Infant Behavior and Development 75 (2024) 101945

Contents lists available at ScienceDirect

Infant Behavior and Development

journal homepage: www.elsevier.com/locate/inbede

Developmental cascades of vocal turn-taking connect prelinguistic vocalizing with early language

Vivian Hanwen Zhang [a], Steven L. Elmlinger [a,b], Michael H. Goldstein [a,*]

[a] Department of Psychology, 211 Uris Hall, Cornell University, Ithaca, NY 14853, USA
[b] Department of Psychology, Peretsman Scully Hall, Princeton University, Princeton, NJ 08540, USA

ARTICLE INFO

Keywords:
Turn-taking
Caregiver-infant interaction
Multimodal communication
Babbling
Language production
Communicative development

ABSTRACT

Conversational turn-taking is ubiquitously found in caregiver-infant interactions, and robustly predictive of infant communicative development. Over the first year, infants take quick adult-like vocal turns with caregivers. Many studies have documented the consistency of caregiver responsiveness and its influence on infant rapid language growth. However, few have examined how caregiver responsiveness facilitates extended vocal turn-taking in real-time with infants over the first year. The influence of prelinguistic vocal turn-taking on the emergence of language has also been under-investigated. We analyzed free-play sessions of 30 caregivers and their infants at both 5 and 10 months, and obtained infant language outcomes at 18 months. We examined the developmental consistency (group-level continuity and dyad-order stability) and change of infant volubility, caregiver responses to babbling in vocal, non-vocal and multimodal modalities, and the influence of modality on caregiver-infant vocal turn-taking. Caregiver contingent responsiveness to infant babbling at 5 months predicted vocal turn-taking at 10 months. Developmental increases in prelinguistic vocalizing and vocal turn-taking from 5 to 10 months predicted infant language outcomes at 18 months. At both 5 and 10 months, caregiver vocal responses were more effective in extending turn-taking than non-vocal or multimodal responses. In summary, prelinguistic vocal turn-taking, facilitated by caregiver vocal responsiveness, is positively related to the emergence of early language.

1. Introduction

Conversational turn-taking, characterized by contingent but non-simultaneous exchanges of vocalizations, is both foundational in caregiver-infant interactions, and crucial for infant communicative, social and cognitive development. During these quick back-and-forth exchanges, the dyads coordinate the timing of their vocalizations in a synchronized way (Feldstein et al., 1993; Jaffe et al., 2001; Gratier et al., 2015). Infants produce vocalizations and receive instantaneous feedback from caregivers (Goldstein & Schwade, 2010; Warlaumont et al., 2014). Vocal turn-taking facilitates the production of mature babbling, supports language comprehension, and predicts infant vocabulary growth from 9 to 24 months (Bloom et al., 1987; Zimmerman et al., 2009; Donnelly & Kidd, 2021). Furthermore, it contributes to infant secure attachment and emotional communication outcomes (Jaffe et al., 2001; Gómez & Strasser, 2021). The long-lasting impact of vocal turn-taking on language and general intelligence remains for years and extends into late childhood (Gilkerson et al., 2018). Early vocal turn-taking is robustly found in caregiver interactions with their 5-month infants across

* Corresponding author.
E-mail address: michael.goldstein@cornell.edu (M.H. Goldstein).

https://doi.org/10.1016/j.infbeh.2024.101945
Received 30 June 2023; Received in revised form 1 March 2024; Accepted 14 March 2024

distinct communities in 11 countries (Bornstein et al., 2015).

What features characterize the development of vocal turn-taking? In what ways is the emergence of vocal turn-taking influenced by social responsiveness to early vocalizations? Turn-taking is present in the first months of life. From as early as 1.5 months until the end of the first year, infants act as active participants in rapid vocal turn-taking with caregivers (Bateson, 1975; Beebe et al., 1988; Elias et al., 1986; Snow, 1977). They initiate turn-taking frequently and are able to produce a turn within a second after the end of caregiver's speech (Gratier et al., 2015; Hilbrink et al., 2015). Such rapid infant response timing remains generally consistent during most of the first year but starts to slow down by the end of it. Along with slower responses, infant interruptions of caregivers also decrease over the first year, which demonstrates gradual improvement in their temporal coordination. In contrast, caregiver response timing as well as interruptions of infants stay consistent during and beyond early infancy. The durations of gaps in between or overlaps with the previous infant turns do not change drastically (Hilbrink et al., 2015).

Although the prevalence, developmental significance and temporal changes of caregiver-infant turn-taking have been well-documented, fewer studies have investigated the sources of both consistency and change in prelinguistic vocal turn-taking. For example, though caregiver response timing is developmentally consistent, does the impact of responsiveness on infant vocal production change over developmental time? Differential influences of responsiveness might be mediated by the modality of the response. Recent findings suggest that caregivers and infants tend to match one anothers modality when responding contingently (van der Klis et al., 2023; Rohlfing et al., 2019). When 9-month infants babble, caregivers' vocal responses to babbling are more likely to elicit a continuous vocal turn-taking sequence than caregiver non-vocal or multimodal responses (Zhang et al., under review). Nevertheless, no studies have investigated how this modality effect look longitudinally during the first year. In the present study, we examined how caregiver responses of different modalities facilitated turn-taking over the first year.

Besides turn-taking, the first year of life is a period of rapid growth in other communicative abilities as well. Around 5 months of age, infants show first signs of word recognition, acknowledging frequent words like their own names (Tincoff & Jusczyk, 1999, but see Kartushina & Mayor, 2019 for Norwegian population). Speechlike consonant-vowel syllables are still rare if not absent in their vocal repertoire (Oller, 2000). By 10 months, however, infants are able to comprehend more than 50 words including ones with abstract meanings (Bergelson & Swingley, 2013) and some even start to produce their first words (Frank et al., 2017). Though there is little doubt that infant language capacities grow substantially during this 5-month time window, conflicting evidence has been found with regard to how infant talkativeness changes. Some investigators found infant volubility to increase with age (e.g., Shapiro et al., 2021; Gilkerson & Richards, 2009), while others observed it to either decrease or remain the same depending on the social context (Iyer et al., 2016). Furthermore, in addition to the direction of change, the impact of prelinguistic volubility on future language development has been under-investigated (Binos & Loizou, 2019; Landon & Sommers, 1979).

Infants are not only sensitive perceivers of caregiver speech, they also actively vocalize and engage caregivers in vocal turn-taking (e.g., Iyer et al., 2016; Gratier et al., 2015). However, compared to caregiver talkativeness, few studies have investigated the relationship between infants' own vocalizing and later language outcomes. A recent bilingual study highlighted a specific pathway to productive language skills through language use. In 30-month-old Spanish-English bilingual children, those who spoke more English than they heard had a faster productive vocabulary growth than those who heard more than they spoke (Ribot et al., 2018). The amount of English children produced predicted productive but not receptive language outcomes. Does prelinguistic vocalizing, before first words are produced, have a similar facilitative effect on later productive language skills? In the present study, we investigated the consistency and change of infant prelinguistic volubility, by examining both its group-level continuity and dyad-level stability (following the framework suggested by Bornstein et al., 2017) from 5 to 10 months, and explored how it predicted infant future receptive versus productive language development.

Despite infant language growth, caregiver language input amount has not been found to increase dramatically during the first year (e.g., Ling & Ling, 1974; Sherrod et al., 1977; Belsky et al., 1984). For example, the group-level frequency of caregiver conversational device (i.e., utterances used to establish or maintain contact) as well as information-salient utterances are consistent from 7 to 11 months (D'Odorico et al., 1999). The mean length of utterance in caregiver speech does not change from 6 to 12 months (Suttora & Salerni, 2011). Across different studies and diverse cultures, caregivers contingently respond to around 60% of babbling when interacting with 3-month (Kärtner et al., 2008), 6-month (Gros-Louis et al., 2014) and 9-month infants (Zhang et al., under review). These convergent reports of caregiver response rates suggest that there is robust group-level continuity in caregiver contingent responsiveness to infant babbling during the first year.

The consistency of caregiver responsiveness during early infancy may be crucial, as the effects of responsiveness on language outcome have been suggested to be predictive rather than concurrent. Caregiver verbal sensitivity at 9 months and 13 months does not predict infant language comprehension at the respective months concurrently. Instead, caregiver verbal sensitivity at 9 months predicts child comprehension outcome at 13 months (Baumwell et al., 1997). Similarly, caregiver handing or pointing to toys at 4 months correlate with infant vocalizations at 12 months but not concurrently at 4 months (Ruddy & Bornstein, 1982). Investigators have proposed that there might be a "sleeper" stage of caregiver stimulation during early infancy (e.g., Beckwith & Cohen, 1989). The consistency in caregiver contingent response rate may allow infants to accumulate information and explore the social functions of their own vocalizations. Understanding the instrumental value of their own behaviors facilitates infants' use of vocalizations in social contexts, and may be crucial for their future social and communicative development (Elmlinger et al., 2023; Goldstein et al., 2009). To probe this hypothesis, we investigated how caregiver contingent responses to babbling predict infant future vocal turn-taking with caregivers as well as the emergence of language production.

1.1. The current study

To understand the consistency and change in infant and caregiver contributions to vocal turn-taking over the first year, and to evaluate the effect of prelinguistic vocal turn-taking on later language outcomes, we video-recorded play sessions between caregivers and their infants at both 5 and 10 months of age. We annotated the timing of infant babbles and measured infant volubility during each play session. We coded the timing as well as modality (vocal, non-vocal, multimodal) of caregiver contingent responses to infant babbling, and measured caregiver response rates. Then, caregiver-infant turn-taking bouts were identified and decomposed into individual turns. We analyzed the likelihood of infant babbling again after caregiver responses of different modalities, and the number of vocal turns in a turn-taking bout following different caregiver response modalities.

We hypothesized that we would find caregiver vocal responses to be more facilitative to vocal turn-taking than non-vocal or multimodal responses at both 5 and 10 months. Furthermore, we hypothesized that caregiver contingent responses to babbling at 5 months would show a predictive effect on infant later vocal-turn-taking at 10 months but not a concurrent effect at 5 months. Lastly, we hypothesized that infant volubility, caregiver responsiveness and caregiver-infant vocal turn-taking would be predictive of infant later language outcomes.

2. Materials and methods

2.1. Participants and procedure

Thirty infants (17 females) and their caregivers, recruited from birth announcements in local newspapers and through advertisements, participated in this study. Caregivers received a T-shirt or a book as a gift for their time. The dyads participated in a 15-minute free play session at 5 months of age (infant mean age = 5 months 10 days, range: 4 months 20 days – 5 months 18 days, *SD* = 6.06 days), and a 30-minute free play session at 10 months of age (infant mean age = 9 months 23 days, range: 8 months 24 days – 10 months 29 days, *SD* = 19.84 days). Only minutes 7.5 to 22.5 of the 10-month play session were analyzed, so that the two sessions being compared across developmental timepoints matched in duration. We selected the middle period of the 10-month play session to avoid infant fussiness at the beginning or end of the session, as stranger anxiety was prevalent in 10-month-olds (Brand et al., 2020). Twenty participants (9 females) returned the MacArthur-Bates Communicative Development Inventory: Words & Gestures (CDI; Fenson et al., 2006) at 18 months of age (mean age 18 months 7 days; range 17 months 9 days – 20 months 1 day; *SD* = 21.86 days).

For both free-play sessions, infants visited the lab on campus with the same caregivers. The dyads were asked to play as at home in a 3.65 m x 4.57 m playroom containing toys, and were video recorded by three disguised wall-mounted cameras. The dyads wore wireless microphones so that we could obtain high-quality audio recordings of caregiver speech and infant babbling.

