabic and English. *Can. J. Linguist*. 63, 633–666. doi: 10.1017/cnj.2018.20

Gleitman, L. (1990). The structural sources of verb meanings. *Lang. Acquis*. 1, 3–55. doi: 10.1207/s15327817la0101_2

Gow, D. W., and McMurray, B. (2007). Word recognition and phonology: The case of English coronal place assimilation. *Pap. Lab. Phonol*. 9, 173–200.

Green, J. R., Nip, I. S. B., Wilson, E. M., Mefferd, A. S., and Yunusova, Y. (2010). Lip movement exaggerations during infant-directed speech. *J. Speech Lang. Hear. Res*. 53, 1529–1542. doi: 10.1044/1092-4388(2010/09-0005)

Gros-Louis, J., West, M. J., Goldstein, M. H., and King, A. P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *Int. J. Behav. Dev*. 30, 509–516. doi: 10.1177/0165025406071914

Guo, L.-Y., McGregor, K. K., and Spencer, L. J. (2015). Are young children with cochlear implants sensitive to the statistics of words in the ambient spoken language? *J. Speech Lang. Hear. Res*. 58, 987–1000. doi: 10.1044/2015_JSLHR-H-14-0135

Han, M., de Jong, N. H., and Kager, R. (2018). Lexical tones in Mandarin chinese infant-directed speech: age-related changes in the second year of life. *Front. Psychol*. 9:434. doi: 10.3389/fpsyg.2018.00434

Harrington, J., Kleber, F., Reubold, U., and Siddins, J. (2015). The relationship between prosodic weakening and sound change: evidence from the German tense/lax vowel contrast. *Lab Phonol*. 6, 87–117. doi: 10.1515/lp-2015-0002

Hart, B., and Risley, T. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*. Baltimore, MD: Paul H. Brookes Publishing.

Hartman, K. M., Bernstein Ratner, N., and Newman, R. S. (2017). Infant-directed speech (IDS) vowel clarity and child language outcomes. *J. Child Lang*. 44, 1140–1162. doi: 10.1017/S0305000916000520

Hlavac, M. (2018). *Stargazer: Well-Formatted Regression and Summary Statistics Tables*. Bratislava; Slovakia: Central European Labour Studies Institute (CELSI).

Hoff, E. (2003). The specificity of environmental influence: socioeconomic status affects early vocabulary development via maternal speech. *Child Dev*. 74, 1368–1378. doi: 10.1111/1467-8624.00612

Hoff, E., Core, C., and Bridges, K. (2008). Nonword repetition assesses phonological memory and is related to vocabulary development in 20- to 24-month-olds. *J. Child Lang*. 35, 903–916. doi: 10.1017/S0305000908008751

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

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

Jones, G. (2016). The influence of children's exposure to language from two to six years: the case of nonword repetition. *Cognition* 153, 79–88. doi: 10.1016/j.cognition.2016.04.017

Kalashnikova, M., and Burnham, D. (2018). Infant-directed speech from seven to nineteen months has similar acoustic properties but different functions. *J. Child Lang*. 45, 1035–1053. doi: 10.1017/S0305000917000629

Kalashnikova, M., Escudero, P., and Kidd, E. (2018). The development of fast-mapping and novel word retention strategies in monolingual and bilingual infants. *Dev. Sci*. 21:e12674. doi: 10.1111/desc.12674

Kern, S., and dos Santos, C. (2018). "Invariance in variation: Frequency and neighbourhood density as predictors of vocabulary size," in *Sources of Variation in First Language Acquisition: Languages, Contexts, and Learners, Trends in Language Acquisition Research*, eds M. Hickmann, E. Veneziano, and H. Jisa (Amsterdam: John Benjamins Publishing Company), 183–200.

Kitamura, C., and Burnham, D. (2003). Pitch and communicative intent in mother's speech: adjustments for age and sex in the first year. *Infancy* 4, 85–110. doi: 10.1207/S15327078IN0401_5

Kitamura, C., Thanavishuth, C., Burnham, D., and Luksaneeyanawin, S. (2001). Universality and specificity in infant-directed speech: pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant Behav. Dev*. 24, 372–392. doi: 10.1016/S0163-6383(02)00086-3

Ko, E.-S. (2012). Nonlinear development of speaking rate in child-directed speech. *Lingua* 122, 841–857. doi: 10.1016/j.lingua.2012.02.005

Krause, J. C., and Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *J. Acoust. Soc. Am*. 115, 362–378. doi: 10.1121/1.1635842

Kuhl, P. K., Andruski, J. E., Chistovich, I., Chistovich, L., Kozhevnikova, E., Ryskina, V. L., et al. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science* 277, 684–686. doi: 10.1126/science.277.5326.684

Kuznetsova, A., Brockhoff, P., and Christensen, R. (2017). lmerTest Package: tests in linear mixed-effects models. *J. Stat. Softw*. 82, 1–26. doi: 10.18637/jss.v082.i13

Lenth, R. V. (2021). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.6.1.

Leung, A., Tunkel, A., and Yurovsky, D. (2020). Parents fine-tune their speech to children's vocabulary knowledge. *PsyArXiv*. doi: 10.31234/osf.io/prz8g

Liu, H.-M., Kuhl, P. K., and Tsao, F.-M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Dev. Sci*. 6, F1–F10. doi: 10.1111/1467-7687.00275

Liu, H.-M., Tsao, F.-M., and Kuhl, P. K. (2009). Age-related changes in acoustic modifications of Mandarin maternal speech to preverbal infants and five-year-old children: a longitudinal study. *J. Child Lang*. 36, 909–922. doi: 10.1017/S030500090800929X

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *J. Acoust. Soc. Am*. 49, 606–608. doi: 10.1121/1.1912396

Luce, P. A., and Pisoni, D. B. (1998). Recognizing spoken words: the neighborhood activation model. *Ear Hear*. 19, 1–36. doi: 10.1097/00003446-199802000-00001

MacDonald, M. C., and Christiansen, M. H. (2002). Reassessing working memory: comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychol. Rev*. 109, 35–54. doi: 10.1037/0033-295X.109.1.35

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk, 3rd Edn*. Mahwah, NJ: Lawrence Erlbaum Associates.

Mahr, T., McMillan, B. T., Saffran, J. R., Ellis Weismer, S., and Edwards, J. (2015). Anticipatory coarticulation facilitates word recognition in toddlers. *Cognition* 142:345–350. doi: 10.1016/j.cognition.2015.05.009

Mann, V., and Repp, B. (1980). Influence of vocalic context on perception of the [sh]-[s] distinction. *Percept. Psychophys*. 28, 213–228. doi: 10.3758/BF03204377

Mattys, S. L., White, L., and Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: a hierarchical framework. *J. Exper. Psychol. Gen*. 134, 477–500. doi: 10.1037/0096-3445.134.4.477

McAuliffe, M., Socolof, M., Mihuc, A., Wagner, M., and Sonderegger, M. (2017). *Montreal Forced Aligner*, 2.0 Edn. (Montreal) Software Program.

McCloy, D. R. (2016). *phonR R package*. Available online at: http://drammock.github.io/phonR/.

McFee, B., Raffel, C., Liang, D., Ellis, D., McVicar, M., Battenberg, E., and Nieto, O. (2015). "Librosa: audio and music signal analysis in python," in *Proceedings of the 14th Python in Science Conference* (Austin, TX), 18–24.

Metsala, J. L. (1997). An examination of word frequency and neighborhood density in the development of spoken-word recognition. *Mem. Cogn*. 25, 47–56. doi: 10.3758/BF03197284

Michalczyk, K., Krajewski, K., Pressler, A.-L., and Hasselhorn, M. (2013). The relationships between quantity-number competencies, working memory, and phonological awareness in 5- and 6-year-olds. *Br. J. Dev. Psychol*. 31, 408–424. doi: 10.1111/bjdp.12016

Munhall, K., Fowler, C., Hawkins, S., and Saltzman, E. (1992). "Compensatory shortening" in monosyllables of spoken English. *J. Phon*. 20, 225–239. doi: 10.1016/S0095-4470(19)30624-2

Munson, B., Edwards, J., and Beckman, M. E. (2005). Relationships between nonword repetition accuracy and other measures of linguistic development in children With phonological disorders. *J. Speech Lang. Hear. Res*. 48, 61–78. doi: 10.1044/1092-4388(2005/006)

Nelson, D. G. K., Hirsh-Pasek, K., Jusczyk, P. W., and Cassidy, K. W. (1989). How the prosodic cues in motherese might assist language learning. *J. Child Lang*. 16, 55–68. doi: 10.1017/S030500090001343X

Newman, R. S., and German, D. J. (2002). Effects of lexical factors on lexical access among typical language-learning children and children with word-finding difficulties. *Lang. Speech* 45, 285–317. doi: 10.1177/00238309020450030401

Newman, R. S., Rowe, M. L., and Bernstein Ratner, N. (2016). Input and uptake at 7 months predicts toddler vocabulary: the role of child-directed speech and infant processing skills in language development. *J. Child Lang*. 43, 1158–1173. doi: 10.1017/S0305000915000446

Oller, D. (2000). *The Emergence of the Speech Capacity*. Mahwah, NJ: Lawrence Erlbaum Associates.

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

Pierce, L. J., Genesee, F., Delcenserie, A., and Morgan, G. (2017). Variations in phonological working memory: linking early language experiences and language learning outcomes. *Appl. Psycholinguist*. 38, 1265–1300. doi: 10.1017/S0142716417000236

Raneri, D., Von Holzen, K., Newman, R., and Bernstein Ratner, N. (2020). Change in maternal speech rate to preverbal infants over the first two years of life. *J. Child Lang*. 47, 1263–1275. doi: 10.1017/S030500091900093X

Renzi, D. T., Romberg, A. R., Bolger, D. J., and Newman, R. S. (2017). Two minds are better than one: cooperative communication as a new framework for understanding infant language learning. *Transl. Issues Psychol. Sci*. 3, 19–33. doi: 10.1037/tps0000088

Rowe, M. (2008). Child-directed speech: relation to socioeconomic status, knowledge of child development, and child vocabulary skill. *J. Child Lang*. 35, 185–205. doi: 10.1017/S0305000907008343

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development: child-directed speech and vocabulary. *Child Dev*. 83, 1762–1774. doi: 10.1111/j.1467-8624.2012.01805.x

RStudio Team (2020). *RStudio: Integrated Development for R*. Boston, MA: RStudio, Inc.

Scarborough, R. (2005). *Coarticulation and the Structure of the Lexicon*. Unpublished doctoral dissertation, University of California, Log Angeles, CA.

Silvey, C., Demir-Lira, E., Goldin-Meadow, S., and Raudenbush, S. (2021). Effects of time-varying parent input on child language outcomes differ for vocabulary and syntax. *Psychol. Sci*. 32, 536–548. doi: 10.1177/0956797620970559

Singh, L., Morgan, J. L., and Best, C. T. (2002). Infants' listening preferences: baby talk or happy talk? *Infancy* 3, 365–394. doi: 10.1207/S15327078IN0303_5

Singh, L., Nestor, S., Parikh, C., and Yull, A. (2009). Influences of infant-directed speech on early word recognition. *Infancy* 14, 654–666. doi: 10.1080/15250000903263973

Sjons, J., Hörberg, T., Östling, R., and Bjerva, J. (2017). "Articulation rate in Swedish child-directed speech increases as a function of the age of the child even when surprisal is controlled for," in *Proceedings of the 18th Annual Conference of the International Speech Communication Association*, ed M. Włodarczak, 1794–1798.

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

Soderstrom, M. (2007). Beyond babytalk: re-evaluating the nature and content of speech input to preverbal infants. *Dev. Rev*. 27, 501–532. doi: 10.1016/j.dr.2007.06.002

Soli, S. D. (1981). Second formants in fricatives: acoustic consequences of fricative-vowel coarticulation. *J. Acoust. Soc. Am*. 70, 976–984. doi: 10.1121/1.387032

Stern, D., Spieker, S., Barnett, R., and MacKain, K. (1983). The prosody of maternal speech: Infant age and context related changes. *J. Child Lang*. 10, 1–15. doi: 10.1017/S0305000900005092

Storkel, H. L. (2004a). Do children acquire dense neighborhoods? An investigation of similarity neighborhoods in lexical acquisition. *Appl. Psycholinguist*. 25, 201–221. doi: 10.1017/S0142716404001109

Storkel, H. L. (2004b). Methods for minimizing the confounding effects of word length in the analysis of phonotactic probability and neighborhood density. *J. Speech Lang. Hear. Res*. 47, 1454–1468. doi: 10.1044/1092-4388(2004/108)

Storkel, H. L., and Hoover, J. R. (2011). The influence of part-word phonotactic probability/neighborhood density on word learning by preschool children varying in expressive vocabulary. *J. Child Lang*. 38, 628–643. doi: 10.1017/S0305000910000176

Szewczyk, J. M., Marecka, M., Chiat, S., and Wodniecka, Z. (2018). Nonword repetition depends on the frequency of sublexical representations at different grain sizes: evidence from a multi-factorial analysis. *Cognition* 179:23–36. doi: 10.1016/j.cognition.2018.06.002

Thiessen, E. D., Hill, E. A., and Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy* 7, 53–71. doi: 10.1207/s15327078in0701_5

Torrington Eaton, C., Newman, R. S., Bernstein Ratner, N., and Rowe, M. L. (2015). Nonword repetition in 2-year-olds: replication of an adapted paradigm and a useful methodological extension. *Clin. Linguist. Phonet*. 29, 523–535. doi: 10.3109/02699206.2015.1029594

Vaden, K. I., Halpin, H., and Hickock, G. (2009). *The Irvine Phonotactic Online Dictionary, Version 2.0*. Available online at: https://www.iphod.com.

Vitevitch, M. S., and Luce, P. A. (1998). When words compete: levels of processing in perception of spoken words. *Psychol. Sci*. 9, 325–329. doi: 10.1111/1467-9280.00064

Vitevitch, M. S., and Luce, P. A. (1999). Probabilistic Phonotactics and Neighborhood Activation in Spoken Word Recognition. *J Mem Lang*., 40, 374–408. doi: 10.1006/jmla.1998.2618

Vitevitch, M. S., and Luce, P. A. (2016). Phonological neighborhood effects in spoken word perception and production. *Ann. Rev. Linguist*. 2, 75–94. doi: 10.1146/annurev-linguistics-030514-124832

Vosoughi, S., and Roy, D. K., (2012). "A longitudinal study of prosodic exaggeration in child-directed speech" in *Speech Prosody, 6th International Conference*. Speech Prosody Special Interest Group (SProSIG).

Wang, L., Kalashnikova, M., Kager, R., Lai, R., and Wong, P. C. (2021). Lexical and prosodic pitch modifications in Cantonese infant-directed speech. *J. Child Lang*. 1–27. doi: 10.1017/S0305000920000707

Warlaumont, A. S., Richards, J. A., Gilkerson, J., and Oller, D. K. (2014). A social feedback loop for speech development and its reduction in Autism. *Psychol. Sci*. 25, 1314–1324. doi: 10.1177/0956797614531023

Watanabe, A. (2001). Formant estimation method using inverse-filter control. *IEEE Trans. Speech Audio Proc*. 9, 317–326. doi: 10.1109/89.917677

Weisleder, A., and Fernald, A. (2013). Talking to children matters: early language experience strengthens processing and builds vocabulary. *Psychol. Sci*. 24, 2143–2152. doi: 10.1177/0956797613488145

Whalen, D. H. (1990). Coarticulation is largely planned. *J. Phon*. 18, 3–35. doi: 10.1016/S0095-4470(19)30356-0

Wickham, H. (2016). *Ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag New York.

Zamuner, T. S., and Thiessen, A. (2018). A phonological, lexical, and phonetic analysis of the new words that young children imitate. *Can. J. Linguist*. 63, 609–632. doi: 10.1017/cnj.2018.10

Zellou, G., and Scarborough, R. (2015). Lexically conditioned phonetic variation in motherese: age-of-acquisition and other word-specific factors in infant- and adult-directed speech. *Lab Phonol*. 6, 305–336. doi: 10.1515/lp-2015-0010

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

# A. APPENDIX

**TABLE A1 |** Real word and non-word stimuli.

| Real word | Non-word (orthography) | Phonetic transcription |
|---|---|---|
| Dog | kog | [ˈkɔg] |
| Juice | buice | [ˈbjus] |
| Cat | jat | [ˈʤæt] |
| Book | dook | [ˈdʊk] |
| Balloon | challoon | [tʃɑ.ˈlun] |
| Cookie | pookie | [ˈpʊ.ki] |
| Puppy | kuppy | [ˈkʌ.pi] |
| Chicken | bicken | [ˈbɪ.kən] |
| Banana | bajapop | [bɑ.ˈjæ.pəp] |
| Telephone | telina | [tɛ.ˈli.nɑ] |
| Lollipop | lolamas | [ˈlɑ.lɑ.mɑs] |
| Pajamas | panaphone | [pə.ˈnæ.foən] |

frontiers in Psychology

ORIGINAL RESEARCH
published: 09 December 2021
doi: 10.3389/fpsyg.2021.719783

Check for updates

# Understanding Child-Directed Speech Around Book Reading in Toddler Classrooms: Evidence From Early Head Start Programs

*Annemarie H. Hindman[1]*, Jean M. Farrow[2], Kate Anderson[3], Barbara A. Wasik[1] and Patricia A. Snyder[4]*

*[1] Department of Teaching and Learning, Temple University, Philadelphia, PA, United States, [2] Graduate School of Education, University of Pennsylvania, Pennsylvania, PA, United States, [3] Center for Social Organization of Schools, Johns Hopkins University, Baltimore, MD, United States, [4] Anita Zucker Center for Excellence in Early Childhood Studies, University of Florida, Gainesville, FL, United States*

Child-directed speech (CDS), which can help children learn new words, has been rigorously studied among infants and parents in home settings. Yet, far less is known about the CDS that teachers use in classrooms with toddlers and children's responses, an important question because many toddlers, particularly in high-need communities, attend group-care settings. This exploratory study examines the linguistic environment during teacher-led book readings in American Early Head Start classrooms serving 2-year-olds from households in poverty. Seven teachers in four classrooms were trained to emphasize target words while reading story and informational books. We first analyzed the nature and quality of their book readings from a macro-level, exploring global instructional quality [Classroom Assessment Scoring System (CLASS)] and linguistic complexity [i.e., diversity of vocabulary (*D*) and sophistication of syntax (MLU-w)], and we also examined micro-level teacher-child talk strategies and use of target words. Compared to prior research, these classrooms had similar global quality and syntactic complexity, although less lexical diversity. Exploratory results also revealed three distinct teacher talk patterns—teachers who emphasized (1) comments, (2) questions, and (3) a balance of the two. Question-focused teachers had more adult and child talk during reading, as well as more repetitions of target words, and stronger CLASS Engaged Support for Learning. However, comment-focused teachers used more diverse vocabulary and had stronger CLASS Emotional and Behavioral Support. Results illuminate the nature and quality of CDS in toddler classrooms, particularly in the context of an intervention emphasizing target vocabulary words, and highlight applications for professional development and questions for further research.

**Keywords: toddler, child-directed speech (CDS), early childhood education, teacher, book reading**

OPEN ACCESS

***Edited by:***
*Chiara Suttora, University of Bologna, Italy*

***Reviewed by:***
*Marta Casla, Autonomous University of Madrid, Spain*
*Mariagrazia Zuccarini, University of Bologna, Italy*

****Correspondence:***
*Annemarie H. Hindman*
*ahindman@temple.edu*

***Specialty section:***
*This article was submitted to Developmental Psychology, a section of the journal Frontiers in Psychology*

***Received:*** *03 June 2021*
***Accepted:*** *07 October 2021*
***Published:*** *09 December 2021*

***Citation:***
*Hindman AH, Farrow JM, Anderson K, Wasik BA and Snyder PA (2021) Understanding Child-Directed Speech Around Book Reading in Toddler Classrooms: Evidence From Early Head Start Programs. Front. Psychol. 12:719783. doi: 10.3389/fpsyg.2021.719783*

## VOCABULARY DEVELOPMENT

One of the most important milestones of the first years of life is learning language, beginning with vocabulary (Samuelsson, 2021). Knowing more words in early childhood facilitates further vocabulary and language development, a virtuous cycle (Peter et al., 2020; Avila-Varela et al., 2021). Children with more vocabulary knowledge have, both immediately and over time, greater success in reading and other content areas (Dickinson et al., 2010; Cristofaro and Tamis-LeMonda, 2012; Morgan et al., 2015), better-adjusted social interactions, and more self-regulation and executive functioning (Winsler, 2009; Manning et al., 2019; Rantalainen et al., 2021). Unfortunately, growing up in poverty is, as early as 18 months of age, associated with less knowledge of vocabulary and slower language processing (Fernald et al., 2013; Suggate et al., 2018), making a focus on these children's early experiences a priority.

## CHILD-DIRECTED SPEECH

Beyond considerable individual differences in how and how quickly children learn words (Fernald et al., 2006; Donnelly and Kidd, 2020), the language input children receive from those around them plays a key role in their language learning (Abend et al., 2017; Golinkoff et al., 2019; Fitch et al., 2020). When communicating with very young children, adults often systematically alter how they talk, using specialized, child-directed speech (CDS) that draws children's attention and highlights the sounds in words, supporting vocabulary and other language outcomes (Bryant and Barrett, 2007; Zauche et al., 2016). Most often studied among pre-verbal infants, CDS is characterized by unusual auditory features such as high pitch, slow pace, exaggerated prosody, and distinct timbre; as well as sparse word volume and frequent repetition of words, focus on concrete ideas, and simple syntactic structure (Rowe, 2008, 2012; Huttenlocher et al., 2010; Longobardi et al., 2016; Quick et al., 2019; Genovese et al., 2020; Rowe and Snow, 2020). Adults' CDS changes as children progress into toddlerhood (e.g., 1–2 years of age) and begin to talk and respond on their own (Durán et al., 2004; Hoff, 2014) using one-, two-, or three-word phrases (i.e., telegraphic speech) (Rice et al., 2010) undergirded by basic syntax and grammar (Hoff et al., 2018; Cadime et al., 2019). Adults' CDS to toddlers employs more standard prosody and longer utterances, with more numerous and complex words and grammatical structures, as well as increased back-and forth through extended adult-child conversations fostered by questions (Rowe, 2012; Longobardi et al., 2016).

However, despite all we know about CDS, most research targets parents and families, with far less work examining interactions between teachers and young children, especially toddlers. In many American communities, particularly those in poverty, approx. 60% of toddlers attend care settings, with 12% in center-based classrooms and 30% in home-based group care (Paschall, 2019). Unfortunately, the quality of teachers' CDS many be low (LaParo et al., 2009), demanding further research and professional development (PD).

## CHILD-DIRECTED SPEECH IN EARLY CHILDHOOD CLASSROOMS

To date, the literature on CDS in toddler classrooms is a patchwork, with a variety of different ways of defining, measuring, and aggregating components of teacher CDS and a mix of observational and PD intervention studies. A critical review of this literature reveals three relatively distinct approaches to conceptualizing/measuring CDS: global conceptual quality, linguistic complexity, and specific teacher (and, occasionally, child) talk strategies. Below, we review key findings from each approach, focusing on toddlers but, because of the small literature, including data from preschool when relevant and necessary.

### Global Quality

Global quality measures assign one score to an entire instructional activity block or day. CDS lies at the heart of global quality scores, but other features of teaching (e.g., materials) and child activities factor in as well. Building on the classic tools such as the Early Childhood Environment Rating Scale (ECERS) and the Early Language and Literacy Classroom Observation (ELLCO), considerable recent research has supported the reliability and validity of the Classroom Assessment Scoring System (CLASS), which can be used in classrooms from infancy through high school (Hamre et al., 2012, 2013). The CLASS-Toddler includes two domains, each with several dimensions: Emotional and Behavioral Support (Positive Climate, Negative Climate, Teacher Sensitivity, Regard for Child Perspectives, Behavior Guidance) and Engaged Support for Learning (Facilitation of Learning and Development, Quality of Feedback, and Language Modeling). At present, we are beginning to understand global quality in toddler settings. National samples in United States toddler classrooms suggest that CDS related to Emotional support is typically high ($M = 5.30$, $SD = 0.07$, on a scale from 1 to 7) while Engaged Support for Learning is modest ($M = 3.60$, $SD = 0.15$) (Bandel et al., 2014). Because CLASS-T scores have been predictive of child vocabulary and language outcomes in prior work (e.g., Aikens et al., 2015), we include CLASS-T in this study.

### Linguistic Complexity

Other approaches have focused on the overall linguistic complexity of teachers' CDS in an instructional activity block or day.

#### Lexical Diversity

Lexical diversity, or the percentage of unique (rather than repeated) words used, has typically been measured using a type-token ratio (i.e., the number of distinct, different words and their inflections and derivations relative to the number of total words). Because CDS features simplified word choice, expanding as children age, studies have generally found values around 0.17 in toddlers' households (Hoff, 2003; Rowe, 2012), meaning that only 17% of the words children hear are words they heard only once during the language sample; most words were repeated multiple times. Interestingly, Girolametto et al. (2003) explored toddler and preschool classroom CDS and found a ratio of 0.44 during book reading, far higher than home settings. More recent work, however (Montag et al., 2018), has found that type-token ratio should be adjusted to account for the length of the language sample, resulting in widespread adoption of a novel measure of lexical diversity referred to as *D*. The construct of *D* has been used in preschool research; for example, Dickinson et al. (2014) found that teacher CDS in Head Start book readings averaged 74.41, aligned with a type/token ratio of 0.58. It is helpful to note that average *D* for children's talk at age 2 is 27.44 and the average at age 3 is 47.83 (Durán et al., 2004). To our knowledge, *D* for teachers' CDS has not been widely gauged in toddler classroom settings, which we explore in the current study.

### Structural Complexity

Another feature of CDS is simple language construction, such as "That's a dog!" (for infants) to "That's a big, furry dog, and it's running down the street" (for toddlers). Complexity can be measured with mean length of utterance (MLU-w), or number of words per statement. In preschool book readings, Dickinson et al. (2014) found an average MLU-w of 8.39, indicating that teachers generally used about 8 words per remark. Less research is available on toddler classrooms, but Girolametto et al. (2003) reported teacher MLU-m (a slightly more liberal measure than MLU-w) around 5.03 during book reading. It is helpful to note that toddlers' average MLU-w is 2.90 (Girolametto and Weitzman, 2002). The current study consequently explores teachers' MLU-w.

More specific still, Justice et al. (2013) examined more nuanced features of the syntax of CDS, adapting Huttenlocher et al. (2010)'s approach to capture constructions that boost MLU-w, including simpler strategies (e.g., verb phrases, prepositional phrases, adjectives/adverbs) and more complex strategies (subordinate clauses, modifying verbal phrases, advanced phrases, compound sentences). They found that, among preschoolers, these constructions are generally used infrequently, but to our knowledge, this study is the first to code toddler teachers' CDS In this way.

## Specific Talk Strategies

A third, highly nuanced approach to understanding CDS is a micro-level examination of teachers' specific words, including the teachers' language modeling (including their conceptual complexity and their use of specific vocabulary words), inviting child talk (including questions that promote conversational turns), and feedback. Relatedly, a few studies have explored child talk, both in response to and independent of teacher talk.

## Teacher Language Modeling

Abundant evidence suggests that an important aspect of teachers' CDS is modeling language through their remarks, which has been understood from two primary perspectives. First, as laid out in Dickinson and Tabors (2001), the *conceptual complexity of teachers' talk* matters, specifically whether it is contextualized (i.e., in reference to information that is immediately apparent, such as on the page of a book, including labeling or describing an illustration) or decontextualized (i.e., in reference to abstract information, such as a synthesis or prediction). Both teacher and/or parent contextualized and decontextualized talk in book reading and other conversational settings supports children's vocabulary outcomes (Pancsofar and Vernon-Feagans, 2006; Rowe, 2012), but decontextualized talk may be particularly supportive of learning for children with stronger language skills (Pellegrini et al., 1990; Reese, 1995; Curenton et al., 2008; Hindman et al., 2008). Book reading studies in preschool have shown that more complex books offer exposure to more complex language via the text and engender more decontextualized talk (Dickinson et al., 2014; Muhinyi et al., 2020).

Another aspect of teacher CDS is *teachers' use of specific vocabulary words*. The language development literature indicates that frequent exposure to language aids in learning it (Ambrose et al., 2015), and one extension is that multiple repetitions of the same vocabulary words, ideally with rich interactions around them, can support children in learning them (Harris et al., 2011; Beck et al., 2013). Adult talk about and repetition of specific target words learn can build child knowledge of those words (Kaiser and Trent, 2007; McLeod et al., 2017), and at least one study with preschoolers (Wasik and Hindman, 2020) suggests that children's standardized receptive and expressive vocabulary skills increase when teachers use target words more frequently. Interestingly, toddler teachers may tend to repeat target words more often than preschool teachers, perhaps to match children's emerging language skills (Girolametto and Weitzman, 2002; Girolametto et al., 2003). This target word focus may have powerful results—one Early Head Start study (Romano and Woods, 2018) trained three teachers in several types of strategies including modeling target phrases (see Roberts et al., 2014), which was ultimately linked to gains in children's talk, including target words. In this study, we track teacher language modeling (complexity) and frequency of target word use.

### Teacher Questions/Back-and-Forth

Beyond what teachers say lies the degree to which teachers invite child talk and foster extended exchanges. *Teacher questions* often start conversations (Rowe et al., 2017; Gilkerson et al., 2018). Questions can be of many formats (Wasik and Hindman, 2013), including open (involving no single correct answer and likely requiring multiple word response) or closed (single correct answer), with the latter including label questions (What's this called?) or yes/no questions (Is this a dog?). There is extensive evidence from preschool that questions in the classroom are a powerful tool for eliciting child language and fostering conversation (Whitehurst, 2004). Open-ended questions may be relatively rare, representing approx. 5% of preschool teachers' prompts (e.g., Siraj-Blatchford and Manni, 2008), but closed questions can help to check understanding (Wasik and Hindman, 2013).

Among toddlers, however, there is little research, but some evidence suggests that patterns of teacher questions are similar. For example, Davis et al. (2015) examined six classrooms serving children ages birth to two and determined that teachers predominately asked the same kinds of questions seen in preschools—closed yes/no questions and closed label questions, with relatively few open-ended questions. Interestingly, O'Brien and Bi (1995) found that toddlers often did not respond to teacher questions, perhaps because their language skills were not yet strong enough. Similarly, Kidd and Rowland (2021) found that, with 2-year-olds as well as 3-year-olds, when presented with conversational opportunities, children contributed just about one-third (37%) of the turns. They hypothesized that more teacher-dominated conversations (such as emphasizing closed questions) with younger children may be supportive of language. In the current study, we track teachers' questions (open/closed).

### Teacher Responsiveness/Feedback

Beyond modeling language and then inviting child talk, research has increasingly focused on a third component of teacher-child exchanges—unpacking how and how much (e.g., conversational turns) teachers *respond to child talk*. Much research has examined the value of teachers' extending children's language (for example, responding to "Dog!" with "That is a dog!") and/or elaborating on children's ideas (for example, responding to "Dog!" with "That dog is brown and furry"), both of which are linked to important gains in a variety of children's language skills (Cabell et al., 2015). Relatively less research (Cabell et al., 2015; Casla et al., 2021) has targeted other contingent teacher remarks; for example, Kidd and Rowland (2021) also highlighted ignoring, copying/repeating, rephrasing, and interpreting, finding some likely benefits of these talk strategies as well. To our knowledge, teachers' feedback to children has not been widely studied in toddler classrooms, and we examine this question in the current study, including this variety of feedback strategies.

### Child Talk

Even though CDS changes as children begin to talk (Rowe and Snow, 2020), it has often been conceptualized as something of a one-way street, with the content and meaning of the exchange exclusively driven by the adult and with less attention to child talk (Golinkoff et al., 2015). Justice et al. (2018) established, at least implicitly, the critical importance of child talk in their discovery that the most important driver of child language in preschool classrooms was teacher invitations for child talk. In one relevant study that included child talk, Wasik and Hindman (2011) offered PD to 19 Head Start teachers and, during teacher-led book readings, coded both *children's responses to teachers' questions* (distinguishing between single- and multiple-word responses) and *children's spontaneous talk*. Findings showed that children's vocabulary growth was uniquely linked to more child talk during book reading. A subsequent study (Champagne, 2019) investigated the *accuracy* of preschoolers' responses and teachers' feedback on errors, finding that teachers tended to call on children whom they presumed would have the correct answer, and that incorrect responses were rare. We build on this work to explore toddlers' responses and spontaneous talk, including accuracy, with a particular focus on discussion of target words.

## INTERRELATIONS AND PATTERNS

Thus, across the literature, teacher CDS has been explored from a collection of macro- and micro-level perspectives. But to date, just two studies have explored these aspects of CDS (and child language) simultaneously; examining how they co-occur and whether there may be "styles" or "registers" has key implications for PD. First, Justice et al. (2018) examined how three ways of capturing CDS—global quality, linguistic complexity, and several kinds of talk strategies (encouraging child language)—uniquely predicted vocabulary in 49 preschool classrooms when considered together. Results showed that overall, these different facets of CDS were uncorrelated with one another, but that only teacher talk strategies (specifically, encouraging child talk) predicted vocabulary learning. Second, Dickinson et al. (2014) examined linguistic complexity and several kinds of talk strategies (modeling language, discussing literacy, social studies, etc.). As in Justice et al. (2018), these different facets of CDS were relatively independent of one another. However, book reading specifically offered evidence of an academic language register or style, wherein more vocabulary focus, greater lexical diversity, and greater structural complexity clustered together. The current study is, to our knowledge, the first to explore the correlations and patterns among the multiple facets of CDS in toddler classrooms, including a wide array of teacher and child talk strategies, to understand their overlap and independence.

## CURRENT STUDY

In sum, teacher CDS has been examined through varied lenses, including global instructional quality, linguistic complexity, and specific talk strategies, but there is great variability in methods and results across studies, and very little work targeting high-need toddler contexts. Because the toddler period is essential for language and vocabulary growth, and because high-quality teacher CDS can foster this growth, we developed and piloted a PD intervention for Early Head Start teachers. *Head Start on Vocabulary* (HSoV) is built on an effective preschool teacher PD model (Wasik et al., 2006; Wasik and Hindman, 2011, 2020) involving training teachers in language modeling, questioning, and feedback, all focused on target words, during book reading. In this study, we explored how toddler teachers used CDS during book reading after *HSoV* training and supports, as well as how children used language, particularly target words.

### Research Questions

We explored several research questions in a small sample of 7 teachers that enabled close examination of classroom CDS:

First, what is the *global quality* of *HSoV* teachers' classroom instruction and CDS? We used a gold-standard measure, the Classroom Assessment Scoring System—Toddler tool.

Second, what is the *linguistic complexity*, including lexical diversity and structural complexity, of teachers' CDS during book reading? We used the well-established CHILDES and CHAT language coding approach.

Third, what is the nature of *teachers' and children's talk strategies* during book reading? We considered several elements of the adult-child exchange: teachers' language modeling, children's responses, and teachers' responses to children. We also explored the nature of child-initiated talk during book reading, and teachers' responses. We examined both central trends and individual differences or patterns across teachers.

Fourth, how frequently are *target words* mentioned in the text, by teachers, and by children during book reading? We employed simple frequency counts.

Finally, to what extent are these *three lenses on CDS correlated or unique* from one another? We employed Pearson zero-order correlations.

Together, these questions explore teachers' CDS and child talk around vocabulary in high-need toddler settings, in the context of the target-word-focused *HSoV* intervention.

# MATERIALS AND METHODS

## Procedure

In fall 2019, we partnered with a local Early Head Start provider in a major urban city in the American northeast to develop *Head Start on Vocabulary (HSoV)*. Program administrators identified a total of four toddler (ages 2–3 year) classrooms from two centers. Each classroom was team-taught by two co-teachers, and all eight consented to participate.

### Head Start on Vocabulary

To our knowledge, there are no widely available interventions to support toddler teachers' CDS with rigorous evidence of effectiveness. In preschool, however, several effective interventions have improved the quality of preschool teachers' talk to children, resulting in gains in preschoolers' knowledge of taught words as well in standardized vocabulary scores (Landry et al., 2011; Weiland and Yoshikawa, 2013), including our own model which we have developed and tested over 20 years (Wasik and Bond, 2001; Wasik et al., 2006; Wasik and Hindman, 2011, 2020). In this project, we adapted our preschool program to address center-based toddler teachers' classroom CDS, developing *HSoV* for Early Head Start.

Like our preschool model, *HSoV* supported toddler teachers over a full academic year, including several distinct components that, taken together, offer a relatively "light touch." First, we offered *group workshops*. Beginning in fall, teachers attended 90-min monthly group workshops for 4 months. Workshops addressed (1) Talking to children using descriptive language, especially regarding new vocabulary; (2) Asking children questions, particularly about new vocabulary; (3) Assisting children in answering questions, if needed; and (4) Providing feedback on what children have said, emphasizing new vocabulary. Each began with an interactive lecture by project staff, presenting the rationale for and explanation of the target strategy and inviting teachers to share their experiences and concerns. We also demonstrated the strategies and shared short (1–3 min) videos of teachers in urban centers using the strategies with fidelity. Finally, we explained the classroom materials teachers would receive (see below) and gave teaching teams time to plan with support from project staff, so they could prepare to use the strategies on their own.

To support the use of the four strategies through book readings and play-based extension activities, we provided one trade book per week, to be read at daily to each child, either in groups or one-on-one. From each book, we selected three words that likely to be unfamiliar to children (Beck et al., 2013). For example, for the book *Little Blue Truck Leads the Way* (Schertle and McElmurry, 2015), we selected the words truck, road, and traffic jam. We then created extension activities for four different classroom areas (e.g., housekeeping, construction) to reinforce the words (e.g., teachers might read and then visit the construction area to act out a traffic jam together). Finally, we provided an 8.5″ × 11″ full-color picture card of each target word, with a child-friendly definition on the back, for the teacher share with children daily.

After each workshop, teachers received *coaching* every other week, including a direct 30-min observation in their classroom by a master teacher and who videotaped the teacher engaging in various activities, including book reading. The coach then watched the videos, took notes, and offered feedback in 45-min one-on-one on-site conferences.

Observations continued until March 2020, at which point the classrooms closed because of COVID-19. This study makes use of the videos collected by the coach in late fall 2019-winter 2020, after teachers had been through the training and were using the strategies in their classrooms. All study procedures were conducted with the approval of our university's IRB.

## Participants

A total of seven teachers participated in the project, as the 8th left the classroom before video collection began, and we did not record the substitutes who temporarily replaced her. All teachers were women of African-American backgrounds. All were native speakers of English. All held a minimum of a high school degree, while two were working toward an associate's degree and two others held an associate's degree. Although this study focuses on teachers, each classroom served 8 children (teacher: child ratio of 1:4), all between 24 and 36 months of age. The sample was evenly divided by gender. All children were of backgrounds that are minoritized in the U.S., with African American (60%), Hispanic-Latino (30%), and/or Asian (10%) heritage. Approximately 50% of children spoke a home language other than English (primarily Spanish).

## Measures

**Table 1** summarizes measures and key variables in the current study.

### Teacher Background

A background survey collected contact data and demographic information (e.g., education, ethnicity). The paper-pencil survey required about 5 min.