Upon the first lab visit, caregivers' mean age was 33.27 years (range: 23 – 44 years, *SD* = 5.92 years). Twenty-eight caregivers were White, one was Middle Eastern, and one was Black. Almost all participants' primary language spoken at home was English, except for one who primarily used Urdu. Four infant participants also regularly heard German, Hebrew, Mandarin and Yiddish in addition to English. Out of all 30 caregiver participants, one had attainted high school diploma or equivalent, seven associate or vocational degree, seven bachelor's degree, twelve master's degree, and three doctorate degree. Participants who returned CDIs all spoke English as their primary language at home.

Table 1
Definitions of Caregiver Response Categories.

Category	Definition	Example	Count per Session Mean (SD)
Vocal	Any sounds made with the mouth	Speech; laughing	5 M: 33.33 (26.56) 10 M: 36.80 (22.79)
Non-vocal	Physical actions	Shaking a toy; handing a toy to the infant; pointing; touching the infant; clapping	5 M: 5.97 (5.36) 10 M: 5.43 (5.44)
Multimodal	*Vocal* and *Non-vocal* co-occurring	Labeling a toy while picking it up	5 M: 13.07 (12.59) 10 M: 12.83 (9.33)

Note. Caregiver vocal responses included speech and non-speech sounds such as laughing, but did not include vegetative sounds like coughing. The onset and offset of a vocal response were marked by the start and end of the sound. Caregiver vocal responses were segmented if they met two out of the three following criteria: 1) if the vocalizations were separated by gaps longer than 2 seconds, 2) if the pitch went down, showing terminal intonation contour (Venker et al., 2015), and 3) if it was a syntactically complete sentence, or made a complete contribution in conversation. Caregiver non-vocal responses did not include ancillary actions like scratching the infant's nose or wiping drool off the infant's face. The onset and offset of a non-vocal response were marked by changes in the action (e.g., the start and stop of a shaking action). A caregiver multimodal response was coded if there was any overlap between a vocal and a non-vocal response, regardless of which came first. The onset and offset of a multimodal response were marked by the earlier onset and later offset between the co-occurring vocal and non-vocal response.

2.2. Data coding and analysis

2.2.1. Infant babbling, caregiver responses, and their turn-taking bouts

We annotated infant babbling as well as caregiver responses to them with accurate onsets and offsets. All coding was conducted using ELAN annotation software (V 5.9) developed by the Language Archive (https://archive.mpi.nl/tla/elan). Infant babbling included all speech-related prelinguistic vocalizations and did not include vegetative sounds like crying, laughing or coughing (Oller, 2000). We segmented infant babbling based on breath groups (Oller & Lynch, 1992). Vocalizations were counted in the same breath group (therefore as one babble), if they occur within one breath of the infant, or if they have pauses shorter than 1 s in between. We measured infant volubility as the number of babbles an infant produced per minute during each play session (Iyer et al., 2016).

Caregiver responses to infant babbling were coded as behaviors that started after the onset of babbling, and within 2 s of the offset of babbling (Van Egeren et al., 2001; Elmlinger et al., 2019). We coded caregiver responses according to their modalities as vocal, non-vocal or multimodal (Table 1). We calculated caregiver response rates in each modality as the proportion of infant babbles which elicited a caregiver response type. To assess reliability, a second coder randomly selected and independently annotated infant babbling and caregiver responses for 20% of all data recordings in each age. We compared the total counts of behaviors annotated by the two coders for each session with intraclass correlation coefficient (ICC). In 5-month data, the ICC was good for infant babbling (0.79), and excellent for caregiver vocal (0.99), non-vocal (0.92), and multimodal (0.93) responses. In 10-month data, the ICC was excellent for infant babbling (0.98) and caregiver vocal responses (0.98), and good for caregiver non-vocal (0.83), and multimodal (0.84) responses (Koo & Li, 2016).

Turn-taking bouts were identified as rounds of caregiver-infant continuous turn-taking. We used an inclusive definition that counted subsequent turns from caregivers and infants beyond the initial serve-and-return, and found that caregivers and infants were similar in their contributions to extended turn-taking (see Results section). Our definition differed from previous studies, that identified a discrete pair of vocal exchange as turn-taking in caregiver-infant interactions (e.g., Romeo et al., 2018; Donnelly & Kidd, 2021; Hilbrink et al., 2015). Nevertheless, naturalistic conversations often consist of longer sequences of turns and loosely balanced contributions from different speakers. Studies of non-human animals defined turn-taking as a vocal bout that included at least one vocal exchange (e.g., Katsu et al., 2019; Koda, 2004), and observed long turn-taking bouts. We also followed the tradition of research in non-human animals in defining turn-taking bouts. In our study, each turn-taking bout started with a pair of infant babble and caregiver response in vocal, non-vocal or multimodal modality. It included the following vocalizations (from caregivers and infants) that occurred within two seconds from each other. Thus, different turn-taking bouts were separated by pauses of at least two seconds before or after any turns (Fig. 1). For each dyad, we measured the frequency of turn-taking bouts in their interaction by calculating the number of bouts (following different caregiver responses) per minute during the play session. For each turn-taking bout, we measured its length by counting the total number of turns in it.

To assess the consistency and change of infant volubility, caregiver response rates, and turn-taking bouts developmentally from 5 to 10 months, we examined their group mean-level continuity and dyad-order stability, following the framework established by Bornstein and colleagues (2017). For continuity, linear mixed-effect regressions were employed to compare these measures at 5 and 10 months (*lme4* package in R; Bates et al., 2015). Infant volubility was tested as a function of infant age with participants as a random effect. Caregiver response rates, the frequency of turn-taking bouts, and the length of turn-taking bouts were tested with infant age and caregiver response modality (and their interaction) as fixed effects, and participants as a random effect. The results of Tukey post-hoc tests comparing 5 and 10 months, as well as the regression model formulas are reported in Table 2. To evaluate the dyad-level stability in infant volubility, caregiver response rates and turn-taking bouts from 5 to 10 months, we conducted Pearson correlations and reported the results in Table 3.

2.2.2. The effect of caregiver response modality on turn-taking

To study the effect of response modality on turn-taking with infants in both 5 and 10 months, we decomposed caregiver-infant turn-taking into three steps. First, when infants babbled, in what modality did caregivers respond to the babbling? We counted the number of infant babbles and caregiver responses in vocal, non-vocal or multimodal modalities, and compared the proportions of babbles

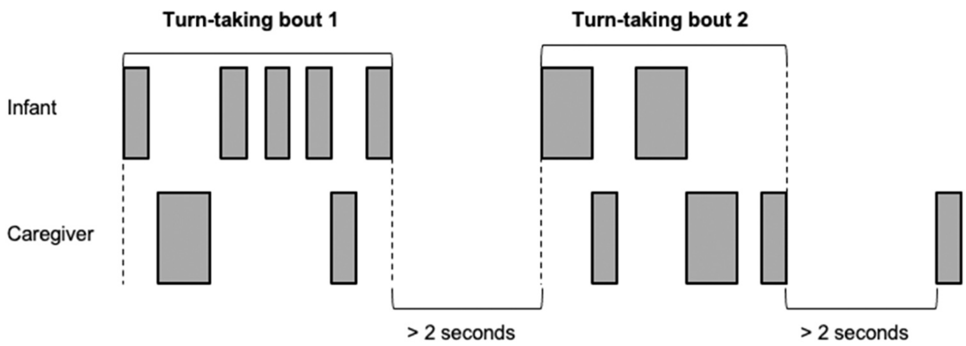


Fig. 1. *Examples of Turn-taking Bouts.* In this illustration, there are two turn-taking bouts separated by pauses of more than 2 s. There are seven turns in bout 1, and five turns in bout 2. Each bout will only count as one even if it contains multiple caregiver responses.

Table 2
Descriptive Statistics and Comparisons Between 5 and 10 Months.

	5 months Mean (SD)	10 months Mean (SD)	Comparison (10 months – 5 months)
A) Infant volubility (babbles/ minute)	2.58 (2.07)	3.34 (1.67)	t (29) = 1.564, p = 0.13
B) Caregiver response rate			
Overall	66.69% (12.65%)	57.10% (20.37%)	t (145) = -1.80, p = 0.07
Vocal	45.73% (15.36%)	40.85% (18.24%)	t (145) = -1.59, p = 0.61
Non-vocal	7.49% (8.03%)	5.35% (5.49%)	t (145) = -0.70, p = 0.98
Multimodal	13.47% (9.35%)	10.90% (9.98%)	t (145) = -0.84, p = 0.96
C) Babbling elicitation rate of caregiver responses			
Overall	0.21 (0.13)	0.24 (0.12)	t (1479) = 1.22, p = 0.22
Vocal	0.14 (0.11)	0.17 (0.10)	t (1479) = 2.12, p = 0.28
Non-vocal	0.02 (0.03)	0.03 (0.04)	t (1479) = 0.55, p = 0.99
Multimodal	0.05 (0.06)	0.05 (0.04)	t (1479) = -0.56, p = 0.99
D) Turn-taking bout frequency (bouts/ minute)			
Overall	1.07 (0.66)	1.22 (0.59)	t (145) = 1.03, p = 0.31
Vocal	0.76 (0.55)	0.88 (0.50)	t (145) = 1.51, p = 0.66
Non-vocal	0.11 (0.12)	0.13 (0.13)	t (145) = 0.21, p = 1.00
Multimodal	0.20 (0.24)	0.21 (0.17)	t (145) = 0.053, p = 1.00
E) Turn-taking bout length (turns/ bout)			
Overall	3.98 (0.95)	3.86 (0.80)	t (856) = -0.14, p = 0.86
Vocal	4.17 (1.11)	4.09 (1.06)	t (970) = -1.30, p = 0.78
Non-vocal	2.95 (1.44)	3.10 (1.12)	t (880) = 0.31, p = 1.00
Multimodal	3.80 (1.25)	3.93 (1.61)	t (766) = 0.002, p = 1.00

Note. The linear mixed-effect model formulas are as follows: A) infant volubility ~ age + (1|subject); B) caregiver response rate ~ age*modality + (1|subject); C) babbling elicitation rate ~ age*modality + (1|subject); D) turn-taking bout frequency ~ age*modality + (1|subject); E) turn-taking bout length ~ age*modality + (1|subject) + (1|subject:modality).

Table 3
Stability of Caregiver-infant Behaviors From 5 to 10 Months.

	Correlation between 5 and 10 months
Infant volubility	r = -0.10, p = 0.61
Caregiver response rate	r = 0.33, p = 0.08
Babbling elicitation rate	r = -0.13, p = 0.50
Turn-taking bout frequency	r = 0.07, p = 0.73
Turn-taking bout length	r = -0.03, p = 0.86

Note. Pearson correlations assessed relations between 5-and 10-month data for each variable. Visualizations of individual developmental changes can be found in Supplementary Figure S1.

responded to in each modality by caregivers.

Second, how did caregiver responses of different modalities compare in eliciting a subsequent infant babble (Fig. 2A)? We measured the babbling elicitation rates of caregiver vocal, non-vocal and multimodal responses. An infant babble was counted as *elicited* if it occurred after caregiver response onset, and within two seconds before or after caregiver response offset. Babbling elicitation rate was calculated as the number of caregiver responses that elicited babbling divided by the total number of caregiver responses. To further understand the time course of babbling elicitation (e.g., were there more babbles elicited 0–0.5 s after caregiver response offset than 1.5–2 s after?), we evenly divided the elicitation time window into bins and obtained babbling elicitation rate for every time bin. There were 9 time bins each with a duration of 0.5 s, ranging from 2.5 s before to 2 s after caregiver response offset. The durations of caregiver responses partially constrained the time course of babbling elicitation, because for a babble to be counted as elicited, it had to occur after caregiver response onset. For example, if a caregiver response was shorter than 1 s, then no babble could possibly be elicited by this response 1 s before its offset. The average duration of caregiver responses was around 1.5 s (5 month: M = 1.51 s, SD = 0.88 s, range: 0.29 s - 8.90 s; 10 month: M = 1.39 s, SD = 1.05 s, range: 0.23 s - 17.05 s). Thus, it was expected that time bin − 2 (−2.5 s ~ −2 s), − 1.5 (−2 s ~ −1.5 s) and − 1 (−1.5 s ~ −1 s) would have low babbling elicitation rates.

Third, how did caregiver vocal, non-vocal and multimodal responses compare in extending further vocal turn-taking with infants (Fig. 2B)? We counted the number of vocalizations in turn-taking bouts, and compared the mean lengths of turn-taking bouts following caregiver vocal, non-vocal or multimodal responses. The analyses of real-time interactions described above were applied separately to both 5- and 10-month-old infants' interactions with caregivers. For each age, we tested the effect of caregiver response modality on babbling elicitation rates across time bins with linear mixed-effect regressions and included participants as a random effect (formula: babbling elicitation rate ~ modality * time bin + (1|subject)). Then a Type III sum of squares test was applied to the regression model

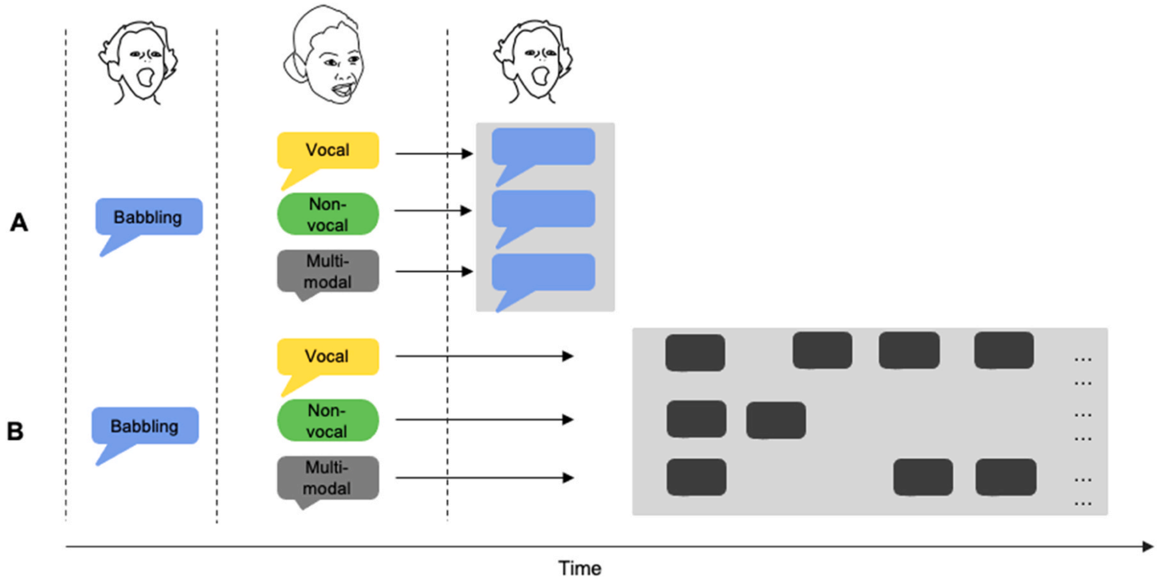


Fig. 2. *Turn-taking Decomposition.* A) We compared the babbling elicitation rate of caregiver responses of different modalities. B) Turn-taking bouts started with an infant babble and a caregiver response in vocal, non-vocal or multimodal modality, and included the subsequent vocal turns. We compared the length (total number of turns) per turn-taking bout following different caregiver response modalities.

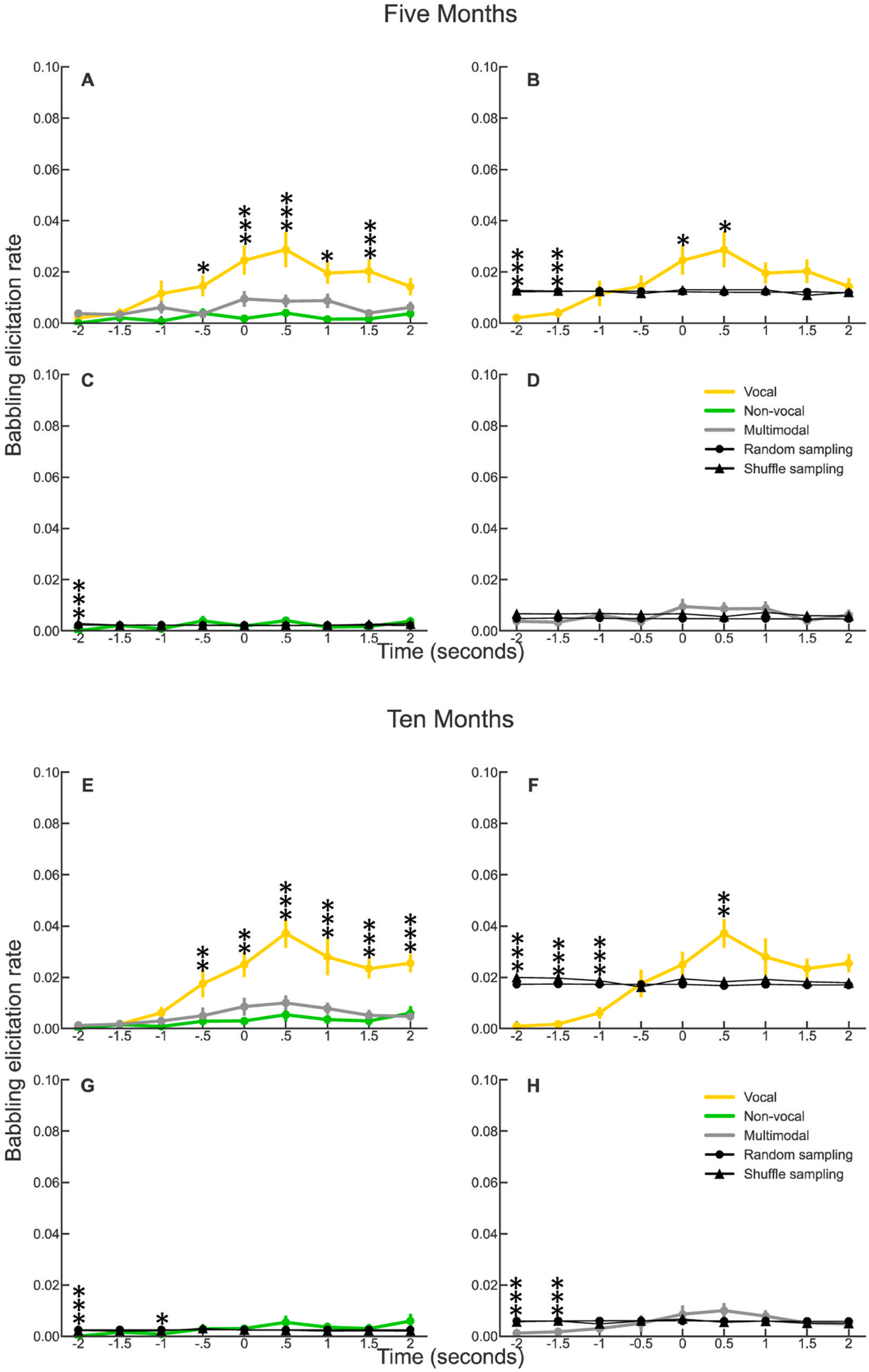
to find the main effects and interaction ("anova" function in R). The effect of caregiver response modality on turn-taking bout length was also tested with a linear mixed-effect regression model, where participants were included as a random effect (formula: turn-taking bout length ~ modality + (1|subject) + (1|subject:modality); *lme4*; Bates et al., 2015). Lastly, to assess the group-level developmental continuity of the effect of caregiver response modality, we added infant age as a fixed effect into the linear mixed-effect regressions. The results of paired t-tests comparing 5- and 10-month elicitation rates and bout lengths for each modality are reported in Table 2.

2.2.3. Determining chance-level babbling elicitation rates

Babbling elicitation rates, as described above, can be influenced by the different frequencies of caregiver vocal, non-vocal and multimodal responses. For example, if caregiver non-vocal responses rarely occurred while vocal responses occurred frequently, by chance alone it would be less likely for a babble to happen following a caregiver non-vocal response than a vocal one. As a result, the babbling elicitation rate of caregiver non-vocal responses would be lower than vocal. To control for chance levels of babbling elicitation, we implemented a random sampling analysis (Suanda et al., 2016; Zhang et al., under review) to determine chance-level babbling elicitation rates of vocal, non-vocal and multimodal responses. We compared the chance-level rates with the elicitation rates calculated from observed experimental data. Significant differences between chance-level and observed rates would suggest that effects of caregiver responses in eliciting babbling were different from those expected by chance.

In the random sampling analysis, we assigned the same amount of caregiver responses (as in our observed data) to random timestamps within the 15-minute play sessions. We kept the amount as well as timing of infant babbling the same as in observed data. Then we calculated the babbling elicitation rates from the randomized dataset. We repeated this sampling process 1000 times and obtained the average chance-level elicitation rate of randomized caregiver responses. We tested the difference between the chance-level and the observed elicitation rates, using one-sample t-tests for each of the nine time bins. A Bonferroni correction for multiple comparisons was then applied to the p-values (corrected p-values = p-values x 9 comparisons).

One potential issue with the random sampling method was that babbling is usually produced in bursts, as opposed to uniformly distributed over time (Abney et al., 2017). Thus, a complete random sampling of timepoints might result in assigning responses in a distribution that did not correspond to the bursty pattern of infant vocalizing. To avoid placing caregiver responses to timepoints when they were realistically unlikely to happen in the play session, we implemented an additional sampling analysis. We retained caregiver response timestamps and shuffled their order across all subjects. Then we calculated the babbling elicitation rates from the shuffled dataset. We repeated this sampling process 1000 times and obtained the average chance-level elicitation rate of shuffled caregiver responses. We conducted one-sample t-tests for each of the nine time bins to test the difference between chance-level (from shuffle sampling) and observed elicitation rates, and applied a Bonferroni correction for multiple comparisons (corrected p-values = p-values x 9 comparisons). As there were more than a thousand caregiver responses across thirty participants, the distribution of timepoints after 1000 shuffles were similar to those from a random assignment (see Supplementary Fig. S2 for an example). As a result, chance-level elicitation rates from both sampling analyses were indistinguishable (Fig. 3). The statistical results of these one-sample t-tests can be found in Supplemental Table S3–6. The code for data processing and sampling, de-identified behavioral and CDI data, and an example ELAN annotation file are available at https://osf.io/g5zrh/?view_only= 2055609c160a429e9da690f58f33f2bc.