### Overall Classroom Quality

We used the Classroom Assessment Scoring System—Toddler (CLASS-T) version (La Paro et al., 2011), targeting two domains, each with several sub-domains: Engaged Support for Learning (i.e., Facilitation of learning and development, Quality of feedback, and Language modeling), as well as Emotional and Behavioral Support (i.e., Positive climate, Negative climate, Teacher sensitivity, Regard for child perspectives, and Behavior guidance). Widely used and considered a gold-standard tool, the CLASS-T has strong psychometric properties, including construct validity and inter-rater reliability (Hamre et al., 2012). A trained rater watched videos and then scored the classroom instruction on a variety of items (all rated 1-very low quality to 7-very high quality), yielding an average for each domain and sub-domain. Because the CLASS is most reliable

**TABLE 1 |** Variables in the current study.

| Construct | Measure | Variables in current study | Possible range of values |
|---|---|---|---|
| **Global quality** | | | |
| | CLASS-Toddler | Emotional and behavioral support<br>Positive climate<br>Recoded negative climate<br>Teacher sensitivity<br>Regard for child perspectives | All items measured on a 1–7 scale; Sub-domains and domains represent the average of all relevant items, so will be scored from 1 to 7. |
| | | Engaged support for learning<br>Facilitation of learning and development<br>Quality of feedback<br>Language modeling | |
| **Linguistic complexity** | | | |
| | CHAT and CLAN standardized coding schemes | Linguistic diversity:<br>*D:* Total word types/total words, adjusted for length of language sample | Minimum = 0, Maximum = 1 |
| | | Structural complexity:<br>Mean length of utterance—w: Average number of words per utterance | Minimum = 0, Maximum unbounded |
| | Justice et al. (2013) coding scheme | Inclusion of specific syntactic constructions<br>Simpler constructions: verb phrases, prepositional phrases, adjectives/adverbs<br>More complex constructions: subordinate clauses, modifying verbal phrases, advanced phrases, compound sentences | Frequency counts Minimum = 0, Maximum unbounded |
| **Talk strategies** | | | |
| | Project-derived coding schemes | Teacher remarks<br>Child responses<br>Teacher responses to child responses<br>Child-initiated Talk<br>Teacher responses to child-initiated talk<br>target words | Frequency counts Minimum = 0, Maximum unbounded |

when more minutes of instruction are coded, we coded all parts of every teacher's video. On average, videos (including but not limited to the book reading segment) coded with CLASS-T were 10.47 min long (*SD* = 2.96, *range* = 4.93–13.58).

## Linguistic Complexity of Teachers' Child-Directed Speech

Videos of book readings were transcribed for analysis in the Codes for Human Analysis of Transcripts (CHAT) from the Child Language Data Exchange System (CHILDES), with analyses of syntactic complexity (MLU-w) and lexical diversity (*D*) of teacher talk conducted using Child Language Analysis (CLAN) program (MacWhinney, 2000). Transcriptions began as soon as the teacher indicated the activity was starting (e.g., "Are all my friends ready to look at our new book?") and ended when the teacher announced the conclusion (e.g., "Okay, you all did a good job today"). The average length of the book reading portion of the videos was 6.00 min (*SD* = 3.07, *range* = 2.42–13.38).

## Transcription

Teachers' speech was parsed into C-units (Loban, 1976), utterances defined as containing one main clause and any modifying phrases and subordinating clauses. Thus, the following example – "The boy is jiggling his ears/and he's shaking his leg" – would be parsed into two separate C-units. We did not code the syntax of teachers' reading of actual text from the book, as the purpose was to analyze the complexity of teachers' CDS. All transcripts were checked twice, in addition to using automated check features within the CLAN program before analysis. An example is provided in **Supplementary Appendix**.

## Lexical Diversity

Teachers' quantity of input was calculated using the FREQ command. CLAN derived the total amount of words and word types within the transcripts of teachers' lessons. We used D as our measure of lexical diversity (McKee et al., 2000), because unlike type-token ratios, D accounts for differences in length of language samples, allowing for comparisons across transcripts, and thus more accurately measuring lexical diversity. The VOCD command in CLAN calculated D for teachers.

## Structural Complexity

We used MLU-w as a proxy for complexity of speech, in that longer word utterances often include words, phrases, and clauses that modify meaning of the main clause (Hoff, 2003). As with lexical diversity, CLAN derived the MLU-w of teachers using the MLU command. We further coded teachers' use of complex language structures in their CDS using a scheme from Justice et al. (2013). We captured relatively simpler strategies (e.g., verb phrases, prepositional phrases, adjectives/adverbs) and more complex strategies (subordinate clauses, modifying verbal phrases, advanced phrases, compound sentences). We calculated the frequency of each kind of structure, as well as the proportion of teachers' total utterances that included one or more of these structures.

## Talk Strategies in Child-Directed Speech

We coded every teacher utterance during the book readings to understand how teachers were using specific conversational strategies in their CDS. Our coding scheme followed previous work in the field (e.g., van Kleeck et al., 2003; Hindman et al., 2008; Wasik and Hindman, 2020) but included new codes as needed. We distinguished among (A) Teacher-initiated talk, (B) Child responses to teacher-initiated talk, (C) Child-initiated talk, and (D) Teacher responses to children. We also coded each sentence that teachers read from the book as (E) Reading Text to track the number of sentences per book. We coded directly from video, without transcription. We coded every video twice to ensure accuracy. An example is provided in **Supplementary Appendix**.

#### *Defining Utterances*

As with the syntax coding, we began coding when teachers announced the start of the activity and ended when teachers moved on to a different activity. Also aligned with the syntax coding, we generally defined an utterance as an independent clause, or a remark that included, at a minimum, a subject and verb. Therefore, a sentence such as, "What's he doing right here?" would be one utterance, whereas "He's dancing and his friend is laughing" would be two utterances. However, children generally did not speak in complete sentences, so stand-alone remarks such as "Red!" (meaning, "That thing is red") were also coded as utterances. Finally, we used the same logic in coding brief, stand-alone teacher responses to child remarks, allowing one-word responses such as "Right!" to be coded.

#### *Teacher-Initiated Talk*

We coded every teacher-initiated remark, focusing on content and format. Regarding content, we coded remarks as either *contextualized* (related to information apparent on the page of the book) or *decontextualized* (involving abstraction or inference) (Dickinson and Tabors, 2001). For example, questions about illustrations (e.g., "What color is her shirt?") were contextualized, but references to real-life experiences, summaries, or predictions (e.g., "What do you think Max will do next?") were decontextualized.

Regarding format, we coded every teacher-initiated remark as a comment, an open prompt, or a closed prompt. *Comments* were statements that did not request a child response (e.g., "There's the cat"). Types of comments included labeling/describing, defining a word, or providing other information. *Open prompts* were those in which more than one correct answer was possible (e.g., "What do you see on this page?"). *Closed prompts* were those in which the correct answer was limited to one option. We coded for several types of closed questions, including (a) label-related (e.g., "What color is this cat?"), (b) choice questions (e.g., "Is this a boat or a car?"), (c) yes/no questions (e.g., "Is she sitting down in this picture?"), (d) point or gesture questions (e.g., "Point to the car"), and (e) requests to repeat (e.g., "Say 'car"'). Notably, prompts included remarks that were technically statements but that functioned as questions (e.g., "Tell me what you see here").

#### *Child Responses to Teacher-Initiated Talk*

We tracked how children responded to each teacher open and closed prompt. We found that content and format of child responses were determined by the initial teacher remark, so within each category of prompt [open, closed (label, choice, yes/no, gesture, repeat)], we focused on whether the child response was correct, incorrect, or no response. We considered both verbal responses and gestures, noting the latter.

We did not track the identities of individual children, and we did not distinguish between responses provided by one child vs. those provided simultaneously by several children (given the quality of our audio and the varied group size within and across videos). Accordingly, when a teacher posed a question to the group and only one child responded (a frequent pattern), we coded only the speaking child's response (correct/incorrect) and excluded the other children's non-response. In addition, we observed one a situation where two children offered a response at the same time, one of which was correct while the other was incorrect. In this case, we marked the response as correct (and the teacher responded to the correct response). Overall, then, our coding scheme privileged correct child responses, and estimates of child talk can be viewed as describing the top end of the possible frequency and accuracy distribution.

#### *Child-Initiated Talk*

We coded spontaneous child remarks without any prompt from the teacher. In the absence of extensive prior research, we adapted our child response codes. We coded for spontaneously *repeating* what a teacher just said; labeling a *target word* ("Boat!"); labeling something that was not a target word ("Dog!" as the child notices a dog in a picture); offering a *description* of a picture, often related to color or ("Red!" when looking at a picture of Elmo or "Ribbit!" when looking at a picture of a frog); and offering a question about a picture ("What he doing?"). We intended to code both child responses and spontaneous remarks as contextualized or decontextualized, but all child-initiated remarks were contextualized.

#### *Teacher Responses to Child Talk*

Finally, we coded teachers' responses to children's talk, based on prior approaches. Codes included *repeating* the child verbatim ("Boat!"), repeating the child and *adding words (*"That's a boat!"), *adding a new idea* ("A boat can sails on the water") and *praising* the child ("Great job!"). We also included, for teacher responses to child incorrect remarks, asking a *rhetorical question* ("You sure that's a car?"), *re-asking* the same question, and *rephrasing/reframing* a question as choice or yes/no. Finally, We coded for giving a *hint* ("Like a car, but it drives on special tracks. . ..") and for giving the *correct answer*.

### Target Word Frequency

In a final round of coding, we tallied the number of times that target words—vocabulary related to the text and/or theme—were used. We separately counted for target words (1) in the text itself, as well as (2) in teacher talk, and (3) in child or children's talk. We counted each instance of use of a target word, even when repetitive; for example, if a teacher said, "That's a tire—there's the tire," we counted both mentions of the target word *tire*. In addition, we counted an individual child using the target word or all children simultaneously using the target word as one instance. Ultimately, we had three values for each classroom—total mentions of target words in the text, teacher talk, and child talks.

# RESULTS

## Question 1: Global Quality of Classroom Interactions

Complete results are presented in **Table 2**. On average, teachers in this sample demonstrated levels of CLASS-T Emotional and Behavioral Support in the moderate/high range ($M$ = 5.00, $SD$ = 0.90, *range* = 3.4–6.4). Within this first domain, the highest average dimension score was on (Recoded) Negative Climate ($M$ = 6.88, $SD$ = 0.33, *range* = 6.0–7.0), indicating very little harshness or negativity. The lowest average score was on Behavior Guidance ($M$ = 3.70, $SD$ = 1.40, *range* = 1.0–6.0), indicating that more effective guidance could be provided in many classrooms.

On average, the domain of Engaged Support for Learning in these classrooms also fell into the moderate range ($M$ = 3.81, $SD$ = 1.43, *range* = 1.3–6.0). The dimension with the highest value was Facilitation of Learning and Development ($M$ = 4.24, $SD$ = 1.46, range = 2.0–6.5), and the lowest was on Quality of Feedback ($M$ = 3.18, $SD$ = 1.59, *range* = 1.0–6.0). We examined Pearson zero-order correlations to explore the degree to which the dimensions and domains were correlated; domains were moderately to highly correlated with one another ($r$ = 0.60, $p$ = 0.10), and dimensions were generally correlated with the domain to which they pertain.

## Question 2: Linguistic Complexity of Teachers' Language

See **Table 3** for complete results. Teachers read an array of different books, including *The Bus for Us* (Bloom, 2001), *Baby Loves Winter* (Katz, 2013), *Froggy Gets Dressed* (London and Remkiewicz, 1994), *Time for a Bath* (Gershator and Walker, 2014), *Ready Set Brush* (Rudko, 2008), and *Shake a Leg!* (Allen and Swanson, 2010). All books were provided by the *Head Start on Vocabulary* intervention and were similar in length and complexity. On average, teachers used 75.86 total utterances ($SD$ = 26.17, range = 44–106) with children during their book readings, comprised of an average of 547.14 words ($SD$ = 229.26). However, the range was wide, with the teacher at the lowest end of the distribution using as few as 317 words while the teacher at the highest end used 944 words, approx. 300% as many words.

**TABLE 2 |** Descriptive statistics for overall classroom quality (CLASS-T).

| CLASS-T domain | M | SD | Range |
|---|---|---|---|
| Emotional and behavioral support | 5.00 | 0.90 | 3.40–6.40 |
| Positive climate | 5.45 | 1.27 | 3.00–7.00 |
| Recoded negative climate | 6.88 | 0.33 | 6.00–7.00 |
| Teacher sensitivity | 4.97 | 1.01 | 3.00–6.00 |
| Regard for child perspectives | 3.90 | 1.72 | 2.00–7.00 |
| Behavior guidance | 3.69 | 1.40 | 1.00–5.00 |
| Engaged support for learning | 3.81 | 1.43 | 1.30–6.00 |
| Facilitation of learning and development | 4.24 | 1.46 | 2.00–6.50 |
| Quality of feedback | 3.18 | 1.59 | 1.00–6.00 |
| Language modeling | 3.97 | 1.47 | 1.00–6.00 |

**TABLE 3 |** Descriptive statistics of linguistic complexity of teachers' talk during book reading.

| Language feature | Mean | SD | Min | Max |
|---|---|---|---|---|
| **CHAT coding** | | | | |
| Total utterances | 75.86 | 26.17 | 44.00 | 106.00 |
| Mean length of utterances-words | 6.14 | 0.73 | 5.11 | 7.06 |
| Total words | 547.14 | 229.26 | 317.00 | 944.00 |
| Total types | 105.14 | 11.61 | 87.00 | 119.00 |
| Type/token ratio | 0.22 | 0.07 | 0.12 | 0.32 |
| Lexical diversity (D) | 32.06 | 7.35 | 19.62 | 39.50 |
| **Proportion of utterances with complex syntactic components** | | | | |
| Complex utterances | 0.20 | 0.09 | 0.11 | 0.34 |
| Simple utterances | 0.80 | 0.09 | 0.66 | 0.89 |
| Embedded clause | 0.14 | 0.07 | 0.09 | 0.28 |
| Verbal phrase | 0.06 | 0.05 | 0.02 | 0.17 |
| Advanced phrase | 0.01 | 0.01 | 0.00 | 0.04 |
| Verb phrase | 0.36 | 0.09 | 0.27 | 0.54 |
| Prepositional phrases | 0.21 | 0.08 | 0.12 | 0.33 |
| Compound structures | 0.05 | 0.03 | 0.01 | 0.09 |
| Adjective or adverb | 0.15 | 0.10 | 0.05 | 0.34 |

### Lexical Diversity

Children heard an average of 105.14 ($SD$ = 11.61, range = 87–119) different word types; framed another way, teachers used many of the same words over and over again). Type/token ratio was, on average, 0.22 ($SD$ = 0.07, *range* = 0.12–0.32). Teachers' lexical diversity ($D$) mean scores, which account for the number of words in the sample (McKee et al., 2000), averaged 32.06 ($SD$ = 7.35, range = 19.62–39.50).

### Structural Complexity

Regarding MLU-w, each utterance, on average, contained 6.14 words ($SD$ = 0.73, *range* = 5.11–7.06). It is helpful to note that children's utterances at age 2 generally involve 1–3 words, meaning that teachers' language was more complex than children's, with twice as many words per utterance, on average. Pearson correlations showed that teachers' lexical diversity ($D$) was not related to syntactic complexity (MLU-w), ($r$ = 0.05, $p$ = 0.908).

Results of additional syntax coding (see Justice et al., 2013) revealed that teachers extended the length of their utterances with simpler constructions. Teachers used verb phrases (e.g., He will go home) in 36% of utterances, prepositional phrases l that specify and describe a noun or event (e.g., "The snow is falling *from the sky*") in 21% of utterances, and phrases with and adjectives or adverbs (e.g., "The truck picks up the *stinky* trash) in 15% of utterances. They also used more complex syntactic structures, including embedded or subordinating clauses (e.g., "Can we pay the driver *so that we can ride our taxi*?") in 15% of utterances, verbal phrases (e.g., "Do you see how happy Froggy is *playing outside with snow coming down*?") in 6% of remarks, and compound structures (e.g., "Froggy wants to go play, *but he has no socks*!") in 5% of utterances, and advanced phrases (e.g., "The children build the base, *the bottom*, of the snowman") in 1% of remarks. It is helpful to note that 2–3 year old children's language rarely includes either these more or less complex syntactic augmentations (Vasilyeva et al., 2008), making teachers' syntax more complex than children's.

## Question 3: Teacher and Child Talk Strategies

Below, we present results in order of their role in the teacher-child exchange: (1) Teacher-initiated talk, (2) Child responses to teacher talk, and (3) Teacher responses to child responses. We also include (4) Child-initiated talk not in response to a teacher remark, and (5) Teacher responses to child-initiated talk. Results are presented in **Table 4**.

### Teacher-Initiated Talk

Teachers initiated 48.86 total book-related utterances during their book readings ($SD$ = 12.40, range = 32–67). Of these, comments predominated ($M$ = 21.71, $SD$ = 11.76, range = 8–46), representing on average 45% of total teacher talk (but the proportion ranged from 15 to 69% across teachers). Closed prompts were less frequent ($M$ = 15.28, $SD$ = 11.02, range = 4–34, accounting for 31% of teacher-initiated talk). Open-ended prompts were rare ($M$ = 1.00, $SD$ = 1.53, range = 0–4, 2%). Attention- and management-related remarks not directly related to the book itself accounted for 9.57 remarks ($SD$ = 5.62, *range* = 4–18), or 19% of overall talk (*range* from 10 to 30%). Praise was offered on average once per reading ($M$ = 1.14, $SD$ = 1.46, *range* = 0–4), representing 3% of teacher talk (*range* = 0–8%).

#### *Teacher Comments*

The vast majority of teacher comments (about 20 per reading) were contextualized, primarily labeling/describing pictures ($M$ = 19.14, $SD$ = 10.51, *range* = 5–40). Decontextualized remarks occurred just twice per reading, on average ($M$ = 2.43, $SD$ = 1.90, *range* = 0–5). There were no instances of teachers defining words. There were no correlations among types of comments; teachers who used one type did not necessarily use other types.

#### *Teacher Closed Prompts*

The vast majority of closed prompts targeted contextualized information, with yes/no questions related to labeling (e.g., "Is this a cat?) used most frequently ($M$ = 5.71, $SD$ = 6.63, *range* = 0–17). Asking children to provide a label (e.g., "What's this?") was also relatively common ($M$ = 3.57, $SD$ = 4.12, *range* = 0–11), as was choosing between two labels ($M$ = 1.86, $SD$ = 4.48, *range* = 0–12) and repeating a label ($M$ = 2.43, 3.15, *range* = 0–9). Many additional strategies were used by only one teacher; for example, the teacher in video #4 used a single decontextualized closed question ("Where do boats drive?"), the teacher in video #16 asked children point to pictures of a target word on seven occasions, and the teacher in video #15 asked children to act out target words five times. Only one correlation among closed prompts was observed: there was a high correlation ($r$ = 0.82, $p$ = 0.02) between asking for labels and asking choice questions.

**TABLE 4** | Descriptive statistics of talk moves.

| Talk move | Mean | *SD* | Min | Max |
|---|---|---|---|---|
| **Teacher remarks** | | | | |
| All teacher-initiated remarks | 48.86 | 12.40 | 32.00 | 67.00 |
| Teacher comments | 21.71 | 11.76 | 8.00 | 46.00 |
| Teacher comment—label | 19.14 | 10.51 | 5.00 | 40.00 |
| Teacher comment—define | 0.00 | 0.00 | 0.00 | 0.00 |
| Teacher comment—decontextualized | 2.43 | 1.90 | 0.00 | 5.00 |
| Teacher open-ended questions | 1.00 | 1.52 | 0.00 | 4.00 |
| Teacher open-ended contextualized | 0.57 | 1.51 | 0.00 | 4.00 |
| Teacher open-ended decontextualized | 0.43 | 0.79 | 0.00 | 2.00 |
| Teacher closed questions | 15.28 | 11.02 | 4.00 | 34.00 |
| Teacher closed—label | 3.57 | 4.12 | 0.00 | 11.00 |
| Teacher closed question—choice | 1.86 | 4.49 | 0.00 | 12.00 |
| Teacher closed question—yes/no | 5.71 | 6.63 | 0.00 | 17.00 |
| Teacher closed question—repeat | 2.43 | 3.15 | 0.00 | 9.00 |
| Teacher closed question—action | 0.71 | 1.89 | 0.00 | 5.00 |
| Teacher closed question—point | 1.00 | 2.65 | 0.00 | 7.00 |
| Teacher closed question—decontextualized | 0.14 | 0.38 | 0.00 | 2.00 |
| **Child responses** | | | | |
| All child responses | 16.57 | 10.94 | 5.00 | 35.00 |
| All child correct responses | 8.71 | 8.26 | 2.00 | 22.00 |
| All child incorrect/no responses | 7.86 | 3.89 | 2.00 | 13.00 |
| Point—correct | 0.71 | 1.89 | 0.00 | 5.00 |
| Point—incorrect | 0.14 | 0.38 | 0.00 | 1.00 |
| Point—no response | 0.14 | 0.38 | 0.00 | 1.00 |
| Repeat—correct | 0.86 | 0.90 | 0.00 | 2.00 |
| Repeat—incorrect | 0.00 | 0.00 | 0.00 | 0.00 |
| Repeat—no response | 1.71 | 2.98 | 0.00 | 8.00 |
| Choice—correct | 1.29 | 3.40 | 0.00 | 9.00 |
| Choice—incorrect | 0.57 | 1.13 | 0.00 | 3.00 |
| Choice—no response | 0.00 | 0.00 | 0.00 | 0.00 |
| Yes/no—correct | 4.28 | 5.59 | 0.00 | 14.00 |
| Yes/no—incorrect | 0.71 | 1.11 | 0.00 | 3.00 |
| Yes/no—no response | 0.71 | 1.25 | 0.00 | 3.00 |
| Word—correct | 0.86 | 1.57 | 0.00 | 4.00 |
| Word—incorrect | 1.28 | 1.25 | 0.00 | 3.00 |
| Word—no response | 1.57 | 1.81 | 0.00 | 4.00 |
| Open-ended—correct | 0.14 | 0.38 | 0.00 | 1.00 |
| Open-ended—incorrect | 0.00 | 0.00 | 0.00 | 0.00 |
| Open-ended—no response | 0.86 | 1.21 | 0.00 | 3.00 |
| Gesture—correct | 0.57 | 1.51 | 0.00 | 4.00 |
| Gesture—incorrect | 0.00 | 0.00 | 0.00 | 0.00 |
| Gesture—no response | 0.14 | 0.38 | 0.00 | 6.00 |
| **Teacher responses, child incorrect** | | | | |
| No response | 3.14 | 3.44 | 0.00 | 10.00 |
| Repeat child | 0.86 | 1.07 | 0.00 | 3.00 |
| Rhetorical question | 0.86 | 1.21 | 0.00 | 3.00 |
| Re-ask same question | 1.57 | 1.72 | 0.00 | 4.00 |
| Reframe as choice question | 0.57 | 1.51 | 0.00 | 4.00 |
| Reframe as Yes/No question | 0.57 | 0.98 | 0.00 | 2.00 |
| Give hint | 1.14 | 1.68 | 0.00 | 4.00 |

*(Continued)*

**TABLE 4 |** (Continued)

| Talk move | Mean | *SD* | Min | Max |
|---|---|---|---|---|
| Explicit no | 0.14 | 0.38 | 0.00 | 1.00 |
| Give correct answer | 2.14 | 1.68 | 0.00 | 5.00 |
| Ask child to repeat their answer | 0.43 | 0.79 | 0.00 | 2.00 |
| Ask follow-up question | 0.28 | 0.75 | 0.00 | 2.00 |
| **Teacher responses, child correct** | | | | |
| No response | 0.85 | 1.21 | 0.00 | 3.00 |
| Repeat child label | 1.14 | 1.77 | 0.00 | 5.00 |
| Rhetorical question | 0.14 | 0.38 | 0.00 | 1.00 |
| Repeat child other | 1.00 | 2.24 | 0.00 | 6.00 |
| Praise | 2.43 | 4.03 | 0.00 | 11.00 |
| Explicit yes | 0.57 | 0.79 | 0.00 | 2.00 |
| Add words to child response | 1.71 | 2.50 | 0.00 | 7.00 |
| Expand child idea | 1.43 | 1.62 | 0.00 | 4.00 |
| Ask follow-up question | 0.00 | 0.00 | 0.00 | 0.00 |
| Child-Initiated Talk | 5.00 | 3.79 | 1.00 | 9.00 |
| All child-initiated talk | 7.00 | 6.16 | 2.00 | 20.00 |
| Voluntarily repeats teacher | 1.71 | 2.21 | 0.00 | 5.00 |
| Volunteers target word | 1.14 | 1.46 | 0.00 | 4.00 |
| Volunteers description of picture | 0.43 | 0.79 | 0.00 | 2.00 |
| Volunteers other information | 1.28 | 0.75 | 0.00 | 2.00 |
| Volunteers question | 0.014 | 0.38 | 0.00 | 1.00 |
| Teacher response, child-initiated talk | | | | |
| All teacher response to child-initiated talk | 7.00 | 6.16 | 2.00 | 20.00 |
| No response | 1.28 | 2.21 | 0.00 | 5.00 |
| Repeat child | 1.14 | 1.07 | 0.00 | 3.00 |
| Rhetorical question | 0.28 | 0.38 | 0.00 | 1.00 |
| Praise | 0.00 | 0.00 | 0.00 | 0.00 |
| Explicit no | 0.00 | 0.00 | 0.00 | 0.00 |
| Explicit yes | 2.28 | 2.21 | 1.00 | 7.00 |
| Add words to child response | 1.57 | 2.23 | 0.00 | 6.00 |
| Expand child idea | 1.14 | 1.35 | 0.00 | 4.00 |
| Ask follow-up question | 0.57 | 0.79 | 0.00 | 2.00 |

### *Teacher Open-Ended Prompts*

Open-ended prompts were used less than once/reading, on average ($M = 0.57$, $SD = 1.51$, *range* = 0–40 and $M = 0.43$, $SD = 0.79$, *range* = 0–2, respectively). Only one teacher (video #6) used a contextualized open question (e.g., "What do you see on the cover of the book?") but did so four times. Two teachers used decontextualized open questions (e.g., "What do you think could happen next?"), one (video #17) just once and the other (video #9) twice. We did not explore correlations, given their infrequency.

### *Patterns in Teacher-Initiated Talk*

To explore patterns in teacher talk, we conducted zero-order Pearson correlations among these variables. Comments and open prompts ($r = -0.19$, $p = 0.968$) were independent of one another, as were open and closed prompts ($r = -0.22$, $p = 0.639$). However, there was an inverse, marginally significant correlation between comments and closed prompts ($r = 0.74$, $p = 0.057$), indicating that teachers who made more comments asked fewer closed questions. When closed and open prompts were combined (i.e., all prompts), there was significant inverse correlation between comments and questions ($r = -0.76$, $p = 0.048$). Thus, teachers in this sample appeared to either favor questions or comments in their book readings.

Descriptive data identified two teachers who used predominately questions rather than comments (41% of their talk was comments whereas 26% was questions and 64% was comments whereas 15% was questions, respectively), whom we termed *Asker* teachers. Conversely, four teachers used predominately comments rather than questions (43% comments to 29% questions, 63% comments to 25% questions, 59% comments to 15% questions, and 69% comments to 6% questions); we termed these teachers *Tellers*. One teacher, termed *Balanced*, used the two types of remarks about equally (44% comments to 38% questions). Descriptive data showed that Teller teachers used more comments ($M = 26.75$, $SD = 13.00$ vs. 12.00, $SD = 5.66$) while Asker teachers used more questions ($M = 30.00$, $SD = 5.66$ vs. $M = 8.50$, $SD = 3.41$). The Balanced teacher fell in between on comments ($n = 21$) and questions ($n = 20$). Exploratory, underpowered ANOVA results comparing Askers and Tellers showed that, while the frequency of comments was not statistically different ($p = 0.216$), Asker teachers employed significantly more questions, $F(1, 4) = 36.80$, $p = 0.004$. In an additional exploratory analysis, we examined whether Askers and Tellers differed in the kinds (rather than just the amount) of comments and questions they made. There were no differences for comments, but Askers used significantly more labeling questions ($M = 8.50$ vs. $M = 0.75$ for Tellers), $F(1, 4) = 16.664$, $p = 0.015$. As these exploratory findings support the distinction between these groups, we employ these categories as we explore the data below.

### *Summary*

Teacher-initiated CDS during toddler book readings predominately focused on contextualized (labeling, describing) information. Among comments, labeling/describing remarks predominated, and among questions, yes/no and labeling questions were most frequent. Three patterns of teacher talk during book reading emerged: Askers, Tellers, and Balanced.

## Child Responses to Teacher Prompts

Complete results are presented in **Table 5** and summarized below.

### *Child Response Frequency*

We observed an average of 16.57 ($SD = 10.94$, *range* = 5–35) child responses to teacher remarks, closely aligned with the number of teacher questions. Beyond this sample average, however, there were differences between Asker and Teller classrooms, in that Asker teachers ($n = 2$) saw 30.5 child responses in their classrooms ($SD = 6.36$), while Teller teachers saw just 8.75 ($SD = 3.30$) child responses. Differences were significant in an (underpowered) ANOVA, $F(1, 4) = 34.44$, $p = 0.004$.

### *Child Response Accuracy*

Overall, across the sample as a whole, children were correct 48% of the time and incorrect or non-responsive 52% of the time, but accuracy varied by question type. Approx. 5 responses per reading ($M = 5.20$) were to yes/no questions, and most were correct (75%). Of the approx. 4 responses to label questions per reading ($M = 3.71$), 23% were correct, 34%

**TABLE 5 |** Correlations among constructs.

| Constructs | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|
| 1. CLASS-T Emotional and behavioral support | 0.60* | 0.14 | 0.65 | 0.14 | −0.61 | −0.63 | 0.02 | |
| 2. CLASS-T Engaged support for learning | 1 | 1.57 | 0.20 | −0.31 | 0.12 | 0.11 | −0.01 | −0.01 |
| 3. Book reading MLU-w | | 1 | 0.05 | 0.02 | 0.09 | 0.09 | −0.32 | 0.15 |
| 4. Book reading *D* | | | 1 | 0.00 | −0.45 | −0.45 | −0.15 | −0.32 |
| 5. All teacher comments | | | | 1 | −0.76* | −0.74∼ | 0.52 | −0.39 |
| 6. All teacher questions | | | | | 1 | 0.99*** | −0.20 | 0.70∼ |
| 7. All child responses | | | | | | 1 | −0.23 | 0.69∼ |
| 8. All child-initiated talk | | | | | | | 1 | −0.02 |
| 9. Book reading length | | | | | | | | 1 |

∼$p < 0.10$ *$p < 0.05$, ***$p < 0.001$.

were incorrect, and 42% involved non-response. Responses to repeating questions (i.e., "Say 'truck'") were offered 2–3 times per reading (*M* = 2.57), and children were correct 33% of the time but did not respond 66% of the time. Responses to choice questions were offered nearly twice per reading (*M* = 1.85), with mostly (69%) correct responses.

#### *Patterns*

Interestingly, when comparing Asker and Teller classrooms, children in Asker classrooms offered significantly more correct answers (*M* = 20.00 and *SD* = 2.83 vs. *M* = 2.75 and *SD* = 0.95), $F(1, 5) = 147.63$, $p < 0.001$, but statistically equivalent numbers of incorrect answers (*M* = 10.50 and *SD* = 3.53 vs. *M* = 6.00 and *SD* = 3.91), $F(1, 4) = 16.64$, $p = 0.015$. Descriptive statistics showed that labeling questions in particular ("What is this?") were substantially more prominent in Asker (*M* = 8.50) classrooms than Teller classrooms (*M* = 0.75).

#### *Summary*

Children responded correctly to about half of teacher questions. Choice and yes/no questions generally resulted in correct responses, while answers to labeling and request for repetition questions were most frequently incorrect.

### Teacher Responses to Child Responses

We coded teacher responses to incorrect vs. correct child answers separately for clarity.

#### *Addressing Incorrect Child Responses*

When children were incorrect, the most common teacher response was non-response (*M* = 3.14, *SD* = 3.44, *range* = 0–10), used at least once by 6 out of 7 teachers, and quite often (10 times) by one teacher (#5). It was also quite rare to tell the child that the answer was not correct (only 1 teacher used this approach, and only once).

Most teachers also used one or more strategies aimed at leading children toward the answer, including re-asking the same question (*M* = 1.57, *SD* = 1.72, *range* = 0–4), used by 5 out of 7 teachers, or giving hints (3 out of 7, *M* = 1.14, *SD* = 1.68, *range* = 0–4). Less common (on average, less than once per reading) were asking the child a rhetorical question ("Are you sure that's a dog?"), used by 3 teachers, or reframing the original question as a yes/no (2 teachers) or choice question (1 teacher). Quite rare was asking a follow-up question (1 teacher, used twice).

Ultimately, most teachers (6 out of 7) also offered the correct answer, on average this twice per reading (*M* = 2.14, *SD* = 1.68, *range* = 0–5).

#### *Addressing Correct Child Responses*

When children were correct, teachers used a different array of responses. Most common was praise (*M* = 2.43, *SD* = 4.03, *range* = 0–11), used by 4 of 7 teachers, one very frequently (11 times). In addition, five teachers repeated what children said verbatim (*M* = 2.14, *SD* = 2.67, *range* = 0–6), four teachers added words to what children said (*M* = 1.71, *SD* = 2.50, *range* = 0–7), and four expanded on their idea (*M* = 1.43, *SD* = 1.62, *range* = 0–4). Rarer strategies (used less than once per reading) included non-response (used once but by three teachers), asking rhetorical questions (used by one teacher, once), explicitly saying yes (used by three teachers), and asking follow-up questions (never used).

#### *Patterns*

Not surprisingly, because they posed more questions and accordingly had more child responses, Askers offered significantly more responses to children than Tellers did, $F(1, 4) = 11.49$, $p = 0.028$. Given the relatively small cell sizes for each type of teacher response, we did not examine further differences in specific responses between Askers and Tellers.

### Child-Initiated Talk

On average 5 times per reading, children volunteered information (*SD* = 3.78, *range* = 1–9). Unlike child responses, child-initiated talk was nearly identical regardless of whether teachers were Tellers (*M* = 5.25, *SD* = 4.35) or Askers (*M* = 5.50, *SD* = 4.95), with no significant differences between groups, $F(1, 4) = 0.01$, $p = 0.952$. In light of small cell sizes, we did not examine group differences further and instead describe the whole sample.

Looking closely at child-initiated talk, with two exceptions (the same child, in one classroom), all volunteered information was accurate. This talk mostly (at least once per classroom) involved children sharing the color or label for an illustration (*M* = 1.71, *SD* = 1.11, *range* = 1–4). In four classrooms, children specifically volunteered a particular target word (*M* = 1.14, *SD* = 1.46, *range* = 0–4). In three classrooms, one child voluntarily repeated what teachers said (*M* = 1.71, *range* = 0–5); moreover, children who repeated the teacher did so more than once (from 3 to 5 times). Only once across all videos did a child volunteer a question.

Zero-order correlations showed that the frequency of child responses in classrooms was unrelated to the frequency of child-initiated talk ($r = -0.23$, $p = 0.627$).

#### *Summary*

In all classrooms, whether Teller or Asker settings, at least one child volunteered information, generally labeling/describing illustrations or repeating; nearly all was accurate. Child-initiated talk was less frequent than responses to teachers.

### Teacher Responses to Child-Initiated Talk

While, as above, the most frequent teacher feedback strategy on child responses to teacher questions was non-response, teacher response to child-initiated talk was different. In fact, only two teachers ever used non-response to child-initiated talk; one did so four times and the other five times ($M$ = 1.29, $SD$ = 2.21, *range* = 0–5). The most common approach, used by all teachers at least once, was to explicitly say "yes" in response to a child-initiated remark; on average, this happened twice per reading (with one offering this feedback 7 times ($M$ = 2.29, $SD$ = 2.21, *range* = 1–7). Five teachers repeated children's remarks, on average once per reading ($M$ = 1.14, $SD$ = 1.60, *range* = 0–3), while four added words to what children said, on average once per reading ($M$ = 1.52, $SD$ = 2.22, *range* = 0–6) and five expanded children's ideas, on average once per reading ($M$ = 1.14, $SD$ = 1.34, *range* = 0–4). Three teachers posed at least one follow-up question ($M$ = 0.57, $SD$ = 0.79, *range* = 0–2).

#### *Patterns*

No clear patterns were apparent. Tellers gave slightly more frequent feedback on child-initiated talk ($M$ = 8.50, $SD$ = 7.77 and $M$ = 5.50, $SD$ = 4.95, respectively), but this difference was not significant, $F(1, 4) = 0.23$, $p = 0.654$.

#### *Summary*

Children volunteered information equally often, whether or not teachers emphasized asking questions. Teachers offered feedback (often a "yes") on this child talk.

## Question 5: Frequency of Target Words in Teacher and Child Talk

Frequency counts of use of target words in text, teacher talk (whether initiations or responses) and child talk (whether initiations or responses) indicated that target words were very common in the text and related conversations. Texts included 24.28 target words ($SD$ = 15.12) but ranged from 12 to 55. Teachers used, on average, 48.57 target words ($SD$ = 26.87), but there was substantial variation (*range* = 22–101). On average, 78% of teacher remarks (whether initiated or responses) included a target word. Children used 8.71 target words ($SD$ = 6.05, *range* = 0–15), so on average, 40% of child remarks (whether initiated or responses) contained a target word.

Correlations among these three variables showed that the number of target words in the text was unrelated to use of target words by teachers ($r = 0.03$, $p = 0.952$) or children ($r = 0.08$, $p = 0.869$). However, when teachers used more target words, children did as well ($r = 0.76$, $p = 0.047$).

### Patterns

Although Askers and Tellers read texts with very similar numbers of target words ($M$ = approx. 20 for both groups), Askers used target words about twice as often ($M$ = 76.50, $SD$ = 34.65) as did Tellers ($M$ = 36.25, $SD$ = 17.52), and children in Asker classrooms used target words about three times as frequently ($M$ = 14.50, $SD$ = 0.71, vs. $M$ = 5.50, $SD$ = 6.03). Differences were not statistically significant ($p > 0.20$).

### Summary

Target words were prominent in teacher extra-textual talk, and to a lesser degree in child talk; they were particularly salient in Asker classrooms.

## Question 6: Global Quality, Linguistic Complexity, and Talk Strategies Correlations

Results are presented in **Table 5**. In general, these constructs were independent of one another. Linguistic complexity was unrelated to frequency of teacher or child talk ($r$ = 0.15–0.35, $p > 0.50$). Global quality was independent of linguistic complexity ($r$ = 0.30–0.60, $p > 0.10$) and frequency of teacher and child talk ($r$ = 0.20–0.60, $p > 0.15$). Interestingly, correlations remained largely non-significant even when exploring individual kinds of talk and very specific, highly related CLASS-T dimensions (e.g., teacher responses to child remarks on our specialized coding scheme and CLASS-T teacher feedback), with four exceptions. First, linguistic complexity was marginally correlated with Positive Climate ($r = 0.71$, $p = 0.069$) and Behavior Guidance ($r = 0.70$, $p = 0.083$); in other words, more diverse teacher vocabulary was linked to stronger support for children's emotional and behavioral well-being. Second, more teacher questions were linked to a less positive climate ($r = -0.81$, $p = 0.027$), as were more child responses ($r = -0.082$, $p = 0.025$); similarly, a marginally significant correlation emerged teacher questions and Recoded Negative Climate ($r = -0.71$, $p = 0.076$). Thus, all three aspects of CDS in this sample were relatively distinct across the sample as a whole, although there were some indications that aspects of Emotional and Behavioral Support were enhanced by more vocabulary-rich teacher talk and less child talk.