(caption on next page)

Fig. 3. *Babbling Elicitation Rates of Caregiver Vocal, Non-vocal and Multimodal Responses at 5 and 10 Months.* The x-axis represents time (seconds) after caregiver response offset, zero indicating caregivers' response offset. Negative values along the x-axis indicate overlap between elicited babbling and caregiver responses, while positive values indicate silent gaps between turns. The y-axis is babbling elicitation rate. A) and E) Comparisons of babbling elicitation rates of different caregiver responses. Tukey post-hoc tests show significant differences between vocal and the other two modalities at multiple time bins, and no difference between non-vocal and multimodal at any time bin. The significance levels of the Tukey tests for each time bin are based on the weaker pairwise difference between either vocal – non-vocal, or vocal – multimodal. For example, at a given time bin, if vocal elicitation rate is higher than that of non-vocal with $p < .001$ and multimodal with $p < .01$, we conservatively indicated $p < .01$ for that time bin. B-D) and F-H) Comparisons of babbling elicitation rates of responses to chance levels obtained from both random and shuffle sampling analyses. The significance levels of the one-sample t-tests for each of the nine time bins are based on comparisons with random sampling analysis. Elicitation rates of vocal responses significantly differ from chance-level rates, at several time bins where there was a significant modality effect. This result showed that the observed effects between modalities were not due to mere chance. P-values were Bonferroni-corrected * *p*-corrected < .05, * * *p*-corrected < .01, * ** *p*-corrected < .001. See Supplementary Table S1–6 for statistical results.

3. Results

3.1. Developmental continuity and stability in measures of caregiver-infant interactions

We assessed the group mean-level continuity and dyad-order stability in infant vocalizing and caregiver responsiveness. On the group level, infant volubility, caregiver response rate, and caregiver-infant turn-taking were continuous and did not differ significantly between 5 and 10 months (Table 2). At both 5 and 10 months, caregiver vocal responses facilitated vocal turn-taking with infants more effectively than non-vocal or multimodal responses. The babbling elicitation rates of different caregiver response modalities did not differ significantly at 5 and 10 months. The number of turns in turn-taking bouts following different response modalities did not differ between 5 and 10 months. Caregivers and infants contributed similarly at both ages. Across turn-taking bouts, the average proportion of infant turns was 41.93% ($SD = 12.96\%$) at 5 months and 43.15% ($SD = 13.02\%$) at 10 months. Although we observed continuity in these measures on the group level, the behaviors of individual dyads were not stable from 5 to 10 months. Infant volubility, caregiver response rate and caregiver-infant turn-taking at 5 and 10 months did not correlate significantly (Table 3). These measures of individual dyads changed in different directions and to different degrees from 5 to 10 months (Supplementary Fig. S1).

3.2. Caregiver vocal responses facilitate turn-taking with infants in both 5 and 10 months

At both 5 and 10 months, we found that caregiver vocal responses facilitated turn-taking more effectively than non-vocal or multimodal responses. When 5- and 10-month-old infants babbled, caregivers responded most frequently with vocal responses (Table 1). We tested the number of caregiver responses as a function of response modality and infant age (formula: count of caregiver responses ~ modality * age + (1|subject)). We found a main effect of caregiver response modality ($F(2, 145) = 58.04, p < 0.0001$), but did not find an effect of infant age or interaction ($ps > 0.59$). Tukey post-hoc tests showed that there were significantly more caregiver vocal responses than non-vocal ($t(145) = 10.10, p < 0.001$) and multimodal responses ($t(145) = 8.29, p < 0.001$).

In turn, caregiver vocal responses were more likely to elicit a following infant babble than non-vocal or multimodal responses (Fig. 3). Type III sum of squares revealed a main effect of caregiver response modality (5-month: $F(2, 58) = 20.67, p < 0.001$; 10-month: $F(2, 58) = 41.83, p < 0.001$), a main effect of time bin (5-month: $F(8, 696) = 6.27, p < 0.001$; 10-month: $F(8, 696) = 11.99, p < 0.001$)), and a significant interaction between the two factors on babbling elicitation rates (5-month: $F(16, 696) = 3.03, p < 0.001$; 10-month: $F(16, 696) = 4.55, p < 0.001$)). Tukey post-hoc tests showed that the elicitation rates were significantly different between vocal and the other two modalities at multiple time bins. There was no significant difference between non-vocal and multimodal modality at any time bins. Infant babbling elicited by caregiver vocal responses occurred most frequently within 0–0.5 s after caregiver response offset, while those elicited by caregiver non-vocal or multimodal responses were distributed more uniformly without a peak in timing (Fig. 3A & 3E). The statistical results of these post-hoc t-tests can be found in Supplementary Table S1–2. We compared the observed babbling elicitation rates with the chance-level rates obtained from random sampling and shuffle sampling analysis. We found that the observed rate of caregiver vocal responses was significantly different from what we would expect by chance in multiple time bins, while those of non-vocal or multimodal responses were generally similar to estimated chance levels (Figs. 3B-D, 3F-H). The statistical results of these one-sample t-tests including the Bonferroni-corrected p-values can be found in Supplementary Table S3–6.

Third, at both 5 and 10 months, extended vocal turn-taking bouts following caregiver vocal responses contained were significantly longer in length than those following caregiver non-vocal responses (Fig. 4; 5-month: $F(2, 461.78) = 4.64, p = 0.009$; 10-month: $F(2, 542.82) = 3.08, p = 0.047$). Post-hoc paired t-tests showed that, at 5-months, bouts following vocal responses ($M = 4.17, SD = 1.11$) were comprised of significantly more turns than those following non-vocal responses ($M = 2.95, SD = 1.44$; $\beta = 1.40, t = 2.93, p = 0.0095$). There was no significant difference between bouts following vocal and multimodal responses ($M = 3.80, SD = 1.25$; $\beta = 0.46, t = 1.26, p = 0.42$), and no significant difference between non-vocal and multimodal ($\beta = 0.94, t = 1.72, p = 0.20$). At 10-months, bouts following vocal responses ($M = 4.09, SD = 1.06$) were comprised of significantly more turns than those following non-vocal ($M = 3.10, SD = 1.12$; $\beta = 0.97, t = 2.47, p = 0.037$). There was no significant difference between bouts following vocal and multimodal responses ($M = 3.93, SD = 1.61$; $\beta = 0.11, t = 0.33, p = 0.94$). There was no significant difference between those following non-vocal and multimodal responses ($\beta = 0.86, t = 1.87, p = 0.15$). In summary, from 5 to 10 months of age, there is continuity in the modality of caregiver responses to babbling, and vocal modality facilitated turn-taking most effectively in both ages.

3.3. Predictive effect of caregiver responsiveness at 5 months on turn-taking at 10 months

Caregiver contingent responses to infant babbling at 5 months predicted caregiver-infant turn-taking at 10 months. We found that caregiver overall response rate (overall responses included vocal, non-vocal and multimodal responses) at 5 months significantly predicted caregiver-infant turn-taking bout frequency ($r = 0.45$, $p = 0.01$), and turn-taking bout length ($r = 0.46$, $p = 0.01$) at 10 months (Fig. 5B, D). The more frequently caregivers responded to infant babbling contingently during the 5-month session, the more vocal turn-taking the dyads produced during the 10-month session. Caregiver response rate at 5 months predicted turn-taking frequency and length at 10 months. However, caregiver response rate at 5 months did not correlate concurrently with those measures at 5 months (Fig. 5A, C; $ps > 0.28$).

The significant predictive effects of caregiver overall response rate were mainly driven by vocal responses (Fig. 6A, D). Caregiver vocal response rate at 5 months significantly and positively predicted the frequency of caregiver-infant turn-taking bouts at 10 months ($r = 0.37$, $p = 0.04$). It trended towards a positive correlation with turn-taking bout length at 10 months ($r = 0.33$, $p = 0.07$). In contrast, caregiver non-vocal or multimodal response rate at 5 months did not significantly predict any turn-taking measures at 10 months (Fig. 6B-C, E-F $ps > 0.54$). We also investigated the relationship between caregiver response rates at 5 months and infant volubility at 5 and 10 months. However, we did not find a significant correlation between caregiver overall, vocal, non-vocal or multimodal response rates and infant volubility ($ps > 0.12$).

3.4. Predicting later receptive and productive language development

Twenty participants returned the CDI at 18 months of age. The average number of spoken words reported in CDI forms was 123.5 ($SD = 88.59$) and understood words was 266 ($SD = 83.93$), which was comparable to previous studies of 18-month children (e.g., Fernald et al., 2013; Frank et al., 2017). We conducted Pearson correlations between infant language outcome at 18 months, and infant volubility, caregiver response rates, and turn-taking bout measures at 5 and 10 months of age. The number of words infants understood at 18 months was predicted by caregiver-infant turn-taking bout length at 10 months ($r = 0.49$, $p = 0.03$). It was also significantly predicted by developmental increases in turn-taking bout length ($r = 0.53$, $p = 0.02$) as well as frequency ($r = 0.49$, $p = 0.03$; Fig. 7). The number of words infants spoke at 18 months was predicted by caregiver overall response rate at 10 months ($r = 0.44$, $p = 0.045$). It was also predicted by developmental increases in infant volubility ($r = 0.46$, $p = 0.04$), and in turn-taking bout frequency ($r = 0.61$, $p = 0.006$) from 5 to 10 months (Fig. 8).

4. Discussion

We examined the developmental changes of prelinguistic vocal turn-taking from 5 to 10 months, and investigated their sources as well as predictive effects on the emergence of language. Caregiver responsiveness to babbling at 5 months facilitated vocal turn-taking

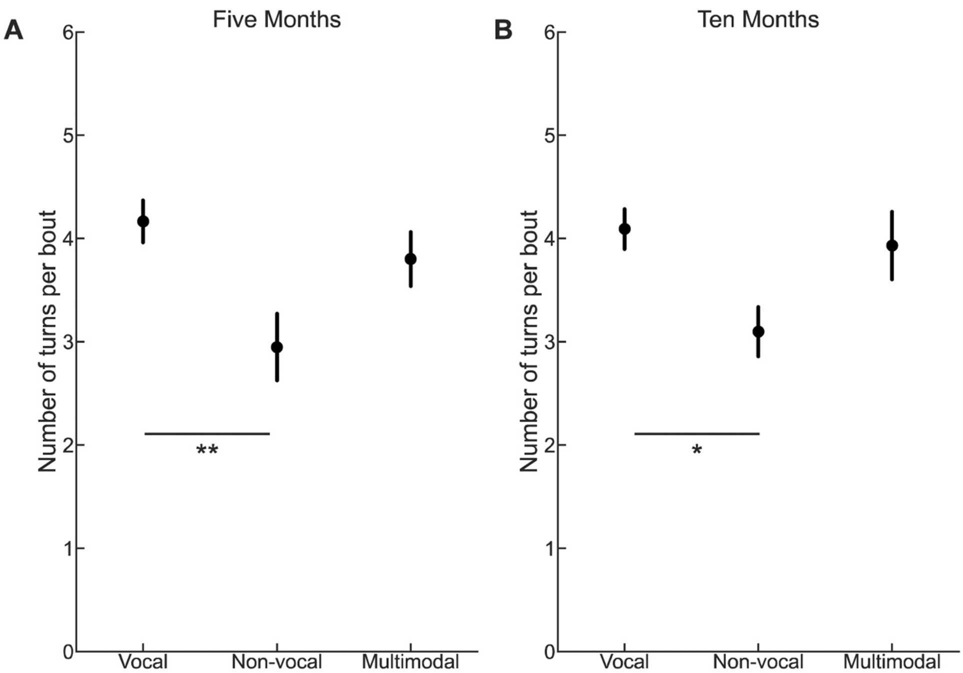


Fig. 4. *Turn-taking Bout Length Following Caregiver Responses at 5 and 10 Months.* At both 5 and 10 months, turn-taking bouts following caregiver vocal responses contained significantly more turns than those following non-vocal responses * p < 0.05, * * p < 0.01, * ** p < 0.001.

at 10 months, and caregiver responsiveness at 10 months predicted infant language production at 18 months. Developmental increases in vocal turn-taking from 5 to 10 months predicted infant receptive and productive language outcomes at 18 months, and increases in infant volubility predicted productive language outcomes. Across participants, measures of caregiver-infant interactions were continuous across the 5-months of development, but there were differential developmental changes on the dyadic level. Furthermore, we analyzed real-time effects of caregiver response modality on vocal turn-taking with infants. At both 5 and 10 months, caregiver vocal responses were the most effective in facilitating vocal turn-taking with infants, compared to non-vocal and multimodal responses. Together, these findings suggest that one of the important sources of developmental growth in prelinguistic turn-taking is caregiver responsiveness, especially vocal responses to babbling. Developmental increases in prelinguistic turn-taking predict the emergence of language.

4.1. *Prelinguistic turn-taking showed high continuity and low stability over development*

Continuity describes consistency in the group mean of a variable across time, while *stability* describes the relative standing of individuals in a group across time (Bornstein et al., 2017). These parameters are independent. We found high continuity and low stability of caregiver-infant turn-taking from 5 to 10 months. On the group level, we did not find significant changes in turn-taking frequency or length from 5 to 10 months. However, by examining interactional behaviors in each individual dyad, we found substantial dyad-level changes from 5 to 10 months. Almost all participants increased or decreased the frequency and length of vocal turn-taking to some degree (Supplementary Fig. S1). However, since there was no systematic change, the group-level statistics appeared to be continuous. This pattern was one of several possibilities described by Bornstein et al. (2017). Why did some dyads produce relatively more and longer turn-taking 5 months later, while others decreased their level of engagement? Developmental increases or decreases in

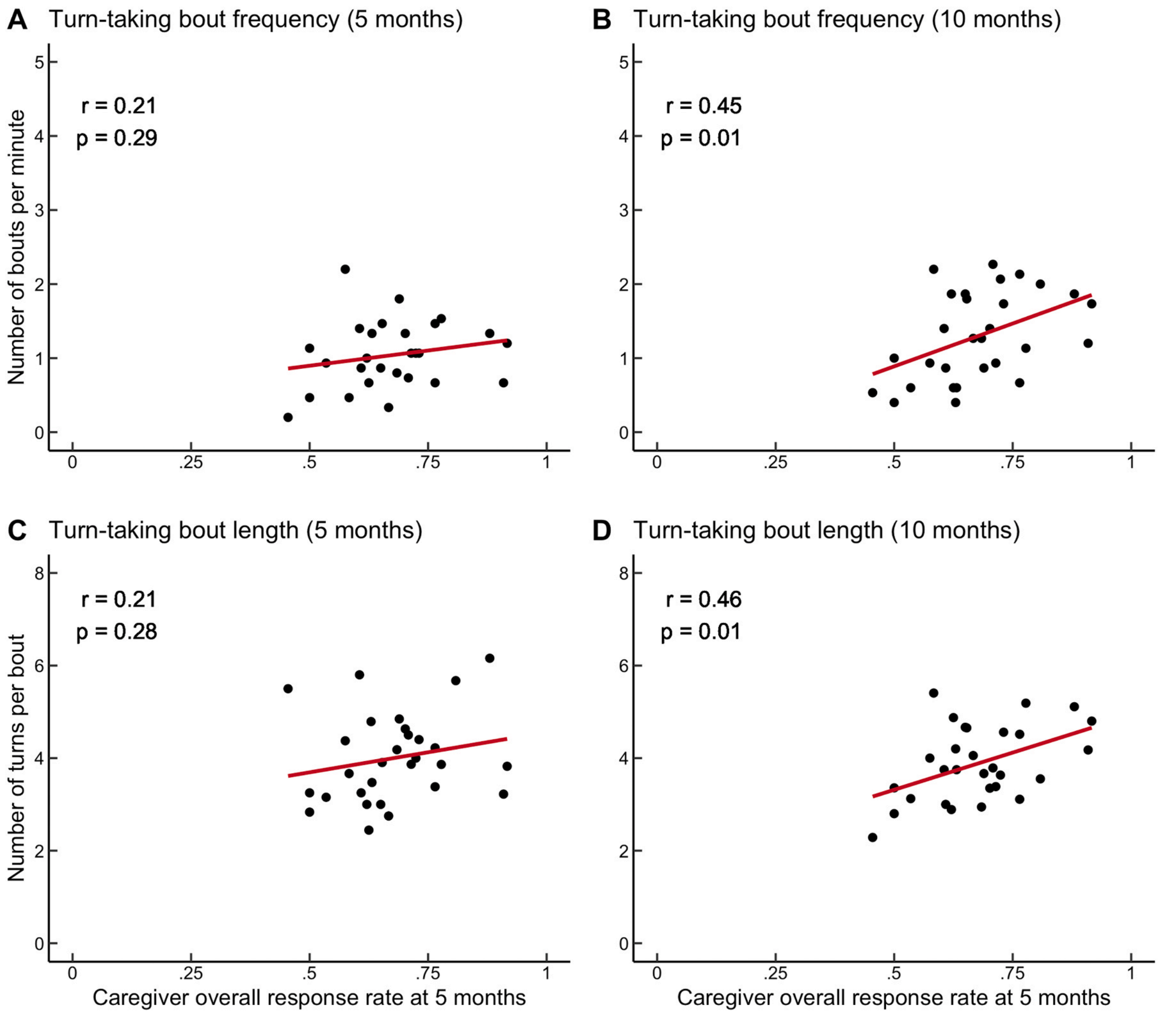


Fig. 5. *Predictive Effects of Caregiver Overall Responses at 5 Months.* Pearson correlations were conducted between 5-month caregiver overall response rate, and turn-taking bout measures (frequency, length) at 5 and 10 months.

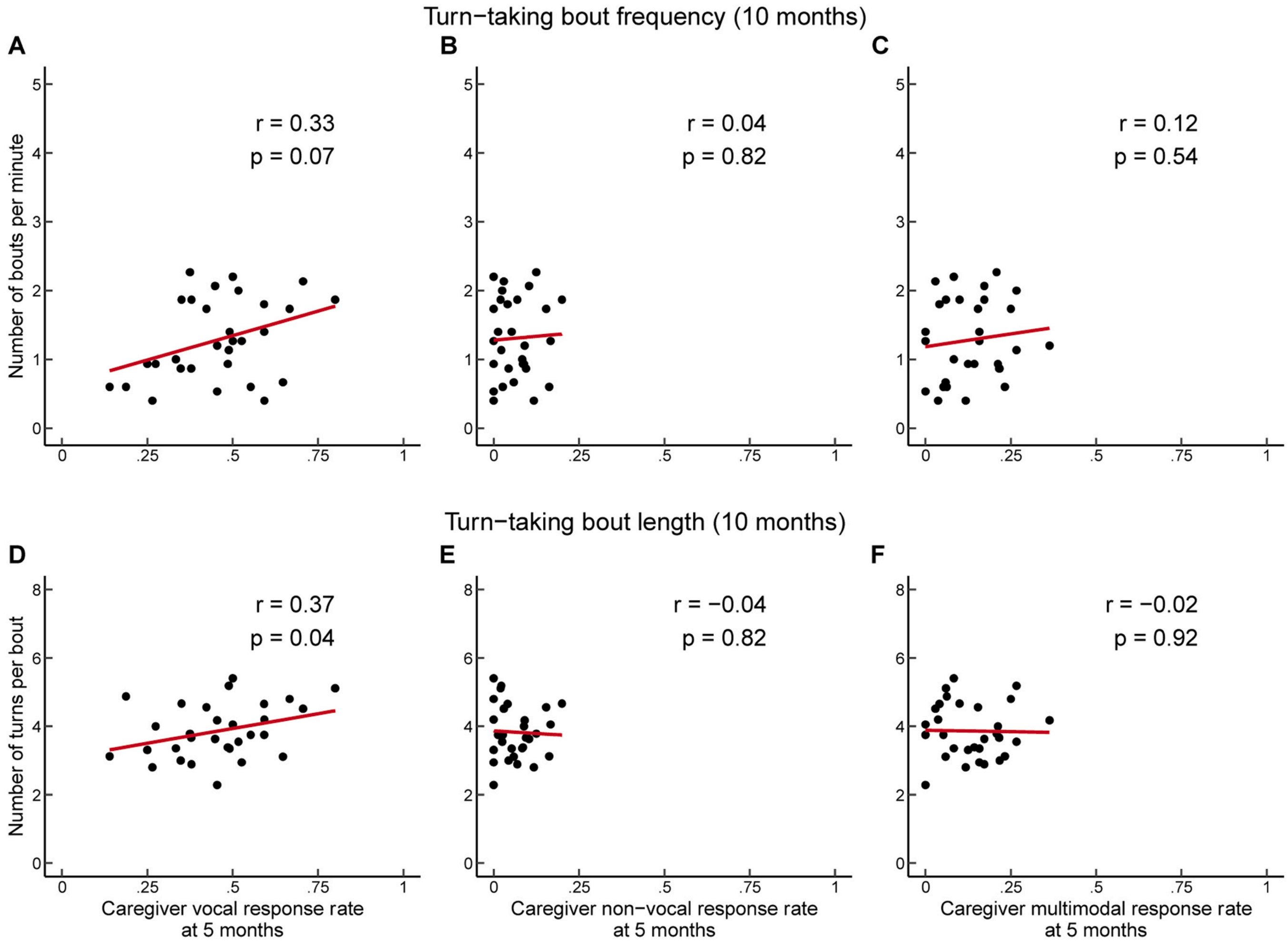


Fig. 6. *Predictive Effects of Caregiver 5-month Responses by Modality.* Pearson correlations were conducted between 5-month caregiver response rates of vocal, non-vocal and multimodal modalities, and turn-taking bout measures (frequency, length) at 10 months.

prelinguistic vocal turn-taking may be shaped by aspects of caregiver-infant everyday interactions, and may influence infant future language development. However, fewer studies have investigated both the group-level continuity and the dyad-level stability in caregiver-infant behaviors. Even when group-level measures appear continuous over development, individual- or dyadic-level changes may lead to infant differential growth in language skills, as discussed below.

4.2. Sources of developmental change in prelinguistic turn-taking: the delayed and modality-specific effect of caregiver responsiveness

Caregiver responsiveness, especially vocal responses to babbling, is an important source of the development (on real-time and longitudinal timescales) of prelinguistic turn-taking. On the longitudinal level, we found that infants who received higher contingent response rates at 5 months, engaged in more frequent and longer vocal turn-taking bouts with their caregivers at 10 months. This significant predictive effect of caregiver responsiveness was mainly driven by vocal responses. Non-vocal or multimodal responses did not positively predict any turn-taking measures. Furthermore, we did not find the same relationship between responsiveness at 5 months and turn-taking measures at 5 months. This finding agreed with previous research suggesting that the effect of caregiver responsiveness in early infancy has a delayed rather than concurrent influence on infant language outcomes (e.g., Baumwell et al., 1997; Beckwith & Cohen, 1989). Since birth, infants vocalize frequently, and caregivers consistently respond to infants and engage them with social interactions. However, infants do not learn that their babbling changes the behaviors of caregivers until 3 to 5 months of age (Elmlinger et al., 2023; Goldstein et al., 2009), which shows that infants build social expectations gradually. They accumulate evidence of the efficacy of their vocalizations from month-long contingent and prompt interactions with caregivers (Ritwika et al., 2020). Our findings suggest that infants who receive more contingent responses to their own babbling, through 5-months of accumulated stimulation, may have a stronger expectation of the instrumental value of their babbling. As a result, they engage in vocal turn-taking with caregivers more frequently, and with more turns, 5 months later.