Analyses of this collection of talk strategies again found differences between Askers and Tellers. In Askers' classrooms, teachers initiated and responded more overall ($M$ = 84.50, $SD$ = 14.85), relative to Tellers ($M$ = 50.00, $SD$ = 19.68), a marginally significant difference, $F(1, 4) = 4.59$, $p = 0.099$, with the Balanced classroom falling in between (65.00). Children in Asker classrooms also talked more (initiations and responses) ($M$ = 36.00, $SD$ = 1.41)—twice as much, on average–than those in Teller classrooms ($M$ = 14.00, $SD$ = 3.74), a significant difference, $F(1, 4) = 58.67$, $p = 0.002$, with the Balanced classroom falling in between (23.00).

Examining global quality and linguistic complexity, Tellers and Askers had very similar structural complexity (MLU-w) to their language ($M$ = 6.11, $SD$ = 0.76 vs. $M$ = 6.21, $SD$ = 1.20.18, respectively). However, echoing the correlations between dimensions of CLASS and CDS, Tellers had higher scores on CLASS Emotional and Behavioral Support domain ($M$ = 5.05, $SD$ = 0.85 vs. $M$ = 4.00, $SD$ = 0.57), while Askers had higher scores on the CLASS Engaged Support for Learning scale ($M$ = 4.70, $SD$ = 1.41, vs. $M$ = 3.65, $SD$ = 1.58). ANOVAs did not find significance for comparisons. Tellers also used more diverse vocabulary ($D$, $M = 36.75$, $SD = 3.50$ vs. $M = 28.90$, $SD = 4.96$), a marginally significant difference ($p = 0.085$).

### Summary

Although these three perspectives on CDS were largely distinct (uncorrelated), Askers' instruction trended toward more alignment with the CLASS teaching-oriented scale, whereas Tellers' use of predominately teacher talk and more different words trended toward higher levels of emotional and behavioral support.

## DISCUSSION

This exploratory, observational study teased apart multiple elements of CDS in the context of toddler classrooms, examining trends and individual differences. Teachers in this study received training from the *Head Start on Vocabulary* (HSoV) model, and we explored the classroom language environment (global quality, linguistic complexity, and teachers' and children's talk strategies) during book reading, a potentially vocabulary-rich part of the early childhood classroom.

Regarding general trends, we found that global classroom quality (CLASS-T) was moderate and generally equivalent to larger Early Head Start samples (Vogel et al., 2015: Emotional and Behavioral Support = 5.3, Engaged Support for Learning = 3.5) and to Head Start preschools (U.S. Department of Health and Human Services [DHHS], 2021: Emotional Support = 6.03, Classroom Organization = 5.78, Instructional Support = 2.94). During book reading, structural complexity (MLU-w) was approximately equal to preschool book reading values (Dickinson et al., 2014) and slightly higher than previous toddler book reading values (Girolametto et al., 2003), with teachers' CDS averaging about 6 words per remark, and about 20% of remarks employing one or more syntactically complex constructions. During book reading, teachers talked about three times as often as toddlers did, and most child talk was in response to teacher talk. Teachers predominately labeled and described illustrations and asked closed questions that invited children to provide a specific label. About half of children's responses were correct. Throughout this talk, target words were frequently used, especially by teachers (on average, in 78% of remarks). In a related finding, lexical diversity ($D$), however, was about half of what has been observed in book readings in preschool (Dickinson et al., 2014) or toddler classrooms (Girolametto et al., 2003), perhaps in part because of the *HSoV* focus on a small set of target words.

Beyond overall trends, three general patterns of CDS during toddler book readings emerged: Teller classrooms ($n = 4$) in which teachers used predominately comments, Asker classrooms ($n = 2$) in which they used predominately questions, and Balanced classrooms ($n = 1$) with an equal mix of both. In Asker classrooms, children talked more overall (frequently with correct answers), and both teachers and children used target words more frequently. Asker classrooms also had (descriptively, although not statistically) higher scores on the CLASS-T Engaged Support for Learning scale, while Teller classrooms performed better on the Emotional and Behavioral Support scale. Teller teachers also used (marginally) more diverse vocabulary. Findings reveal new information about CDS during book reading toddler classrooms, at least in the context of this target-word-focused *HSoV* intervention, and potential patterns of talk have implications future research in PD with this population of teachers.

### Head Start on Vocabulary Teachers' Child-Directed Speech Emphasized Target Words

The field has not determined optimal teacher lexical diversity and syntactic complexity for 2-year-olds. On one hand, there is clear evidence from the field that exposure to an extensive selection of vocabulary, including abstract words, is predictive of longer term benefits (Hart and Risley, 1995; Dickinson and Tabors, 2001), as is exposure to complex syntax (Hoff, 2003). However, in these classrooms, consistent with *HSoV* aims, teachers repeatedly labeled concrete items/images (e.g., truck, toothbrush), likely limiting lexical diversity but increasing repetitions of focal words and in some cases, allowing for moderately sophisticated descriptions of pictures in the texts.

Interestingly, when teachers asked closed, contextualized questions (mostly labeling), children were incorrect in half of their responses. This result is particularly intriguing given that the words selected for *HSoV* were relatively simple and concrete. This finding may indicate that a high degree of focused repetition, with few distractors, is very appropriate for 2-year-old children in high-need settings. Indeed, this focused talk could, over time, help children build target word knowledge, which in turn would offer a foundation for more sophisticated understandings and more extensive contributions to conversations. However, if both complex and simpler language inputs matter, future research might carefully track the appropriate balance of the two, perhaps through tracing teachers' introduction of new words and children's adoption of those words to better understand the key pathways by which new words become salient for and familiar to children.

### Children Have Different Experiences With Tellers vs. Askers

A major finding from this work involved preliminary evidence that some teachers, at least during book readings with toddlers, emphasize questions (Askers) while others emphasize comments (Tellers), and yet others have a Balance. There were some clear advantages for children of Asker classrooms—teachers and children talked substantially more, particularly about target words, linked at least marginally to higher global quality around teaching. Although we do not have data on child language learning, children in Asker teacher classrooms tended to answer questions correctly, even though Asker teachers used more labeling questions, which were, overall, one of the more challenging kinds of questions.

On the other hand, there were positive aspects of Teller classrooms, in that teachers used more diverse vocabulary, which was linked to a more Positive Climate and to higher Behavior Guidance. At the same time, teacher questions and child responses, both of which were less frequent in Teller classrooms, were inversely correlated with CLASS-T Positive Climate. Future research on this point is needed, but it may be the case that more teacher-managed discussion supports a smoothly run classroom and/or showcases teachers' support for children in ways that the CLASS-T is particularly sensitive to notice.

As only one classroom fell into the Balanced category, conclusions about this group are limited. However, a general implication of these broader findings is that there are potentially benefits of various patterns for children, and that teachers could potentially be trained to use different sets of strategies together. In addition, if these patterns or styles are widespread among teachers, coaches may need to gauge how teachers approach CDS during book reading and tailor PD to optimize that approach and include other talk strategies as well.

## LIMITATIONS AND FUTURE DIRECTIONS

A number of limitations to the current work highlight future directions for research. Most critically, as is often the case in professional development research, the total sample of teachers ($n$ = 7) is small to accommodate the costs of teacher training, observation, and coaching (Schacter, 2015). In turn, this limits the representativeness of the findings and the power of inferential analyses. Although a small sample allows for detailed coding and rich descriptive data, gathering parallel information from a larger array of classrooms would support firm statistical conclusions and, ultimately, generalizability to the Early Head Start population and beyond.

A second concern is that these seven teachers read different books to children, which can foster different kinds of talk. For example, some books were narratives (such as *Baby Loves Winter*), whereas others were more informational or participatory (*Shake a Leg! Elmo*). In our experience, allowing teachers to choose the book they would like to read helps to ensure that observations are representative of what typically happens in a class day (i.e., ecological validity); however, using the same book across all settings facilitates classroom comparisons. Future research might include both approaches and directly compare them. As a related point, more examination of the texts that teachers read to/with toddlers is needed. In particular, we found that some texts mentioned target words more than others, and some included questions (e.g., the refrain, "Is this the bus for us, Gus?") that could potentially support children's talk. Careful analyses of the teacher-child discourse around a wider array of texts in a larger sample could be helpful.

A third issue is that it was beyond the scope of this implementation-focused study to include more detailed child data. The most important dimension of any classroom-based program is its role in improving children's outcomes, ideally explored through rigorous methods that allow for causal conclusions. One future approach would be to include rigorous background information (e.g., exposure to English and other languages at home), as well as standardized and/or project-based measures of children's skills. Also of value, however, would be tracking individual children during observations to gauge which/how many children offer the most responses, are the most accurate, or even never responds. Individual-level data of this nature would potentially elucidate why teachers ask the specific questions they choose and would also help teachers better tailor their instruction to children's skills.

## CONCLUSION

Little is known about the CDS that teachers use to communicate with toddlers in classrooms. This study examined CDS as well as child talk during book readings in American Early Head Start classrooms serving 2-year-olds from households in poverty, as teachers piloted a new, light-touch PD model. Overall, the environment was similar in global quality and linguistic complexity to prior observations of toddler book readings, but close analysis of teacher and child talk strategies revealed potential differences between Askers (more focused on closed questions), Tellers (more focused on comments), and those with a Balanced approach to questions and comments. Overall, this study weaves together a more comprehensive perspective on toddler classroom language, illuminating pathways for future research.

## DATA AVAILABILITY STATEMENT

The dataset presented in this article is not readily available because these videos are protected by Temple University's IRB. Requests to access the datasets should be directed to AH (annemarie.hindman@temple.edu).

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Temple University IRB. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

AH was the PI of the grant and led the data analysis and manuscript writing. JF conducted the coding and analysis of syntax. KA conducted the coding with the CLASS. BW and PS are project PIs as well, helping to conceptualize the project, and design and implement the teacher professional development. All authors reviewed and approved the manuscript.

## FUNDING

This research was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R324A180192 to Temple University.

## REFERENCES


Abend, O., Kwiatkowski, T., Smith, N. J., Goldwater, S., and Steedman, M. (2017). Bootstrapping language acquisition. *Cognition* 164, 116–143. doi: 10.1016/j.cognition.2017.02.009

Aikens, N., Xue, Y., Bandel, E., Caronongan, P., Vogel, C., and Boller, K. (2015). *Early Head Start Home Visits and Classrooms: Stability, Predictors, and Thresholds of Quality* (OPRE Report No. 2015-34). Washington, DC: U.S. Administration for Children and Families, Office of Planning, Research and Evaluation.

Allen, C., and Swanson, M. (2010). *Shake A Leg!*. New York, NY: Random House.

Ambrose, S. E., Walker, E. A., Unflat-Berry, L. M., Oleson, J. J., and Moeller, M. P. (2015). Quantity and quality of caregivers' linguistic input to 18-month and 3-year-old children who are hard of hearing. *Ear Hear.* 36, 48S–59S. doi: 10.1097/AUD.0000000000000209

Avila-Varela, D. S., Arias-Trejo, N., and Mani, N. (2021). A longitudinal study of the role of vocabulary size in priming effects in early childhood. *J. Exp. Child Psychol.* 205:105071. doi: 10.1016/j.jecp.2020.105071

Bandel, E., Aikens, N., Vogel, C. A., Boller, K., and Murphy, L. (2014). *Observed Quality and Psychometric Properties of the CLASS-T in the Early Head Start FACES, OPRE Technical Brief 2014-34*. Available online at: https://econpapers.repec.org/paper/mprmprres/1017ea36a7044079911975a2123bc79e.htm (accessed June 1, 2021).

Beck, I. L., McKeown, M. G., and Kucan, L. (2013). *Bringing Words to Life: Robust Vocabulary Instruction*, 2nd Edn. New York, NY: Guilford.

Bloom, S. (2001). *The Bus for Us*. New York, NY: Boyds Mills Press.

Bryant, G. A., and Barrett, H. C. (2007). Recognizing intentions in infant-directed speech. *Psychol. Sci.* 18, 746–751. doi: 10.1111/j.1467-9280.2007.01970.x

Cabell, S. Q., Justice, L. M., McGinty, A. S., DeCoster, J., and Forston, L. D. (2015). Teacher-child conversations in preschool classrooms: contributions to children's vocabulary development. *Early Child. Res. Q.* 30, 80–92. doi: 10.1016/j.ecresq.2014.09.004

Cadime, I., Moreira, C. S., Santos, A. L., Silva, C., Ribeiro, I., and Viana, F. L. (2019). The development of vocabulary and grammar: a longitudinal study of European Portuguese-speaking toddlers. *J. Child Lang.* 46, 653–681. doi: 10.1017/S0305000919000060

Casla, M., Mendez-Cabezas, C., Montero, I., Murillo, E., Nieva, S., and Rodriguez, J. (2021). Spontaneous verbal repetition in toddler-adult conversations: a longitudinal study with Spanish-speaking two- year-olds. *J. Child Lang.* 1–36. doi: 10.1017/S0305000921000015 [Epub ahead of print].

Champagne, C. (2019). *Teacher-Child Interactions Around Academic Errors in Preschool.* Unpublished Doctoral dissertation. Philadelphia, PA: Temple University.

Cristofaro, T. N., and Tamis-LeMonda, C. S. (2012). Mother–child conversations at 36 months and at prekindergarten: relations to children's school readiness. *J. Early Child. Lit.* 12, 68–97. doi: 10.1177/1468798411416879

Currenton, S. M., Craig, M. J., and Flanigan, N. (2008). Use of decontextualized talk across story contexts. How oral storytelling and emergent reading can scaffold children's development. *Early Educ. Dev.* 19, 161–187. doi: 10.1080/10409280701839296

Davis, B., Torr, J., and Degotardi, S. (2015). Infants and toddlers: how visible are they in the early years learning framework? *Int. J. Child Care Educ. Policy* 9, 1–14. doi: 10.1186/s40723-015-0014-y

Dickinson, D. K., Hofer, K. G., Barnes, E. M., and Grifenhagen, J. F. (2014). Examining teachers' language in Head Start classrooms from a systemic linguistics approach. *Early Child. Res. Q.* 29, 231–244. doi: 10.1016/j.ecresq.2014.02.006

Dickinson, D., and Tabors, P. (2001). *Beginning Literacy with Language: Young Children Learning at Home and School*. Baltimore, MD: P.H. Brookes Pub.

Dickinson, D. K., Golinkoff, R. M., and Hirsh-Pasek, K. (2010). Speaking out for language: why language is central to reading development. *Educ. Res.* 39, 305–310. doi: 10.3102/0013189X10370204

Donnelly, S., and Kidd, E. (2020). Individual differences in lexical processing efficiency and vocabulary in toddlers: a longitudinal investigation. *J. Exp. Child Psychol.* 192, 104781. doi: 10.1016/j.jecp.2019.104781

Durán, P., Malvern, D., Richards, B., and Chipere, N. (2004). Developmental trends in lexical diversity. *Appl. Linguist.* 25, 220–242. doi: 10.1093/applin/20.2.220

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

Fernald, A., Performs, A., and Marchman, V. A. (2006). Picking up speed in understanding: speech processing efficiency and vocabulary growth across the 2nd year. *Dev. Psychol.* 42, 98–116. doi: 10.1037/0012-1649.42.1.98

Fitch, A., Lieberman, A. M., Luyster, R. J., and Arunachalam, S. (2020). Toddlers' word learning through overhearing: Others' attention matters. *J. Exp. Child Psychol.* 193:104793. doi: 10.1016/j.jecp.2019.104793

Genovese, G., Spinelli, M., Romero Lauro, L. J., Aureli, T., Castelletti, G., and Fasolo, M. (2020). Infant-directed speech as a simplified but not simple register: a longitudinal study of lexical and syntactic features. *J. Child Lang.* 47, 22–44. doi: 10.1017/S0305000919000643

Gershator, P., and Walker, D. (2014). *Time for a Bath*. New York, NY: Sterling.

Gilkerson, J., Richards, J. A., Warren, S. F., Oller, K., Russo, R., and Vohr, B. (2018). Language experience in the second year of life and language outcomes in late childhood. *Pediatrics* 142:e20174276. doi: 10.1542/peds.2017-4276

Girolametto, L., and Weitzman, E. (2002). Responsiveness of child care providers in interactions with toddlers and preschoolers. *Lang. Speech Hear. Serv. Sch.* 33, 268–281. doi: 10.1044/0161-1461(2002/022)

Girolametto, L., Weitzman, E., and Greenberg, J. (2003). Training day care staff to facilitate children's language. *Am. J. Speech Lang. Pathol.* 12, 299–311. doi: 10.1044/1058-0360(2003/076)

Golinkoff, R. M., Can, D. D., Soderstrom, M., and Hirsh-Pasek, K. (2015). (Baby)Talk to me: the social context of infant-directed speech and its effects on early language acquisition. *Curr. Direct. Psychol. Sci.* 24, 339–344. doi: 10.1177/0963721415595345

Golinkoff, R. M., Hoff, E., Rowe, M. L., Tamis-LeMonda, C. S., and Hirsh-Pasek, K. (2019). Language matters: Denying the existence of the 30-million-word gap has serious consequences. *Child Dev.* 90, 985–992. doi: 10.1111/cdev.13128

Hamre, B., Pianta, R., Downer, J., DeCoster, J., Mashburn, A. J., Jones, S. M., et al. (2013). Teaching through interactions: testing a developmental framework of teacher effectiveness in over 4,000 classrooms. *Element. Sch. J.* 113, 461–487. doi: 10.1086/669616

Hamre, B. K., Pianta, R. C., Burchinal, M., Field, S., Locasale-Crouch, J. L., Downer, J. T., et al. (2012). A course on effective teacher-child interactions: effects on teacher beliefs, knowledge, and observed practice. *Am. Educ. Res. J.* 49, 88–123. doi: 10.3102/0002831211434596

Harris, J., Golinkoff, R. M., and Hirsh-Pasek, K. (2011). "Lessons from the crib for the classroom: How children really learn vocabulary," in *Handbook of Early Literacy Research*, eds S. B. Neuman and D. K. Dickinson (New York, NY: Guilford Press).

Hart, B., and Risley, T. R. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*. Baltimore, MD: Brookes.

Hindman, A. H., Connor, C. M., Jewkes, A. M., and Morrison, F. J. (2008). Untangling the effects of shared book reading: multiple factors and their associations with preschool literacy outcomes. *Early Child. Res. Q.* 23, 330–350. doi: 10.1016/j.ecresq.2008.01.005

Hoff, E. (2003). "Causes and consequences of SES-related differences in parent-to-child speech," in *Socioeconomic Status, Parenting, and Child Development*, eds M. H. Bornstein and R. H. Bradley (Mahwah, NJ: Erlbaum), 147–160.

Hoff, E. (2014). *Language Development*, 5th Edn. New York, NY: Cengage.

Hoff, E., Quinn, J. M., and Giguere, D. (2018). What explains the correlation between growth in vocabulary and grammar? New evidence from latent change score analyses of simultaneous bilingual development. *Dev. Sci.* 21:e12536. doi: 10.1111/desc.12536

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

Justice, L. M., Jiang, H., and Strasser, K. (2018). Linguistic environment of preschool classrooms: What dimensions support children's language growth? *Early Child. Res. Q.* 42, 79–92. doi: 10.1016/j.ecresq.2017.09.003

Justice, L. M., McGinty, A., Zucker, T., Cabell, S., and Piasta, S. (2013). Bi-directional dynamics underlie the complexity of talk in teacher–child play-based conversations in classrooms serving at-risk pupils. *Early Child. Res. Q.* 28, 496–508. doi: 10.1016/j.ecresq.2013.02.005

Kaiser, A. P., and Trent, J. A. (2007). "Communication intervention for young children with disabilities: naturalistic approaches to promoting development," in *Handbook of Developmental Disabilities*, eds S. L. Odom, R. H. Horner, M. E. Snell, and J. Blacher (New York, NY: Guilford Press), 224–246.

Katz, K. (2013). *Baby Loves Winter*. New York, NY: Little Simon.

Kidd, L., and Rowland, C. F. (2021). The effect of language-focused professional development on the knowledge and behaviour of preschool practitioners. *J. Early Child. Lit.* 21, 27–59. doi: 10.1177/1468798418803664

La Paro, K. M., Hamre, B. K., and Pianta, R. C. (2011). *Classroom Assessment Scoring System Toddler Manual*. Charlottesville, VA: Teachstone.

Landry, S. H., Swank, P. R., Anthony, J. L., and Assel, M. A. (2011). An experimental study evaluating professional development activities within a state funded pre-kindergarten program. *Read. Writ.* 24, 971–1010. doi: 10.1007/s11145-010-0243-1

LaParo, K. M., Hamre, B. K., Locasale-Crouch, J., Pianta, R. C., Bryant, D., Early, D., et al. (2009). Quality in kindergarten classrooms: observational evidence for the need to increase children's learning opportunities in early education classrooms. *Early Educ. Dev.* 20, 657–692. doi: 10.1080/10409280802541965

Loban, W. (1976). *Language Development: Kindergarten Through Grade Twelve*. Urbana, IL: National Council of Teachers of English.

London, J., and Remkiewicz, F. (1994). *Froggy Gets Dressed*. New York, NY: Puffin.

Longobardi, E., Spataro, P., Putnick, D. L., and Bornstein, M. H. (2016). Noun and verb production in maternal and child language: continuity, stability, and prediction across the second year of life. *Lang. Learn. Dev.* 12, 183–198.

MacWhinney, B. (2000). The CHILDES project: Tools for analyzing talk. *Child Lang. Teach. Ther.* 8, 217–218. doi: 10.1177/026565909200800211

Manning, B. L., Roberts, M. Y., Estabrook, R., Petitclerc, A., Burns, J. L., Briggs-Gowan, M., et al. (2019). Relations between toddler expressive language and temper tantrums in a community sample. *J. Appl. Dev. Psychol.* 65:101070. doi: 10.1016/j.appdev.2019.101070

McKee, G. T., Malvern, D. D., and Richards, B. J. (2000). Measuring vocabulary diversity using dedicated software. *Lit. Linguist. Comput.* 15, 323–337. doi: 10.1093/llc/15.3.323

McLeod, R. H., Hardy, J. K., and Kaiser, A. P. (2017). The effects of play-based intervention on vocabulary acquisition by preschoolers at risk for reading and language delays. *J. Early Interv.* 39, 147–160. doi: 10.1177/1053815117702927

Montag, J. L., Jones, M. N., and Smith, L. B. (2018). Quantity and diversity: simulating early word learning environments. *Cogn. Sci.* 42, 375–412. doi: 10.1111/cogs.12592

Morgan, P. L., Farkas, G., Hillemeier, M. M., Hammer, C. S., and Maczuga, S. (2015). 24-month-old children with larger oral vocabularies display greater academic and behavioral functioning at kindergarten entry. *Child Dev.* 86, 1351–1370. doi: 10.1111/cdev.12398

Muhinyi, A., Hesketh, A., Stewart, A. J., and Rowland, C. F. (2020). Story choice matters for caregiver extra-textual talk during shared reading with preschoolers. *J. Child Lang.* 47, 633–654. doi: 10.1017/S0305000919000783

O'Brien, M., and Bi, X. (1995). Language learning in context: teacher and toddlers speech in three classroom play areas. *Top. Early Child. Spec. Educ.* 15, 148–163. doi: 10.1177/027112149501500202

Pancsofar, N., and Vernon-Feagans, L. (2006). Mother and father language input to young children: contributions to later language development. *J. Appl. Dev. Psychol.* 27, 571–587. doi: 10.1016/j.appdev.2006.08.003

Paschall, K. (2019). *Nearly 30 Percent of Infants and Toddlers Attend Home-Based Child Care as their Primary Arrangement*. Washington, DC: Child Trends.

Pellegrini, A. D., Perlmutter, J. C., Galda, L., and Brody, G. H. (1990). Joint reading between black Head Start children and their mothers. *Child Dev.* 61, 443–453. doi: 10.1111/j.1467-8624.1990.tb02791.x

Peter, M. S., Durrant, S., Jessop, A., Bidgood, A., Pine, J. M., and Rowland, C. F. (2020). Does speed of processing or vocabulary size predict later language growth in toddlers? *Cogn. Psychol.* 115:101238. doi: 10.1016/j.cogpsych.2019.101238

Quick, N., Erickson, K., and McCright, J. (2019). The most frequently used words: comparing child-directed speech and young children's speech to inform vocabulary selection for aided input. *Augment. Altern. Commun.* 35, 120–131. doi: 10.1080/07434618.2019.1576225

Rantalainen, K., Paavola-Ruotsalainen, L., Alakortes, J., Carter, A. S., Ebeling, H. E., and Kunnari, S. (2021). Early vocabulary development: relationships with prelinguistic skills and early social-emotional/behavioral problems and competencies. *Infant Behav. Dev.* 62:101525. doi: 10.1016/j.infbeh.2020.101525

Reese, E. (1995). Predicting children's literacy from mother-child conversations. *Cogn. Dev.* 10, 381–405. doi: 10.1016/0885-2014(95)90003-9

Rice, M. L., Smolik, F., Perpich, D., Thompson, T., Rytting, N., and Blossom, M. (2010). Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments. *J. Speech Lang. Hear. Res.* 53, 333–349. doi: 10.1044/1092-4388(2009/08-0183)

Roberts, M. Y., Kaiser, A. P., Wolfe, C. E., Bryant, J. D., and Spidalieri, A. M. (2014). Effects of the teach-model-coach-review instructional approach on caregiver use of language support strategies and children's expressive language skills. *J. Speech Lang. Hear. Res.* 57, 1851–1869. doi: 10.1044/2014_JSLHR-L-13-0113

Romano, M., and Woods, J. (2018). Collaborative coaching with early head start teachers using responsive communication strategies. *Top. Early Child. Spec. Educ.* 38, 30–41. doi: 10.1177/0271121417696276

Rowe, M. L. (2008). Child-directed speech: relation to socioeconomic status, knowledge of child development and child vocabulary skill. *J. Child Lang.* 35, 185–205. doi: 10.1017/S03000907008343

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Dev.* 83, 1762–1774. doi: 10.1111/j.1467-8624.2012.01805.x

Rowe, M. L., Leech, K. A., and Cabrera, N. (2017). Going beyond input quantity: wh-questions matter for toddlers' language and cognitive development. *Cogn. Sci.* 41, 162–179. doi: 10.1111/cogs.12349

Rowe, M. L., and Snow, C. E. (2020). Analyzing input quality along three dimensions: interactive, linguistic, and conceptual. *J. Child Lang.* 47, 5–21. doi: 10.1017/S0305000919000655

Rudko, C. (2008). *Ready, Set, Brush*. New York, NY: Studio Fun International.

Samuelsson, L. K. (2021). Toward a precision science of word learning: Understanding individual vocabulary pathways. *Child Dev. Perspect.* 15, 117–124. doi: 10.1111/cdep.12408

Schacter, R. (2015). An analytic study of the professional development research in early childhood education. *Early Educ. Dev.* 26, 1057–1085. doi: 10.1080/10409289.2015.1009335

Schertle, A., and McElmurry, J. (2015). *Little Blue Truck Leads the Way*. New York, NY: HMH Books for Young Readers.

Siraj-Blatchford, I., and Manni, L. (2008). 'Would you like to tidy up now?' An analysis of adult questioning in the English Foundation Stage. *Early Years* 28, 5–22. doi: 10.1080/09575140701842213

Suggate, S., Schaughency, E., McAnally, H., and Reese, E. (2018). From infancy to adolescence: the longitudinal links between vocabulary, early literacy skills, oral narrative, and reading comprehension. *Cogn. Dev.* 47, 82–95. doi: 10.1016/j.cogdev.2018.04.005

U.S. Department of Health and Human Services [DHHS] (2021). *A national overview of grantee CLASS® scores in 2020*. Available online at: https://eclkc.ohs.acf.hhs.gov/data-ongoing-monitoring/article/national-overview-grantee-class-scores-2020 (accessed June 1, 2021).

van Kleeck, A., Stahl, S. A., and Bauer, E. B. (2003). *On Reading Books to Children: Parents and Teachers*. New York, NY: Routledge.

Vasilyeva, M., Waterfall, H., and Huttenlocher, J. (2008). Emergence of syntax: commonalities and differences across children. *Dev. Sci.* 11, 84–97. doi: 10.1111/j.1467-7687.2007.00656.x

Vogel, C. A., Caronongan, P., Thomas, J., Bandel, E., Xue, Y., Henke, J., et al. (2015). *Toddlers in Early Head Start: A Portrait of 2-Year-Olds, their Families, and the Programs Serving them (OPRE Report #2015-10)*. Washington, DC: U.S. Department of Health and Human Services.

Wasik, B. A., and Bond, M. A. (2001). Beyond the pages of a book: Interactive book reading and language development in preschool classrooms. *J. Educ. Psychol.* 93, 243–250. doi: 10.1037/0022-0663.93.2.243

Wasik, B. A., Bond, M. A., and Hindman, A. H. (2006). The effects of a language and literacy intervention on Head Start children and teachers. *J. Educ. Psychol.* 98, 63–74. doi: 10.1037/0022-0663.98.1.632

Wasik, B. A., and Hindman, A. H. (2011). Improving vocabulary and pre-literacy skills of preschoolers in poverty through teacher professional development model. *J. Educ. Psychol.* 103, 455–469. doi: 10.1037/a0023067

Wasik, B. A., and Hindman, A. H. (2013). Realizing the promise of open-ended questions. *Read. Teach.* 67, 302–311. doi: 10.1002/TRTR.1218

Wasik, B. A., and Hindman, A. H. (2020). Increasing preschoolers' vocabulary development through a streamlined teacher professional development intervention. *Early Child. Res. Q.* 50, 101–113. doi: 10.1016/j.ecresq.2018.11.001

Weiland, C., and Yoshikawa, H. (2013). Impacts of a pre-kindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Dev.* 84, 2112–2130.

Whitehurst, G. J. (2004). *Making Education Evidence-Based: Premises, Principles, Pragmatics, and Politics (IPR Distinguished Public Policy Lecture Series, 2003-2004)*. Evanston, IL: Northwestern University, Institute for Policy Research.

Winsler, A. (2009). "Still talking to ourselves after all these years: a review of current research on private speech," in *Private Speech, Executive Functioning, and the Development of Verbal Self-Regulation*, eds A. E. Winsler, C. E. Fernyhough, and I. E. Montero (London: Cambridge University Press), 3–41. doi: 10.1017/CBO9780511581533.003

Zauche, L. H., Thul, T. A., Mahoney, A. E. D., and Stapel-Wax, J. L. (2016). Influence of language nutrition on children's language and cognitive development: an integrated review. *Early Child. Res. Q.* 36, 318–333. doi: 10.1016/j.ecresq.2016.01.015

**Author Disclaimer:** The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

ORIGINAL RESEARCH
published: 17 February 2022
doi: 10.3389/fpsyg.2022.800568

# Semantic Contingency of Maternal Verbal Input Directed at Very Preterm and Full-Term Children

*Nicoletta Salerni*[1]* and *Chiara Suttora*[2]

[1]Department of Psychology, University of Milano - Bicocca, Milano, Italy, [2]Department of Psychology "Renzo Canestrari", University of Bologna, Bologna, Italy

Several studies have testified to the importance of a responsive linguistic input for children's language acquisition and development. In particular, maternal use of expansions, imitations, interpretations, and labels has been shown to promote both children's language comprehension and production. From this perspective, the present study examined the semantically contingent linguistic input addressed to very preterm children's comparing it to that directed to full-term children observed during a semi-structured play session when the children were 24 months of age. The relationships between maternal contingent utterances and children's communicative repertoires were also investigated. The main results showed that mothers of full-term children produced a higher proportion of semantically contingent utterances than those of very preterm children; moreover, this variable was associated with children's more advanced communicative-linguistic outcomes. Overall, this study supports the interdependence between mothers' use of certain linguistic strategies and children's communicative-linguistic repertoire, extending this evidence to children born very preterm and suggesting the importance of considering the semantic contingency aspect of child-directed speech to support the communicative and linguistic development of these children.

**Keywords: very preterm children, language development, responsiveness, maternal semantically contingent input, child-directed speech**

OPEN ACCESS

***Edited by:***
*Martina Smorti,*
*University of Pisa, Italy*

***Reviewed by:***
*Lynn K. Perry,*
*University of Miami, United States*
*Caroline Junge,*
*Utrecht University, Netherlands*

***Correspondence:***
*Nicoletta Salerni*
*nicoletta.salerni@unimib.it*

***Specialty section:***
*This article was submitted to Developmental Psychology, a section of the journal Frontiers in Psychology*

***Received:*** *23 October 2021*
***Accepted:*** *24 January 2022*
***Published:*** *17 February 2022*

***Citation:***
*Salerni N and Suttora C (2022) Semantic Contingency of Maternal Verbal Input Directed at Very Preterm and Full-Term Children. Front. Psychol. 13:800568. doi: 10.3389/fpsyg.2022.800568*

## INTRODUCTION

### Child-Directed Speech: the Relevance of Maternal Input Contingency in Children's Communicative Development

It is now widely accepted that the linguistic input young children are exposed to contribute to influencing both their language acquisition and development.

In this area of investigation, many studies focused on the quantitative aspects of child-directed speech (CDS) showing that one source of variability in language growth is represented by different amount of exposure to parents' speech directly addressed to the infant (Huttenlocher et al., 1991, 2010; Hoff, 2003; Weisleder and Fernald, 2013).

Nonetheless, other empirical evidence pointed out that the quality of CDS plays an equally crucial role in fostering the development of early language skills. As recently argued by Rowe and Snow (2020) in their review on language input, three partially overlapping dimensions can be identified as facilitating language learning at different developmental stages, referring, respectively, to interactional features (i.e., responsiveness, shared attention, and discussion of child interest), linguistic aspects (i.e., phonological, lexical and grammatical levels of complexity, and redundancy), and conceptual content (i.e., topics of conversation).

In this regard, the social-interactionist perspective states that language learning is made easier if the content of the adult's CDS corresponds to the child's own processing mechanisms, namely, if it is responsive (Girolametto et al., 2002). Particularly, during the vocabulary development stage and up to the onset of the first word combinations, is well documented that maternal responsive behaviors, defined as contingent, appropriate, and prompt responses to a child's communicative initiations (Bornstein et al., 2008), can be considered a valid predictor of children's achievement of several language milestones (Tamis-LeMonda et al., 2001).

Although verbal responsiveness is considered a multidimensional construct susceptible to various operationalizations (McGillion et al., 2013), one aspect that has been focal in several studies is semantic contingency, in which the conversational partner verbally responds to the child's focus of utterances or attention (Harkness, 1988). In other words, semantic contingency may be broadly defined as the maintenance, by the adult, of the communicative exchange topic, or interest, promoted by the child, while respecting conversational turn taking and providing verbal information specifically related to it. In this sense, the linguistic input that is both conversationally (temporally connected) and conceptually related to children's utterances and actions (Rowe and Snow, 2020) is thought to increase the saliency of the input itself and the likelihood that words will be bound to real-world referents, as it reduces the cognitive loads the child needs to process linguistic information (McGillion et al., 2013; Tamis-LeMonda et al., 2014) and fosters the associations among language content, form, and use (Girolametto et al., 1999).

Different caregivers' linguistic behaviors are contingent to the child's attentional focus and topic of conversation, incorporating them into one's own immediately following turn, such as the labeling of objects or events to which the child is paying attention to, the imitation of his/her verbal productions, the expansion of the child's verbal initiatives, and the interpretation of child's prelinguistic utterances (Conti-Ramsden, 1990; Girolametto et al., 1999). In other words, during interactive exchanges, the adult adopts a series of pragmatic strategies, sometimes referred as recasts, to scaffold child's language production, by providing information about words meaning and use, correcting and reformulating incomplete or ungrammatical utterances, or repeating and extending them to encourage the production of more complex linguistic units (Clarke et al., 2017).

All of these strategies adopted by the caregiver—which are also aimed to ensure mutual involvement and active participation of the child in interaction through fine-tuning to his/her language level (Cross, 1977; Snow, 1989; Sokolov, 1993)—support children's transition from prelinguistic to linguistic communication and promote accelerated language development, as demonstrated by several studies focusing on different lexical and morphosyntactic aspects, including vocabulary size, early vocabulary composition, word type growth, spontaneous productions of newly acquired semantic relations, acquisition of grammatical morphemes, and syntactic structures (Scherer and Olswang, 1984; Tomasello and Farrar, 1986; Farrar, 1990, 1992; Pine, 1994; Nelson et al., 1996; Tamis-LeMonda et al., 2001; McGillion et al., 2013; Taumoepeau, 2016).

At the same time, the importance of such input in promoting language development has also been highlighted in children with language delays or difficulties. In this regard, Girolametto and colleagues (Girolametto et al., 1999) pointed out that maternal responsive behaviors classified as expansions and imitations represented the best predictors of children's expressive language outcomes (i.e., number of children's utterances, number of different words, vocabulary size, and word combinations) in a group of male late talkers observed longitudinally. In line with these findings, in a later study, Girolametto et al. (2002) found that mothers' semantically contingent language measures (i.e., imitations, interpretations, and expansions) were related to higher levels of verbal productivity in children exhibiting delays in expressive vocabulary development but age-appropriate cognitive and receptive language skills. More recently, a study aimed at determining, in a community-based sample of slow-to-talk toddlers, the extent to which specific maternal responsive behaviors at 24 months predict child language both concurrently and at 36 months (Levickis et al., 2014) showed that expansions, imitations, and responsive questions were strongly associated with better receptive and expressive language outcomes at both ages; moreover, expansions were the only maternal linguistic behavior that predicted language improvement by the end of children's second year of life. The emphasis placed on expansions and labeling in predicting later language development is also confirmed in a study that points out that these parental responsive utterances, in addition to child's earlier language skills, increase the ability to predict language outcomes at age 4 in a sample of children who are late to talk (Levickis et al., 2018).

## Maternal Verbal Input Directed to Children Born Preterm

Overall, empirical evidence seems to support that some of the variability observed in language abilities of both typically and atypically developing children can be explained by differences in features of language input, including responsiveness to children's early focus of attention and communicative attempts.

For this reason, it may be relevant to investigate these characteristics of maternal input also in children born preterm since, from the early stages of development, they exhibit delays that tend to persist over time, with cascading effects on more sophisticated skills developing later in their life (De Schuymer et al., 2011). A large body of literature indicates the negative impact of premature birth on the development of language abilities, especially in children born extremely and very preterm (i.e., those born, respectively, before 28 and between 28 and 32 weeks of gestation). Studies involving very preterm children found that they show a slower acquisition in both word comprehension and production with an increasing divergence, with respect to full-terms, from 12 to 24 months (Sansavini et al., 2011); moreover, at 2 years of age, a considerable proportion of them do not produce word combinations or are characterized by a small expressive vocabulary, below the 10th percentile on the MB-CDI questionnaire (Sentenac et al., 2020). Similarly, some evidence suggests that even low-risk preterm score significantly lower than term infants on the BSID-III language scales at 24 months, considering both corrected and chronological age (Ionio et al., 2016), and that their language development appears suboptimal compared to that of their full-term peers, performing less favorably even as young as 2 years of age (Nepomnyaschy et al., 2012).