The predictive impact of caregiver responsiveness was the most robust on the dyadic level, instead of on the behavior of either the infant or caregiver alone. Caregiver responsiveness at 5 months only weakly predicted itself at 10 months, and did not predict infant volubility. Nevertheless, responsiveness at 5 months significantly predicted vocal turn-taking between the two parties at 10 months. Similarly, findings from older children showed that dyadic-level vocal turn-taking had a stronger impact than individual measures on children's language development (e.g., Romeo et al., 2018).

Besides longitudinal influence, caregiver responsiveness also facilitated turn-taking in real-time. With fine-grained temporal

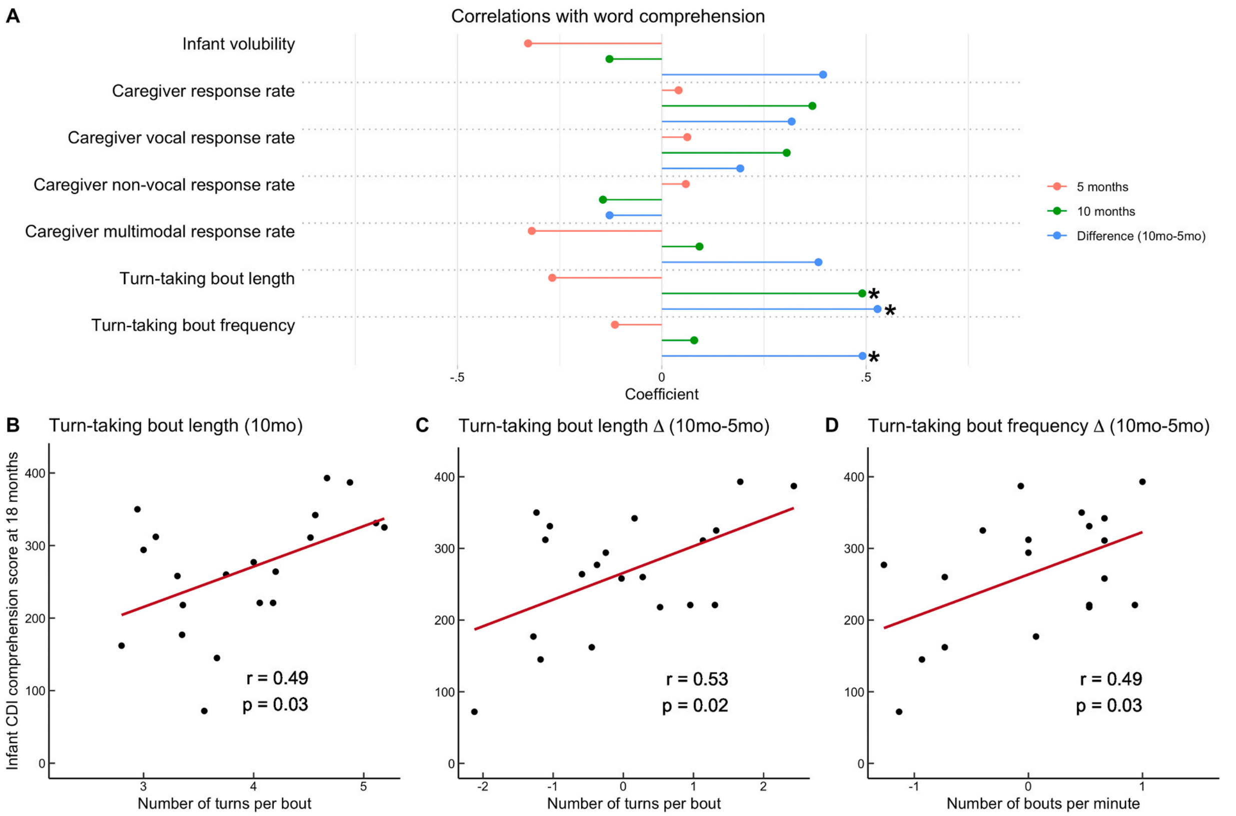


Fig. 7. *Predicting Infant Word Comprehension at 18 Months.* Pearson correlations were conducted between infant CDI comprehension scores at 18 months, and caregiver and infant measures at 5 and 10 months of age. A) Correlation coefficients and significance levels are indicated by the direction and length of the lines. B-D) Scatter plots of significant correlations are graphed * $p < 0.05$, * * $p < 0.01$, * ** $p < 0.001$.

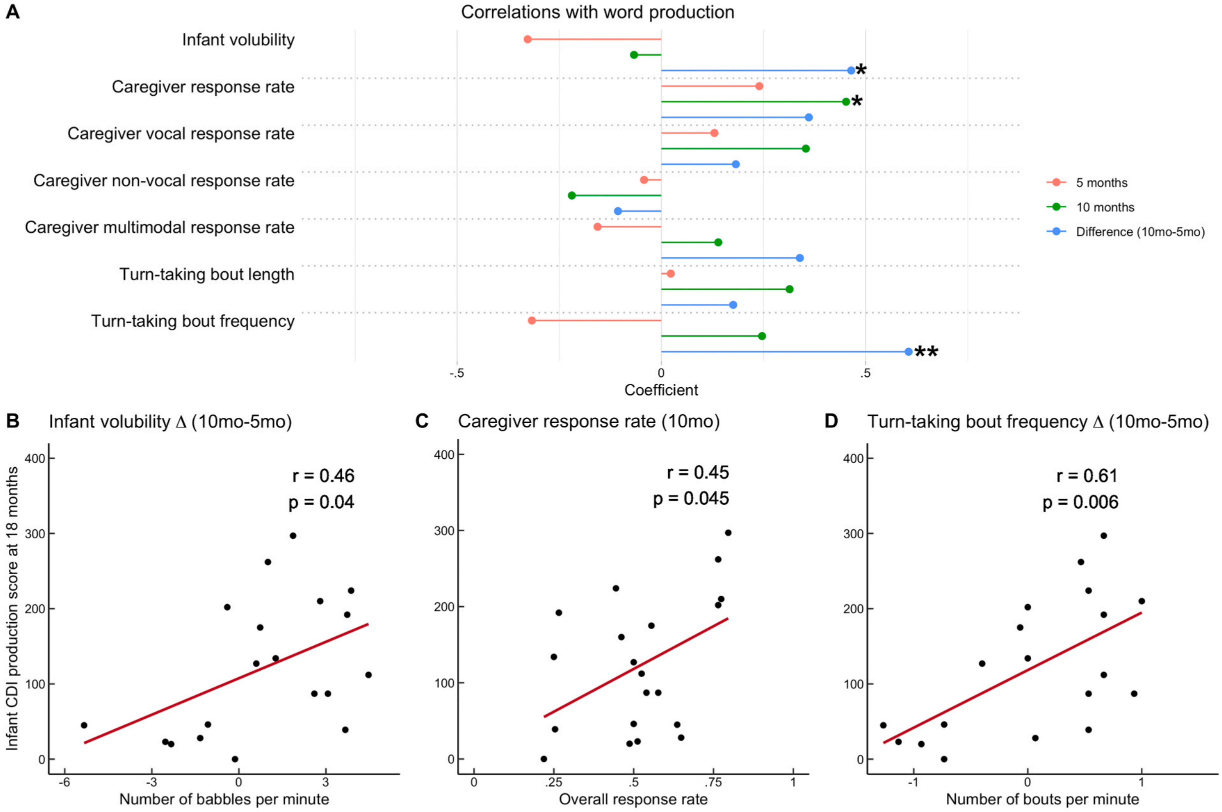


Fig. 8. *Predicting Infant Word Production at 18 Months.* Pearson correlations were conducted between infant CDI production scores at 18 months, and caregiver and infant measures at 5 and 10 months of age. A) Correlation coefficients and significance levels are indicated by the direction and length of the lines. B-D) Scatter plots of significant correlations are graphed * $p < 0.05$, * * $p < 0.01$, * ** $p < 0.001$.

analysis, we found that caregiver contingent vocal responses to babbling were the most effective in extending prelinguistic vocal turn-taking with infants. At both 5 and 10 months, caregivers responded to infant babbling most frequently with vocal responses. In turn, their vocal responses were the most likely to elicit subsequent babbling and lead to the longer turn-taking with infants. When infants act, caregivers tend to respond to infant non-vocal behaviors with non-vocal responses, and to infant bimodal behaviors with multimodal responses (van der Klis et al., 2023; Yurkovic et al., 2021). Similarly, we found that caregivers respond to infant vocal responses. Caregiver-infant contingent exchanges seemed to match each other's modality. As a result, their turn-taking interactions stayed within one modality.

Previous research proposed that the effect of caregiver responsiveness is specific to its response target, rather than globally influential (Baumwell et al., 1997; Paavola et al., 2006). For example, caregiver contingent responses to babbling predicted infant language outcomes but not play outcomes, while in turn caregiver contingent responses to play behaviors predicted infant play but not language outcomes (Tamis-LeMonda et al., 1996). Similarly, maternal responsiveness to infant activities at 5 months predicted infant symbolic play and attention span, but not infant language comprehension at 13 months (Bornstein & Tamis-Lemonda, 1997). Furthermore, caregiver mental knowledge of child language and play behaviors appeared to be unrelated to each other and domain-specific (Tamis-LeMonda et al., 1998). These studies provided evidence suggesting that the influence of caregiver responsiveness is modality- or context-specific. Nevertheless, they focused on different domains of development (e.g., language, play behaviors), rather than on modalities (e.g., vocal, non-vocal) in caregiver responsiveness directly. Our study, among other recent findings (van der Klis et al., 2023; Yurkovic et al., 2021), further revealed the specific effects of caregiver responsiveness at the perceptual level.

What might be the mechanism behind this modality-matching phenomenon? Van der Klis and colleagues (2023) suggested that modality matching might be a signal of high synchrony between caregiver-infant dyads (Abney et al., 2014). Behavioral synchrony in caregiver-infant play was found to have an impact on infant cognitive development (Thompson & Trevathan, 2009), word learning (Matatyaho & Gogate, 2008), general intelligence (Feldman, 2007), and school adjustment (Harrist et al., 1994; Leclère et al., 2014). Although this proposal explained that modality matching might be a result of behavioral synchrony, it does not explain how behavioral synchrony was achieved.

Another potential mechanism for the modality effect may be through infant attentional constraints (Zhang et al., under review). Caregiver non-vocal or multimodal responses to infant babbling were less likely to elicit a following infant babble, compared to vocal responses. The reduction in babbling might be a result of the interference from infant increased attentional loads when they process a change of modality in turn-taking. In adult vocal turn-taking interactions, when conversation partners' voices or voice locations changed, interactants had to process this difference with increased attentional loads. As a result, their performance in designated cognitive tasks decreased (Lin & Carlile, 2015, 2019). Like processing a change in voice or voice location, perceiving caregiver responses in a different modality other than vocal might induce a similar attentional demand in infants. As a result, infant subsequent coordination of babbling was inhibited, and infants were unlikely to babble again after caregiver non-vocal or multimodal responses. We did not directly test the attentional constraint hypothesis in the present study. Future studies should further explore the effect of modality in turn-taking interactions, and investigate the underlying mechanisms for the modality matching phenomenon.