Although the difficulties observed in these children appear related to biologically determined factors, part of the observed variability in language skills can also be associated with external, environmental variables (Howard et al., 2011; Wild et al., 2013). In this area of investigation, studies addressing the potential role of caregiver input provided to preterm children do not seem to be many, thus highlighting a gap in the research aimed at exploring the interrelationships between the abovementioned factor and language skills in this population. Moreover, much of this research has primarily focused on structural-linguistic and/or quantitative aspects of maternal verbal input. For instance, some authors found that during the first months of life, mothers of premature infants, in comparison with those of term infants, follow significantly more frequently their infants' non-cry vocalizations with an utterance directed at the child and initiate conversational turns more often (Reissland and Stephenson, 1999); nevertheless, similar quantities of words and utterances produced have been observed in the two groups of mothers, and no differences have been detected regarding lexical and syntactical complexity of linguistic input directed to preterm and full-term children (Salerni et al., 2007), even when considering a sample of low-risk preterm and full-term children with language delay (Suttora et al., 2020). In addition, findings suggested that changes over time in the structural characteristics of language input directed to preterm infants are also substantially similar to those described in the literature concerning typically developing infants: major increases in both lexical complexity and productivity were observed in the transition from 6 months to the second year, whereas the syntactic complexity of maternal speech, measured in terms of mean length of utterance, showed a significant increase during the second year of life (Suttora and Salerni, 2011).

As regards the functional characteristics of linguistic input, some studies highlighted that mothers of preterm children produced more directive and controlling utterances than mothers of term children (Menyuk et al., 2014), although others failed to detect such differences from 6 months of age onward (Hebert et al., 2004; Salerni et al., 2007). Despite these partially contradictory results, it seems, however, that an input characterized by maintaining the child's attention using non-directive strategies, including verbal ones, and focused on the activity or object on which the child himself/herself is currently engaged has a positive influence on language skills, especially for high-risk children and at earlier ages (Hebert et al., 2004).

More recently Benassi et al. (2018) focused on a specific feature of parental linguistic input, considering maternal responses to babies' communicative behaviors; the authors classified them according to their temporal contingency (i.e., whether they occurred within 5 s from the end of the infant's communicative production) and the degree of relevance (i.e., whether they focused on the infant's communicative behavior providing meaning to it). The data collected showed that maternal contingent relevant responses with a repeated label at 12 months were concurrently related to infants' communication skills (i.e., pointing and giving gestures, words, and receptive communication) and predicted infants' expressive communication skills at 24 months, even after controlling for neonatal status and 12-month expressive communication and cognitive skills.

Taken together, then, this empirical evidence supports the importance of considering the influence of both early infants' abilities and quality of maternal input they receive for a deeper understanding of language development in preterm infants.

## Aims of the Study

Moving from the above considerations, the present study was designed to achieve two primary goals. The first was to deepen the analysis of semantically contingent linguistic input addressed to children born very preterm and to compare it to that directed to full-term children.

As this type of investigation has usually considered children with typical development and, in some cases, those with language delay, the identification of any similarities and/or differences in the strategies adopted by the two groups of mothers may extend our knowledge about the communicative environment to which preterm infants are exposed to. Since it is possible to hypothesize that these infants show a lower level of communicative-linguistic development, or at least a different repertoire of communicative behaviors, than full-term infants, the interaction strategies that mothers use to encourage their verbal spontaneous participation may also differ from those commonly adopted with typically developing children.

The second goal was to investigate synchronic interrelations between maternal semantically contingent utterances and infant's communicative and linguistic skills in each of the two groups of dyads. From the literature, positive associations between maternal utterances and child language productivity are expected. However, assuming that the potential vulnerability of preterm infants makes them particularly susceptible to environmental factors, it is reasonable to hypothesize the presence of a pattern of associations that may be different from that found in full-term dyads. In addition, this examination can also help confirm the validity of some parental language strategies in supporting the communicative-language development of these children, leading to the identification of those that are most effective.

# MATERIALS AND METHODS

## Participants

Thirty-six monolingual Italian mothers and their children participated in the study, including 16 very preterm infants

(VPT; eight females) recruited at two neonatology follow-up services in Milan, and the remaining 20 born at term (FT; 10 females; groups did not differ for gender, $\chi^2 = 0.00$, $p = 0.631$) enrolled through notices posted in pediatric clinics in the same town. The main criteria for the selection of the very preterm children were a birthweight of less than 2000 grams ($M = 1333.75$, $SD = 338.35$), a gestational age under 32 weeks ($M = 29.94$, $SD = 2.25$) and the absence of genetic abnormalities, severe neurofunctional impairment, and/or sensorineural disabilities. Thus, the preterm children can be considered at low risk nevertheless their degree of prematurity. All full-term children were healthy and typically developing. Children born very preterm were mostly firstborn (VPT; 15 firstborn), while full-term children were more equally distributed between first and second born (FT; 11 firstborn; groups differed significantly for birth order, $\chi^2 = 6.65$, $p = 0.010$). Three very preterm participants were twins from different set of twins. Children in the very preterm and full-term groups did not differ for their mothers' level of education that was overall medium-high ($\chi^2 = 0.90$, $p = 0.710$).

Each dyad was observed at children's 24 months of age (VPT corrected age: $M = 24.16$ months, $SD = 12.35$ days, range = 23.50–24.97 months; chronological age: $M = 26.16$ months, $SD = 18.01$ days, range = 25.51–27.65 months; FT: $M = 24.07$ months, $SD = 7.82$ days; range: 23.53–24.53 months; $U = 139.50$, $p = 0.519$) and mothers were asked to fill out the questionnaire Primo Vocabolario del Bambino—scheda Parole e Frasi (Caselli and Casadio, 1995) which is the Italian version of the MacArthur-Bates Communicative Development Inventories—Words and Sentences form (Fenson et al., 1993, 2007). Data collected showed that the productive vocabulary size of VPT children ($M = 179.54$; $SD = 192.62$) was significantly lower than FT ($M = 366.95$; $SD = 134.96$), as attested by the statistical comparison carried out on the respective average values ($U = 200.50$; $p = 0.009$).

The study met ethical guidelines for human subject protections, including adherence to the legal requirements of Italy, and received formal approval by the local Research Ethical Committee of the University of Milano - Bicocca. All parents were informed about both the research procedure and general aims and gave informed written consent for study participation, data analysis, and data publication.

## Procedure

All mother-child dyads participated in a video-recorded semi-structured play session lasting approximately 20 min. VPT dyads were observed in a quiet room designed for observation at each of the two follow-up services involved, while full-term children and their mothers attended the session in the Early Childhood Observation Laboratory of the University of Milano - Bicocca.

In both observational contexts, which can be considered structurally equivalent, mothers and children sat together on a mat and mothers were encouraged to interact and play with their children freely as they would at home. The play materials have been selected with the specific goal of stimulating communicative and interactive exchanges in the dyads and consist of a series of age-appropriate toys (i.e., a toy farm with miniature people and animals, a telephone toy, a doll with clothes, and a kitchen set with pretend fruit and vegetables) and picture books. Each one was made available to mothers and children according to a fixed sequential order and it was not removed from the play area to ensure that each dyad had the opportunity to play with the toys that were most likely to stimulate their exchanges.

### Coding and Measures

Maternal linguistic input and all spontaneous children's communicative behaviors recorded during the observation sessions were entirely transcribed in CHAT format (CHILDES system; MacWhinney, 2000) by a trained observer, organizing them into separate utterances defined as any sequences of words, and/or prelinguistic sounds, and/or gestures preceded or followed by an auditory pause (1 s or more of non-speech), a change in the conversational turn, or an understandable modification in the intonation pattern.

Unclear speech in the transcriptions was reviewed by a second observer and resolved by checking the video-recorded material again.

#### *Maternal Speech*

From the transcripts, all maternal utterances that matched the ongoing topic of dyadic interaction were preliminarily identified as semantically contingent if: (a) they were produced in a joint attention situation and included a content word and/or (b) they were both temporally contiguous and linguistically and content-wise aligned to an infant production.

These utterances were, then, classified by means of a coding system developed by Girolametto et al. (2002) and partially modified for the aims of this study, which includes the following categories:

- Responsive labels: utterances that reflect the child's focus of attention and highlight a content word referring to people, events, and objects (e.g., the mother says: *"C'è un cagnolino lì!"*/*"There's a little dog there!"* while the child is looking at a picture book; *"Quanti frutti ci sono nel cesto!"*/*"So many fruits in the basket!,"* while the child is exploring a play food set).
- Interpretations: labels or short utterances aiming to disambiguate a child vocalization that is not clearly recognizable as a word, but it is produced with a communicative intention (e.g., the child vocalizes opening his arms and the mother says: *"È caduto!"*/*"It fell down!"*; the child vocalizes and points, and the mother says: *"Un pomodoro per il sugo"*/*"A tomato for the sauce"*).
- Reformulations: utterances that reproduce the child verbal production in a correct phonological form (e.g., the child says *"a falalla"* and the mother replies *"la farfalla"*/*"the butterfly"*; the child says *"etto è piccoino"* and the mother replies *"questo è piccolino"*/*"this is tiny"*).
- Imitations: partial or complete reproductions of the child's preceding preverbal or verbal production (e.g., the child says *"palla blu"*/*"blue ball"* and the mother repeats *"palla blu"*/*"blue*

ball"; the child says *"gira molto veloce"/"it turns very fast"* and the mother repeats *"gira veloce"/"it run fast"*).

– Expansions: utterances containing the repetition of an immediately preceding child's verbalization and the addition of one or more morphemes or words (e.g., the child says *"un cappello"/"a hat"* and the mother replies *"un cappello come quello del nonno"/"a hat like grandpa's"*; the child says *"è fiore"/"it's flower"* and the mother repeats *"è un fiore"/"it's a flower"*).

The following measures were then calculated: (a) the frequency per minute of utterances (as an index of speech productivity); (b) the proportion of semantically contingent utterances out of the total number of utterances. The proportions associated with each type of semantically contingent utterances were also computed. However, because children vary not only in the total number of spontaneous communicative behaviors they produce but also in their quality, mothers' opportunities to respond contingently to children's productions by using the various strategies considered also vary. For this reason, the proportional values were calculated by varying the denominator according to the meaning of each category of maternal contingent utterances. Consequently, the following measures were considered: (c) responsive labels on the total number of maternal utterances; (d) interpretations on the total number of children's preverbal productions; (e) reformulations on the total number of children's verbal productions; (f) imitations, and (g) expansions on the total number of children's productions.

#### *Children's Prelinguistic and Verbal Behavior*

In order to assess children's prelinguistic and verbal repertoire, each transcribed utterance, excluding crying vocalizations and sounds of distress, was classified according to its complexity. Therefore, the following categories were considered:

– Preverbal productions: vocalizations containing a vowel or a syllable composed of a glottal or a glide consonant, single-syllable speech sounds, reduplicated and variegated babbling, and onomatopoeic sounds (e.g., "*ba," "dada," "bati,"/"bu-bu"*; Stoel-Gammon, 1989).
– One-word utterances: verbal productions consistently used to signal the same meaning and that approximate the sound of the conventional word used by adults (e.g., *"cappe" [scarpe]/"shoes"*; *"bello,"/"beautiful"*; Vihman and McCune, 1994).
– Transitional forms: utterances composed of two or more vocal elements, in which at least one is a word, but that cannot be considered true multi-word utterances, including: content words anticipated by a single not meaningful vocalic element (e.g., *"/e/ fiore"/"/e/ flower"*); chained words, which are two words uttered with close temporal contiguity (less than 1 s) and which perform two separate speech acts (e.g., the child sees two figures on a book and says *"cane, gatto"/"dog, cat"*); horizontal repetitions, in which the child produces a word and repeats it in close temporal contiguity within the same turn (e.g., *"casa, casa," "home, home"*; *"alto, alto,"/"tall, tall"*); frozen phrases, in which a number of words are uttered as a single word, since the distinct words are not used in other occurrences (e.g., the child uses the expression *"chi è?"/"who's it?,"* but the function word *"chi"/"who,"* is produced only in this context); and non-word combinations, in which a content word is anticipated or followed by a multi-syllabic form not recognizable as a word in the adult lexicon (e.g., *"bimbo bedi"/"baby bedi,"* where *"bedi"* is a non-word; Fasolo and D'Odorico, 2012).
– Word combinations: utterances consisting of two or more words semantically and prosodically related to each other and uttered in a close temporal succession (e.g., *"questo latte" "this milk"*; *"voglio io" "I want"*; *"io siedo qui"/"I sit here"*; D'Odorico and Carubbi, 2003).

For each category, the proportional frequency was calculated out of the total number of utterances. The number of child utterances per minute was also computed as a measure of child talkativeness.

Interrater reliability was calculated using the intraclass correlation coefficient (ICC) obtaining high levels of agreement for all maternal (ICC: ≥0.95) and children measures (ICC: ≥0.96).

## RESULTS

### Statistical Analyses

IBM SPSS Statistics 25 was used to conduct data analyses. Tests were bilateral with a statistical significance set at 0.05. Preliminary analyses have been performed to assess data distribution for normality. Results showed that most of the variables were not normally distributed with Kolmogorov-Smirnov and Shapiro-Wilk tests reporting $ps < 0.01$. For this reason, non-parametric tests were used.

A first set of Mann-Whitney tests was carried out to assess our first aim regarding the presence of differences in semantic contingency in mothers of very preterm and full-term children.

Secondly, we used Spearman correlations to explore the concurrent associations among maternal input variables and children's linguistic measures, separately for very preterm and full-term participants. Before performing these analyses, a set of Mann-Whitney tests were performed to assess differences between very preterm and full-term children's speech production.

### Semantic Contingency of Maternal Verbal Input in Very Preterm and Full-Term Mother-Child Dyads

The descriptive statistics for the measures of maternal input for all participants, and for mothers of very preterm and full-term children, are reported in **Table 1**. During the mother-child interactive session, mothers addressed their children with approximately 15 utterances per minute, with no difference being observed between mothers of very preterm and full-term children. 26% of maternal utterances were semantically contingent to children's communicative initiatives, with mothers of full-term children using such utterances significantly more than mothers in the very preterm group. Furthermore, results indicated that mothers of full-term children also produced a significantly

**TABLE 1** | Descriptive data for maternal speech input measures and Mann-Whitney tests comparing mothers of very preterm and full-term children.

| | All participants (*n* =36) | | | Very preterm group (*n* =16) | | | Full-term group (*n* =20) | | | *U* | *p* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | ***M*** | ***SD*** | **Range** | ***M*** | ***SD*** | **Range** | ***M*** | ***SD*** | **Range** | | |
| Total Utterances per Minute | 15.54 | 3.89 | 8.50–23.64 | 16.35 | 4.83 | 8.90–23.64 | 14.89 | 2.91 | 8.50–19.66 | 134.00 | 0.408 |
| Contingent Utterances | 0.26 | 0.08 | 0.05–0.38 | 0.21 | 0.07 | 0.05–0.34 | 0.30 | 0.06 | 0.18–0.38 | 48.50 | <0.001 |
| Responsive Labels | 0.15 | 0.06 | 0.05–0.32 | 0.13 | 0.05 | 0.05–0.21 | 0.17 | 0.06 | 0.06–0.32 | 97.00 | 0.045 |
| Interpretations | 0.05 | 0.09 | 0.00–0.30 | 0.06 | 0.11 | 0.00–0.30 | 0.04 | 0.07 | 0.00–0.30 | 149.00 | 0.708 |
| Imitations | 0.09 | 0.10 | 0.00–0.51 | 0.11 | 0.14 | 0.00–0.51 | 0.07 | 0.05 | 0.01–0.17 | 157.00 | 0.924 |
| Expansions | 0.12 | 0.07 | 0.00–0.27 | 0.08 | 0.06 | 0.00–0.17 | 0.14 | 0.07 | 0.03–0.27 | 94.00 | 0.036 |
| Reformulations | 0.08 | 0.06 | 0.00–0.24 | 0.08 | 0.08 | 0.00–0.24 | 0.07 | 0.05 | 0.01–0.17 | 160.00 | 1.000 |

**TABLE 2** | Descriptive data for children's speech measures and Mann-Whitney tests comparing mothers of very preterm and full-term children.

| | All participants (*n* =36) | | | Very preterm group (*n* =16) | | | Full-term group (*n* =20) | | | *U* | *p* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | ***M*** | ***SD*** | **Range** | ***M*** | ***SD*** | **Range** | ***M*** | ***SD*** | **Range** | | |
| Productions per Minute | 5.86 | 3.39 | 0.00–13.05 | 4.63 | 3.4 | 0.00–11.04 | 6.83 | 3.13 | 0.00–13.05 | 107.5 | 0.090 |
| Preverbal Productions | 0.35 | 0.27 | 0.07–1.00 | 0.59 | 0.24 | 0.28–1.00 | 0.17 | 0.08 | 0.07–0.33 | 2.50 | <0.001 |
| One-word Utterances | 0.36 | 0.17 | 0.00–0.78 | 0.29 | 0.17 | 0.00–0.53 | 0.42 | 0.16 | 0.22–0.78 | 108.00 | 0.102 |
| Transitional Forms | 0.15 | 0.10 | 0.00–0.41 | 0.09 | 0.07 | 0.00–0.21 | 0.20 | 0.09 | 0.03–0.41 | 53.50 | <0.001 |
| Word Combinations | 0.13 | 0.13 | 0.00–0.42 | 0.03 | 0.04 | 0.00–0.14 | 0.21 | 0.13 | 0.02–0.42 | 26.00 | <0.001 |

greater proportion of responsive labels over the total utterances produced. The same pattern emerged regarding the use of expansions, as full-terms mothers were observed to expand their children's verbal productions more than mothers of very preterm children. Concerning the use of interpretations, imitations, and reformulations no significant differences due to birth condition were observed (Mann-Whitney test results are also reported in **Table 1**).

## Associations Between Input Semantic Contingency Measures and Children's Speech Production

**Table 2** summarizes the descriptive statistics of children's speech measures collected during mother-child interactive sessions, for all participants and for each group separately. At a descriptive level, children produced almost six communicative productions per minute with no differences between very preterm and full-term participants. The communicative repertoires of very preterm children were mostly constituted by preverbal productions, followed by one-word utterances, transitional and combinational forms. Contrariwise, full-term children exhibited mostly verbal utterances. Comparing very preterm and full-term children, results showed that children born preterm exhibited more preverbal production and less transitional forms and word combinations than their peers born at term (Mann-Whitney test results are also reported in **Table 2**).

Results of correlational analyses carried out to assess synchronic associations between maternal speech measures and children's communicative repertoire variables within each group are shown in **Table 3**. Maternal use of semantically contingent utterances was associated with better children's outcomes both in very preterm and full-term children. As a matter of fact, contingent utterances correlated negatively with children's use of preverbal productions in both groups, and positively with one-word, transitional and combinational utterances in very preterm children, and with word combinations in full-term participants. Moreover, mothers' use of expansions of children's productions was significantly and positively associated with the use of one-word and transitional utterances in the very preterm group and, similarly, to the use of one-word utterances in children born at term. Maternal expansions were also negatively associated with the amount of preverbal productions in very preterm children. Finally, mothers' use of imitations was observed as positively associated with very preterm children's use of one-word utterances and negatively with their preverbal productions.

**TABLE 3** | Results of Spearman correlation performed between speech input measures and children's speech measures for the very preterm and the full-term group.

| | Contingent utterances | Responsive labels | Interpretations | Imitations | Expansions | Reformulations |
|---|---|---|---|---|---|---|
| **Very Preterm Children** | | | | | | |
| Preverbal Productions | −0.635** | −0.135 | 0.090 | −0.634** | −0.700** | −0.350 |
| One-word Utterances | 0.588* | 0.018 | 0.209 | 0.815** | 0.768** | 0.370 |
| Transitional Forms | 0.519* | 0.094 | −0.224 | 0.280 | 0.522* | 0.415 |
| Word Combinations | 0.543* | 0.132 | −0.441 | 0.366 | 0.353 | 0.414 |
| **Full-term Children** | | | | | | |
| Preverbal Productions | −0.546* | −0.269 | 0.111 | 0.006 | 0.066 | 0.080 |
| One-word Utterances | −0.146 | −0.223 | 0.370 | −0.320 | 0.466* | 0.414 |
| Transitional Forms | −0.081 | −0.011 | −0.411 | 0.250 | −0.280 | −0.292 |
| Word Combinations | 0.647** | 0.346 | −0.265 | 0.119 | −0.242 | −0.421 |

*$p<0.05$; **$p<0.01$.*

## DISCUSSION

This observational study represents one of the first attempts to examine maternal semantically contingent input directed to very preterm children. More specifically, the first aim was to investigate whether mothers differed in the verbal strategies they adopted in response both to their children focus of attention and spontaneous utterances depending on birth condition.

Overall, the results showed that, although mothers of both groups exhibited similar verbal productivity, those of very preterm children produced proportionally fewer semantically contingent utterances during interactions, compared to full-term mothers. Thus, even though the two groups of children were exposed to a similar maternal amount of maternal language, the linguistic input directed at very preterm children appeared to match to a lesser extent the ongoing topic of the interaction. Furthermore, this maternal interactive style did not appear to depend directly on children's talkativeness, as no significant difference emerged between the two groups of children in the frequency of utterances per minute. This finding seems to be in line with some studies showing that mothers of preterm infants appear to be less able to establish an interactive symmetry, especially documented in the early stages of development. On this matter, mothers of preterm infants have been described as more intrusive and less synchronous during social exchanges with their infants, demonstrating poor competence in coordinating their interactive behaviors with the infant's alertness (Greenberg and Crnic, 1988; Feldman and Eidelman, 2007; Salerni et al., 2007; Provasi, 2019).

However, very preterm children participating in the study, regardless of their verbal productivity, showed a different spontaneous communicative-linguistic repertoire than term children, mainly characterized by preverbal productions, while transitional forms and word combinations were less frequently produced. Therefore, their less advanced language skills may have affected the mothers' opportunities to use some specific types of semantically contingent utterances in a relevant way. In fact, if we consider the various categories of semantically contingent utterances controlling, when appropriate, for the communicative and linguistic behaviors spontaneously produced by children, the only differences between the two groups of mothers are those referring to responsive labels and expansions. Specifically, mothers of preterm infants were less likely to provide verbal information that were semantically related to their children attentional focus and to follow infants' preverbal or verbal productions by repeating them and adding syntactic or semantic information. Taken together, these findings suggest that the language environment to which preterm infants are exposed to has certain characteristics that make it less optimal for language development, as evidenced by studies that have found positive associations between measures of maternal semantically contingent language and infant language outcomes (Tamis-LeMonda et al., 2001, 2014; Vigil et al., 2005; Roseberry et al., 2014; Masek et al., 2021).

A similar pattern of relationships also emerged in the present study, in that mothers' greater use of semantically contingent utterances corresponds to a children's more advanced communicative-linguistic repertoire. In both very preterm and full-term children, this sort of maternal input is concurrently related to a lower spontaneous production of preverbal communicative behaviors and to a greater use of combinatorial utterances. Moreover, positive associations were found, in very preterm dyads only, between maternal semantically contingent utterances and one-word and transitional forms produced by children during interaction.

From the observed pattern of correlations, a particular role in this type of interrelations seems to be played by maternal utterances aimed at expanding children's productions and, limited to very preterm children, imitating them, a result, the latter, in line with that found for Italian late talkers (Girolametto et al., 2002; Suttora et al., 2021). Both these linguistic strategies emphasize the communicative value of the child's vocal/verbal productions and, at the same time, lead the child to focus his/her attention on a new stimulus inserted in an immediate and familiar linguistic context, thus facilitating the extraction of relevant information and enhancing child language development (Girolametto et al., 1999).

The lack of associations found between maternal responsive labeling and child language measures deserves a special consideration. Such a finding appears to be in contrast with previous empirical evidence in which labels produced within joint attention episodes promoted vocabulary learning, both in typically developing children and in those born preterm (Goldin-Meadow et al., 2007; Olson and Masur, 2015; Benassi et al., 2018).

In this regard, it is possible to hypothesize that in the developmental stage considered in this study, expressive language skills are more supported by contingent maternal responses to child productions, rather than by the simple exposure to the naming of objects or events that fall within the child's attentional focus. In other words, although labeling may be considered a relevant strategy in the period when children need to learn the first vocabulary, its effectiveness may be reduced when the stage-specific task is to expand the productive vocabulary and verbally express semantic relations. In this case, children may benefit from a richer linguistic input.

Taken together, the results of this study provide evidence for an effect of mothers' semantically contingent input on the variability of linguistic competences of children who are expanding their expressive lexicon and acquiring early combinatorial skills, confirming previous studies conducted with typically developing children and extending this evidence to children born very preterm (Tamis-LeMonda et al., 2001; Masek et al., 2021). From a theoretical point of view, then, this study supports a perspective of language acquisition that relies on an interaction between children's processing mechanisms and the content of CDS they are exposed to.

In this sense, it is also important to acknowledge that the associations detected in this study do not necessarily represent direct influences of maternal input on the child's language abilities. The results, in fact, can also be interpreted from the perspective of how the child influences the parent (Conti-Ramsden, 1990). Recent studies showed that very preterm infants are characterized by difficulties and delays in several areas of development, including psychomotor skills (Fuentefria et al., 2017), temperamental traits (Cassiano et al., 2020), and social-emotional competences (Yaari et al., 2018), which can impact on the overall quality of mother-infant interaction from early on. These possible sources of variation along with the continuous transactions, over time, between the child and his/her proximal environment might result in specific maternal behaviors and attitudes which reflected in different communicative and interactive styles.

At the same time, the pattern of highlighted associations suggests the relevance of considering maternal linguistic responsivity as a modular construct, as particular linguistic strategies may serve different functions and exert specific effects on children's language at different stages of development (Bornstein et al., 2008). In this sense, considering maternal language responsiveness in a more differentiated way also allows for a better understanding of it in different populations, leading to the identification of specificities that would remain unnoticed if more global measures were used.

## Implication for Clinical Practice

The interventions directed at promoting language development in populations of children with language delay (Heidlage et al., 2020) and of children with secondary linguistic issues (Pennington and Thomson, 2007; Weitzman, 2013) have been proven to be effective in supporting children language development. One basic aim of such interventions is to enhance parents' responsiveness, making their input much more attuned to their children's focus of attention and responsive to their communicative bids. This research suggests that special attention should be paid to parents' provision of semantically contingent linguistic input, beyond the importance that structural features of language may take on. Furthermore, the findings point at the input provided to very preterm children, encouraging clinicians to set up projects and interventions aimed at ameliorating their language environment.

## Limitations and Strengths of the Study

The study presents some limitations that should be considered. The first is that the sample is limited in terms of size which can lead to issues of generalization of the findings. However, given the peculiarity of the population in exam, i.e., very preterm children, the study should be considered exploratory in its aims. Another issue concerns the different distribution of firstborn children in the groups, with preterm children being mostly firstborn and full-term being equally distributed between first and laterborn. Even if we cannot rule out the effect of parity from the interpretation of our findings, previous literature reports that being firstborn represents a protective factor and not a risk for language development and for input exposure, as first-time mothers usually spend more quality time with their children than mothers of laterborn children. Findings documented that firstborns are often favored in their language development, showing an earlier lexical onset and greater lexical acquisition than laterborns (Hoff, 2003; Nafissi and Vosoughi, 2015). From these data, we can hypothesize that the lower level of semantically contingency found in mothers of very preterm children is most probably due to the peculiarity of this birth condition, rather than by parity, but still, new evidence is needed in support of this thesis.

Another limit that has to be acknowledged refers to the correlational design adopted which not allowed to disambiguate the direction of the associations between maternal and children's measures. In this regard, longitudinal studies are needed to also investigate whether semantically contingent maternal input changes over time and to understand the different impact that specific indices may have at different developmental stages.

This study has some relevant strengths too. The first is represented by the detailed analysis of maternal and child speech performed, which provided rich information on the types of maternal input contingency and on the quality and quantity of children's verbal utterances. Another point of strength is the choice of the target population of very preterm children. Literature highlights that children with higher degrees of immaturity, such as very preterm children, are at risk of developing a language delay (Sansavini et al., 2010, 2011);

identifying those aspects that can support or rather further hinder their language development is pivotal for research and clinic.

## CONCLUSION

The present work contributes with new data to the understanding of the role of maternal verbal input contingency in children communicative and linguistic development, with a specific focus on the effects of preterm birth on this topic. Overall, the study highlights that mothers of children born very preterm differ from full-term in the way they respond to their children's communicative bids and use their verbal input to attune to their children's ongoing focus of interest. In addition, our findings reveal that a higher use of semantically contingent utterances is associated with children's more mature communicative repertoires. These findings suggest the opportunity to address parental speech contingency to implement intervention targeted at foster communicative and language development in very preterm children.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary files, and further inquiries can be directed to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of University of Milano - Bicocca. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## AUTHOR CONTRIBUTIONS

NS: conceptualization and methodology. NS and CS: data collection, data transcription, coding, analysis, writing original draft, review, and editing. All authors have read and approved the manuscript.

## REFERENCES


Benassi, E., Guarini, A., Savini, S., Iverson, J. M., Caselli, M. C., Alessandroni, R., et al. (2018). Maternal responses and development of communication skills in extremely preterm infants. *First Lang.* 38, 175–197. doi: 10.1177/0142723717736449

Bornstein, M. H., Tamis-LeMonda, C. S., Hahn, C. S., and Haynes, O. M. (2008). Maternal responsiveness to young children at three ages: longitudinal analysis of a multidimensional, modular, and specific parenting construct. *Dev. Psychol.* 44, 867–874. doi: 10.1037/0012-1649.44.3.867

Caselli, M. C., and Casadio, P. (1995). *Il primo vocabolario del bambino. Guida all'uso del questionario MacArthur per la valutazione della comunicazione e del linguaggio nei primi anni di vita*. Milano: Franco Angeli.

Cassiano, R. G. M., Provenzi, L., Linhares, M. B. M., Gaspardo, C. M., and Montirosso, R. (2020). Does preterm birth affect child temperament? A meta-analytic study. *Infant Behav. Dev.* 58:101417. doi: 10.1016/j.infbeh.2019.101417

Clarke, M. T., Soto, G., and Nelson, K. (2017). Language learning, recasts, and interaction involving AAC: background and potential for intervention. *Augment. Altern. Commun.* 33, 42–50. doi: 10.1080/07434618.2016.1278130

Conti-Ramsden, G. (1990). Maternal recasts and other contingent replies to language-impaired children. *J. Speech Hear. Disord.* 55, 262–274. doi: 10.1044/jshd.5502.262

Cross, T. G. (1977). "Mothers' speech adjustments: the contribution of selected child listener variables," in *Talking to Children*. eds. C. Snow and C. Ferguson (United Kingdom: Cambridge University Press), 151–188.

D'Odorico, L., and Carubbi, S. (2003). Prosodic characteristics of early multi-word utterances in Italian children. *First Lang.* 23, 97–116. doi: 10.1177/0142723703023001005

De Schuymer, L., De Groote, I., Beyers, W., Striano, T., and Roeyers, H. (2011). Preverbal skills as mediators for language outcome in preterm and full-term children. *Early Hum. Dev.* 87, 265–272. doi: 10.1016/j.earlhumdev.2011.01.029

Farrar, M. J. (1990). Discourse and the acquisition of grammatical morphemes. *J. Child Lang.* 17, 607–624. doi: 10.1017/S0305000900010904

Farrar, M. J. (1992). Negative evidence and grammatical morpheme acquisition. *Dev. Psychol.* 28, 90–98. doi: 10.1037/0012-1649.28.1.90

Fasolo, M., and D'Odorico, L. (2012). Gesture-plus-word combinations, transitional forms, and language development. *Gesture* 12, 1–15. doi: 10.1075/gest.12.1.01fas

Feldman, R., and Eidelman, A. I. (2007). Maternal postpartum behavior and the emergence of infant-mother and infant-father synchrony in preterm and full-term infants: the role of neonatal vagal tone. *Dev. Psychobiol.* 49, 290–302. doi: 10.1002/dev.20220

Fenson, L., Dale, P., Reznick, J. S., Thal, D., Bates, E., Hartung, J., et al. (1993). *MacArthur Communicative Development Inventories: User's Guide and Technical Manual*. San Diego, CA: Singular Publishing.

Fenson, L., Marchman, V. A., Thal, D. J., Reznick, J. S., and Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories: User's Guide and Technical Manual. 2nd Edn*. Baltimore: Brookes.

Fuentefria, R. D. N., Silveira, R. C., and Procianoy, R. S. (2017). Motor development of preterm infants assessed by the Alberta infant motor scale: systematic review article. *J. Pediatr.* 93, 328–342. doi: 10.1016/j.jped.2017.03.003

Girolametto, L., Bonifacio, S., Visini, C., Weitzman, E., Zocconi, E., and Pearce, P. S. (2002). Mother-child interactions in Canada and Italy: linguistic responsiveness to late-talking toddlers. *Int. J. Lang. Commun. Disord.* 37, 153–171. doi: 10.1080/13682820110116794

Girolametto, L., Weitzman, E., Wiigs, M., and Pearce, P. S. (1999). The relationship between maternal language measures and language development in toddlers with expressive vocabulary delays. *Am. J. Speech Lang. Pathol.* 8, 364–374. doi: 10.1044/1058-0360.0804.364

Goldin-Meadow, S., Goodrich, W., Sauer, E., and Iverson, J. (2007). Young children use their hands to tell their mothers what to say. *Dev. Sci.* 10, 778–785. doi: 10.1111/j.1467-7687.2007.00636.x

Greenberg, M. T., and Crnic, K. A. (1988). Longitudinal predictors of developmental status and social interaction in premature and full-term infants at age two. *Child Dev.* 59, 554–570. doi: 10.1111/J.1467-8624.1988.TB03216.X

Harkness, S. (1988). The cultural construction of semantic contingency in mother-child speech. *Lang. Sci.* 10, 53–67. doi: 10.1016/0388-0001(88)90005-8

Hebert, H., Swank, P., Smith, K., and Landry, S. (2004). Maternal support for play and language across early childhood. *Early Educ. Dev.* 15, 93–116. doi: 10.1207/s15566935eed1501_6

Heidlage, J. K., Cunningham, J. E., Kaiser, A. P., Trivette, C. M., Barton, E. E., Frey, J. R., et al. (2020). The effects of parent-implemented language interventions on child linguistic outcomes: a meta-analysis. *Early Child. Res. Q.* 50, 6–23. doi: 10.1016/j.ecresq.2018.12.006

Hoff, E. (2003). The specificity of environmental influence: socioeconomic status affects early vocabulary development *via* maternal speech. *Child Dev.* 74, 1368–1378. doi: 10.1111/1467-8624.00612

Howard, K., Roberts, G., Lim, J., Lee, K. J., Barre, N., Treyvaud, K., et al. (2011). Biological and environmental factors as predictors of language skills in very preterm children at 5 years of age. *J. Dev. Behav. Pediatr.* 32, 239–249. doi: 10.1097/DBP.0b013e31820b7882

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., and Lyons, T. (1991). Early vocabulary growth: relation to language input and gender. *Dev. Psychol.* 27, 236–248. doi: 10.1037/0012-1649.27.2.236

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

Ionio, C., Riboni, E., Confalonieri, E., Dallatomasina, C., Mascheroni, E., Bonanomi, A., et al. (2016). Paths of cognitive and language development in healthy preterm infants. *Infant Behav. Dev.* 44, 199–207. doi: 10.1016/j.infbeh.2016.07.004

Levickis, P., Reilly, S., Girolametto, L., Ukoumunne, O. C., and Wake, M. (2014). Maternal behaviors promoting language acquisition in slow-to-talk toddlers: prospective community-based study. *J. Dev. Behav. Pediatr.* 35, 274–281. doi: 10.1097/DBP.0000000000000056

Levickis, P., Reilly, S., Girolametto, L., Ukoumunne, O. C., and Wake, M. (2018). Associations between maternal responsive linguistic input and child language performance at age 4 in a community-based sample of slow-to-talk toddlers. *Child Care Health Dev.* 44, 776–783. doi: 10.1111/cch.12600

MacWhinney, B. (2000). *The CHILDES Project: The Database*. *Vol. 2.* United Kingdom: Psychology Press.

Masek, L. R., McMillan, B. T. M., Paterson, S. J., Tamis-LeMonda, C. S., Golinkoff, R. M., and Hirsh-Pasek, K. (2021). Where language meets attention: how contingent interactions promote learning. *Dev. Rev.* 60:100961. doi: 10.1016/j.dr.2021.100961

McGillion, M. L., Herbert, J. S., Pine, J. M., Keren-Portnoy, T., Vihman, M. M., and Matthews, D. E. (2013). Supporting early vocabulary development: what sort of responsiveness matters? *IEEE Trans. Auton. Ment. Dev.* 5, 240–248. doi: 10.1109/TAMD.2013.2275949

Menyuk, P., Liebergott, J. W., and Schultz, M. C. (2014). *Early Language Development in Full-Term and Premature Infants*. United Kingdom: Psychology Press.

Nafissi, Z., and Vosoughi, M. (2015). A critical meta-analytic exploration of birth order effect on L1 onset time of speaking and language development progression; is the pointer towards first or later Borns? *Theory Pract. Lang. Stud.* 5:1960. doi: 10.17507/tpls.0509.28

Nelson, K. E., Camarata, S. M., Welsh, J., Butkovsky, L., and Camarata, M. (1996). Effects of imitative and conversational recasting treatment on the acquisition of grammar in children with specific language impairment and younger language-normal children. *J. Speech Lang. Hear. Res.* 39, 850–859. doi: 10.1044/jshr.3904.850

Nepomnyaschy, L., Hegyi, T., Ostfeld, B. M., and Reichman, N. E. (2012). Developmental outcomes of late-preterm infants at 2 and 4 years. *Matern. Child Health J.* 16, 1612–1624. doi: 10.1007/s10995-011-0853-2

Olson, J., and Masur, E. F. (2015). Mothers' labeling responses to infants' gestures predict vocabulary outcomes. *J. Child Lang.* 42, 1289–1311. doi: 10.1017/S0305000914000828

Pennington, L., and Thomson, K. (2007). It takes two to talk - The Hanen program® and families of children with motor disorders: a UK perspective. *Child Care Health Dev.* 33, 691–702. doi: 10.1111/j.1365-2214.2007.00800.x

Pine, J. M. (1994). Environmental correlates of variation in lexical style: interactional style and the structure of the input. *Appl. Psycholinguist.* 15, 355–370. doi: 10.1017/S0142716400004495

Provasi, J. (2019). "Early interaction and developmental psychopathology," in *Early Interaction and Developmental Psychopathology*. eds. G. Apter, E. Devouche and M. Gratier (Germany: Springer), 123–149.