4.3. Predictive effects of developmental increases in prelinguistic turn-taking

Developmental increases in the frequency of prelinguistic turn-taking from 5 to 10 months predicted both language comprehension and production abilities. Developmental increases in the length of prelinguistic turn-taking predicted infant language comprehension, and increases in infant volubility predicted later language production. Though previous studies have reported the significance of developmental changes in maternal responsiveness (e.g., Bornstein et al., 1999; Hirsh-Pasek & Burchinal, 2006; Landry et al., 2006), our study is one of the few that connects changes in prelinguistic and dyadic-level behaviors (i.e., turn-taking) with later language development.

We identified turn-taking bouts based on fine-grained temporal structures, and counted the number of turns in each continuous bout. Across turn-taking bouts, caregivers and infants contributed similar amounts of turns. Our definition of turn-taking bouts is distinct from previous studies, which define caregiver-infant turn-taking as a discrete pair of vocal exchange between caregivers and infants, or any switch from one speaker to another (e.g., Romeo et al., 2018; Donnelly & Kidd, 2021; Hilbrink et al., 2015). By focusing on only one turn transition, these strict definitions enable researchers to analyze the timing of turn-taking and ensure equal contribution from the two speakers. However, naturalistic turn-taking, conversations or dialogues usually contain longer sequences of utterances and uneven contributions from different parties. Studies of non-human animals defined turn-taking as a vocal bout that included at least one vocal exchange (e.g., Katsu et al., 2019; Koda, 2004). They found that Japanese macaques, a highly social primate species, took a median of 5 turns per bout with a range of 2- 16 turns per bout. We followed the tradition of research in non-human animals in defining vocal turn-taking bouts. By looking beyond pairs of serve-and-return, we were able to observe substantial dyadic variation in frequency and length of turn-taking bouts. For example, at 5 months, thirty pairs of participants produced a range of 3 – 47 bouts during the play session. Each individual bout contained a range of 2 – 29 turns. At 10 months, participants produced a range of 6 – 34 no. of bouts, with 2- 23 turns in each bout. We found significant predictive effects in the frequency as well as length of vocal turn-taking from mere 15-minute-long sessions. The effect of these substantial individual differences could accumulate and result in significant influences over infant communicative development. Our measurements of turn-taking captured these accumulated developmental changes, and showed their significant predictive effects on the emergence of language abilities. By identifying extended turn-taking bouts, researchers can further understand the structure as well as developmental changes of caregiver-infant turn-taking interactions. In addition, they will also be able to investigate how specific aspects of turn-taking (e.g., modality, frequency, length, duration) influence the development of language.

The connection between vocal responsiveness in early infancy and later cognitive skills has been reported by previous research. At as early as 3 months of age, infant differential vocal responsiveness to their mother versus a stranger predicted their vocabulary size, reading and arithmetic abilities at 12 years old (Roe et al., 1982; Roe, 1978). In turn, maternal responsiveness to infants at 5 months significantly predicted their representational competence (i.e., language comprehension and pretense play) at 13 months (Bornstein & Tamis-LeMonda, 1989). These studies provided some of the developmentally earliest evidence linking vocal interactions with later development. However, they did not investigate the direct relationship between infant's own prelinguistic vocal use and later development in vocal production, by quantifying the frequency of infant vocalizations.

We found that increases in infant prelinguistic volubility were only predictive of language production, but not language comprehension outcomes. This finding parallels previous research on bilingualism. In children who hear and speak differing levels of two languages, their frequency of language production only predicts productive but not receptive outcomes in that language (Ribot et al., 2018), independent of how much they hear that language. Our finding suggests that prelinguistic vocal use serves a similar role as language use, in facilitating the development of language production.

Such a connection between prelinguistic vocalizing and the emergence of language production has been understudied, as babbling is traditionally viewed as spontaneous and unrelated to actual language formation (Jakobson, 1968). However, some previous work supports continuity between prelinguistic vocalizing and the acquisition of productive language (McCune & Vihman, 2001). For example, the same consonants infants produce in prelinguistic vocalizing, also predominantly appear in their first words. Infants use the same consonants in their vocal repertoire before and after they start to produce language, which suggests that these two stages are not utterly delimited (McCune & Vihman, 2001). Furthermore, in infants who receive cochlear implants, those who received earlier implantation produce more babbling than those who did later, suggesting that prelinguistic vocalizing is guided by auditory input. Thus, researchers have proposed that prelinguistic volubility should be used as a prognostic marker of language development (Binos & Loizou, 2019). Our finding also suggests that developmental increases in prelinguistic vocal use is predictive of the emergence of later productive language.

4.4. Limitations

The current study has several limitations. First, the sample size of this study is relatively small. Thirty dyads participated in both sessions at 5 and 10 months, and twenty of them returned CDIs at 18 months. Since there are considerable individual differences in behavioral factors like infant volubility and caregiver response rate, our correlational results between these factors and later language outcomes may be influenced by the small sample size. Future large-sample studies should incorporate assessments of extended turn-taking and further investigate its predictive effects on later language comprehension and production. A second limitation is that our measurement of infant language outcome at 18 months was based on caregivers' reports. However, the predictive validity of CDI has been examined and supported by previous studies (e.g., Can et al., 2013; Feldman et al., 2005), and our results were comparable to previous studies (e.g., Fernald et al., 2013; Frank et al., 2017). Lastly, our measures of turn-taking bouts did not focus on pairs of caregiver-infant exchanges, but included extended sequences of caregiver and infant vocalizations. Turn-taking bouts started with at least two exchanges between caregivers and infants, but subsequent consecutive turns by the same party were counted when we calculated the length of turn-taking. Thus, we captured developmental changes in turn-taking on the dyadic level, rather than distinguish changes in individual contributions. Nevertheless, the effect of caregiver-infant interactions appears to be the most robust on the dyadic level. For example, we found that caregiver responsiveness at 5 months predicted vocal turn-taking more strongly than itself at 10 months. Thus, we believe that our quantification of turn-taking as a dyadic system affords novel analyses to further examine the impact of caregiver-infant extended turn-taking.

5. Conclusion

We observed that prelinguistic caregiver-infant turn-taking exhibited high continuity and low stability during the first year, and investigated the sources as well as predictive effects of developmental changes in turn-taking. The more frequently caregivers responded contingently to babbling at 5 months, the more infants engage in vocal turns with caregivers at 10 months. Developmental increases in prelinguistic vocal turn-taking predicted later language development at 18 months. Furthermore, this study examined the real-time structure of caregiver-infant turn-taking beyond the initial serve-and-return. At both 5 and 10 months, caregiver vocal responses were the most effective in facilitating vocal turn-taking with infants. We integrated multiple timescales by investigating first-year prelinguistic turn-taking with fine-grained temporal analyses, and brought new insights into the significance and development of early vocal turn-taking.

Funding

This work was supported by the National Science Foundation (BCS-0844015).

CRediT authorship contribution statement

Vivian Hanwen Zhang: Conceptualization, Methodology, Formal analysis, Data curation, Writing – original draft, Visualization. **Steven L Elmlinger:** Conceptualization, Methodology, Formal analysis, Data curation, Writing – review & editing, Visualization. **Michael H Goldstein:** Conceptualization, Investigation, Writing – review & editing, Funding acquisition.

Declaration of Competing Interest

The authors report no conflicts of interest.

Data Availability

Data will be made available on request.

Acknowledgments

We thank undergraduate members of the Eleanor J. Gibson Laboratory for Developmental Psychology for their assistance with data collection and coding, especially Sarra Al-Zayer, Cherrie Lan, Elshadai Melkam and Megan Rait. We thank all families who participated in this research.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.infbeh.2024.101945.

References

Abney, D. H., Paxton, A., Dale, R., & Kello, C. T. (2014). Complexity matching in dyadic conversation. *Journal of Experimental Psychology: General, 143*(6), 2304.

Abney, D. H., Warlaumont, A. S., Oller, D. K., Wallot, S., & Kello, C. T. (2017). Multiple coordination patterns in infant and adult vocalizations. *Infancy, 22*(4), 514–539.

Bates, D., M.chler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed- effects models using lme4. *Journal of Statistical Software, 67*, 1–48.

Bateson, M. C. (1975). Mother-infant exchanges: the epigenesis of conversational interaction. *Annals of the New York Academy of Sciences, 263*(1), 101–113.
Baumwell, L., Tamis-LeMonda, C. S., & Bornstein, M. H. (1997). Maternal verbal sensitivity and child language comprehension. *Infant Behavior and Development, 20*(2), 247–258.
Beckwith, L., & Cohen, S. E. (1989). Maternal responsiveness with preterm infants and later competency. *New Directions for Child and Adolescent Development, 1989*, 75–87.
Beebe, B., Alson, D., Jaffe, J., Feldstein, S., & Crown, C. (1988). Vocal congruence in mother-infant play. *Journal of Psycholinguistic Research, 17*, 245–259.
Belsky, J., Rovine, M., & Taylor, D. G. (1984). The Pennsylvania Infant and Family Development Project, III: The origins of individual differences in infant-mother attachment: Maternal and infant contributions. *Child Development, 55*(3), 718–728.
Bergelson, E., & Swingley, D. (2013). The acquisition of abstract words by young infants. *Cognition, 127*(3), 391–397.
Binos, P., & Loizou, E. (2019). Vocalization frequency as a prognostic marker of language development following early cochlear implantation. *Audiology Research, 9*(1), 217.
Bloom, K., Russell, A., & Wassenberg, K. (1987). Turn taking affects the quality of infant vocalizations. *Journal of child language, 14*(2), 211–227.
Bornstein, M. H., Putnick, D. L., Cote, L. R., Haynes, O. M., & Suwalsky, J. T. (2015). Mother-infant contingent vocalizations in 11 countries. *Psychological Science, 26*(8), 1272–1284.
Bornstein, M. H., Putnick, D. L., & Esposito, G. (2017). Continuity and stability in development. *Child Development Perspectives, 11*(2), 113–119.
Bornstein, M. H., & Tamis-Lemonda, C. S. (1997). Maternal responsiveness and infant mental abilities: Specific predictive relations. *Infant Behavior and Development, 20*(3), 283–296.
Bornstein, M. H., Tamis-LeMonda, C. S., & Haynes, O. M. (1999). First words in the second year: Continuity, stability, and models of concurrent and predictive correspondence in vocabulary and verbal responsiveness across age and context. *Infant Behavior and Development, 22*(1), 65–85.
Bornstein, M. H., & Tamis-LeMonda, C. S. (1989). Maternal responsiveness and cognitive development in children. *New Directions for Child and Adolescent Development, 1989*(43), 49–61.
Brand, R. J., Escobar, K., & Patrick, A. M. (2020). Coincidence or cascade? The temporal relation between locomotor behaviors and the emergence of stranger anxiety. *Infant Behavior and Development, 58*, Article 101423.
Can, D. D., Ginsburg-Block, M., Golinkoff, R. M., & Hirsh-Pasek, K. (2013). A long-term predictive validity study: Can the CDI Short Form be used to predict language and early literacy skills four years later? *Journal of Child Language, 40*(4), 821–835.
D'Odorico, L., Salerni, N., Cassibba, R., & Jacob, V. (1999). Stability and change of maternal speech to Italian infants from 7 to 21 months of age: A longitudinal study of its influence on early stages of language acquisition. *First Language, 19*(57), 313–346.
Donnelly, S., & Kidd, E. (2021). The longitudinal relationship between conversational turntaking and vocabulary growth in early language development. *Child Development, 92*(2), 609–625.
Elias, G., Hayes, A., and, & Broerse, J. (1986). Maternal control of co-vocalization and inter-speaker silences in mother-infant vocal engagements. *Journal of Child Psychology and Psychiatry, 27*, 409–415.
Elmlinger, S. L., Schwade, J. A., & Goldstein, M. H. (2019). The ecology of prelinguistic vocal learning: Parents simplify the structure of their speech in response to babbling. *Journal of Child Language, 46*(5), 998–1011.
Elmlinger, S. L., Schwade, J. A., Vollmer, L., & Goldstein, M. H. (2023). Learning how to learn from social feedback: The origins of early vocal development. *Developmental Science, 26*(2), Article e13296.
Feldman, H. M., Dale, P. S., Campbell, T. F., Colborn, D. K., Kurs-Lasky, M., Rockette, H. E., & Paradise, J. L. (2005). Concurrent and predictive validity of parent reports of child language at ages 2 and 3 years. *Child Development, 76*(4), 856–868.
Feldman, R. (2007). Parent–infant synchrony and the construction of shared timing; physiological precursors, developmental outcomes, and risk conditions. *Journal of Child Psychology and Psychiatry, 48*(3-4), 329–354.
Feldstein, S., Jaffe, J., Beebe, B., Crown, C. L., Jasnow, M., Fox, H., & Gordon, S. (1993). Coordinated interpersonal timing in adult-infant vocal interactions: A cross-site replication. *Infant Behavior and Development, 16*, 455–470.
Fenson, L., Marchman, V. A., Thal, D., Dale, P. S., Reznick, J. S., & Bates, E. (2006). *MacArthur-Bates Communicative Development Inventories: User's guide and technical manual 2*. Baltimore, MD: Brookes.
Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science, 16*(2), 234–248.
Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language, 44*(3), 677–694.
Gilkerson, J., & Richards, J.A. (2009). The power of talk: Impact of adult talk, conversational turns, and TV during the critical 0–4 years of child development. In the power of talk (2nd ed.). LENA Foundation.
Gilkerson, J., Richards, J. A., Warren, S. F., Oller, D. K., Russo, R., & Vohr, B. (2018). Language experience in the second year of life and language outcomes in late childhood. *Pediatrics, 142*(4).
Goldstein, M. H., & Schwade, J. A. (2010). From birds to words: Perception of structure in social interactions guides vocal development and language learning. In M. S. Blumberg, J. H. Freeman, & S. R. Robinson (Eds.), *Oxford handbook of developmental behavioral neuroscience* (pp. 708–729). Oxford University Press.
Goldstein, M. H., Schwade, J. A., & Bornstein, M. H. (2009). The value of vocalizing: Five-month-old infants associate their own noncry vocalizations with responses from caregivers. *Child Development, 80*(3), 636–644.
Gómez, E., & Strasser, K. (2021). Language and socioemotional development in early childhood: The role of conversational turns. *Developmental Science, 24*(5), Article e13109.
Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., & Parlato-Oliveira, E. (2015). Early development of turn-taking in vocal interaction between mothers and infants. *Frontiers in Psychology, 6*(1167), 236–245.
Gros-Louis, J., West, M. J., & King, A. P. (2014). Maternal responsiveness and the development of directed vocalizing in social interactions. *Infancy, 19*(4), 385–408.
Harrist, A. W., Pettit, G. S., Dodge, K. A., & Bates, J. E. (1994). Dyadic synchrony in mother-child interaction: Relation with children's subsequent kindergarten adjustment. *Family Relations*, 417–424.
Hilbrink, E. E., Gattis, M., & Levinson, S. C. (2015). Early developmental changes in the timing of turn-taking: a longitudinal study of mother–infant interaction. *Frontiers in Psychology, 6*(1492), 246–257.
Hirsh-Pasek, K., & Burchinal, M. (2006). Mother and caregiver sensitivity over time: Predicting language and academic outcomes with variable-and person-centered approaches. *Merrill-Palmer Quarterly*, (1982), 449–485.
Iyer, S. N., Denson, H., Lazar, N., & Oller, D. K. (2016). Volubility of the human infant: Effects of parental interaction (or lack of it). *Clinical Linguistics & phonetics, 30*(6), 470–488.
Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., & Jasnow, M. (2001). Rhythms of dialogue in infancy. *Monographs of the Society for Research in Child Development, Series, 264, 66*(2), 1–132.
Jakobson, R. (1968). Child language, aphasia and phonological universals. de Gruyter.
Kärtner, J., Keller, H., Lamm, B., Abels, M., Yovsi, R. D., Chaudhary, N., & Su, Y. (2008). Similarities and differences in contingency experiences of 3-month-olds across sociocultural contexts. *Infant Behavior and Development, 31*(3), 488–500.
Kartushina, N., & Mayor, J. (2019). Word knowledge in six-to nine-month-old Norwegian infants? Not without additional frequency cues. *Royal Society Open Science, 6*(9), Article 180711.
Katsu, N., Yamada, K., Okanoya, K., & Nakamichi, M. (2019). Temporal adjustment of short calls according to a partner during vocal turn-taking in Japanese macaques. *Current Zoology, 65*(1), 99–105.
Koda, H. (2004). Flexibility and context-sensitivityduring the vocal exchange of coo calls in wild Japanese macaques *Macaca fuscata yakui. Behaviour, 141*, 1279–1296.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163.
Landon, S. J., & Sommers, R. K. (1979). Talkativeness and children's linguistic abilities. *Language and Speech, 22*(3), 269–275.
Landry, S. H., Smith, K. E., & Swank, P. R. (2006). Responsive parenting: establishing early foundations for social, communication, and independent problem-solving skills. *Developmental Psychology, 42*(4), 627.
Leclère, C., Viaux, S., Avril, M., Achard, C., Chetouani, M., Missonnier, S., & Cohen, D. (2014). Why synchrony matters during mother-child interactions: a systematic review. *PloS One, 9*(12), Article e113571.
Lin, G., & Carlile, S. (2015). Costs of switching auditory spatial attention in following conversational turn-taking. *Frontiers in Neuroscience, 9*(124), 1–11.
Lin, G., & Carlile, S. (2019). The effects of switching non-spatial attention during conversational turn taking. *Scientific Reports, 9*(1), 1–9.
Ling, D., & Ling, A. H. (1974). Communication development in the first three years of life. *Journal of Speech and Hearing Research, 17*(1), 146–159.
Matatyaho, D. J., & Gogate, L. J. (2008). Type of maternal object motion during synchronous naming predicts preverbal infants' learning of word–object relations. *Infancy, 13*(2), 172–184.
McCune, L., & Vihman, M. M. (2001). Early phonetic and lexical development: a productivity approach. *Journal of Speech, Language, and Hearing Research, 44*(3), 670–684.
Oller, D. K. (2000). The emergence of the speech capacity. *Laurence Erlbaum and Associates*. Mahwah, N.J.: Psychology Press,.
Oller, D. K., & Lynch, M. P. (1992). Infant vocalizations and innovations in infraphonology: toward a broader theory of development and disorders. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: models, research, implications* (pp. 509–536). Timonium, MD: York Press.
Paavola, L., Kemppinen, K., Kumpulainen, K., Moilanen, I., & Ebeling, H. (2006). Maternal sensitivity, infant co-operation and early linguistic development: Some predictive relations. *European Journal of Developmental Psychology, 3*(1), 13–30.
Ribot, K. M., Hoff, E., & Burridge, A. (2018). Language use contributes to expressive language growth: Evidence from bilingual children. *Child Development, 89*(3), 929–940.
Ritwika, V. P. S., Pretzer, G. M., Mendoza, S., Shedd, C., Kello, C. T., Gopinathan, A., & Warlaumont, A. S. (2020). Exploratory dynamics of vocal foraging during infant-caregiver communication. *Scientific Reports, 10*(1), 1–14.
Roe, K. V. (1978). Infants' mother–stranger discrimination at 3 months as a predictor of cognitive development at 3 and 5 years. *Developmental Psychology, 14*(2), 191.
Roe, K. V., McClure, A., & Roe, A. (1982). Vocal interaction at 3 months and cognitive skills at 12 years. *Developmental Psychology, 18*(1), 15.
Rohlfing, K. J., Leonardi, G., Nomikou, I., Rączaszek-Leonardi, J., & Hullermeier, E. (2019). Multimodal turn-taking: motivations, methodological challenges, and novel approaches. *IEEE Transactions on Cognitive and Developmental Systems, 12*(2), 260–271.
Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M. L., & Gabrieli, J. D. (2018). Beyond the 30-million-word gap: Children's conversational exposure is associated with language-related brain function. *Psychological Science, 29*(5), 700–710.
Ruddy, M. G., & Bornstein, M. H. (1982). Cognitive correlates of infant attention and maternal stimulation over the first year of life. *Child Development*, 183–188.
Shapiro, N. T., Hippe, D. S., & Ramírez, N. F. (2021). How chatty are daddies? An exploratory study of infants' language environments. *Journal of Speech, Language, and Hearing Research, 64*(8), 3242–3252.
Sherrod, K. B., Friedman, S., Crawley, S., Drake, D., & Devieux, J. (1977). Maternal language to prelinguistic infants: syntactic aspects. *Child Development*, 1662–1665.
Snow, C. E. (1977). Mothers' speech research: From input to interaction. In C. E. Snow, & C. A. Ferguson (Eds.), *Talking to children: Language input and acquisition* (pp. 31–49). London: Cambridge University Press.
Suanda, S. H., Smith, L. B., & Yu, C. (2016). The multisensory nature of verbal discourse in parent–toddler interactions. *Developmental neuropsychology, 41*(5-8), 324–341.
Suttora, C., & Salerni, N. (2011). Maternal speech to preterm infants during the first 2 years of life: stability and change. *International Journal of Language & Communication Disorders, 46*(4), 464–472.
Tamis-LeMonda, C. S., Chen, L. A., & Bornstein, M. H. (1998). Mothers' knowledge about children's play and language development: short-term stability and interrelations. *Developmental Psychology, 34*(1), 115.
Tamis-LeMonda, C. S., Bornstein, M. H., Baumwell, L., & Melstein Damast, A. (1996). Responsive parenting in the second year: Specific influences on children's language and play. *Early Development and Parenting. An International Journal of Research and Practice, 5*(4), 173–183.
Thompson, L. A., & Trevathan, W. R. (2009). Cortisol reactivity, maternal sensitivity, and infant preference for mother's familiar face and rhyme in 6-month-old infants. *Journal of Reproductive and Infant Psychology, 27*(2), 143–167.
Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science, 10*(2), 172–175.
van der Klis, A., Adriaans, F., & Kager, R. (2023). Infants' behaviours elicit different verbal, nonverbal, and multimodal responses from caregivers during early play. *Infant Behavior and Development, 71*(101828), 1–17.
Van Egeren, L. A., Barratt, M. S., & Roach, M. A. (2001). Mother–infant responsiveness: Timing, mutual regulation, and interactional context. *Developmental Psychology, 37*(5), 684–697.
Venker, C. E., Bolt, Dm, Meyer, A., Sindberg, H., Ellis Weismer, S., & Tager-Flusberg, H. (2015). Parent telegraphic speech use and spoken language in preschoolers with ASD. *Journal of Speech, Language, and Hearing Research, 58*(6), 1733–1746.
Warlaumont, A. S., Richards, J. A., Gilkerson, J., & Oller, D. K. (2014). A social feedback loop for speech development and its reduction in autism. *Psychological Science, 25*(7), 1314–1324.
Yurkovic, J. R., Kennedy, D. P., & Yu, C. (2021). Multimodal behaviors from children elicit parent responses in real-time social interaction. *In Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 43)(No. 43).
Zimmerman, F. J., Gilkerson, J., Richards, J. A., Christakis, D. A., Xu, D., Gray, S., & Yapanel, U. (2009). Teaching by listening: The importance of adult-child conversations to language development. *Pediatrics, 124*(1), 342–349.