Reissland, N., and Stephenson, T. (1999). Turn-taking in early vocal interaction: a comparison of premature and term infants' vocal interaction with their mothers. *Child Care Health Dev.* 25, 447–456. doi: 10.1046/j.1365-2214.1999.00109.x

Roseberry, S., Hirsh-Pasek, K., and Golinkoff, R. M. (2014). Skype me! Socially contingent interactions help toddlers learn language. *Child Dev.* 85, 956–970. doi: 10.1111/cdev.12166

Rowe, M. L., and Snow, C. E. (2020). Analyzing input quality along three dimensions: interactive, linguistic, and conceptual. *J. Child Lang.* 47, 5–21. doi: 10.1017/S0305000919000655

Salerni, N., Suttora, C., and D'Odorico, L. (2007). A comparison of characteristics of early communication exchanges in mother-preterm and mother-full-term infant dyads. *First Lang.* 27, 329–346. doi: 10.1177/0142723707081654

Sansavini, A., Guarini, A., Justice, L. M., Savini, S., Broccoli, S., Alessandroni, R., et al. (2010). Does preterm birth increase a child's risk for language impairment? *Early Hum. Dev.* 86, 765–772. doi: 10.1016/j.earlhumdev.2010.08.014

Sansavini, A., Guarini, A., Savini, S., Broccoli, S., Justice, L., Alessandroni, R., et al. (2011). Longitudinal trajectories of gestural and linguistic abilities in very preterm infants in the second year of life. *Neuropsychologia* 49, 3677–3688. doi: 10.1016/j.neuropsychologia.2011.09.023

Scherer, N. J., and Olswang, L. B. (1984). Role of mothers' expansions in stimulating children's language production. *J. Speech Lang. Hear. Res.* 27, 387–396. doi: 10.1044/jshr.2703.387

Sentenac, M., Johnson, S., Charkaluk, M. L., Sëppanen, A. V., Aden, U., Cuttini, M., et al. (2020). Maternal education and language development at 2 years corrected age in children born very preterm: results from a European population-based cohort study. *J. Epidemiol. Community Health* 74, 346–353. doi: 10.1136/jech-2019-213564

Snow, C. E. (1989). "Understanding social interaction and language acquisition; sentences are not enough," in *Crosscurrents in Contemporary Psychology. Interaction in Human Development*. eds. M. H. Bornstein and J. S. Bruner (NJ: Lawrence Erlbaum Associates Inc.), 83–103.

Sokolov, J. L. (1993). A local contingency analysis of the fine-tuning hypothesis. *Dev. Psychol.* 29:1008, –1023. doi: 10.1037/0012-1649.29.6.1008

Stoel-Gammon, C. (1989). Prespeech and early speech development of two late talkers. *First Lang.* 9, 207–223. doi: 10.1177/014272378900900607

Suttora, C., Guarini, A., Zuccarini, M., Aceti, A., Corvaglia, L., and Sansavini, A. (2020). Speech and language skills of low-risk preterm and full-term late talkers: the role of child factors and parent input. *Int. J. Environ. Res. Public Health* 17, 1–22. doi: 10.3390/ijerph17207684

Suttora, C., and Salerni, N. (2011). Maternal speech to preterm infants during the first 2 years of life: stability and change. *Int. J. Lang. Commun. Disord.* 46, 464–472. doi: 10.1111/j.1460-6984.2011.00007.x

Suttora, C., Zuccarini, M., Aceti, A., Corvaglia, L., Guarini, A., and Sansavini, A. (2021). The effects of a parent-implemented language intervention on late-talkers' expressive skills: the mediational role of parental speech contingency and dialogic reading abilities. *Front. Psychol.* 12:723366. doi: 10.3389/fpsyg.2021.723366

Tamis-LeMonda, C. S., Bornstein, M. H., and Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Dev.* 72, 748–767. doi: 10.1111/1467-8624.00313

Tamis-LeMonda, C. S., Kuchirko, Y., and Song, L. (2014). Why is infant language learning facilitated by parental responsiveness? *Curr. Dir. Psychol. Sci.* 23, 121–126. doi: 10.1177/0963721414522813

Taumoepeau, M. (2016). Maternal expansions of child language relate to growth in children's vocabulary. *Lang. Learn. Dev.* 12, 429–446. doi: 10.1080/15475441.2016.1158112

Tomasello, M., and Farrar, M. J. (1986). Joint attention and early language. *Child Dev.* 57:1454. doi: 10.2307/1130423

Vigil, D. C., Hodges, J., and Klee, T. (2005). Quantity and quality of parental language input to late-talking toddlers during play. *Child Lang. Teach. Ther.* 21, 107–122. doi: 10.1191/0265659005ct284oa

Vihman, M. M., and McCune, L. (1994). When is a word a word? *J. Child Lang.* 21, 517–542. doi: 10.1017/S0305000900009442

Weisleder, A., and Fernald, A. (2013). Talking to children matters: early language experience strengthens processing and builds vocabulary. *Psychol. Sci.* 24, 2143–2152. doi: 10.1177/0956797613488145

Weitzman, E. (2013). More Than words—The Hanen program for parents of children with autism Spectrum disorder: A teaching model for parent-implemented language intervention. *Perspect. Lang. Learn. Educ.* 20, 96–111. doi: 10.1044/lle20.3.86

Wild, K. T., Betancourt, L. M., Brodsky, N. L., and Hurt, H. (2013). The effect of socioeconomic status on the language outcome of preterm infants at toddler age. *Early Hum. Dev.* 89, 743–746. doi: 10.1016/J.EARLHUMDEV.2013.05.008

Yaari, M., Rotzak, N. L., Mankuta, D., Harel-Gadassi, A., Friedlander, E., Eventov-Friedman, S., et al. (2018). Preterm-infant emotion regulation during the still-face interaction. *Infant Behav. Dev.* 52, 56–65. doi: 10.1016/j.infbeh.2018.05.008

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 09 May 2022
doi: 10.3389/fpsyg.2022.804792

# Infant-Directed Speech From a Multidimensional Perspective: The Interplay of Infant Birth Status, Maternal Parenting Stress, and Dyadic Co-regulation on Infant-Directed Speech Linguistic and Pragmatic Features

*Maria Spinelli[1]*, Francesca Lionetti[1], Maria Concetta Garito[1], Prachi E. Shah[2], Maria Grazia Logrieco[1], Silvia Ponzetti[1], Paola Cicioni[3], Susanna Di Valerio[3] and Mirco Fasolo[1]*

[1] Department of Neuroscience, Imaging and Clinical Sciences, G. d'Annunzio University of Chieti-Pescara, Chieti, Italy, [2] Division of Developmental Behavioral Pediatrics, Department of Pediatrics, University of Michigan, Ann Arbor, MI, United States, [3] Department of Maternal and Child Health, Santo Spirito Hospital, Pescara, Italy

OPEN ACCESS

**Edited by:**
Iris Nomikou,
University of Portsmouth,
United Kingdom

**Reviewed by:**
Theano Kokkinaki,
University of Crete, Greece
Anne Henning,
SRH Hochschule für Gesundheit,
Germany

***Correspondence:**
Maria Spinelli
maria.spinelli@unich.it

**Specialty section:**
This article was submitted to
Developmental Psychology,
a section of the journal
Frontiers in Psychology

**Received:** 29 October 2021
**Accepted:** 14 April 2022
**Published:** 09 May 2022

**Citation:**
Spinelli M, Lionetti F, Garito MC,
Shah PE, Logrieco MG, Ponzetti S,
Cicioni P, Di Valerio S and Fasolo M
(2022) Infant-Directed Speech From
a Multidimensional Perspective:
The Interplay of Infant Birth Status,
Maternal Parenting Stress,
and Dyadic Co-regulation on
Infant-Directed Speech Linguistic
and Pragmatic Features.
Front. Psychol. 13:804792.
doi: 10.3389/fpsyg.2022.804792

Infant-directed speech (IDS), the particular form of spontaneous language observed in interactions between parents and their infants, is a crucial aspect of the mother-infant interaction and an index of the attunement of maternal linguistic input to her infant communicative abilities and needs during dyadic interactions. The present study aimed to explore linguistic and pragmatic features of IDS during mother-infant interactions at 3-month of infant age. The effects of infant (birth status: preterm vs. full-term birth), maternal (perceived parenting stress) and dyadic (dyadic co-regulation) factors on IDS were explored. Results evidenced few differences between the groups on IDS linguistic characteristics. Moreover, observing the interaction of birth status and dyadic co-regulation, full-term mothers varied their IDS pragmatic features according to the quality of co-regulation while preterm mothers did not. Parenting stress was associated to specific linguistic IDS features independently from the birth status. Findings are discussed underling implications for the study of preterm dyads interactions and the importance to consider the interplay of several factors in affecting the quality of IDS.

**Keywords: infant-directed speech, preterm birth, parenting stress, dyadic co-regulation, mother-infant interaction**

## INTRODUCTION

Infant-directed speech (IDS) is the particular form of spontaneous language observed in interactions between parents and their infants (Ferguson, 1964; Saxton, 2009; Saint-Georges et al., 2013). Compared to adult-directed speech (ADS), this type of verbal interaction is characterized by a simplification of speech demonstrated by fewer utterances, lexical and syntactic simplification, specific pragmatic functions, and emphasized prosody (see Soderstrom et al., 2008, for a review; Fernald and Simon, 1984; Genovese et al., 2020).

Characteristics of IDS and associations with infant development have been studied extensively over the last few decades. The specific vocal patterns and linguistic features of IDS are an important part of mother-infant interactions, and play a role in regulating and attracting infant attention, making linguistic input more apparent and salient to infants, and helping infant interpretation of the emotional signals of adult speakers (Saint-Georges et al., 2013; Spinelli et al., 2017). Decades of research have further elucidated the role of infant directed speech on facilitating co-regulated attention and affect during mother-child interactions, and the potential help to foster early language development (Ferguson, 1964; Soderstrom, 2007).

Patterns of child-directed speech have been observed from infancy to childhood (see for example Kitamura et al., 2001; Genovese et al., 2020), and have demonstrated patterns of stability and instability over time. Some characteristics of IDS, such as the prosodic features tend to be more stable, with minor variations over time (Kitamura and Burnham, 2003). Other features, (e.g., linguistic characteristics), appear to change over time (Bornstein et al., 1992; Ko, 2012), with mothers adjusting their speech to make it more complex and variable, in line with their infants and children's increasing cognitive and communicative abilities (Genovese et al., 2020).

The specific features of IDS, its important role for infant and child development, its variations with infant age, make IDS a crucial aspect of the mother-infant interaction and an index of the attunement of maternal linguistic input to her infant and child communicative abilities and needs during dyadic interactions (Saint-Georges et al., 2013). While it is believed that IDS varies *within* the dyadic interaction according to the communicative needs and aims of the mother and infant communication, as well as *among* mothers according to maternal ability to attune to infant's needs, most studies of IDS have focused on descriptive features of IDS and/or its associations with infants outcomes. The intra and inter-individual differences in IDS have been under-explored and is a current gap in the research.

The main aim of the present study is to contribute to this field of research by exploring how select individual and dyadic factors, previously associated with the quality of the mother-infant interaction, are also associated with IDS features. We examined these associations during the preverbal age of 3 months of age. This age is crucial for several reasons, first because the growing communicative abilities of the infant, i.e., vocalizations, smiles and movements, make him/her contribution to the interaction very relevant, second because at this age maternal voice is considered to be one of the main interactive modalities during dyadic face-to-face interactions, all of this resulting in moments of dyadic shared affect and attention (Stern et al., 1983; Lavelli and Fogel, 2013). We explored infant characteristics [i.e., infant's birth condition, represented by preterm (PT) vs. full-term (FT) birth], maternal characteristics (i.e., maternal well-being, represented by mother's levels of parenting stress), and dyadic characteristics such as the quality of dyadic co-regulation, and their associations with IDS characteristics (i.e., linguistic and pragmatic aspects).

Preterm birth–birth before the 37th week of gestational age—is a non-normative birth experience, associated with perturbations in several areas of newborn development. One area of developmental vulnerability centers on the communicative abilities of preterm infants. Compared to infants born full-term, preterm infants, especially during the first months of life, show different communicative abilities. PT infants are reported to be less reactive to social cues than FT infants, and manifest interactions characterized by lower attentional control (Bhutta et al., 2002; Clark et al., 2008), diminished alertness and responsivity (Goldberg, 1978; Goldberg and DiVitto, 1995), greater passivity, less initiation (Sajaniemi et al., 1998), increased irritability (Hughes et al., 2002; Larroque et al., 2005), and fewer expressions of positive affect (Garcia Coll et al., 1992). In addition, compared to their FT counterparts, PT infants are less interactive, and vocalize less in response to the utterances of their mother (Reissland and Stephenson, 1999).

These interactive difficulties complicate the social and affective exchanges between PT infants and their mothers, who may find the infant's cues and reactions difficult to understand (Loi et al., 2017). There is some evidence to suggest that mothers of PT infants tend to look, smile, vocalize, affectionately touch them less often, and appear to be less competent at coordinating their social behaviors with the infant's signals, compared to mothers of FT newborns (Forcada-Guex et al., 2006; Olafsen et al., 2006; Feldman and Eidelman, 2007; De Schuymer et al., 2011). However, there is also variability within the pattern of PT dyadic interactions, with not all mothers and PT infants showing the same interactional difficulties (Agostini et al., 2014; Sansavini et al., 2015; Neri et al., 2017). Underscoring this finding, a recent meta-analysis examining studies on maternal sensitivity of PT and FT mothers reported a general lack of evidence of group differences (Bilgin and Wolke, 2015).

Similar mixed findings have been reported by the few studies exploring differences in IDS between PT and FT mothers during dyadic interactions. Some studies found that mothers of PT infants demonstrated contingent vocalizations more frequently (Reissland and Stephenson, 1999), used more complex interrogatives (Reissland et al., 1999), and tended to interrupt silent pauses in the conversation more often (Salerni et al., 2007) than FT mothers. Other studies failed to find differences in IDS linguistic features between FT and PT dyads (i.e., mean lenght of utterance (MLU), type/token ratio, quantity of tokens and types per minute and frequency of utterances per minute (Salerni et al., 2007; Suttora et al., 2020a) and the total amount of speech (Adams et al., 2018).

It is thought that mothers of children born PT may need to modulate their interactions to the appropriate level of linguistic stimulation to avoid over- or underwhelming the PT infant's communication and arousal regulation capacities (Suttora and Salerni, 2011). Some mothers might be more able to do that, others less able, not only because of the infant prematurity condition, but also because of specific maternal characteristics.

It is well known that the interactive difficulties of PT mothers are not homogeneous, but vary by conditions related to neonatal and maternal health. Interactional differences are attributed to both the level of neonatal risk, e.g., mothers of more at risk infants have demonstrated lower sensitivity to infants cues (Agostini et al., 2014; Bilgin and Wolke, 2015), and also to the level of maternal wellbeing. Preterm birth is a non-normative transition to motherhood characterized by the sudden interruption of pregnancy, the subsequent fear for and worry about the infant's survival and health condition, and the experience of caregiving in the highly technological environment of the Neonatal Intensive Care Unit (NICU) for an extended time. These experiences can result in preterm mothers experiencing high levels of psychological distress (Coppola et al., 2007; Feldman and Eidelman, 2007; Spinelli et al., 2016). This heightened psychological distress might alter mothers' perceptions and attitudes toward the infant, rendering her experience of parenting stressful and demanding. There is evidence suggesting that the majority of mothers of PT infants experience high levels of parenting stress (Brummelte et al., 2011; Gray et al., 2012), with associated negative impacts on the quality of dyadic interactions (Spinelli et al., 2013). Higher parenting stress has been associated with less attuned, less positive interactions, and more intrusive behaviors (Spinelli et al., 2013; Suttora et al., 2020b). To our knowledge, the associations between maternal emotional difficulties (e.g., parenting stress) and the quality of IDS in mothers of PT infants have not been examined, a gap in the science which this research will address.

## The Present Study

While it is well described that PT birth poses a risk for suboptimal mother-infant interactions, the impact of preterm birth on the quality of dyadic interactions and maternal wellbeing varies. As a consequence, the quality of maternal verbal communication varies. Considering the role of IDS in infant development, there is a growing need to explore the interactive effects of preterm birth with other potential sources of variability in the quality of mother-infant interactions and their associations with maternal communicative behavior in IDS.

The aim of the present study was to explore the interactive effects of PT birth, maternal parenting stress, and the quality of dyadic co-regulation on the linguistic and pragmatic features of IDS. We considered dyadic co-regulation as a form of dyadic process that consider both the infant and the mother behaviors. Co-regulation refers to a form of coordinated action between participants that involves a continuous mutual adjustment of actions and intentions (Fogel and Thelen, 1987). We expected to find variability in IDS features which was associated with PT birth and maternal psychological characteristics. Specifically, since PT mothers are more at risk for psychological difficulties, we hypothesized that PT mothers IDS would be more affected by parenting stress, with PT mothers with higher parenting stress demonstrating IDS less appropriate to infant age characterized by, i.e., low quantity and variety of verbal interaction, more control sentences. Moreover, we expected to find PT and FT mothers using different patterns of IDS during moments of shared and un-shared co-regulated interactions, i.e., more control sentences and more complex speech during un-shared co-regulated patterns.

# MATERIALS AND METHODS

## Participants

One hundred and one mothers and their 3 month-old (corrected age for PT) infants (PT = 56 and FT = 55) participated in the study. Among those, 14 mothers of PT infants and 12 mothers of FT infants were excluded due to failure to complete study questionnaires (i.e., Parenting Stress Index). The final sample consisted of 86 dyads (PT = 42 and FT = 44).

Preterm infants born <37 weeks gestational age were included. Exclusion criteria for both preterm and full-term groups were the presence of genetic abnormalities, severe neurodevelopmental impairment, and/or neurosensory disabilities (e.g., blindness or deafness).

Most mothers (mean age: PT = 33.63, SD = 5.05; FT = 35.23, SD = 4.84) had a middle or high school level of education: 38.8% had a high school degree; 55.3% graduated college or had a master's degree; 5.9% had less than a high school education. Preterm infants (43% Males; 78% First born) and FT infants (59% Males; 66% First born) had a mean gestational age of 30.71 (SD = 2.63) and 39.41 (SD = 1.18) weeks, and a mean birth weight of 1,379 (SD = 437.92) and 3,397 (SD = 406.97) grams, respectively. All infants were singletons.

## Procedure

Mothers were invited to participate with their infant in a videotaped observational session when their infant was 3 months old (for preterm infants the corrected age was used). FT dyads were recruited via public services and advertisements, PT dyads were recruited by nurses and doctors of the hospital where they were born. All mothers completed and signed a consensus form before participation. The session consisted of 3 min of face-to-face interaction with the infant seated on an infant seat, and the mother seated directly in front of the infant, facing a mirror which was located behind the infant's seat so that both partners' faces could be clearly seen. After a brief introduction, mothers were asked to interact with their infants, as they would do at home.

After the interactive episode, mothers were asked to complete the short version of the Parenting Stress Index. Each session was entirely transcribed according to the CHAT transcription system (Codes for the Human Analysis of Transcripts) of the CHILDES computational system (Child Language Data Exchange System) (MacWhinney, 2000).

The study was approved by the Ethical Committee of the Department.

## Coding and Measures

### Dyadic Co-regulation

Mother–infant interactions were coded using the Revised Relational Coding System (Fogel et al., 2003) to capture the quality of the interactive involvement between mothers and infants. The whole interaction was coded. The quality of dyadic behaviors ranges from the absence of orientation of one partner to the other, to the mutual and continuous adjustment of their respective actions.

The coding system includes five global categories of communicative interactions: unilateral, asymmetrical, symmetrical, disruptive, and unengaged. In the present study we considered the global categories of *symmetrical co-regulation* (characterized by both partners adjusting their communicative actions to the continuously changing actions of the partner, and engaging in active, mutual engagement, and sharing experience via vocal and non-vocal behaviors) and *unilateral co-regulation* (characterized by only one partner trying to engage the other, while the other is absorbed in their own activity and failing to pay attention to the partner, or respond to the partner's initiations).

The co-regulation patterns were coded every second from the videotapes by a trained coder, using the Mangold Interact 18 software. The relative total duration of each pattern was computed. An independently trained coder processed 25% of the sessions to compute inter-observer reliability. The Kappa values were 0.86 for symmetrical co-regulation and 0.94 for unilateral co-regulation.

### Parenting Stress

Mothers were asked to complete the PSI-Short Form questionnaire (PSI-SF; (Abidin, 1995)). The PSI-SF is a commonly used questionnaire designed to measure stress in the parent-child system and to identify caregivers that are most in need of support. The PSI-SF includes 36 items rated from 1 to 5 on a Likert scale (1 = strongly disagree; 5 = strongly agree), and consists of three subscales, of 12 items each: Parental Distress (PSI-PD), Parent-Child Dysfunctional Interaction (PSI-P-CDI) and Difficult Child (PSI-DC). High values indicate more parenting stress. For the present study, the Parental Distress subscale was used (Cronbach's $\alpha$: PD = 0.86). This scale explores the stress related to the parent's perception of her/his child-rearing competences, the level of spousal conflicts or support, and the restrictions placed by parental role. Item mean scores were calculated by dividing the sum of item scores by the number of items comprising that scale.

### Infant-Directed Speech: Linguistic Features

Maternal vocal productions were coded in order to analyze:

- Verbosity: Rate per minute of utterances, word types and tokens (Phillips, 1973; Henning et al., 2005).
- Lexical variability: Type/token ratio (*TTR*), which is a ratio of the number of types to tokens (Johnson, 1944; Broen, 1972; Phillips, 1973).
- Syntactic complexity: *MLU,* which is a ratio of the total number of words spoken to the total number of utterances (Snow, 2009).

### Infant-Directed Speech: Pragmatic Features

In order to classify the pragmatic meaning of maternal productions, these were divided into the following categories and the percentages for each category over the total number of utterances considered were calculated (Longobardi, 1992):

- Conversational: Sentences used to promote and maintain the conversation with the infant (i.e., emphatic sentences and comments "You look happy," open questions "What are you looking at?", comments on the present/past activity of the infant "We are playing together").
- Control: Sentences used to re-orient infant attention, to direct infant attention toward something (i.e., direct requests "Speak to me," claiming infant attention "Ehy, look at me").
- Preverbal: Sounds and sentences using typical baby-talk words, repetition of infant's sounds.

## Analyses Plan

We first computed descriptive statistics and bivariate correlations among study variables in the full sample and separately in the two groups. The two birth groups were compared for mean values along the investigated variables. Afterward, to explore the single and additive role of birth status, dyadic co-regulation and parenting stress, and the interplay between dyadic co-regulation and parenting stress with birth status on IDS characteristics, we estimated and compared several multivariate models and then explored parameters of the best selected model More specifically, pertaining to predictors, first birth status was included in the model. Then, we considered the additive role of the psychological variables investigated (co-regulation and parenting stress). Finally, we included the interaction term between birth status and each of the dyadic/maternal variables considered to see if co-regulation and parenting stress differently predicted IDS depending on birth status. The Akaike Information Criterion (AIC) was used for model comparison, with lower values providing more support to a model against the others. The first group of multivariate models included as outcome variables IDS linguistic characteristics (verbosity, TTR and MLU), and then we considered IDS pragmatic characteristics (conversational, control and preverbal sentences) Regression parameters were explored for the best fitting model. Analyses were run using the statistical software R, using Lavaan package.

# RESULTS

## Descriptive Statistics

Means, SDs, and correlation values among variables of interest in the full sample are reported in **Table 1**. Within the IDS features, we observed significant correlations as expected. Symmetrical co-regulation was negatively associated with unilateral co-regulation as expected ($r = -0.56$). Parenting stress was positively associated with TTR, and MLU ($r = 0.28$ and $r = 0.22$, respectively).

Exploration of bivariate associations among investigated variables run separately for the two birth status groups (see **Table 2**) suggested that IDS linguistic characteristics were strongly associated with Parenting Stress in the FT group, but not in the PT group. More stressed FT mothers were observed to speak less (verbosity: $r = -0.37$) and demonstrated higher lexical variability and syntactic complexity (TTR: $r = 0.34$, MLU: $r = 0.36$). Similarly, IDS pragmatic characteristics were significantly associated with Unilateral and Symmetrical co-regulation only in the FT group. Mothers of FT infants pronounced less conversational and more controlled sentences when the dyad spent more time in unilateral co-regulation

**TABLE 1 |** Descriptive and bivariate correlations for the full sample.

| | | Mean (SD) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Verbosity | 26.16 (8.52) | – | | | | | | | | |
| 2 | TTR | 0.40 (0.09) | −0.46* | – | | | | | | | |
| 3 | MLU | 4.07 (0.94) | −0.19 | 0.10 | – | | | | | | |
| 4 | % Conversational | 56.84 (15.36) | −0.19 | 0.01 | 0.46** | – | | | | | |
| 5 | % Control | 30.89 (13.97) | 0.25* | −0.22* | −0.30** | −0.68** | – | | | | |
| 6 | % Preverbal | 7.59 (10.77) | −0.08 | 0.25* | −0.33** | −0.42** | −0.30** | – | | | |
| 7 | Parenting stress | 1.83 (0.56) | −0.16 | 0.28** | 0.22* | 0.08 | −0.14 | 0.15 | – | | |
| 8 | Unilateral co-regulation | 0.31 (0.26) | −0.09 | 0.21 | −0.11 | −0.09 | 0.19 | −0.05 | 0.08 | – | |
| 9 | Symmetrical co-regulation | 0.21 (0.16) | 0.18 | −0.16 | 0.06 | 0.02 | −0.18 | 0.15 | 0.03 | −0.56** | – |

*$p > 0.05$, **$p < 0.01$.

**TABLE 2 |** Descriptive and bivariate correlations in the preterm (above the diagonal, *n* = 42) and full-term (below the diagonal, *n* = 44) groups.

| | | Mean (SD) PT | Mean (SD) FT | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Verbosity | 21.91 (6.32) | 30.20 (8.43) | – | −0.43** | −0.01 | −0.28 | 0.19 | 0.03 | −0.10 | −0.11 | 0.10 |
| 2 | TTR | 0.41 (0.09) | 0.39 0.09) | −0.55** | – | −0.14 | 0.02 | −0.19 | 0.33* | 0.24 | 0.33* | −0.19 |
| 3 | MLU | 4.33 (0.91) | 3.83 0.91) | −0.12 | 0.31* | – | 0.47** | −0.43** | −0.17 | 0.18 | −0.17 | 0.18 |
| 4 | % Conversational | 60.77 (14.13) | 53.09 (15.69) | 0.05 | −0.03 | 0.39** | – | −0.77** | −0.29 | 0.17 | 0.10 | −0.08 |
| 5 | % Control | 29.02 (13.47) | 32.68 (14.35) | 0.24 | −0.24 | −0.13 | −0.59** | – | −0.22 | −0.14 | −0.08 | 0.06 |
| 6 | % Preverbal | 5.27 (6.67) | 9.80 (13.30) | −0.31* | 0.26 | −0.37* | −0.45** | −0.41** | – | 0.07 | 0.21 | −0.13 |
| 7 | Parenting Stress | 1.75 (0.54) | 1.90 (0.58) | −0.37* | 0.34* | 0.36* | 0.08 | −0.18 | 0.16 | – | 0.30 | −0.06 |
| 8 | Unilateral co-regulation | 0.34 (0.28) | 0.29 (0.23) | 0.01 | 0.06 | −0.11 | −0.34* | 0.52** | −0.17 | −0.12 | – | −0.54** |
| 9 | Symmetrical co-regulation | 0.19 (0.17) | 0.23 (0.16) | 0.18 | −0.13 | 0.01 | 0.18 | −0.44** | 0.27 | 0.08 | −0.58** | – |

*$p > 0.05$, **$p < 0.01$.

($r = -0.34$ and $r = 0.52$, respectively) and mothers used less control sentences when the dyad spent more time in symmetrical co-regulation ($r = -0.44$).

The One Way ANOVA evidenced only few differences between PT and FT groups. When speaking to their infants, mothers of PT infants, talked less [verbosity: $F_{(1,84)} = 26.45$, $p < 0.001$], demonstrated higher syntactic complexity [MLU: $F_{(1,84)} = 6.37$, $p = 0.01$], more conversational and less preverbal sentences [$F_{(1,84)} = 5.67$, $p = 0.02$ and $F_{(1,84)} = 3.91$, $p = 0.05$, respectively] than FT mothers. Concerning parenting stress [$F_{(1,84)} = 0.49$, $p = 0.21$] and dyadic co-regulation [unilateral: $F_{(1,84)} = 0.78$, $p = 0.38$; symmetrical: $F_{(1,84)} = 1.27$, $p = 0.26$], no differences emerged between the groups.

## Multivariate Regression Models

### Birth Status and Unilateral Co-regulation on Infant-Directed Speech Linguistic Characteristics

Comparison of multivariate regression models (see **Table 3**) demonstrated that model 2 (which included birth status and dyadic unilateral co-regulation as single effects) outperformed the other models. Standardized estimates of model 2 are reported in **Table 4**. Only regression parameters of the effect of birth status were significant at $p < 0.05$, except for the role of birth status on TTR. Results showed that mothers of PT infants spoke less (verbosity: $\beta = 0.97$, $p < 0.001$), and with lower syntactic complexity (MLU: $\beta = -0.52$, $p = 0.007$) than mothers of FT infants. None of the parameters regarding the effect of unilateral co-regulation were significant.

### Birth Status and Unilateral Co-regulation on Infant-Directed Speech Pragmatic Characteristics

Comparison of multivariate regression models (see **Table 3**) demonstrated that model 3 (which included birth status, dyadic unilateral co-regulation and their interaction) outperformed the other models. Standardized estimates of model 3 are reported in **Table 4**. Regression parameters of the interaction effect were significant at $p < 0.05$ for conversational ($\beta = -28.20$, $p = 0.021$) and control sentences ($\beta = 35.95$, $p = 0.001$) (see **Table 4**). As represented in **Figure 1**, the more time the dyad spends in unilateral co-regulation, the more likely FT mothers are to reduce the quantity of conversational sentences while PT mothers continue using a high percentage of conversational sentences. Conversely, the more time the dyad spends time in unilateral co-regulation, the more likely FT mothers are to use control sentences, while PT mothers do not vary in the amount of control sentences pronounced (see **Figure 2**).

### Birth Status and Symmetrical Co-regulation on Infant-Directed Speech Linguistic Characteristics

Comparison of multivariate regression models (see **Table 3**) demonstrated that model 1 (which included only birth status as a single effect) outperformed the other models. Standardized estimates of model 1 are reported in **Table 5**. Regression parameters of the effect of birth status were significant at $p < 0.05$,

**TABLE 3 |** Model comparison, effects of birth status, and dyadic co-regulation: AIC.

| | IDS linguistic | IDS pragmatic |
|---|---|---|
| **Model** | **AIC** | **AIC** |
| Model 1: Birth status | 262.51 | 1876.5 |
| Model 2: Birth status, unilateral co-regulation | **262.29** | 1876.3 |
| Model 3: Birth status, unilateral co-regulation, birth status × unilateral co-regulation | 266.98 | **1868.5** |
| Model 1: Birth status | **262.51** | 1876.5 |
| Model 2: Birth status, symmetrical co-regulation | 264.63 | 1877.5 |
| Model 3: Birth status, symmetrical co-regulation, birth status × symmetrical co-regulation | 269.31 | **1874.3** |

*In bold are highlighted models receiving more support for each set of outcome variables considered (IDS linguistic and IDS pragmatic).*

**TABLE 4 |** Multivariate analysis on IDS linguistic and pragmatic characteristics: Standardized estimated parameters of models 2 and 3, respectively.

| | Verbosity | *TTR* | MLU | Conversational | Control | Preverbal |
|---|---|---|---|---|---|---|
| | β *(p)* | β *(p)* | β *(p)* | β *(p)* | β *(p)* | β *(p)* |
| Birth status | 0.97 (<0.001) | −0.01 (0.712) | −0.52 (0.007) | 0.67 (0.889) | −6.85 (0.113) | 8.98 (0.011) |
| Unilateral co-regulation | −0.15 (0.682) | 0.07 (0.05) | −0.49 (0.189) | 33.33 (0.068) | −39.68 (0.015) | 19.55 (0.139) |
| Birth status × Unilateral co-regulation | – | – | – | −28.20 (0.021) | 35.95 (0.001) | −14.68 (0.096) |

except for the role of birth status on TTR. Results showed that mothers of PT infants spoke less (verbosity: $\beta = 0.97$, $p < 0.001$), and with lower syntactic complexity (MLU: $\beta = -0.50$, $p = 0.011$) compared with mothers of FT infants. None of the parameters regarding the effect of symmetrical co-regulation were significant.

## Birth Status and Symmetrical Co-regulation on Infant-Directed Speech Pragmatic Characteristics

Comparison of multivariate regression models (see **Table 3**) demonstrated that model 3 (which included birth status, dyadic symmetrical co-regulation and their interaction) outperformed the other models. Standardized estimates of model 3 are reported in **Table 5**. Regression parameters of the interaction effect were significant at $p < 0.05$ for control ($\beta = -44.74$, $p = 0.010$) and preverbal sentences ($\beta = 27.44$, $p = 0.042$) (see **Table 5**). As represented on **Figure 3**, the more time the dyad spends time in symmetrical co-regulation, the less FT mothers use control sentences, while PT mothers do not vary in the amount of use of control sentences. Conversely, the more time the dyad spends time in symmetrical co-regulation, the more likely FT mothers are to increase the quantity of preverbal sentences, while PT mothers continue using a low percentage of preverbal sentences (see **Figure 4**).

## Birth Status and Parenting Stress on Infant-Directed Speech Linguistic Characteristics

Comparison of multivariate regression models (see **Table 6**) showed model 2 (which included birth status and parenting stress as single additive effects) outperformed the other models.

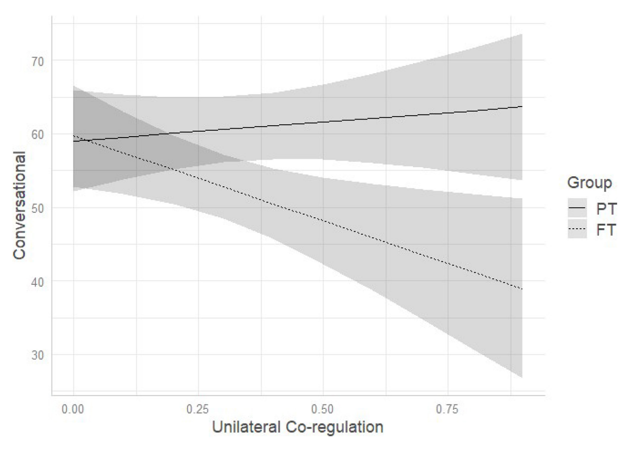



**FIGURE 1 |** Interaction among birth status and unilateral co-regulation on IDS conversational pragmatic sentences. PT, preterm; FT, full-term.

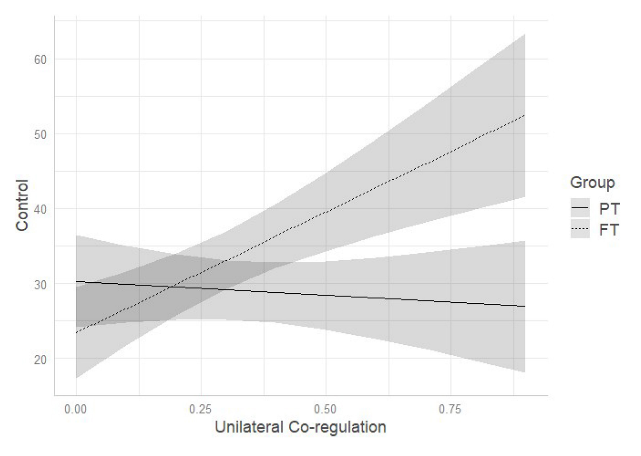



**FIGURE 2 |** Interaction among birth status and unilateral co-regulation on IDS control pragmatic sentences. PT, preterm; FT, full-term.

**TABLE 5 |** Multivariate analysis on IDS linguistic and pragmatic characteristics: Standardized estimated parameters of models 1 and 3, respectively.

| | Verbosity | *TTR* | MLU | Conversational | Control | Preverbal |
|---|---|---|---|---|---|---|
| | β (*p*) | β (*p*) | β (*p*) | β (*p*) | β (*p*) | β (*p*) |
| Birth status | 0.97 (<0.001) | −0.01 (0.586) | −0.50 (0.011) | −12.87 (0.013) | 13.67 (0.003) | −1.52 (0.672) |
| Symmetrical co-regulation | – | – | – | −30.36 (0.319) | 49.33 (0.068) | −32.64 (0.123) |
| Birth status × Symmetrical co-regulation | – | – | – | 23.91 (0.220) | −44.74 (0.010) | 27.44 (0.042) |

Standardized estimates of model 2 are reported in **Table 7**. All regression parameters were significant at $p < 0.05$ except for the role of birth status on TTR. Results showed that mothers of PT infants spoke less (verbosity: β = 1.03, $p < 0.001$), and demonstrated lower syntactic complexity (MLU: β = −0.56, $p = 0.003$) than mothers of FT infants. Moreover, for both PT and FT groups, mothers with higher levels of parenting stress spoke less (verbosity: β = −0.40, $p = 0.012$), and demonstrated higher lexical variability (TTR: β = 0.045, $p = 0.005$) and higher syntactic complexity (MLU: β = 0.443, $p = 0.008$).

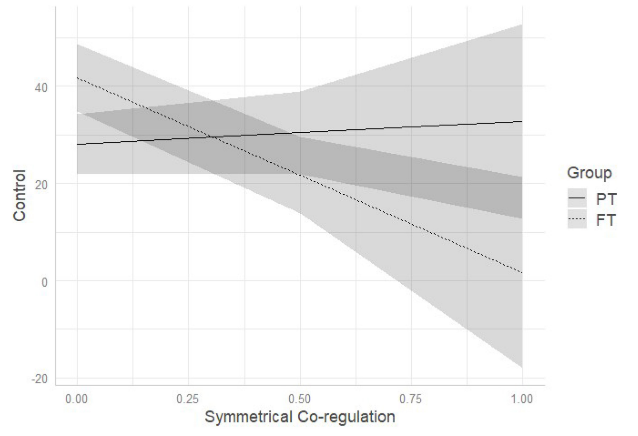



**FIGURE 3 |** Interaction among birth status and symmetrical co-regulation on IDS control pragmatic sentences. PT, preterm, FT, full-term.

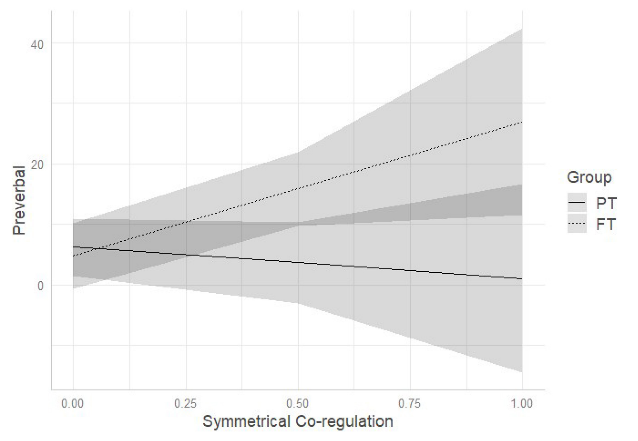



**FIGURE 4 |** Interaction among birth status and symmetrical co-regulation on IDS preverbal pragmatic sentences. PT, preterm, FT, full-term.

### Birth Status and Parenting Stress on Infant-Directed Speech Pragmatic Characteristics

Comparison of multivariate regression models (see **Table 6**) demonstrated that model 1 (which included birth status as single effect) outperformed the other models. Standardized estimates of model 1 are reported in **Table 7**. All regression parameters were significant at $p < 0.05$ except for the role of birth status on control sentences. Results showed that for mothers of PT infants, IDS was characterized by more conversational (β = −7.68, $p = 0.016$) and less preverbal (β = 4.52, $p = 0.045$) sentences compared with mothers of FT infants.

## DISCUSSION

The present study aimed to explore the intra- and inter-individual differences of the linguistic and pragmatic features of IDS, directed to 3-month-old full-term and preterm infants. We found evidence of interactive effects between individual (infant and maternal) and dyadic factors, on IDS characteristics. Recognizing that birth status (i.e., preterm birth) is an important condition affecting the infant, mother and dyad (Spinelli et al., 2013, 2016; Poehlmann-Tynan et al., 2015), we chose to examine patterns of IDS of mothers of PT and FT infants, and their associated predictors. It is well described that in the first year of life, PT infants manifest interactive difficulties, especially related to the communicative and regulatory aspects of infant's interaction with the environment (Goldberg and DiVitto, 1995; Sajaniemi et al., 1998). For this reason, mothers of PT infants are expected to adapt their interactive style so as not to over- or under-stimulate the infant (Feldman and Eidelman, 2007). Our findings regarding group differences evidenced that, during face-to-face

**TABLE 6 |** Model comparison, effects of birth status and parenting stress: AIC.

| | IDS linguistic | IDS pragmatic |
|---|---|---|
| **Model** | **AIC** | **AIC** |
| Model 1: Birth status | 262.51 | **1876.5** |
| Model 2: Birth status, parenting stress | **253.31** | 1877.4 |
| Model 3: Birth status, parenting stress, birth status × parenting stress | 256.26 | 1882.9 |

*In bold are highlighted models receiving more support for each set of outcome variables considered (IDS linguistic and IDS pragmatic).*

**TABLE 7 |** Multivariate analysis on IDS linguistic and pragmatic characteristics: Standardized estimated parameters of models 2 and 1, respectively.

| | Verbosity | *TTR* | MLU | Conversational | Control | Preverbal |
|---|---|---|---|---|---|---|
| | β *(p)* | β *(p)* | β *(p)* | β *(p)* | β *(p)* | β *(p)* |
| Birth status | 1.03 (<0.001) | −0.02 (0.346) | −0.56 (0.003) | −7.68 (0.016) | 3.66 (0.217) | 4.52 (0.045) |
| Parenting stress | −0.40 (0.012) | 0.04 (0.005) | 0.44 (0.008) | – | – | – |

interactions, mothers of PT infants spoke less, and vocalized with higher syntactic complexity, with the use of more conversational sentences and less preverbal sentences than mothers of FT infants. The linguistic and pragmatic features of IDS in mothers of PT infants are suggestive of a more complex pattern of IDS, evidenced by a less-talkative interactive style. This syntactic complexity is typically manifest in IDS directed to older infants, because this pattern of communication is more difficult to follow for a younger infant (Suttora and Salerni, 2011; Genovese et al., 2020). At 3 months, when infants have a limited ability to follow conversational exchanges, more complex sentences might result in less proto-conversational dyadic exchanges because the infant has less opportunities to vocalize in response to the sound of maternal voice. We have some possible explanations for these findings. Regarding PT dyads, one possibility is that this interactive style is related to the PT infant's communicative difficulties, resulting in a lower responsiveness to the mother's vocalizations. When experiencing less feedback from their PT infant, these mothers might therefore speak in a more complex way, because they don't expect a consistent participation of the infant during their vocal exchanges. At the same time, they leave more silent moments in which the infant has the space he/she need to respond to IDS stimulation. To better examine this possibility, future studies should consider the reciprocal influence of PT infants vocal and interactive responses to maternal IDS.

Regarding the role of dyadic co-regulation and IDS in PT versus FT infants, we found no differences between the groups concerning the quality of the dyadic, symmetrical and unilateral, co-regulation. Expressed differently, our results suggest that different patterns of IDS (between FT and PT) are not associated with fewer moments of co-regulated attention and affect. This is relevant because many studies considered the interactive qualities of PT mothers to be under- or over-stimulating, which was presumed to be suboptimal compared to what was observed in typical FT mothers (Forcada-Guex et al., 2006). However, the lack of differences in co-regulation between FT and PT groups suggests that the interactive vocal communication of PT mothers may be just as effective in contributing to the creation of dyadic shared moments. We hypothesize that this pattern of maternal vocalizations is a part of a specific interactive style that is attuned to the communicative and interactive abilities of PT infants, although this is an area in need of further research. While most studies have focused on difference between PT and FT dyads, with the aim to evaluate the adequacy with FT dyads as comparisons, future research should focus on describing the specific characteristics of PT dyads as probably mothers' adaptations to the specificities of premature birth condition, and on exploring within PT dyads differences in associations with later child development outcomes (Poehlmann-Tynan et al., 2015).

One notable difference emerged between the groups when exploring how mothers vary their IDS with respect to the time spent in co-regulated interaction. We observed that the duration of shared versus un-shared moments was associated with different pragmatic features of IDS in FT, but not PT, mothers. In FT dyads, the greater time spent in moments of un-shared attention and affect (i.e., unilateral co-regulation) was associated with an IDS characterized by reduced conversational and increased control sentences. Conversely, the greater time spent in shared moments of attention and affect (i.e., symmetrical co-regulation) was associated with increased use of preverbal, and decreased use of control sentences. Conversational sentences have, as its primary purpose, to keep open the communicative channels between two individuals when they are engaged in the same subject or are having a shared emotional experience. This is manifest by, for example, making comments, offering compliments, and asking open-ended questions. In contrast, control sentences are used to redirect, modify, and capture the attention of another when the other individual is focused on something different. This is manifested by, for example, calling or giving orders (Longobardi, 1992). Consistently, FT mothers who lose more the attention of their infants, reduce more the quantity of comments and open questions, and use a conversational style to try to elicit the infant's attention again. On the other hand, when experiencing more moments of co-regulated attention and affect, FT mothers tend to use fewer control sentences and more preverbal sentences, with IDS characterized by repeating the infant's vocalizations, singing, or making animal sounds. What we observed is that FT mothers demonstrate the ability to adapt their IDS to the quality of dyadic co-regulation, whereas this adaptability is not present in PT mothers. Of note, the pragmatic features of PT mothers' IDS did not vary according to dyadic co-regulation. One possible explanation for this finding is that PT mothers are less flexible in using verbal communication as an interactive modality to elicit or maintain infant attention, and preferentially use other interactive modalities (e.g., touch) instead (Wigley et al., 2022). This may be partially attributed to the more ambiguous and less frequent vocal feedback received from the infant. Before labeling this lack of variation of IDS as suboptimal, it would be helpful to observe whether this pattern of interaction is observed at other ages, when infants are expected to be more vocally interactive, and whether there are associations with later infant development outcomes. Further qualitative investigation may be needed on the comparison between PT and FT dyadic communication to better describe their specificities. Moreover, moment by moment analyses of IDS features as well as sequential analyses would help understanding how mothers adapt their IDS over time and according to changes in the quality of co-regulation patterns.

Consistent with previous studies, we did not find differences in parenting stress between the two groups (Gray et al., 2012; Suttora et al., 2020b). The experience of a preterm delivery, even if it was potentially traumatic for mothers, did not result in higher self-reported parenting stress in mothers of PT infants compared to mothers of FT infants at 3 months of age (Gray et al., 2012). When exploring the effect of parenting stress and birth status on IDS, the multivariate models demonstrated no interactive effects of stress with birth status, suggesting that the perceived parenting stress has similar effects in FT and PT groups. In both FT and PT groups, mothers who reported perceiving their parenting role as a stressful experience, demonstrated lower verbosity, higher lexical variability, and higher syntactic complexity. This less simple IDS is more typical of conversations directed to adults or to older children and might be considered a lower ability to connect and attune to the infant needs and communicative abilities (Genovese et al., 2020). Parenting stress might therefore be considered a maternal wellbeing risk factor which affects the quality of linguistic input. Long-term consequences of this effect should be examined in future studies. Since the linguistic characteristics of IDS have been associated with infants' and children's language development (Soderstrom, 2007), this raises the possibility that higher levels of parenting stress may also reduce the positive impact of IDS on language development.

We would like to acknowledge some limitations of the current study. First, while our results were relevant, our sample sizes were not big and we included only 3-month-old infants. A larger sample followed longitudinally would have allowed us to explore additional interactive effects. An additional limitation is that our PT sample was quite homogeneous, composed of low-risk PT infants and of well educated low-risk mothers. Additional research should examine these associations in a more at-risk population of preterms to identify differential effects both at maternal, infant, and dyadic levels. With a larger and more at-risk sample, also the associations of IDS with the other co-regulation patterns, i.e., asymmetrical and unengaged, could be explored. Further studies should also consider the paternal dyadic communication and explore if these findings are replicable in father-infant dyads. Lastly, we did not explore infant vocal communication during the interaction. This information, as well as the inclusion of non-verbal infant and maternal cues, would help interpret our findings and should be the focus of future studies.

Despite these limits, this study presents several strengths. To the best of our knowledge, this is the first study to examine infant, maternal, and dyadic factors and their associations with characteristics of IDS. This study highlights the need to go beyond exploring IDS effects on language development, and to consider its potential importance when exploring the quality of dyadic interaction and its role in sharing attention, affect and meaning between the mother and the infant (Saint-Georges et al., 2013). Our results could be useful in structuring interventions aimed to promote PT dyads quality of interaction. Knowing the specific characteristic of PT mothers' vocal communication could help defining more appropriated and well-designed interventions by helping mothers adapt their IDS to the specificities of infant communicative abilities in order to promote positive linguistic, attentive and affective outcomes.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the Department of Neuroscience, Imaging and Clinical Sciences of the University G. d'Annunzio Chieti-Pescara. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## AUTHOR CONTRIBUTIONS

MS and MF conceptualized the study and organized the data collection. MS and FL wrote the first draft of the manuscript. FL ran the analyses and wrote the results section. MG, ML, and SP conducted the data collection and coding. SD and PC coordinated preterm data collection. MF and PS revised the manuscript. All authors contributed to revision of the final version of the manuscript.

## FUNDING

This work was partially supported with PON-AIM funding (1811283-3) and by the "Departments of Excellence 2018-2022" initiative of the Italian Ministry of Education, University and Research for the Department of Neuroscience, Imaging and Clinical Sciences of the G. d'Annunzio University of Chieti-Pescara.

## ACKNOWLEDGMENTS

We thank the families who participated in the study and the students who contributed to the data collection.

## REFERENCES


Abidin, R. R. (1995). *The Parenting Stress Index Professional Manual 3rd ed*. Odessa, FL: Psychological Assessment Resources.

Adams, K. A., Marchman, V. A., Loi, E. C., Ashland, M. D., Fernald, A., and Feldman, H. M. (2018). Caregiver talk and medical risk as predictors of language outcomes in full term and preterm toddlers. *Child Dev.* 89, 1674–1690. doi: 10.1111/cdev.12818

Agostini, F., Neri, E., Dellabartola, S., Biasini, A., and Monti, F. (2014). Early interactive behaviours in preterm infants and their mothers: influences of maternal depressive symptomatology and neonatal birth weight. *Infant Behav. Dev.* 37, 86–93. doi: 10.1016/j.infbeh.2013.12.003

Bhutta, A. T., Cleves, M. A., Casey, P. H., Cradock, M. M., and Anand, K. (2002). Cognitive and behavioral outcomes of school-aged children who were born preterm: a meta-analysis. *JAMA* 288, 728–737. doi: 10.1001/jama.288.6.728

Bilgin, A., and Wolke, D. (2015). Maternal sensitivity in parenting preterm children: a meta-analysis. *Pediatrics* 136, e177–e193. doi: 10.1542/peds.2014-3570

Bornstein, M. H., Tal, J., Rahn, C., Galperín, C. Z., Pêcheux, M.-G., Lamour, M., et al. (1992). Functional analysis of the contents of maternal speech to infants of 5 and 13 months in four cultures: Argentina, France, Japan, and the United States. *Dev. Psychol.* 28:593. doi: 10.1037/0012-1649.28.4.593

Brummelte, S., Grunau, R. E., Synnes, A. R., Whitfield, M. F., and Petrie-Thomas, J. (2011). Declining cognitive development from 8 to 18 months in preterm children predicts persisting higher parenting stress. *Early Hum. Dev.* 87, 273–280. doi: 10.1016/j.earlhumdev.2011.01.030

Broen, P. A. (1972). *The Verbal Environment of the Language-Learning Child. ASHA Monographs, No. 17.* Washington, DC: American Speech-Language-Hearing Association.

Clark, C. A., Woodward, L. J., Horwood, L. J., and Moor, S. (2008). Development of emotional and behavioral regulation in children born extremely preterm and very preterm: biological and social influences. *Child Dev.* 79, 1444–1462. doi: 10.1111/j.1467-8624.2008.01198.x

Coppola, G., Cassibba, R., and Costantini, A. (2007). What can make the difference?: Premature birth and maternal sensitivity at 3 months of age: the role of attachment organization, traumatic reaction and baby's medical risk. *Infant Behav. Dev.* 30, 679–684. doi: 10.1016/j.infbeh.2007.03.004

De Schuymer, L., De Groote, I., Striano, T., Stahl, D., and Roeyers, H. (2011). Dyadic and triadic skills in preterm and full term infants: a longitudinal study in the first year. *Infant Behav. Dev.* 34, 179–188. doi: 10.1016/j.infbeh.2010.12.007

Feldman, R., and Eidelman, A. I. (2007). Maternal postpartum behavior and the emergence of infant–mother and infant–father synchrony in preterm and full-term infants: the role of neonatal vagal tone. *Dev. Psychobiol.* 49, 290–302. doi: 10.1002/dev.20220

Ferguson, C. A. (1964). Baby talk in six languages. *Am. Anthropol.* 66, 103–114. doi: 10.2307/668164

Fernald, A., and Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Dev. Psychol.* 20, 104–113. doi: 10.1037/0012-1649.20.1.104

Fogel, A., and Thelen, E. (1987). Development of early expressive and communicative action: reinterpreting the evidence from a dynamic systems perspective. *Dev. Psychol.* 23:747. doi: 10.1037/0012-1649.23.6.747

Fogel, A., De Koyer, I., Secrist, C., Sipherd, A., Hafen, T., and Fricke, M. (2003). *The Revised Relational Coding System*. Salt Lake City, UT: University of Utah.

Forcada-Guex, M., Pierrehumbert, B., Borghini, A., Moessinger, A., and Muller-Nix, C. (2006). Early dyadic patterns of mother–infant interactions and outcomes of prematurity at 18 months. *Pediatrics* 118, e107–e114. doi: 10.1542/peds.2005-1145

Garcia Coll, C. T., Halpern, L. F., Vohr, B. R., Seifer, R., and Oh, W. (1992). Stability and correlates of change of early temperament in preterm and full-term infants. *Infant Behav. Dev.* 15, 137–153. doi: 10.1016/0163-6383(92)80020-u

Genovese, G., Spinelli, M., Lauro, L. J. R., Aureli, T., Castelletti, G., and Fasolo, M. (2020). Infant-directed speech as a simplified but not simple register: a longitudinal study of lexical and syntactic features. *J. Child Lang.* 47, 22–44. doi: 10.1017/S0305000919000643

Goldberg, S. (1978). Prematurity: effects on parent-infant interaction. *J. Pediatr. Psychol.* 3, 137–144. doi: 10.1093/jpepsy/3.3.137

Goldberg, S., and DiVitto, B. (1995). "Parenting children born preterm," in *Handbook of Parenting Vol 1 Children and Parenting*, ed. M. H. Bornstein (Mahwah, NJ: Lawrence Erlbaum Associates, Inc), 328–354.

Gray, P. H., Edwards, D. M., O'Callaghan, M. J., and Cuskelly, M. (2012). Parenting stress in mothers of preterm infants during early infancy. *Early Hum. Dev.* 88, 45–49. doi: 10.1016/j.earlhumdev.2011.06.014

Henning, A., Striano, T., and Lieven, E. V. (2005). Maternal speech to infants at 1 and 3 months of age. *Infant Behav. Dev.* 28, 519–536.

Hughes, M. B., Shults, J., McGrath, J., and Medoff-Cooper, B. (2002). Temperament characteristics of premature infants in the first year of life. *J. Dev. Behav. Pediatr.* 23, 430–435. doi: 10.1097/00004703-200212000-00006

Johnson, W. (1944). Studies in language behavior: I. A program of research. *Psychol. Monogr.* 56, 1–15. doi: 10.1037/h0093508

Kitamura, C., and Burnham, D. (2003). Pitch and communicative intent in mother's speech: adjustments for age and sex in the first year. *Infancy* 4, 85–110. doi: 10.1207/s15327078in0401_5

Kitamura, C., Thanavishuth, C., Burnham, D., and Luksaneeyanawin, S. (2001). Universality and specificity in infant-directed speech: pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant Behav. Dev.* 24, 372–392. doi: 10.1016/S0163-6383(02)00086-3

Ko, E.-S. (2012). Nonlinear development of speaking rate in child-directed speech. *Lingua* 122, 841–857. doi: 10.1016/j.lingua.2012.02.005

Larroque, B., N'Guyen The Tich, S., Guedeney, A., Marchand, L., Burguet, A., and Group, T. E. S. (2005). Temperament at 9 months of very preterm infants born at less than 29 weeks' gestation: the epipage study. *J. Dev. Behav. Pediatr.* 26, 48–55.

Lavelli, M., and Fogel, A. (2013). Interdyad differences in early mother–infant face-to-face communication: real-time dynamics and developmental pathways. *Dev. Psychol.* 49:2257. doi: 10.1037/a0032268

Loi, E. C., Vaca, K. E., Ashland, M. D., Marchman, V. A., Fernald, A., and Feldman, H. M. (2017). Quality of caregiver-child play interactions with toddlers born preterm and full term: antecedents and language outcome. *Early Hum. Dev.* 115, 110–117. doi: 10.1016/j.earlhumdev.2017.10.001

Longobardi, E. (1992). Funzione comunicativa del comportamento materno e sviluppo comunicativo-linguistico del bambino nel secondo anno di vita. *G. Ital. Psicol.* 3, 425–440.

MacWhinney, B. (2000). *The CHILDES Project: The Database*, Vol. 2. Hove: Psychology Press.

Neri, E., Agostini, F., Perricone, G., Morales, M. R., Biasini, A., Monti, F., et al. (2017). Mother-and father-infant interactions at 3 months of corrected age: the effect of severity of preterm birth. *Infant Behav. Dev.* 49, 97–103. doi: 10.1016/j.infbeh.2017.08.001

Olafsen, K. S., Rønning, J. A., Kaaresen, P. I., Ulvund, S. E., Handegård, B. H., and Dahl, L. B. (2006). Joint attention in term and preterm infants at 12 months corrected age: the significance of gender and intervention based on a randomized controlled trial. *Infant Behav. Dev.* 29, 554–563. doi: 10.1016/j.infbeh.2006.07.004

Phillips, J. R. (1973). Syntax and vocabulary of mothers' speech to young children: age and sex comparisons. *Child Dev.* 44, 182–185.

Poehlmann-Tynan, J., Gerstein, E. D., Burnson, C., Weymouth, L., Bolt, D. M., Maleck, S., et al. (2015). Risk and resilience in preterm children at age 6. *Dev. Psychopathol.* 27, 843–858. doi: 10.1017/S095457941400087X

Reissland, N., and Stephenson, T. (1999). Turn-taking in early vocal interaction: a comparison of premature and term infants' vocal interaction with their mothers. *Child Care Health Dev.* 25, 447–456. doi: 10.1046/j.1365-2214.1999.00109.x

Reissland, N., Shepherd, J., and Stephenson, T. (1999). Maternal verbal interaction in different situations with infants born prematurely or at term. *Infant Child Dev. Int. J. Res. Pract.* 8, 39–48. doi: 10.1002/(sici)1522-7219(199903)8:1<39::aid-icd189>3.0.co;2-#

Saint-Georges, C., Chetouani, M., Cassel, R., Apicella, F., Mahdhaoui, A., Muratori, F., et al. (2013). Motherese in interaction: at the cross-road of emotion and cognition?(A systematic review). *PLoS One* 8:e78103. doi: 10.1371/journal.pone.0078103

Sajaniemi, N., Salokorpi, T., and Von Wendt, L. (1998). Temperament profiles and their role in neurodevelopmental assessed preterm children at two years of age. *Eur. Child Adolesc. Psychiatry* 7, 145–152. doi: 10.1007/s007870050060

Salerni, N., Suttora, C., and D'Odorico, L. (2007). A comparison of characteristics of early communication exchanges in mother-preterm and mother-full-term infant dyads. *First Lang.* 27, 329–346. doi: 10.1177/0142723707081654

Sansavini, A., Zavagli, V., Guarini, A., Savini, S., Alessandroni, R., and Faldella, G. (2015). Dyadic co-regulation, affective intensity and infant's development at 12 months: a comparison among extremely preterm and full-term dyads. *Infant Behav. Dev.* 40, 29–40. doi: 10.1016/j.infbeh.2015.03.005

Saxton, M. (2009). "The inevitability of child directed speech," in *Language Acquisition*, ed. S. Foster-Cohen (Berlin: Springer), 62–86. doi: 10.1057/9780230240780_4

Snow, C. E. (1972). Mothers' speech to children learning language. *Child Dev.* 43, 549–565.

Soderstrom, M. (2007). Beyond babytalk: re-evaluating the nature and content of speech input to preverbal infants. *Dev. Rev.* 27, 501–532. doi: 10.1016/j.dr.2007.06.002

Soderstrom, M., Blossom, M., Foygel, R., and Morgan, J. L. (2008). Acoustical cues and grammatical units in speech to two preverbal infants. *J. Child Lang.* 35, 869–902. doi: 10.1017/S0305000908008763

Spinelli, M., Fasolo, M., and Mesman, J. (2017). Does prosody make the difference? A meta-analysis on relations between prosodic aspects of infant-directed speech and infant outcomes. *Dev. Rev.* 44, 1–18. doi: 10.1016/j.dr.2016.12.001

Spinelli, M., Frigerio, A., Montali, L., Fasolo, M., Spada, M. S., and Mangili, G. (2016). 'I still have difficulties feeling like a mother': the transition to motherhood of preterm infants mothers. *Psychol. Health* 31, 184–204. doi: 10.1080/08870446.2015.1088015

Spinelli, M., Poehlmann, J., and Bolt, D. (2013). Predictors of parenting stress trajectories in premature infant–mother dyads. *J. Fam. Psychol.* 27:873. doi: 10.1037/a0034652

Stern, D. N., Spieker, S., Barnett, R. K., and MacKain, K. (1983). The prosody of maternal speech: infant age and context related changes. *J. Child Lang.* 10, 1–15. doi: 10.1017/S0305000900005092

Suttora, C., and Salerni, N. (2011). Maternal speech to preterm infants during the first 2 years of life: stability and change. *Int. J. Lang. Commun. Disord.* 46, 464–472. doi: 10.1111/j.1460-6984.2011.00007.x

Suttora, C., Guarini, A., Zuccarini, M., Aceti, A., Corvaglia, L., and Sansavini, A. (2020a). Speech and language skills of low-risk preterm and full-term late talkers: the role of child factors and parent input. *Int. J. Environ. Res. Public Health* 17:7684. doi: 10.3390/ijerph17207684

Suttora, C., Spinelli, M., Aureli, T., Fasolo, M., Lionetti, F., Picciolini, O., et al. (2020b). Mind-mindedness and parenting stress: a cross-sectional study in a cohort of mothers of 3-month-old full-term and preterm infants. *Int. J. Environ. Res. Public Health* 17:7735. doi: 10.3390/ijerph17217735

Wigley, I. L. C. M., Mascheroni, E., Bonichini, S., and Montirosso, R. (2022). Epigenetic protection: maternal touch and DNA-methylation in early life. *Curr. Opin. Behav. Sci.* 43, 111–117. doi: 10.1016/j.cobeha.2021.09.004

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 13 September 2022
DOI 10.3389/fpsyg.2022.805447



OPEN ACCESS

EDITED BY
Maria Spinelli,
University of Studies G. d'Annunzio Chieti and Pescara, Italy

REVIEWED BY
Azza Al-Kendi,
Sultan Qaboos University, Oman
Monja Knoll,
University of Edinburgh,
United Kingdom

*CORRESPONDENCE
Hannah S. Sarvasy
h.sarvasy@westernsydney.edu.au

SPECIALTY SECTION
This article was submitted to Developmental Psychology, a section of the journal Frontiers in Psychology

RECEIVED 30 October 2021
ACCEPTED 30 June 2022
PUBLISHED 13 September 2022

CITATION
Sarvasy HS, Li W, Elvin J and Escudero P (2022) Vowel acoustics of Nungon child-directed speech, adult dyadic conversation, and foreigner-directed monologues. *Front. Psychol.* 13:805447.
doi: 10.3389/fpsyg.2022.805447



# Vowel acoustics of Nungon child-directed speech, adult dyadic conversation, and foreigner-directed monologues

Hannah S. Sarvasy[1,2]*, Weicong Li[1,2], Jaydene Elvin[3] and Paola Escudero[1,2]

[1]MARCS Institute for Brain, Behaviour, and Development, Western Sydney University, Penrith, NSW, Australia, [2]Australian Research Council (ARC) Centre of Excellence for the Dynamics of Language, Canberra, ACT, Australia, [3]Department of Linguistics, California State University, Fresno, CA, United States

In many communities around the world, speech to infants (IDS) and small children (CDS) has increased mean pitch, increased pitch range, increased vowel duration, and vowel hyper-articulation when compared to speech directed to adults (ADS). Some of these IDS and CDS features are also attested in foreigner-directed speech (FDS), which has been studied for a smaller range of languages, generally major national languages, spoken by millions of people. We examined vowel acoustics in CDS, conversational ADS, and monologues directed to a foreigner (possible FDS, labeled MONO here) in the Towet dialect of the Papuan language Nungon, spoken by 300 people in a remote region in northeastern Papua New Guinea. Previous work established that Nungon CDS entails optional use of consonant alteration, special nursery vocabulary, and special morphosyntax. This study shows that Nungon CDS to children aged 2;2–3;10 lacks vowel hyper-articulation, but still displays other common prosodic traits of CDS styles around the world: increased mean pitch and pitch range. A developmental effect was also attested, in that speech to 2-year-olds contained vowels that were significantly longer than those in speech to 3-year-olds, which in turn had vowels of similar duration to those in Nungon ADS. We also found that Nungon FDS vowel triangles, measured from monologues primarily directed to a non-native speaker, were significantly larger than those of either CDS or conversational ADS, indicating vowel hyper-articulation. The Nungon pattern may align with the patterns of vowels in Norwegian IDS, CDS, and FDS, where hyper-articulation is found in FDS, but not CDS or IDS. The languages of the New Guinea area constitute 20% of the world's languages, but neither an acoustic comparison of vowels in CDS and ADS, nor an acoustic study of FDS, has previously been completed for any language of New Guinea. The function of an FDS style in a small, closed community like those of much of New Guinea may differ from that in larger societies, since there are very few non-native speakers of Nungon. Thus, this study uses monologues recorded with a foreign researcher as interlocutor to study Nungon FDS.

KEYWORDS

**Nungon, acoustics, vowel, child-directed speech, foreigner-directed speech, hypo-articulation, hyper-articulation, prosody**

## Introduction

In many communities around the world, speech directed at infants (IDS) and young children (CDS) involves special acoustic and prosodic features, compared with adult-directed speech (ADS). Among the special acoustic and prosodic features frequently attested in IDS/CDS styles are: increased mean pitch and increased pitch range (Fernald et al., 1989), longer vowel durations (Swanson et al., 1992 found this especially for English content words, not function words), and vowel hyper-articulation, usually understood to involve an expanded vowel space (most often calculated using the first and second formant frequencies of the vowels /i/, /a/, and /u/; Kuhl et al., 1997; Burnham et al., 2002; Uther et al., 2007; Lam and Kitamura, 2008). These acoustic and prosodic modifications may change with children's age and development, as seen, for instance, in differences in caregiver vowel spaces in Cantonese IDS to children of different ages (Stern et al., 1983; Kitamura et al., 2002; Kitamura and Burnham, 2003; Englund and Behne, 2006; Rattanasone et al., 2013). The magnitude of increase in CDS mean pitch and pitch range, relative to ADS, may decrease after the first year or two of life (Garnica, 1977; Stern et al., 1983; Warren-Leubecker and Bohannon, 1984), but Warren-Leubecker and Bohannon (1984) found that American English-speaking mothers (but not fathers) still used elevated pitch in CDS to 5-year-old children.

Perhaps the most controversial proposed feature of IDS/CDS, with the greatest number of counter-examples in the literature, is vowel hyper-articulation (Cristia and Seidl, 2014). The IDS/CDS vowel space has been found to be larger than in adult-directed speech (ADS) for: American, Australian, and British English (Kuhl et al., 1997; Burnham et al., 2002; Uther et al., 2007; but see Green et al., 2010), Russian (Kuhl et al., 1997), Mandarin (Liu et al., 2003), Spanish (García-Sierra et al., 2021), Swedish (Kuhl et al., 1997; but see Van de Weijer, 2001), and Japanese (Andruski et al., 1999; but see Martin et al., 2015; Miyazawa et al., 2017). However, for other languages, not only has no expansion of the vowel space in IDS/CDS relative to ADS been demonstrated, but rather a reduction of the vowel space has been shown. For instance, Rattanasone et al. (2013) found that the vowel triangle (formed from the three "corner" vowels, /i/, /a/, and /u/) for Cantonese speech addressed to 3-month-old infants was significantly smaller than that for ADS, suggesting hypo-articulation of IDS vowels at that stage. A marked reduction of the vowel space in IDS/CDS relative to ADS is also attested for Dutch (Benders, 2013), German (Audibert and Falk, 2018), and Norwegian (Englund and Behne, 2006).

Postulated functions of the special acoustic and prosodic features of IDS and CDS can be divided into three main categories (Grieser and Kuhl, 1988; Cooper et al., 1997; Singh et al., 2002; Uther et al., 2007): (a) obtaining infants'/children's attention (especially through expanded pitch range; Fernald and Simon, 1984); (b) expressing positive affect and establishing an emotional bond (especially through increased mean pitch: Werker et al., 1994; Trainor et al., 2000; Singh et al., 2002); (c) aiding children in the task of language learning (especially through vowel hyper-articulation: Kemler Nelson et al., 1989; Singh et al., 2008; Song et al., 2010). Further investigation of vowel hyper-articulation has demonstrated that its application often relates to the speaker's perception of the listener's linguistic abilities (Burnham et al., 2010; Rice and Burnham, 2010; Lam and Kitamura, 2012); indeed, Xu et al. (2013) found that vowel space area in the speech of Australian English-speaking mothers increased from ADS to an unfamiliar adult to speech directed to a dog, to speech directed to a parrot with the ability to repeat speech, to IDS.

Burnham et al. (2002) and Uther et al. (2007) explored the possibility that a speaker's relationship with various types of interlocutors, including children and unfamiliar adults, but also foreign adults and even pets, could predict which types of special acoustic and prosodic features were applied. Burnham et al. (2002) showed that Australian English-speaking mothers addressed their pets and infants with similar degrees of increased pitch and affect, compared with when they addressed unfamiliar adults, but that only the mothers' speech to infants (not to pets or unfamiliar adults), featured vowel hyper-articulation (presumably a didactic feature of IDS). Uther et al. (2007) followed Burnham et al. (2002), but replaced pets with non-native English speakers. They showed that British English-speaking mothers addressed infants and foreigners with similar degrees of vowel hyper-articulation (presumably for didactic effect), but that speech to foreigners (FDS) lacked the increased mean pitch of IDS, indicating an absence of affective prosodic modification, relative to speech to native-speaker adults.

Indeed, since at least the 1970s, IDS and CDS have been compared with foreigner-directed speech styles (FDS: Ferguson, 1975; Freed, 1981a,b), and some authors even considered FDS to "derive" from IDS/CDS (DePaulo and Coleman, 1986). Similar acoustic and prosodic features have been claimed for FDS as for IDS/CDS, including longer vowel durations and/or slower speech rates, and vowel hyper-articulation. These features are also often claimed for a more general "clear speech" style that speakers may produce in noisy environments, when interacting with the hearing-impaired, when reading out loud, and when asked to enunciate clearly (Uchanski, 2005). FDS in both spoken and signed languages has further been described as involving fewer sandhi effects than speech to native-speaking adults (Henzl, 1979; Swisher, 1984; Tweissi, 1990). Gesture accompanying spoken French and Dutch FDS in Belgium has been shown to involve modifications, relative to gesture to native-speaking adults (Provė et al., 2022). But the acoustic and prosodic properties of FDS have not been studied comprehensively for all languages for which IDS/CDS data are available.

Comparisons of FDS with speech to native-speaking adults have shown slower FDS speech rates, sometimes with more pauses, for: English (Henzl, 1979; Ulichny, 1979; Warren-Leubecker and Bohannon, 1982; Wesche and Ready, 1985; Bobb et al., 2019; but see Arthur et al., 1980; Kühnert and Antolík, 2017), Czech (Henzl, 1973, 1979), French (Kühnert and Antolík, 2017, but see Wesche and Ready, 1985), German (Henzl, 1979), and Jordanian Arabic (Tweissi, 1990). Several studies found that English foreigner-directed speech (in the U.S., Scotland and England) featured vowel hyper-articulation, relative to speech directed to native-speaker adults (Knoll and Scharrer M., 2007; Scarborough et al., 2007; Uther et al., 2007; Knoll et al., 2009; Hazan et al., 2015; Bobb et al., 2019), but other studies on English did not yield this result (Knoll and Scharrer M. A., 2007; Knoll et al., 2015). In Norwegian (Sikveland, 2006) and Omani Arabic (Al-Kendi and Khattab, 2019; Al-Kendi, 2020, pp. 221–233), FDS has been found to feature increased first formant (F1) values for corner vowels, indicating a more open mouth and/or increased vocal effort (Ferguson and Kewley-Port, 2002). Exploration of pitch modifications in FDS compared with native-speaker ADS has yielded mixed results: from no modifications in British English FDS (Uther et al., 2007) to increased mean pitch in Omani Arabic FDS (Al-Kendi and Khattab, 2019).

Several studies have compared acoustic and prosodic features of IDS or CDS with those of FDS and native-speaker-directed ADS, resulting in mixed findings. In an early study in which Mandarin speakers simulated speech to babies, foreigners, and native-speaker adults, pitch contours in the simulated IDS and FDS differed (Papoušek and Hwang, 1991). Uther et al. (2007) showed that British FDS lacked the increased mean pitch, wider pitch range, and longer vowel durations evinced by British IDS relative to ADS, but that IDS and FDS had similar degrees of vowel hyper-articulation (also found by Kangatharan, 2015). Bobb et al. (2019) found that simulated IDS and FDS by speakers of American English showed significant differences in mean pitch (IDS > FDS), but not pitch range. Biersack et al. (2005) found that British English speakers produced longer vowels when addressing imaginary children, but longer pauses, rather than vowels, when addressing imaginary foreign adults, but this finding was not replicated by Uther et al. (2007), using naturalistic data. Knoll and Scharrer M. A. (2007) were unable to replicate the vowel hyper-articulation finding of Uther et al. (2007) using simulated IDS, ADS, and FDS interactions by undergraduate native speakers of British English, but did successfully replicate Uther et al.'s findings when the simulation was by actresses (Knoll et al., 2009, 2011). Lorge and Katsos (2019) compared simulated recipe explanations by English monolinguals and bilinguals directed to a 10-year-old child, a native English speaker, and a non-native English speaker; CDS to the 10-year-old child featured higher mean pitch, increased pitch range, and vowel hyper-articulation than ADS, while FDS featured a slower speech rate than ADS. Only the bilinguals also used vowel hyper-articulation in FDS.

IDS and CDS styles that differ lexically, acoustically, and/or prosodically from ADS styles are widespread and found in diverse communities around the world. But it has been claimed that some communities employ no special speech styles when addressing infants and young children; one of these purported exceptions is the Kaluli speech community of Papua New Guinea (Schieffelin, 1990). The languages of the New Guinea region constitute 20% of languages in the world today (Palmer, 2017), and therein, those of Papua New Guinea constitute at least 10% of languages in the world (Kik et al., 2021). If the absence of a special IDS or CDS style is an areal feature, a sizable proportion of the world's speech communities could lack such a style. Recent research into child language development in other communities of Papua New Guinea (listed in Hellwig et al., Forthcoming) has successfully identified various special qualities of IDS or CDS, including nursery lexicon (Aikhenvald, 2008; Stebbins, 2011; Sarvasy, 2017, 2019), consonant alteration (Schieffelin and San Roque, forthcoming; Rumsey, 2017; Sarvasy, 2019) and pitch modification (Frye aus Schwerte, 2019). To date, however, IDS or CDS in any language of Papua New Guinea has not been thoroughly analyzed acoustically, relative to conversational ADS in the same language (Sarvasy et al., 2019 is a pilot study on which the current study builds). Indeed, since Kaluli adults are said to maintain that children should hear only well-formed speech, to learn to speak correctly (Schieffelin, 1990), it could be the case that Kaluli adults, although possibly not employing modified lexical or other grammatical features in speech to children, do practice vowel hyper-articulation, to ensure that children are exposed to proper speech sounds.

Most of the literature on FDS, including its acoustic and prosodic features, targets major world languages, spoken in industrialized countries by literate communities. The notion that vowel hyper-articulation is a basic feature of FDS remains to be validated with data from small speech communities without traditions of literacy, where daily life may involve interactions only with known members of a small, closed community (*cf.* Wray and Grace, 2007). The languages of Papua New Guinea, which represent an outsized portion of the world's total, are overwhelmingly spoken by fewer than 10,000 speakers each, often in this type of closed community (Kik et al., 2021). If people rarely interact with non-native speakers of their language, one might assume that they might not employ a uniform FDS style.

In this study, we examine acoustic features of vowels in CDS, ADS, and speech that may be termed FDS, since it was directed at a non-native speaker (but could also be considered to exemplify a more general story-telling performative style), in the Towet dialect of the Papuan language Nungon, spoken by about 300 people in remote Towet village in the Uruwa River valley in NE Papua New Guinea (Sarvasy, 2017).

Previous work established that Nungon CDS involves several optional features that differentiate it from ADS: nursery lexicon, consonant alteration, and special morphosyntax (Sarvasy, 2019, 2020, 2021a). These features are evident in speech to infants and in speech to children of up to 3 years and even older. But the degree to which Nungon CDS vowels can be said to differ acoustically from ADS was not examined in previous studies, which also did not examine prosodic features of CDS, relative to ADS. Case studies of child acquisition of Nungon morphosyntax established that morphosyntactic complexity (verbal inflections, length of complex sentences, use of complex predicates) in the speech of two children increased substantially from about 3 years of age. By the same measures, morphosyntactic complexity in CDS to the same children also increased from when the children were about 3 years old (Sarvasy, 2019, 2020, 2021a). Further, a special Nungon morphosyntactic alteration found only in young children's speech, IDS, and CDS (not ADS) declines markedly in frequency in maternal CDS to children of about 3 years of age or more (Sarvasy, 2019). Nungon-speaking parents thus seem to exhibit "fine-tuning"—adjusting their own speech to the child's perceived cognitive and linguistic sophistication (Bohannon III and Marquis, 1977; Soderstrom, 2007)—in the morphosyntactic domain. It is as yet unknown whether Nungon CDS to children under 3 years also features acoustic differences from Nungon CDS to children over 3 years.

Nungon speakers communicate in a world of intimates and classificatory relatives, without strangers (Wray and Grace, 2007). They rarely interact with non-native speakers of dialects of Nungon or of the closely related language Yau. A handful of people, mostly women, marry into the region from adjoining regions, where distantly related languages are spoken. Apart from the first author and the biologist Gabriel Porolak, no outsiders not married into local families have learned Nungon in at least the past two decades or so. Outsiders who travel fleetingly through the region usually speak the unrelated lingua franca Tok Pisin, which most Nungon speakers under about 40 understand and can speak.

Although true outsiders (people originating outside the Uruwa River valley) rarely learn Nungon, the linguistic situation among speakers of Nungon is complex. Each of the six Nungon-speaking villages has its own distinct dialect, with a particular accent, and which shares <90% of key vocabulary with other villages (Sarvasy, 2013, 2017). People marry into other villages, travel between them, and interact with speakers of other Nungon dialects on a regular basis. If Nungon has an identifiable FDS style, this could be rooted in modes of cross-dialect interaction, rather than norms for speaking Nungon to true foreigners (such as the first author, Gabriel Porolak, and the few in-married foreigners). Sarvasy (2017) also notes that there seem to be conventional ways to interact with Nungon speakers who have speech impediments or are intellectually disabled, which involve increased loudness of speech, increased use of conventionalized gestures, and possibly increased lip movements; this could represent another conventionalized Nungon clear speech style, potentially related to FDS.

We investigated vowel space size, vowel duration, mean pitch, and pitch range in Nungon CDS in the Towet village dialect to children aged 2;2–3;10. We compared the Nungon CDS acoustic results to results from dyadic adult Nungon conversations (conversational ADS) by the same speakers from the CDS data. We then compared these CDS and conversational ADS results to results for the same measures from Nungon monologues that had been recorded as semi-performances, with a non-native Nungon speaker (the first author) as primary listener ("MONO"; potential FDS). We aimed to evaluate the following, for the Towet village dialect of Nungon: (a) whether Nungon CDS vowels were hyper-articulated, relative to conversational ADS; (b) whether Nungon CDS featured increased mean pitch, expanded pitch range, longer vowel durations, and/or an enhanced contrast between phonologically short and phonologically long vowels, relative to ADS; and (c) how these acoustic measures patterned in the Nungon monologues with a non-native listener. Additionally, we evaluated (d) whether acoustic features of Nungon CDS vary with children's age, and whether (f) women and men differed in their use of CDS.

## Methods

Nungon has six phonemic vowels: an unrounded high front /i/, an unrounded mid front /e/, an unrounded low central /a/, a rounded high back /u/, a rounded mid-high back /o/ with extra lip protrusion, and a rounded mid-low back /ɔ/ (Sarvasy, 2017; Sarvasy et al., 2020). Nungon is slightly unusual typologically in having more phonemic back vowels than front vowels. All the Nungon vowels can occur as either phonologically short or long; this is lexically determined. Details on vowel trajectories, using a multi-point analysis technique, are in Sarvasy et al. (2020).

We examined vowel tokens from three corpora for this study: CDS, ADS, and what we term MONO (monologual narratives directed at a non-native Nungon speaker with foreign appearance). We targeted the first two vowel formant values (F1 and F2), duration, and mean and range of fundamental frequency (F0: an objective measurement that is closely related to pitch).

## CDS dataset

Child-directed speech was extracted from a corpus of child-caregiver conversation from a longitudinal study of eight children acquiring Towet Nungon (Sarvasy, 2021c). This study occurred in two parts, with two different cohorts. The first cohort of five children were recorded for 1 h monthly for 2 years between 2015 and 2017. The second cohort of three children were recorded for 4 h within a single week each month over a 5-month period in 2019. Parents were paid an honorarium of 1,000–1,500 Papua New Guinean kina for their participation in the study, plus a gift, and research assistants were paid for each recording session and transcription they completed.

In each recording session, the target child sat with one or both parents and, sometimes, a sibling, in a natural indoor or outdoor setting. The parents understood that the purpose of the recording sessions was to elicit speech from the child in a relatively natural manner, to be able to study how the child's language development transpired. Parents and children usually looked at picture books together, or discussed events or activities the child had or would participate in, punctuated by occasional discussion of what they observed from their seated location (people walking by, etc.). The sessions were audio- and video-recorded by a local research assistant, usually the classificatory aunt or uncle of the child. Interaction was videorecorded using a Canon IXUS 190 digital camera held by the research assistant or mounted on a tripod. Interaction was audio-recorded with a Zoom H5 recorder mounted on a tripod and pointed toward the target child. Recordings were done at a 44.1 kHz sampling rate, in WAV format. Recordings were transcribed in Mid-CHAT format (MacWhinney, 2000) by native Towet Nungon speakers on Lenovo laptops in Towet village, using solar power.

Twelve recording sessions, involving six target children (three girls and three boys) and one or both of their parents, were used for analysis here. Ten sessions were selected according to whether their digital transcriptions, originally completed by Nungon speakers in the villages, had been finalized and checked; the other two were selected because the adult interlocutors had previously also recorded a monologue in the MONO dataset, to be analyzed here (see below). At the time of recording, three children's ages were in the range 2;2–2;9 (a girl, TO, recorded at ages 2;2 and 2;9; a girl, MA, recorded at age 2;7, and a boy, AB, recorded at age 2;7), and the other three were aged 3;0–3;10 (a girl, AR, recorded at age 3;10; a boy, NI, recorded at ages 3;0, 3;1, and 3;2; and a boy, DA, recorded at ages 3;5, 3;6, 3;7, and 3;8). The adults whose speech was analyzed here were all in their twenties or thirties at the time of the recordings, and all were native speakers of the Towet Nungon dialect.

## Dyadic adult conversational dataset (ADS)

We commissioned six 15-min recordings of free conversation between two Towet Nungon-speaking adults each, specifying that the dyads should include parents from the CDS dataset. In the sessions, two adult Nungon speakers, usually classificatory relatives, sat close to each other and spoke about shared past experiences, or related anecdotes from their separate experiences. Note that because of the nature of these small communities (Wray and Grace, 2007), it is impossible to find speakers of the same Nungon dialect (of which there are 100-350 speakers) who are not classificatorily related to each other and do not know each other, so the usual practice in studies comparing CDS with a "baseline" ADS in which an adult addresses an unfamiliar adult was impossible to achieve here. The sessions were recorded using a Zoom H5 recorder mounted on a tripod, placed between the two speakers. Recordings were done at a 44.1 kHz sampling rate, in WAV format. All participants received an honorarium of 50 Papua New Guinean kina each, and the assistants who ran the recording sessions were also compensated for their time. All speakers but one who were recorded conversing here also feature in the CDS sample. The only speaker who does not also feature in the CDS dataset, DI, was in her thirties at the time of the recording.

## Adult performed monologues with non-native listener (MONO)

The final dataset from which vowel tokens were extracted involved monologual narratives that adult Towet Nungon speakers recorded individually between September 2011 and March 2013, with the first author, a non-native speaker of Nungon, as primary listener (wearing headphones and usually holding a digital recorder in one hand). The first author's fieldwork was performed monolingually (Sarvasy, 2016), meaning that she always spoke only Nungon to Nungon speakers, without recourse to a contact language such as the Papua New Guinean lingua franca Tok Pisin or English. Nungon speakers would thus have observed her language development over a series of field trips (generally 1.5–2.5 months in length), from very early stages in mid-2011 to near-fluency in 2013. The recordings used here were made at different stages of the author's linguistic development, but even for the earlier recordings, the author was fluent enough to be able to discuss the recording context and protocols with each speaker herself. Elsewhere, Sarvasy (2021b) notes that the degree of intimacy she had with a speaker could be a factor in determining the length and amount of detail in a recorded monologue.

The speakers usually framed each recording as a *hat* "story," and most had approached the first author in advance of the recording session to notify her that they intended to record one or more stories for her; the topics were either chosen by the speakers in this way, or were responses to questions from the first author regarding, for instance, life in the olden days. Speakers were aware that these recordings would support the first author's research into the Nungon language grammar (Sarvasy, 2017). They had further been warned by the local government Councillor, early in the first author's research into the grammar, that he expected community members to provide the author with only *maa orogo* "good language," by which he intended that *maa muyam* "cursing" should not be used, but which highlights a type of pressure some speakers might have felt to produce "good" speech.

Although we could categorize these recordings as FDS, simply because of the identity of the interlocutor, we will refer to them here as MONO because at least two other factors in the recording context could have led speakers to produce speech differently than in normal conversation. First, with the recording's purpose in mind, speakers could have aimed to speak clearly to produce a clear record for posterity. Second, the performative aspect of the recording context, in which speakers entered a hut with the express purpose of telling a narrative well, could have led to extra care in speech: they were conscious that they were performing a storytelling act, for audiences (present and future). Future work may attempt to separate these effects from the effect of non-native interlocutor in a controlled way, but we are unable to do this definitively with the present data.

Of a corpus of 221 such recorded and transcribed texts (Sarvasy, 2015), 15 were analyzed for the present study. All of the monologues analyzed were narratives describing personal experiences of the speakers, except for two local legends, and a personal introduction, where the speaker described her family origin. These were recorded in close range using the built-in twin microphones of a Zoom H4N Handy audio recorder with no external microphone, at a 44.1 kHz sampling rate, in WAV format. Speakers were paid 10–20 Papua New Guinean kina for a storytelling session. These were digitally transcribed by the first author in close consultation with each speaker him- or herself in Towet village in 2011–2013.

## Selection of tokens and acoustic analysis

In each dataset, transcriptions and/or audio recordings were searched for words that included the six phonemic Nungon vowels in word-initial, word-medial and word-final environments, and not adjacent to nasals (to avoid coarticulation and prosodic effects). While the ideal was to find the same words (e.g., *agep* "firm" for the vowels /a/ and /e/) in all corpora and uttered by all speakers, the uncontrolled nature of the corpora made this difficult, so the words extracted varied slightly from speaker to speaker. If tokens of a word exceeded 20 in a single session by one speaker, only the first 20 tokens were used. The corresponding audio was hand-screened for obviously poor recording quality and obscuring background noise. Tokens of all six vowels were used to examine mean pitch and pitch range, and vowel duration; tokens of the corner vowels /i/, /a/, and /u/ were used to determine vowel space size, following the method in García-Sierra et al. (2021).

For the CDS dataset, we selected adult vowel tokens from the 12 sample child-caregiver interaction recordings, coded according to the identity and sex of the adult. In analyses, we do not differentiate between parents and the three Nungon-speaking assistants who occasionally interact with the children in the recording sessions (LY, JA and ST in Table 1). The reason for this is that the assistants were classificatory close relatives of the children and habitually interacted with them, occasionally caring for them, in the close-knit, 30-household, Towet village community, where child-rearing has a communal character. In selecting vowel tokens, we avoided tokens that were adjacent to nasal segments, since Sarvasy et al. (2020) showed that coarticulation effects are present in Nungon vowels adjacent to nasals. We also tried to extract vowels from a fixed set of words, as much as possible, for all speakers. For instance, the word *agep* "firm, loud" occurs frequently in these recordings, addressed to the target child, when the parent wants the child to speak louder. This was a good source for vowel tokens of /a/ and /e/. Table 1 shows the CDS vowel tokens used. As mentioned above, Nungon distinguishes lexically determined phonological vowel length contrasts, such that each vowel takes both a phonologically short and long form. These length contrasts are not shown in Tables 1–3.

Table 1 shows that one woman and three men addressed the 2-year-olds (2;2–2;9), while four women and two men addressed 3-year-olds (3;0–3;10). We will compare features of CDS to these two different age groups in the Results.

For the ADS dataset, sections of the conversations were transcribed by the first author. From these sections, we extracted a subset of vowel tokens not adjacent to nasal consonants for analysis. Table 2 gives the number of vowel tokens per speaker used from the conversational ADS dataset here.

Finally, for the MONO dataset, vowel tokens not adjacent to nasal segments were extracted. Five of the speakers in the MONO dataset also feature in the CDS and ADS corpora. Of the remaining three, one, NK, was in her twenties at the time of the recording, one, RO, was in her forties, and the third, OR, was in her late forties or early fifties.

Comparison of Tables 1–3 shows that, while eight of nine ADS speakers also feature in the CDS dataset, only five MONO speakers feature in the CDS dataset. Of these, one (ST) does not feature in the conversational ADS dataset, yielding just two male and two female speakers who feature in all three corpora. For this reason, we performed three sets of analyses: one using just tokens from the four speakers who featured in all three datasets (comparison of CDS, ADS, and MONO); one using tokens from the seven speakers who featured in both CDS and ADS datasets (comparison of just CDS and ADS); and one with all 14 speakers (comparison of CDS, ADS, and MONO); results will be summarized for the three groups separately in Table 8.

For CDS and MONO, existing transcriptions were manually aligned at utterance level before the use of Munich Automatic Segmentation System (MAUS) for forced alignment at the word and phonetic level. WebMAUS (Kisler et al., 2017) was used following two steps: grapheme-to-phoneme conversion,

TABLE 1 Number of tokens for each vowel, by speaker, in the CDS dataset.

| Speaker | Gender | Age of child | /i/ | /e/ | /a/ | /ɔ/ | /o/ | /u/ |
|---|---|---|---|---|---|---|---|---|
| YI | Female | 3;5–3;8 | 29 | 37 | 66 | 53 | 22 | 21 |
| LY | Female | 3;5–3;10 | 5 | 5 | 32 | 13 | 9 | 11 |
| NU | Female | 3;1–3;2 | 11 | 44 | 76 | 100 | 24 | 30 |
| AM | Female | 3;10 | 7 | 33 | 37 | 10 | 8 | 15 |
| TM | Female | 2;2, 2;9 | 74 | 52 | 92 | 76 | 14 | 25 |
| DE | Male | 2;7 | 17 | 25 | 49 | 33 | 6 | 19 |
| BO | Male | 2;7 | 8 | 11 | 24 | 21 | 2 | 16 |
| MA | Male | 2;9 | 14 | 8 | 20 | 33 | 6 | 7 |
| JA | Male | 3;5–3;6 | 7 | 6 | 29 | 51 | 6 | 1 |
| ST | Male | 3;0–3;10 | 11 | 17 | 50 | 33 | 4 | 15 |
| Total | | | 183 | 238 | 475 | 423 | 101 | 160 |

TABLE 2 Number of tokens for each vowel, by speaker, in the conversational ADS dataset.

| Speaker | Gender | /i/ | /e/ | /a/ | /ɔ/ | /o/ | /u/ |
|---|---|---|---|---|---|---|---|
| YI* | Female | 21 | 19 | 50 | 85 | 19 | 13 |
| LY* | Female | 23 | 25 | 53 | 45 | 10 | 15 |
| NU* | Female | 12 | 15 | 29 | 23 | 8 | 4 |
| DI | Female | 9 | 6 | 9 | 18 | 10 | 2 |
| DE* | Male | 1 | 0 | 7 | 1 | 1 | 4 |
| BO* | Male | 5 | 5 | 6 | 7 | 4 | 2 |
| MA* | Male | 11 | 9 | 14 | 20 | 7 | 9 |
| JA* | Male | 10 | 14 | 22 | 22 | 3 | 11 |
| Total | | 92 | 93 | 190 | 221 | 62 | 60 |

*Indicates speakers who also appear in the CDS dataset.

TABLE 3 Number of tokens for each vowel, by speaker, in the MONO dataset.

| Speaker | Gender | /i/ | /e/ | /a/ | /ɔ/ | /o/ | /u/ |
|---|---|---|---|---|---|---|---|
| YI* | Female | 24 | 24 | 45 | 44 | 14 | 16 |
| LY* | Female | 2 | 25 | 8 | 47 | 1 | 7 |
| NK | Female | 51 | 8 | 50 | 61 | 30 | 26 |
| OR | Female | 9 | 22 | 28 | 62 | 13 | 3 |
| RO | Female | 51 | 12 | 88 | 73 | 47 | 41 |
| DE* | Male | 5 | 5 | 15 | 18 | 1 | 8 |
| BO* | Male | 39 | 26 | 45 | 118 | 31 | 14 |
| ST* | Male | 29 | 53 | 46 | 75 | 17 | 30 |
| Total | | 210 | 175 | 325 | 498 | 154 | 145 |

*Indicates speakers who also appear in the CDS dataset.

followed by alignment in the "language independent" mode that does not require language training in advance, which seems to be a good option for under-described languages (Kisler et al., 2012; Jones et al., 2019). Subsequently, the segmentation by MAUS at the phonetic level was manually checked and adjusted, as there were a large number of cases in which misalignment occurred. For conversational ADS, as described above, there were no existing transcriptions, so vowels were segmented directly by hand by the first author.

The vowels' duration, F0, and formant values were extracted using the analysis techniques of previous similar studies (e.g., Williams and Escudero, 2014; Kashima et al.,

TABLE 4 Number of tokens and formant values for the six Nungon vowels in CDS.

| | | | /i/ | /e/ | /a/ | /ɔ/ | /o/ | /u/ |
|---|---|---|---|---|---|---|---|---|
| 2-year-olds | F | *N* | 52 | 119 | 211 | 176 | 63 | 77 |
| | | F1 | 440 (160) 22 | 443 (87) 8 | 764 (220) 15 | 525 (134) 10 | 512 (155) 20 | 424 (98) 11 |
| | | F2 | 2,133 (379) 22 | 2,177 (255) 8 | 1,797 (158) 15 | 1,465 (381) 10 | 1,263 (360) 20 | 1,617 (324) 11 |
| | M | *N* | 18 | 23 | 79 | 84 | 10 | 16 |
| | | F1 | 351 (87) 21 | 453 (114) 24 | 590 (156) 18 | 529 (148) 16 | 478 (79) 25 | 414 (88) 22 |
| | | F2 | 1,976 (203) 21 | 1,814 (304) 24 | 1,559 (234) 18 | 1,378 (291) 16 | 1,209 (261) 25 | 1,432 (388) 22 |
| 3-year-olds | F | *N* | 52 | 119 | 211 | 176 | 63 | 77 |
| | | F1 | 440 (160) 22 | 443 (87) 8 | 764 (220) 15 | 525 (134) 10 | 512 (155) 20 | 424 (98) 11 |
| | | F2 | 2,133 (379) 22 | 2,177 (255) 8 | 1,797 (158) 15 | 1,465 (381) 10 | 1,263 (360) 20 | 1,617 (324) 11 |
| | M | *N* | 18 | 23 | 79 | 84 | 10 | 16 |
| | | F1 | 351 (87) 21 | 453 (114) 24 | 590 (156) 18 | 529 (148) 16 | 478 (79) 25 | 414 (88) 22 |
| | | F2 | 1,976 (203) 21 | 1,814 (304) 24 | 1,559 (234) 18 | 1,378 (291) 16 | 1,209 (261) 25 | 1,432 (388) 22 |

Formant values (Hz) averaged for the six vowels produced by 10 speakers: mean (standard deviation) standard error.

2016; Sarvasy et al., 2020): 30 evenly distributed points starting from the 20% point to the 80% point of the vowel duration in Praat (Boersma and Weenink, 2021). We excluded the initial and final 20% of the vowel token in order to minimize coarticulation influences (Williams and Escudero, 2014). Following the method described in Williams and Escudero (2014), we then smoothed the series of formant values (or trajectories) for each vowel to reduce the influence of noisy formants by fitting each vowel token with parametric curves, specifically, the discrete cosine transform (DCT) in MATLAB. Then the vowel duration and formants values were averaged over the speakers and across the different positions. The pitch range was calculated by finding the maximum and minimum F0 values among the 30 evenly distributed points. Linear mixed modeling and *post-hoc* analysis were performed in R using the lmerTest and emmeans packages (Searle et al., 1980; Kuznetsova et al., 2017).

## Results

We first present a comparison of acoustic characteristics of vowels in CDS addressed to the two age-divided groups of children, three children per group (2;2–2;9 vs. 3;0–3;10). We found that the only acoustic feature that differed significantly in CDS to the younger and older cohorts was vowel duration. Thus, in the remaining analyses, we compare vowel formants, mean pitch, and pitch range, for CDS to all children, but the two age cohorts are separated when examining vowel duration and duration contrasts. We then compare the acoustic characteristics of vowels in Nungon CDS with those of conversational ADS and monologue narrative (MONO).

TABLE 5 Duration (ms), mean F0, and F0 range (Hz) for vowels in Nungon CDS by 10 speakers: mean (standard deviation) standard error.

| | | CDS to 3-year-olds | CDS to 2-year-olds |
|---|---|---|---|
| Long | *N* | 69 | 81 |
| | Duration | 128 (46) 6 | 158 (173) 19 |
| Short | *N* | 859 | 571 |
| | Duration | 73 (47) 2 | 114 (105) 4 |
| Female | *N* | 698 | 333 |
| | Mean F0 | 220 (55) 2 | 207 (63) 3 |
| | F0 range | 18 (21) 2 | 18 (23) 3 |
| Male | *N* | 230 | 319 |
| | Mean F0 | 140 (41) 3 | 130 (38) 2 |
| | F0 range | 10 (22) 3 | 9 (10) 2 |

### CDS to children of different age ranges compared

Tables 4, 5 give an overview of vowel token numbers, first and second formant values, durations, mean pitch values, and pitch ranges, in CDS to the two age cohorts of children: "2-year-olds" (2;2–2;9), and "3-year-olds" (3;0–3;10).

### Vowel triangles in CDS

Figure 1 presents the vowel triangles from data for the 10 speakers in the CDS dataset. For women, a shift of the vowel triangle toward the bottom left (i.e., increase of both F1 and F2) can be observed for the 2-year-olds, compared with the 3-year-olds. Recall, however, from Table 1, that only one woman

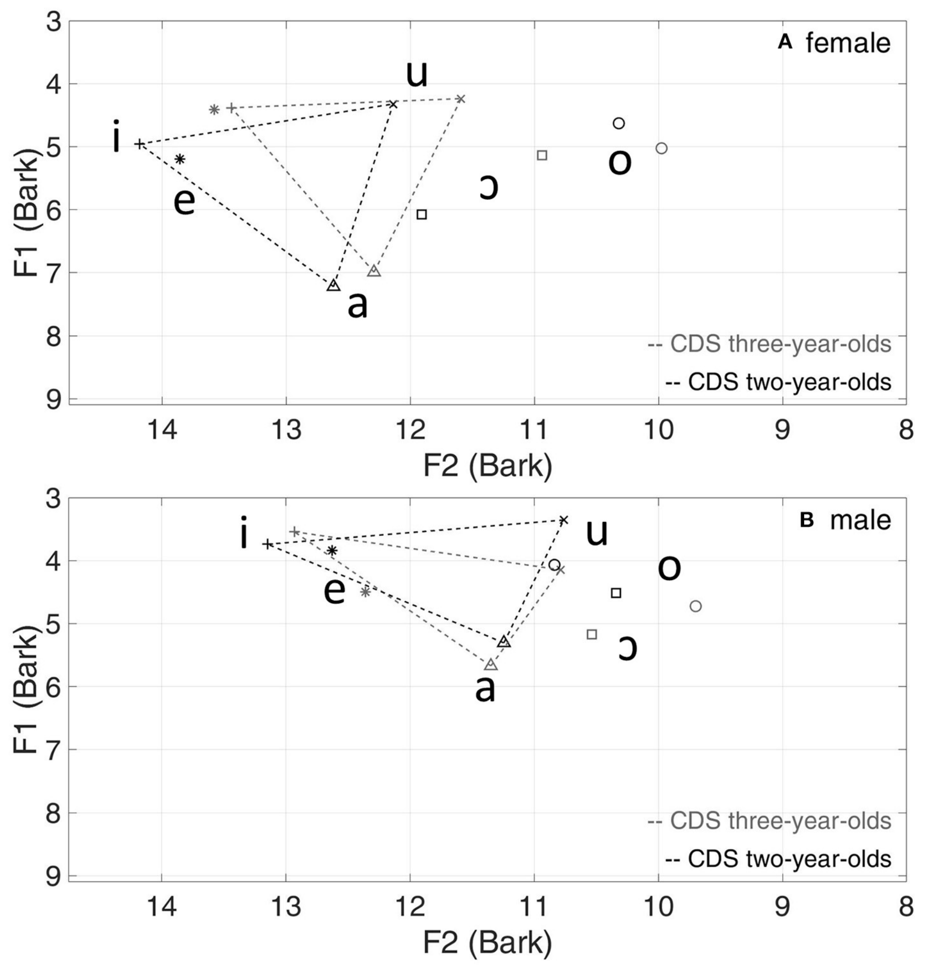



FIGURE 1
Mean vowel formants (Bark) and triangles of the 2-year-olds vs. 3-year-olds in CDS for women **(A)** and men **(B)**.

features in the 2-year-olds group, compared with four women in the 3-year-olds group. For men, there is little change in the placement of the vowel triangle relating to age of the children.

From the vowel triangle plots, the vowel space area (VSA) was then derived based on the vowels /i/, /a/, and /u/, following the method used by García-Sierra et al. (2021), for further comparison between CDS to the 2-year-olds and CDS to the 3-year-olds. A linear mixed model was run with VSA as a dependent variable, child's age (2- and 3-year-olds) and speaker gender (women and men) as independent variables, and speaker as a random intercept. There was no significant interaction between child's age and gender [$F_{(1,8)} = 0.09$, $p = 0.78$]. The difference in VSA between men and women was also not significant for both younger ($p = 0.97$; $N = 4$) and older groups ($p = 0.67$; $N = 6$), suggesting that men and women could be collapsed for analysis. Thus, a simpler linear model with child's age (2- and 3-year-olds) as one independent variable was used. *Post-hoc* analysis showed that VSA for CDS to 2-year-olds (Mean = 0.084 $LnHz^2$, SE = 0.020) was not significantly different from that for CDS to 3-year-olds (Mean = 0.072 $LnHz^2$, SE = 0.009, $p = 0.54$; $N = 10$), indicating no hyper- or hypo-articulation in CDS to the 2-year-olds compared with CDS to the 3-year-olds, for both women and men.

## Pitch mean and range in CDS

Linear mixed effects models were run with mean F0 and F0 range as dependent variables, child's age (2- and 3-year-olds) and gender (women and men) as fixed effects, and speaker as a random intercept. For mean F0, there was no significant interaction between child's age and gender [$F_{(1,1578)} = 0.0006$, $p = 0.98$]. As would be expected based on physical differences, women's mean F0 was higher than that of men ($\hat{\beta} = 70.8$, $p =$

TABLE 6 Vowel token data for Nungon ADS and MONO.

| | | | /i/ | /e/ | /a/ | /ɔ/ | /o/ | /u/ |
|---|---|---|---|---|---|---|---|---|
| ADS | F | *N* | 65 | 65 | 141 | 171 | 47 | 34 |
| | | F1 | 428 (54) 7 | 499 (52) 6 | 789 (139) 12 | 595 (80) 6 | 549 (71) 10 | 439 (56) 10 |
| | | F2 | 2,313 (306) 7 | 2,194 (303) 6 | 1,743 (171) 12 | 1,440 (274) 6 | 1,201 (356) 10 | 1,639 (266) 10 |
| | M | *N* | 27 | 28 | 49 | 50 | 15 | 26 |
| | | F1 | 343 (48) 9 | 413 (64) 12 | 576 (94) 13 | 488 (55) 8 | 475 (68) 18 | 347 (42) 8 |
| | | F2 | 2,058 (229) 9 | 2,069 (111) 12 | 1,581 (191) 13 | 1,359 (292) 8 | 1,136 (256) 18 | 1,378 (376) 8 |
| MONO | F | *N* | 137 | 91 | 219 | 287 | 105 | 93 |
| | | F1 | 411 (60) 5 | 537 (76) 8 | 907 (191) 13 | 684 (106) 6 | 564 (107) 10 | 442 (79) 8 |
| | | F2 | 2,381 (268) 5 | 2,235 (182) 8 | 1,737 (133) 13 | 1,332 (186) 6 | 1,039 (230) 10 | 1,402 (388) 8 |
| | M | *N* | 73 | 84 | 106 | 211 | 49 | 52 |
| | | F1 | 338 (48) 6 | 444 (52) 6 | 656 (87) 8 | 522 (62) 4 | 432 (42) 6 | 332 (38) 5 |
| | | F2 | 2,242 (251) 6 | 2,038 (286) 6 | 1,486 (156) 8 | 1,146 (179) 4 | 927 (196) 6 | 1,177 (383) 5 |

Number of tokens and formant values (Hz) averaged for the six vowels: mean (standard deviation) standard error.

TABLE 7 Duration (ms), mean F0 and F0 range (Hz) for vowels in Nungon ADS and MONO: mean (standard deviation) standard error.

| | | ADS | MONO |
|---|---|---|---|
| Long | *N* | 76 | 153 |
| | Duration | 134 (44) 5 | 175 (79) 6 |
| Short | *N* | 642 | 1,354 |
| | Duration | 82 (39) 2 | 89 (46) 1 |
| Female | *N* | 523 | 932 |
| | Mean F0 | 201 (32) 1 | 213 (38) 1 |
| | F0 range | 10 (12) 1 | 13 (16) 1 |
| Male | *N* | 195 | 575 |
| | Mean F0 | 117 (14) 1 | 123 (16) 1 |
| | F0 range | 5 (4) 1 | 5 (7) 1 |

0.003). Pairwise comparisons showed no significant difference between mean F0 in CDS to 2-year-olds vs. CDS to 3-year-olds for women ($p = 0.61$; $N = 1{,}031$) and men ($p = 0.56$; $N = 549$). For F0 range, there was no significant interaction between child's age and gender [$F_{(1,1,578)} = 0.475$, $p = 0.52$]. Women showed a larger F0 range than men ($\hat{\beta} = 7.56$, $p = 0.003$), but neither sex showed a significant difference between F0 range in CDS to 2-year-olds and F0 range in CDS to 3-year-olds.

## Vowel durations and duration contrast in CDS

Linear mixed effects models were run with vowel duration as a dependent variable, vowel length (long and short) and child's age (2- and 3-year-olds) as fixed effects, and speaker as a random intercept. *Post-hoc* analysis showed no significant interaction between child's age and vowel length [$F_{(1,1578)} = 0.388$, $p = 0.53$]. Pairwise comparisons showed that the duration contrast between long and short vowels was significant in both CDS to 2-year-olds ($\hat{\beta} = 46.0$, $p < 0.001$) and CDS to 3-year-olds ($\hat{\beta} = 54.9$, $p < 0.001$), demonstrating that in neither variety is the usual Nungon phonological duration contrast minimized or lost. The duration of short vowels in CDS to 2-year-olds was significantly longer than that in CDS to 3-year-olds ($\hat{\beta} = 43.1$, $p < 0.001$), and the same was found for long vowels ($\hat{\beta} = 34.2$, $p = 0.02$). In sum, vowels in CDS to the 2-year-olds were found to last significantly longer than vowels in CDS to the 3-year-olds.

## CDS, ADS, and narrative monologues compared

Vowel token numbers, first and second formant values, durations, mean pitch, and pitch range for the ADS and MONO datasets are summarized in Tables 6, 7.

We ran three sets of analyses to compare CDS with the two types of adult-directed speech: conversational (the "ADS" dataset) and performative storytelling, directed at a non-native speaker (the "MONO" dataset). The ideal here would be to have obtained data for all three speech contexts from each speaker, such that all comparisons would be within-speaker. Unfortunately, our datasets include only four speakers who produced CDS, ADS and MONO (allowing for true within-speaker comparisons across those three contexts), and seven who produced both CDS and ADS (allowing for within-speaker comparisons across those two contexts). We thus also ran a further analysis that included all 14 speakers who feature in at least one of these three recording contexts, and compared the

TABLE 8 Comparisons of vowel space area (VSA), F0 mean and range, and vowel duration for CDS, ADS and MONO dataset.

| | | All 14 speakers | 7 overlapping speakers in CDS and ADS | 4 overlapping speakers in CDS, ADS and MONO |
|---|---|---|---|---|
| VSA | | MONO > ADS ∼ CDS | ADS ∼ CDS | MONO > ADS ∼ CDS |
| Mean F0 | F | MONO > CDS > ADS | CDS > ADS | MONO > CDS > ADS |
| | M | CDS > ADS > MONO | CDS > ADS | CDS > MONO > ADS |
| F0 range | F | CDS > MONO > ADS | CDS > ADS | MONO ∼ CDS > ADS |
| | M | CDS > ADS ∼ MONO | CDS > ADS | CDS ∼ ADS ∼ MONO (CDS > MONO) |
| Duration | Short | CDS_2-year-olds > MONO > ADS ∼ CDS_3-year-olds | CDS_2-year-olds > ADS ∼ CDS_3-year-olds | CDS_2-year-olds > MONO > ADS > CDS_3-year-olds |
| | Long | MONO ∼ CDS_2-year-olds > ADS ∼ CDS_3-year-olds | CDS_2-year-olds > ADS ∼ CDS_3-year-olds | CDS_2-year-olds > MONO > ADS ∼ CDS_3-year-olds |

Symbols '>' and '∼' stand for 'higher/larger than' with and without significance, respectively.

results with those of the smaller, within-speaker, analyses. That is: First, we compared only tokens from the four speakers who featured in CDS, ADS, and MONO; then we compared only tokens from the seven speakers who featured in both CDS and ADS, comparing those two datasets; then we compared tokens from all 14 speakers, comparing CDS, ADS, and MONO. The results from these three sets of analyses are in Table 8.

Overall, the results for smaller, within-speaker analyses are similar to those for the entire 14-speaker group, so the results from that group are the ones presented in depth in the following sections.

## Vowel triangles in CDS, ADS, and MONO

Figure 2 presents the vowel triangles from data for all 14 speakers in CDS, ADS and MONO corpora. MONO speech involves by far the largest vowel space, evaluated in terms of F1 and F2 of the three corner vowels. For both women and men, the vowel triangle for CDS to 3-year-olds is similar to that for ADS.

Since there was no significant difference in VSA between the CDS to 2-year-olds and CDS to 3-year-olds, the two groups were collapsed as "CDS" for the following comparison with ADS and MONO. There was no significant interaction between child's age and gender [$F_{(2,23)} = 0.22$, $p = 0.80$]. The difference in VSA between men and women was not significant for all three corpora (CDS: $p = 0.75$, $N = 10$; ADS: $p = 0.55$, $N = 8$; MONO: $p = 0.85$, $N = 8$), suggesting that men and women could be collapsed when the three corpora were compared. A linear model with dataset (CDS, ADS and MONO) as one independent variable showed that MONO has the largest VSA value (Mean = 0.236 LnHz$^2$, SE = 0.029), significantly larger than ADS (Mean = 0.105 LnHz$^2$, SE = 0.016; $\hat{\beta} = 0.131$, $p < 0.001$) and CDS (Mean = 0.076 LnHz$^2$, SE = 0.009; $\hat{\beta} = 0.159$, $p < 0.001$), but the difference between ADS and CDS was not significant ($p = 0.53$; $N = 18$). The results for analyses including only overlapping speakers (those who feature in all three datasets) were similar to those with all speakers, as seen in Table 8.

## Pitch mean and range in CDS, ADS, and MONO

Since there was no significant difference in either mean pitch or pitch range between CDS to 2-year-olds and CDS to 3-year-olds, the two groups were collapsed as "CDS" in the following pitch analysis. Linear mixed effects models similar to those in 3.1.2 were used for pairwise comparisons of the three corpora. For mean F0, there was a significant interaction between dataset and gender [$F_{(2,3125)} = 102.1$, $p < 0.001$], and the trend for different corpora varied with gender. For women, F0 mean for MONO was significantly higher than that for CDS ($\hat{\beta} = 28.7$, $p < 0.001$) and ADS ($\hat{\beta} = 52.4$, $p < 0.001$). For men, CDS showed the highest mean F0, followed by ADS ($\hat{\beta} = 11.5$, $p = 0.002$) and MONO ($\hat{\beta} = 22.6$, $p < 0.001$). For F0 range, there was also a significant interaction between dataset and gender [$F_{(2,3125)} = 8.1$, $p < 0.001$]. For women, F0 range for CDS was the largest, but not significantly different from MONO ($p = 0.12$; $N = 1{,}576$); both CDS and MONO were significantly larger than ADS ($\hat{\beta} = 8.0$, $p < 0.001$ for CDS and $\hat{\beta} = 5.8$, $p < 0.001$ for MONO). For men, F0 range for CDS was also the largest, and significantly different from ADS ($\hat{\beta} = 3.3$, $p < 0.001$) and MONO ($\hat{\beta} = 4.8$, $p < 0.001$).

There were seven speakers (three women and four men) who featured in both the CDS and ADS corpora. As shown in Table 8, the results for these overlapping speakers were consistent with those for all speakers: for both women and men, mean F0 for CDS was significantly higher than that for ADS, and F0 range for CDS was also significantly larger than that for ADS. Results for the four overlapping speakers (two women and two men) featuring in all three corpora were also consistent with those for all speakers in general.

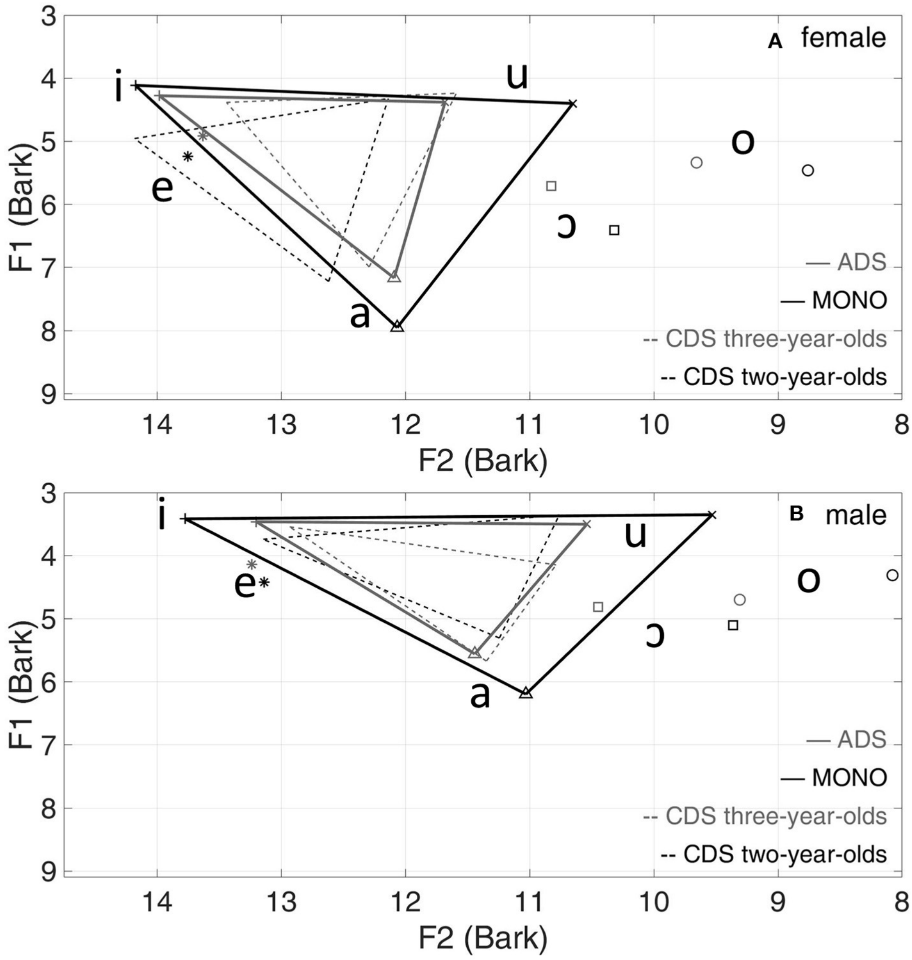



FIGURE 2
Mean vowel formants (Bark) and triangles of all speakers in CDS, ADS and MONO for women **(A)** and men **(B)**.

## Vowel durations and duration contrast in CDS, ADS, and MONO

Both phonologically short and long vowels in CDS to 2-year-olds were significantly longer in duration than those in CDS to 3-year-olds. Thus, for comparison of duration with ADS and MONO, CDS to 2-year-olds and CDS to 3-year-olds were separated. Short vowels in CDS to 2-year-olds were significantly longer than those in MONO ($\hat{\beta} = 23.4$, $p < 0.001$), ADS, and CDS to 3-year-olds; those in MONO were significantly longer than those in ADS ($\hat{\beta} = 24.9$, $p < 0.001$) and CDS to 3-year-olds, but those in ADS were similar to those in CDS to 3-year-olds ($p = 0.06$; $N = 1{,}501$). Long vowels in MONO were similar to those in CDS to 2-year-olds ($p = 0.50$; $N = 234$); those in CDS to 2-year-olds were significantly longer than those in ADS ($\hat{\beta} = 45.0$, $p < 0.001$) and CDS to 3-year-olds, but those in ADS were similar to those in CDS to 3-year-olds ($p = 0.98$, $N = 145$). So for both short and long vowels, duration was exaggerated in CDS to 2-year-olds and in MONO compared with ADS, while duration in CDS to 3-year-olds was similar to that in ADS. In general, these results were consistent with those for overlapping speakers.

The duration contrast between phonologically long and short vowels was significant in all four corpora ($\hat{\beta} > 46.8$, $p < 0.001$). For overlapping speakers (both across ADS, CDS and across ADS, CDS and MONO), the contrast remained significant ($\hat{\beta} > 46.3$, $p \leq 0.001$). That is, in none of these genres is the duration contrast between phonologically short and long vowels neutralized.

## Discussion

In this paper, we present a first study comparing acoustic and prosodic properties of vowels in CDS (to children aged 2;2-3;10), ADS (to familiar adults), and monologual speech directed at a foreigner, for the Towet dialect of the Papuan language Nungon, spoken by about 300 people in a remote mountain village of Papua New Guinea (of the 1,000 speakers of Nungon across all dialects). Our results show that Nungon CDS does not feature expansion of the vowel space, compared to conversational Nungon ADS. The Nungon CDS vowel triangle area is neither significantly smaller nor larger than that of conversational ADS. As seen in Table 8, this result applies for both women and men, and holds true when analyzed for three different cross-sections of our dataset: first, the seven speakers who produced both CDS and conversational ADS; second, the four speakers who produced CDS, conversational ADS, and narrative monologues (MONO); and third, the total 14 speakers who feature in at least one cross-section of our dataset: CDS, conversational ADS, and/or MONO. The size of the vowel triangle does not differ significantly for either men or women in CDS to children aged 2;2–2;9, compared with CDS to children aged 3;0–3;10. Since we did not examine data on Nungon vowel acoustics in IDS to infants under 12 months, or CDS to children under 2 years of age, we cannot rule out the possibility that Nungon IDS or CDS to children aged 0–25 months does indeed involve vowel hyper-articulation.

While we found that Nungon CDS to the children in our sample did not exhibit vowel hyper-articulation, Nungon MONO (our measure of FDS in Nungon, from monologual narratives told to a non-native listener holding a recording device) did. This pattern holds both for the four individual speakers who produced tokens of all three speech varieties (CDS, conversational ADS, and MONO), and for the vowel tokens of the entire group of 14 speakers across the three datasets. In terms of vowel space area, then, the Nungon pattern could be similar to that suggested by studies of Norwegian IDS (Englund and Behne, 2005, 2006; Englund, 2018) and CDS (Steen and Englund, 2022), and Norwegian FDS (Sikveland, 2006). Norwegian IDS and CDS (to children in "kindergartens," 10–34 months of age) do not show vowel hyper-articulation relative to ADS, but Norwegians addressing second-language learners do display expansion of the vowel space (perhaps, in part, due to an overall more open mouth during speech).

Nungon-speaking adults, both male and female, used higher mean F0 and expanded F0 range in CDS than in ADS, even though the ADS samples here were conversations between adults who had known each other very well for many years, rather than strangers. Male caregivers have been shown to differ from female caregivers in pitch modifications in American English CDS (Warren-Leubecker and Bohannon, 1984, although see Fernald et al., 1989, where both men and women used increased mean pitch and expanded pitch range in American English IDS). At least in speech addressed to children aged 2;2–3;10, we found no such differences between Nungon-speaking women and men. Women and men both used higher mean F0 in Nungon CDS than in conversational ADS, and both used greater F0 range in CDS than in conversational ADS, as found for mothers and fathers speaking a range of major world languages by Fernald et al. (1989).

Women and men do differ in mean F0 and F0 range within the MONO dataset, and in the relationship between F0 in the MONO dataset and in the other two datasets. For women, MONO features the highest mean F0, followed by CDS, then ADS; men's highest mean F0 occurs in CDS, followed by MONO, then ADS. Women have similar F0 ranges in CDS and MONO, which are both higher than the F0 range in ADS. In contrast, men have a greater F0 range in CDS than in MONO and ADS, which have similar F0 ranges. This implies that men's performative FDS monologue style is similar in pitch features to their ADS, albeit with slightly higher pitch overall, while CDS is the most divergent for both pitch mean and range. For women, in contrast, the performative FDS monologue style is the outlier, rather than CDS; MONO exceeds CDS in mean pitch and pitch range. Previous studies have yielded mixed results in terms of increased mean pitch or pitch range in FDS, relative to native-speaker-directed speech; increased mean pitch was found for Omani Arabic FDS (Al-Kendi and Khattab, 2019) and for English FDS by Knoll and Scharrer M. A. (2007), but not by Biersack et al. (2005) or Uther et al. (2007). Increased pitch range has been found in Mandarin FDS (Papoušek and Hwang, 1991) and French FDS (Smith, 2007).

In sum, when Nungon CDS is compared to conversational ADS, women and men show very similar patterns of mean F0 and F0 range. For both women and men, Nungon CDS to children aged 2;2–3;10 features significantly higher mean F0 and greater F0 range than conversational ADS. It is when CDS and conversational ADS are further compared to a third genre, MONO, which probably conflates elements of performative monologue with FDS, that sex differences in pitch use emerge. These differences could relate to: a) sex-based differences in speaking to a non-native interlocutor; b) sex-based differences in perfomative storytelling style; and/or c) possibly, heightened nervousness of the female speakers in the recorded performative context, interacting with the foreigner, which could induce higher mean pitch and greater pitch range (Fairbanks and Pronovost, 1938; Bonner, 1943; Apple et al., 1979; Laukka et al., 2008).

No significant differences for either women or men in mean F0 or F0 range between CDS to children aged 2;2–2;9 and CDS to children aged 3;0–3;10 were evident. In other words, CDS to both 2-year-olds and 3-year-olds featured higher pitch and greater pitch range than conversational ADS. Previous research (Sarvasy, 2019) suggested that a special morpho-syntactic alteration found in Nungon CDS is most prevalent in speech to children under 3;0, with phasing out thereafter; earlier studies also showed that the morpho-syntactic complexity of Nungon CDS increases from when the child is 3 years old (Sarvasy, 2019, 2020, 2021a). The current results imply that elevated mean pitch and increased pitch range may be evident in CDS even after caregivers no longer use special morpho-syntactic alterations, and have already increased the morpho-syntactic complexity of their CDS.

Much research on IDS and CDS has found that features of IDS/CDS relate to children's ages and developmental stages (Stern et al., 1983; Kitamura et al., 2002; Kitamura and Burnham, 2003; Englund and Behne, 2006; Rattanasone et al., 2013). We found no evidence that the size of the vowel triangle differs in Nungon speech directed to children in their third year of life (2;2–2;9), compared with Nungon speech directed to children in their fourth year of life (3;0–3;10). We further found no evidence that mean pitch and pitch range differ in speech to these two different age cohorts. But we did find that speech directed to the younger cohort features significantly longer vowel durations than speech directed to the older cohort, which appears similar to conversational ADS in vowel durations and duration contrasts. When only tokens from speakers featured in all three datasets are examined, both phonologically short and long vowels in CDS to 2-year-olds are significantly longer than those in MONO, which are in turn longer than those in conversational ADS and CDS to 3-year-olds. This said, we did not examine speech from a single caregiver to a single child over time to confirm these results: this longitudinal investigation remains a desideratum for future work.

Uther et al. (2007) found that vowels in British English FDS were shorter than those in IDS, as with Nungon MONO, but they found no difference in vowel length between FDS and ADS, while this does seem to exist between Nungon MONO and ADS. Again, it is as yet impossible to ascertain whether the foreigner-directed aspect of the MONO dataset induced longer vowel durations in MONO than in ADS, or whether another aspect, such as the performative context, induced this.

Burnham et al. (2002), Uther et al. (2007), Lam and Kitamura (2012), and Xu et al. (2013), among others, demonstrated, for two varieties of English, that some acoustic and prosodic modifications of speech may be predictable according to the interlocutor's relationship to the speaker and perceived linguistic abilities. For Australian and British English, the interlocutor's increased ability to provide linguistic feedback (accompanied by an apparent need for didactic intervention by the speaker) led to increased degrees of vowel hyper-articulation, while increased affect in the relationship led to increased mean pitch. The Nungon findings here from CDS and ADS do not follow the same pattern as these English studies, since Nungon CDS shows no vowel hyper-articulation, but does show increased mean pitch and pitch range. That said, the Nungon children studied here were older than the infants in the English studies. The MONO dataset shows marked vowel hyper-articulation, relative to both ADS and CDS, but it is unclear from the present study whether this stems from the identity of the interlocutor (non-native Nungon speaker with a foreign appearance) or from the particular performative context for speech.

FDS has been studied acoustically in a relatively small number of languages. The present study of acoustic properties of vowels in Nungon monologues addressed to a foreigner, compared with those of conversational, native-speaker-directed, adult speech, represents the first of which we are aware for any language of the Pacific, and for any speech community of fewer than 1 million people anywhere in the world. Like members of other small, remote speech communities (Wray and Grace, 2007), Nungon speakers do not interact with strangers on a daily basis; the very few non-native speakers of Nungon they encounter are women from nearby communities who married Nungon-speaking men. Each of the six Nungon-speaking villages has its own dialect, and Nungon further belongs to a longer dialect continuum that encompasses six additional villages, whose languages are grouped under the umbrella term "Yau" (Sarvasy, 2017). The MONO data here were addressed to one of the very few truly foreign learners of Nungon or Yau who are not native speakers of other Papuan languages spoken in Morobe Province, Papua New Guinea. One might therefore question how systematic an FDS style could be among Nungon speakers. While they do not regularly interact with foreign speakers of Nungon, Nungon speakers do regularly speak to people whose Nungon and Yau varieties differ from their own. It is unclear whether their strategies in so doing should be termed FDS. For instance, the first author observed that some speakers of the Towet village dialect of Nungon would actually assume a Kotet village-like accent and use some Kotet lexicon in interacting with speakers of the Kotet village dialect (Sarvasy, 2017).

As noted earlier, the MONO data may be hyper-articulated for a conglomeration of reasons: non-native interlocutor, high-importance recording context, and performative, rather than conversational, speech. It could be that the MONO data should be considered to represent "clear speech," primarily (Picheny et al., 1986; Moon and Lindblom, 1994; Ferguson and Kewley-Port, 2002, 2007), and FDS only secondarily. Future work will aim to disentangle these factors. If we take the MONO results to represent FDS, however, they seem to support the notion that FDS often involves vowel hyper-articulation, and add to the complicated general picture of mean pitch and pitch range in FDS, since Nungon-speaking women and men pattern differently in how they use pitch in MONO, compared with ADS and CDS.

Acoustic research on under-described languages spoken by remote communities entails different research conditions than research with speech communities who live in proximity to university campuses with laboratory facilities. In remote areas, speech must be recorded in an outdoor or sound-permeable indoor environment, not a laboratory. Field trips to visit such communities are of limited duration and frequency. If a field trip is not devoted to a particular research question, evidence must often be culled after the trip, from an existing, multi-purpose natural speech corpus. Thus, for a study such as the present one, vowel tokens may have to be culled from uncontrolled, natural interactions in an existing speech corpus, and it may be hard to ensure that the same speakers produce vowel tokens in all three interactive contexts under study. In studying child language development in such a community, researchers may end up devoting all resources into building a corpus of child-caregiver speech, without also constructing a counterpart corpus of the caregivers addressing other adults in a similar, informal and conversational, vein. A child-caregiver corpus will yield ample tokens of CDS vowels and consonants, but the researcher may not have access to optimal ADS tokens for comparison.

In recent work, some researchers have chosen to compare features of conversational CDS to a corpus of non-conversational adult speech. For instance, Frye aus Schwerte (2019) compares prosodic features of CDS in the Papuan language Qaqet to those in adult "Pear Story" narratives, rather than truly conversational adult speech. Our first exploratory comparative study of Nungon CDS and ADS compared CDS directly to the narrative monologue data in the MONO dataset here (Sarvasy et al., 2019). We resorted to this because we had not yet commissioned the dyadic ADS conversations, with no foreigners present, described in the present study.

The results here indicate that, difficult as it may be to source conversational ADS to compare to CDS in languages spoken by small, remote communities, this is just as important for these languages as it is for languages like Japanese (Miyazawa et al., 2017). The marked contrast in vowel space area between MONO, ADS, and CDS shows that even in a community without a tradition of literacy, vowel acoustics differ in different contexts and settings: here, speech addressed to children, conversation among familiars, and a more presentative mode with a non-native interlocutor. Future research would do well to consider this carefully.

## Conclusion

This study has shown for the Towet village dialect of the Papuan language Nungon that vowels in speech directed to children aged 2;2–3;10 are not hyper-articulated, compared with speech directed to adults. Nungon speech directed to children of this age by both women and men features higher mean pitch and increased pitch range, compared with conversational speech directed to adults. Speech directed to 2-year-olds (2;2–2;9) features similar vowel space areas (vowel triangle sizes), mean pitch, and pitch range, to speech directed at 3-year-olds (3;0–3;10); this holds for both men and women. Speech directed at 2-year-olds features significantly longer vowels than speech directed at 3-year-olds, in which vowel length is similar to that of conversational adult-directed speech. The duration distinction between phonologically short and phonologically long Nungon vowels is upheld in speech to 2-year-olds, 3-year-olds, and adults. Compared with both child-directed speech and conversational adult-directed speech, monologues directed at a non-native Nungon speaker feature marked vowel hyper-articulation. The relationship between mean pitch and pitch range in these foreigner-directed monologues and those in child-directed and adult-directed conversational speech differed for men and women. This has been the first study comparing vowel acoustics and pitch in child-directed speech and adult-directed speech for a language of the New Guinea area, where 20% of the world's languages are spoken; it is further the first study comparing vowels in child-directed speech, adult-directed speech, and another clear speech style for a small language community.

## Data availability statement

The datasets analyzed for this study can be found in the Open Science Framework site at: https://osf.io/6pr8s/.

## Ethics statement

The studies involving human participants were reviewed and approved by the Australian National University and Western Sydney University Human Research Ethics Committees. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## Author contributions

HSS collected data, drafted, and revised the manuscript. WL ran analyses, helped draft, and revised the manuscript. JE assisted with revising the manuscript. PE assisted with revising the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

Funding was received from: the Australian Research Council (Grants CE140100041 and DE180101609 to HSS and FT160100514 to PE), the MARCS Institute for Brain, Behavior and Development, and the ARC Centre of Excellence for the Dynamics of Language (Language Documentation grant to HSS).

## Acknowledgments

This project depended on the generosity and goodwill of Nungon speakers for participating in speech data collection, and the expert transcription work of Nungon speakers Lyn Ögate, Stanly Girip, James Jio, Nathalyne Ögate, and Yongwenwen

Hessy. Jason Peed and Nicole Traynor assisted with the manual segmentation and adjustment of the data.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

Aikhenvald, A. Y. (2008). *The Manambu Language of East Sepik, Papua New Guinea.* Oxford: Oxford University Press.

Al-Kendi, A. (2020). *Foreigner-directed speech and L2 speech learning in an understudied interactional setting: the case of foreign-domestic helpers in Oman* (Doctoral dissertation, Newcastle University).

Al-Kendi, A., and Khattab, G. (2019). "Acoustic properties of foreigner directed speech," in *Proc. 19th International Congress of Phonetic Sciences*, Melbourne, Australia.

Andruski, J. E., Kuhl, P. K., and Hayashi, A. (1999). Point vowels in Japanese mothers' speech to infants and adults. *J. Acoust. Soc. Am.* 105, 1095–1096. doi: 10.1121/1.425135

Apple, W., Streeter, L. A., and Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *J. Personal. Soc. Psychol.* 37, 715–727. doi: 10.1037/0022-3514.37.5.715

Arthur, B., Wemer, R., Culver, M., Lee, Y. J., and Thomas, D. (1980). "The Register of Impersonal Discourse to Foreigners: Verbal Adjustments to a Foreign Accent," in *Discourse Analysis in Second Language Research*, ed D. Larsen-Freeman (Rowley, MA: Newbury House), 111–124.

Audibert, N., and Falk, S. (2018). "Vowel space and f0 characteristics of infant-directed singing and speech," in *Proc. 9th Int. Conf. Speech Prosody 2018* (Poland: Poznań), 153–157.

Benders, T. (2013). Mommy is only happy! Dutch mothers' realisation of speech sounds in infant-directed speech expresses emotion, not didactic intent. *Infant Behav. Dev.* 36, 847–862. doi: 10.1016/j.infbeh.2013.09.001

Biersack, S., Kempe, V., and Knapton, L. (2005). "Fine-tuning speech registers: a comparison of the prosodic features of child-directed and foreigner-directed speech," in *Proc. 9th European Conference on Speech Communication and Technology*, Lisboa, Portugal.

Bobb, S. C., Mello, K., Turco, E., Lemes, L., Fernandez, E., and Rothermich, K. (2019). Second language learners' listener impressions of foreigner-directed speech. *J. Speech Lang. Hear. Res. JSLHR* 62, 3135–3148. doi: 10.1044/2019_JSLHR-S-18-0392

Boersma, P., and Weenink, D. (2021). Praat: doing phonetics by computer [Computer program]. Version 6.1.53. Available online at: http://www.praat.org/ (accessed September 8, 2021).

Bohannon III, J. N., and Marquis, A. L. (1977). Children's control of adult speech. *Child Dev.* 148, 1002–1008. doi: 10.2307/1128352

Bonner, M. R. (1943). Changes in the speech pattern under emotional tension. *Am. J. Psychol.* 56, 262–273. doi: 10.2307/1417508

Burnham, D., Joeffry, S., and Rice, L. (2010). "d-o-e-s-not-c-o-m-p-u-t-e": vowel hyperarticulation in speech to an auditory-visual avatar." In *Proceedings of the 9th International Conference on Auditory-Visual Speech Processing (AVSP-2010),* Hakone, Japan, September 30–October 3, 2010.

Burnham, D., Kitamura, C., and Vollmer-Conna, U. (2002). What's new, pussycat? on talking to babies and animals. *Science.* 296, 1435–1435. doi: 10.1126/science.1069587

Cooper, R. P., Abraham, J., Berman, S., and Staska, M. (1997). The development of infants' preference for motherese. *Infant Behav. Dev.* 20, 477–488. doi: 10.1016/S0163-6383(97)90037-0

Cristia, A., and Seidl, A. (2014). The hyper-articulation hypothesis of infant-directed speech. *J. Child Lang.* 41, 913–934. doi: 10.1017/S0305000912000669

DePaulo, B. M., and Coleman, L. M. (1986). Talking to children, foreigners, and retarded adults. *J. Pers. Soc. Psychol.* 51, 945–959. doi: 10.1037/0022-3514.51.5.945

Englund, K., and Behne, D. (2006). Changes in infant directed speech in the first 6 months. *Infant Child Dev.* 15, 139–160. doi: 10.1002/icd.445

Englund, K. T. (2018). Hypoarticulation in infant-directed speech. *Appl. Psycholing.* 39, 67–87. doi: 10.1017/S0142716417000480

Englund, K. T., and Behne, D. M. (2005). Infant directed speech in natural interaction: Norwegian vowel quantity and quality. *J. Psycholing. Res.* 34, 259–280. doi: 10.1007/s10936-005-3640-7

Fairbanks, G., and Pronovost, W. (1938). Vocal pitch during simulated emotion. *Science* 88, 382–383. doi: 10.1126/science.88.2286.382

Ferguson, C. A. (1975). Toward a characterization of english foreigner talk. *Anthropological Linguistics* 17, 1–14.

Ferguson, S. H., and Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* 112, 259–271. doi: 10.1121/1.1482078

Ferguson, S. H., and Kewley-Port, D. (2007). Talker differences in clear and conversational speech: acoustic characteristics of vowels. *J. Speech Lang.* 50, 1241–1255. doi: 10.1044/1092-4388(2007/087)

Fernald, A., and Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Dev. Psychol.* 20, 104–113. doi: 10.1016/j.dr.2007.06.002

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., and Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *J. Child Lang.* 16, 477–501. doi: 10.1017/S0305000900010679

Freed, B. (1981a). Foreigner talk, baby talk, native talk. *Int. J. Sociol. Lang.* 28, 19–40.

Freed, B. (1981b). "Talking to foreigners vs. talking to children: similarities and differences," in *Research in Second Language Acquisition*, eds R. Scarcella and S. Kiashen (Rowley, MA: Newbury House), 19–27.

Frye aus Schwerte, H. (2019). *Child-directed speech in Qaqet, a language of East New* Britain, Papua New Guinea [PhD dissertation]. University of Cologne.

García-Sierra, A., Ramírez-Esparza, N., Wig, N., and Robertson, D. (2021). Language learning as a function of infant directed speech (IDS) in Spanish: testing neural commitment using the positive-MMR. *Brain Lang.* 212, 104890. doi: 10.1016/j.bandl.2020.104890

Garnica, O. (1977). "Some prosodic and paralinguistic features of speech to young children," in *Talking to Children: Language Input and Acquisition*, eds C. E. Snow and C. A. Ferguson Cambridge, England: Cambridge University Press.

Green, J., Nip, I., Wilson, E., Mefferd, A., and Yunusova, Y. (2010). Lip movement exaggerations during infant-directed speech. *J. Speech Lang. Hear. Res.* 53, 1529–1542. doi: 10.1044/1092-4388(2010/09-0005)

Grieser, D. L., and Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: support for universal prosodic features in motherese. *Dev. Psychol.* 24, 14–20. doi: 10.1037/0012-1649.24.1.14

Hazan, V., Uther, M., and Granlund, S. (2015). *How does foreigner-directed speech differ from other forms of listener-directed clear speaking styles*? *ICPhS* 2015.

Hellwig, B., Sarvasy, H. S., Casillas, M. (Forthcoming). *Language Acquisition*, in *Oxford Guide to Papuan Languages*, eds Evans, N. and Fedden, S. Oxford: Oxford University Press.

Henzl, V. M. (1973). Linguistic register of foreign language instruction. *Lang. Learn.* 23, 207–222. doi: 10.1111/j.1467-1770.1973.tb00656.x

Henzl, V. M. (1979). Foreigner talk in the classroom. *IRAL* 17, 159. doi: 10.1515/iral.1979.17.1-4.169

Jones, C., Li, W., Almeida, A., and German, A. (2019). Evaluating cross-linguistic forced alignment of conversational data in north Australian Kriol, an under-resourced language. *Lang. Doc. Conserv*. 13, 281–299.

Kangatharan, J. (2015). *The role of vowel hyperarticulation in clear speech to foreigners and infants* (Doctoral dissertation, Brunel University London).

Kashima, E., Williams, D., Ellison, M., T., Schokkin, D., and Escudero, P. (2016). Uncovering the acoustic vowel space of a previously undescribed language: the vowels of Nambo. *J. Acoust. Soc. Am.* 139, EL252–EL256. doi: 10.1121/1.4954395

Kemler Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W., and Cassidy, K. W. (1989). How the prosodic cues in motherese might assist language learning. *J. Child Lang.* 16, 5–68. doi: 10.1017/S030500090001343X

Kik, A., Adamec, M., Aikhenvald, A. Y., Bajzekova, J., Baro, N., Bowern, C., et al. (2021). Language and Ethnobiological skills decline precipitously in Papua New Guinea, the world's most linguistically diverse nation. *Proc. Nat. Acad. Sci.* 1, 98. doi: 10.1073/pnas.2100096118

Kisler, T., Reichel, U. D., and Schiel, F. (2017). Multilingual processing of speech *via* web services. *Comput. Speech Lang.* 45, 326–347. doi: 10.1016/j.csl.2017.01.005

Kisler, T., Schiel, F., and Sloetjes, H. (2012). Signal processing *via* web services: the use case WebMAUS. in *Proc. Digital Humanities Conference*, Hamburg, Germany.

Kitamura, C., and Burnham, D. (2003). Pitch and communicative intent in mother's speech: adjustments for age and sex in the first year. *Infancy* 4, 85–110. doi: 10.1207/S15327078IN0401_5

Kitamura, C., Thanavishuth, C., Burnham, D., and Luksaneeyanawin, S. (2002). Universality and specificity in infant-directed speech: pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant Behav. Develop.* 24, 372–392. doi: 10.1016/S0163-6383(02)00086-3

Knoll, M., and Scharrer, L. (2007). "Acoustic and affective comparisons of natural and imaginary infant-, foreigner-and adult-directed speech," in *Eighth Annual Conference of the International Speech Communication Association.*

Knoll, M., Scharrer, L., and Costall, A. (2009). Are actresses better simulators than female students? the effects of simulation on prosodic modifications of infant-and foreigner-directed speech. *Speech Commun.* 51, 296–305. doi: 10.1016/j.specom.2008.10.001

Knoll, M., Scharrer, L., and Costall, A. (2011). "Look at the shark": Evaluation of student-and actress-produced standardised sentences of infant-and foreigner-directed speech. *Speech Commun.* 53, 12–22. doi: 10.1016/j.specom.2010.08.004

Knoll, M. A., Johnstone, M., and Blakely, C. (2015). "Can you hear me? Acoustic modifications in speech directed to foreigners and hearing-impaired people," in *Sixteenth Annual Conference of the International Speech Communication Association.*

Knoll, M. A., and Scharrer, L. (2007). Acoustic and affective comparisons of natural and imaginary infant-, foreigner- and adult-directed speech. *Proc. Interspeech* 2007, 4. doi: 10.21437/Interspeech.2007-29

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., et al. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science* 277, 684–686. doi: 10.1126/science.277.5326.684

Kühnert, B., and Antolík, T. (2017). "Patterns of articulation rate variation in English/French tandem interactions," in *Pronunciation of English by Speakers of Other Languages*, eds J. Volin and R. Skarnitzl (Cambridge: Cambridge Scholars Publishing). Available online at: https://halshs.archives-ouvertes.fr/halshs-01505928 (accessed January 8, 2022).

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmertest package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13

Lam, C., and Kitamura, C. (2008). "Your baby can't hear you: How mothers talk to infants with simulated hearing loss," in *9th Annual Conference of the International Speech Communication Association* Brisbane, QLD).

Lam, C., and Kitamura, C. (2012). Mommy, speak clearly: Induced hearing loss shapes vowel hyperarticulation. *Develop. Sci.* 152, 212–221. doi: 10.1111/j.1467-7687.2011.01118.x

Laukka, P., Linnman, C., Åhs, F., Pissiota, A., Frans, Ö., Faria, V., et al. (2008). In a nervous voice: Acoustic analysis and perception of anxiety in social phobics' speech. *J. Nonverb. Behav.* 32, 195–214. doi: 10.1007/s10919-008-0055-9

Liu, H. M., Kuhl, P. K., and Tsao, F. M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Dev. Sci.* 6, F1–F10, doi: 10.1111/1467-7687.00275

Lorge, I., and Katsos, N. (2019). Listener-adapted speech: Bilinguals adapt in a more sensitive way. *Linguistic Approach. Bilingualism* 9, 376–397. doi: 10.1075/lab.16054.lor

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. 3rd Ed. Mahwah, NJ: Lawrence Erlbaum Associates.

Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., et al. (2015). Mothers speak less clearly to infants than to adults: a comprehensive test of the hyperarticulation hypothesis. *Psychol. Sci.* 26, 341–347. doi: 10.1177/0956797614562453

Miyazawa, K., Shinya, T., Martin, A., Kikuchi, H., and Mazuka, R. (2017). Vowels in infant-directed speech: more breathy and more variable, but not clearer. *Cognit*. 166, 84–93. doi: 10.1016/j.cognition.2017.05.003

Moon, S.-J., and Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *J. Acoust. Soc. Am.* 96, 40–55. doi: 10.1121/1.410492

Palmer, B. (2017). "Language families of the New Guinea area," in *The Languages and Linguistics of the New Guinea Area: A Comprehensive Guide*, ed B. Palmer (Berlin: De Gruyter Mouton), 1–19.

Papoušek, M., and Hwang, S.-F. C. (1991). Tone and intonation in Mandarin babytalk to presyllabic infants: comparison with registers of adult conversation and foreign language instruction. *Appl. Psycholinguist.* 12, 481–504. doi: 10.1017/S0142716400005889

Picheny, M. A., Durlach, N. L., and Braida, L. D. (1986). Speaking clearly for the hard of hearing. *II:* acoustic characteristics of clear and conversational speech. *J. Speech Lang. Hear. Res.* 29, 434–446. doi: 10.1044/jshr.2904.434

Prové, V., Oben, B., and Perrez, J. (2022). Foreigner directed gesture: larger, faster, longer. *Leuven Working Papers in Linguistics* 8, 1–31.

Rattanasone, N. X., Burnham, D., and Reilly, R. G. (2013). Tone and vowel enhancement in Cantonese infant-directed speech at 3, 6, 9, and 12 months of age. *J. Phon.* 41, 332–343. doi: 10.1016/j.wocn.2013.06.001

Rice, L., and Burnham, D. (2010). "Speak clearly mummy: Speech to 6- and 9-month old infants with degraded input," in *Combined Abstracts of 2010 Australian Psychology Conferences*, 34, ed. N. Voudouris and V. Mrowinski. Melbourne: The Australian Psychological Society Ltd.

Rumsey, A. (2017). "Dependency and relative determination in language acquisition: the case of Ku Waru," in *Dependencies in Language*, ed N. J. Enfield (Berlin: Language Science Press), 97–116.

Sarvasy, H., Elvin, J., Li, W., and Escudero, P. (2019). "Vowel acoustics of Nungon, Papua New Guinea," in *Proc. 19th Int. Cong. Phon. Sci.* Melbourne, Australia. pp. 1714–1718.

Sarvasy, H., Elvin, J., Li, W., and Escudero, P. (2020). An acoustic phonetic description of Nungon vowels. *J. Acoust. Soc. Am.* 147, 2891–2900. doi: 10.1121/10.0001003

Sarvasy, H. S. (2013). Across the great divide: how birth-order terms scaled the Saruwaged Mountains in Papua New Guinea. *Anthropologic. Linguistics* 55, 234–255. doi: 10.1353/anl.2013.0012

Sarvasy, H. S. (2015). *Nungon oral Texts*. Corpus deposited with the Firebird Foundation.

Sarvasy, H. S. (2016). "Monolingual fieldwork in and beyond the classroom," in *Chicago Linguistic Society,* eds K. Ershova, J. Falk, J. Geiger, Z. Hebert, R. Lewis, P. Munoz, et al. pp. 471–484.

Sarvasy, H. S. (2017). *A Grammar of Nungon, a Papuan Language of Northeast New Guinea*. Leiden: Brill.

Sarvasy, H. S. (2019). The root nominal stage: a case study of early Nungon verbs. *J. Child Lang.* 46, 1073–1101. doi: 10.1017/S0305000919000357

Sarvasy, H. S. (2020). Acquisition of clause chains in Nungon. *Front. Psychol.* 11:1456. doi: 10.3389/978-2-88966-291-3

Sarvasy, H. S. (2021a). Acquisition of multi-verb predicates in Nungon. First *Lang.* 41, 478–503. doi: 10.1177/01427237211026764

Sarvasy, H. S. (2021c). *Nungon Narratives by Roslyn Ögate and Fooyu (Finisterre-Huon, Morobe Province, Papua New Guinea).* Volume 3 of Texts in the Languages of the Pacific series, Language and Linguistics in Melanesia. Available online at: https://c-cluster-110.uploads.documents.cimpress.io/v1/uploads/6a9a5644-ae30-4d3b-abc6-be8722b4b05b~110/original?tenant=vbu-digital (accessed January 8, 2022).

Sarvasy, H. S. (2021b). *Sarvasy Nungon corpus*. Available online at: https://childes.talkbank.org/access/Other/Nungon/Sarvasy.html (accessed November 8, 2021).

Scarborough, R., Dmitrieva, O., Hall-Lew, L., Zhao, Y., and Brenier, J. (2007). An acoustic study of real and imagined foreigner-directed speech. *J. Acoustic. Soc. Am.* 121, 3044. doi: 10.1121/1.4781735

Schieffelin, B. B. (1990). *The Give and Take of Everyday Life: Language, Socialization of Kaluli Children (No. 9).* CUP Archive.

Schieffelin, B. B. and San Roque, L. (forthcoming). "Language socialization," in *To appear: Oxford Guide to Papuan Languages*, eds N. Evans and S. Fedden, Oxford: Oxford University Press.

Searle, S. R., Speed, F. M., and Milliken, G. A. (1980). Population marginal means in the linear model: an alternative to least squares means. *Amer. Statist.* 34, 216–221. doi: 10.1080/00031305.1980.10483031

Sikveland, R. O. (2006). "How do we speak to foreigners?—phonetic analyses of speech communication between L1 and L2 speakers of Norwegian," in *Proc. Fonetik 2006*, Lund, Sweden, 109–112.

Singh, L., Morgan, J. L., and Best, C. T. (2002). Infants' listening preferences: baby talk or happy talk? *Infancy* 3, 365–394. doi: 10.1207/S15327078IN0303_5

Singh, L., White, K., and Morgan, J. L. (2008). Building a word-form lexicon in the face of variable input: influences of pitch and amplitude on early spoken word recognition. *Lang. Learn. Dev.* 4, 157–178. doi: 10.1080/15475440801922131

Smith, C. (2007). "Prosodic accommodation by French speakers to non-native interlocutor," in *Proc. 16th Int. Cong. Phon. Sci.* Saarbrücken, Germany. *pp*. 313–348.

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Dev. Rev.* 27, 501–532.

Song, J. Y., Demuth, K., and Morgan, J. (2010). Effects of the acoustic properties of infant-directed speech on infant word recognition. *J. Acoust. Soc. Am.* 128, 389–400. doi: 10.1121/1.3419786

Stebbins, T. (2011). *Mali (Baining) grammar: A language of the East New Britain Province, Papua New Guinea*. Canberra: Pacific linguistics.

Steen, V. B., and Englund, N. (2022). Child-directed speech in a Norwegian Kindergarten Setting. *Scand. J. Educ. Res.* 66, 1–14. doi: 10.1080/00313831.2021.1897873

Stern, D. N., Spieker, S., Barnett, R. K., and MacKain, K. (1983). The prosody of maternal speech: Infant age and context related changes. *J. child lang.* 10, 1–15. doi: 10.1017/S0305000900005092

Swanson, L. A., Leonard, L. B., and Gandour, J. (1992). Vowel duration in mothers' speech to young children. *J. Speech Lang. Hear. Res.* 35, 617–625.

Swisher, M. (1984). Signed input of the hearing mothers to deaf children. *Lang. Learn.* 34, 69–85. doi: 10.1111/j.1467-1770.1984.tb01004.x

Trainor, L. J., Austin, C. M., and Desjardins, R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychologic. Sci.* 11, 188–195. doi: 10.1111/1467-9280.00240

Tweissi, A. (1990). Foreigner talk in Arabic: Evidence for the universality of language simplification. *Perspect. Arabic Linguistic.* 9, 296–326. doi: 10.1075/cilt.72.16twe

Uchanski, R. M. (2005). Clear speech. *Handbook Speech Percept.* 2, 708. doi: 10.1002/9780470757024.ch9

Ulichny, P. (1979). Adult foreigner talk: Input language to l 2 Learners. *Studi Italiani di Linguistica Teorica ed Applicata Bologna.* 8, 187–200.

Uther, M., Knoll, M. A., and Burnham, D. (2007). Do you speak E-NG-LI-SH? a comparison of foreigner-and infant-directed speech. *Speech commun.* 49, 2–7. doi: 10.1016/j.specom.2006.10.003

Van de Weijer, J. (2001). Vowels in infant- and adult-directed speech. *Lund Univ. Work. Papers Linguistics* 49, 172–175.

Warren-Leubecker, A., and Bohannon, J. N. (1982). The effects of expectation and feedback on speech to foreigners. *J. Psycholinguist. Res.* 11, 207–215. doi: 10.1007/BF01067564

Warren-Leubecker, A., and Bohannon, J. N. (1984). Intonation patterns in child-directed speech: mother-father differences. *Child Dev.* 55, 1379–1385. doi: 10.2307/1130007

Werker, J. F., Pegg, J. E., and McLeod, P. J. (1994). A cross-language investigation of infant preference for infant-directed communication. *Infant Behav. Dev.* 17, 323–333. doi: 10.1016/0163-6383(94)90012-4

Wesche, M. B., and Ready, D. (1985). "Foreigner talk in the university classroom," in *Input in Second Language Acquisition*, eds S. M. Gass and C. G. Madden (Cambridge, MA: Newbury House), 89–114.

Williams, D., and Escudero, P. (2014). A cross-dialectal acoustic comparison of vowels in Northern and Southern British English. *J. Acoust. Soc. Am.* 136, 2751–2761. doi: 10.1121/1.4896471

Wray, A., and Grace, G. W. (2007). The consequences of talking to strangers: evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua* 117, 543–578. doi: 10.1016/j.lingua.2005.05.005

Xu, N., Burnham, D., Kitamura, C., and Vollmer-Conna, U. (2013). Vowel hyperarticulation in parrot-, Dog- and infant-directed speech. *Anthrozoös.* 26, 373–380. doi: 10.2752/175303713X13697429463592

# Frontiers in Psychology

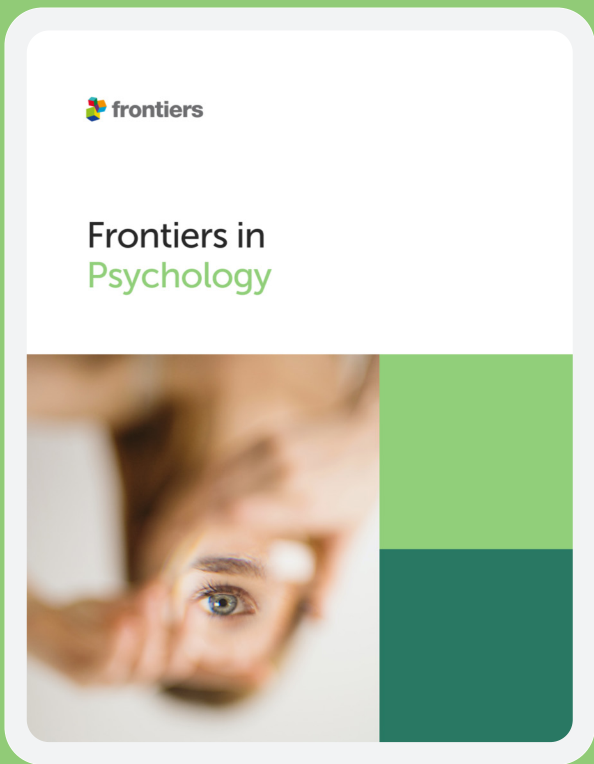
**Paving the way for a greater understanding of human behavior**

The most cited journal in its field, exploring psychological sciences - from clinical research to cognitive science, from imaging studies to human factors, and from animal cognition to social psychology.

---

## Discover the latest Research Topics

See more →




Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact