Infant Behavior and Development 57 (2019) 101325

Contents lists available at ScienceDirect

Infant Behavior and Development

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

Full length article

Infant-adult vocal interaction dynamics depend on infant vocal type, child-directedness of adult speech, and timeframe

Gina M. Pretzer[a,c,*], Lukas D. Lopez[a], Eric A. Walle[a], Anne S. Warlaumont[b]

[a] University of California, Merced, United States
[b] University of California, Los Angeles, United States
[c] HopeSparks Children's Developmental Services, Tacoma, WA, United States

ARTICLE INFO

Keywords:
Temporal contingencies
Infant vocalizations
Infant-directed speech
Other-directed speech
Infant-adult interactions
LENA

ABSTRACT

This study explored the temporal contingencies between infant and adult vocalizations as a function of the type of infant vocalization, whether adult caregivers' vocalizations were infant-directed or other-directed, and the timescale of analysis. We analyzed excerpts taken from day-long home audio recordings that were collected from nineteen 12- to 13-month-old American infants and their caregivers using the LENA system. Three 5-minute sections having high child vocalization rates were identified within each recording and coded by trained researchers. Infant and adult vocalizations were sequenced and defined as contingent if they occurred within 1 s, 2 s, or 5 s of each other. When using 1 s or 2 s definitions of temporal adjacency, infant vocalizations generally predicted subsequent infant-directed adult vocalizations. A reflexive vocalization (i.e. a cry or a laugh) was the strongest predictor. Likewise, within 1–2 s timeframes, infant-directed adult speech generally predicted infant vocalizations with reflexive vocalizations being particularly predictive. Infant vocalizations predicted fewer subsequent other-directed adult vocalizations and were less likely following other-directed adult vocalizations when considering up to 5 s lags. This suggests an understudied communicative role for infants of non-infant-directed adult speech. These results demonstrate the importance of timescale in studying infant-adult interactions, support the communicative significance of reflexive infant vocalizations and other-directed adult speech in addition to more commonly studied vocalization types, and highlight the challenges of determining direction(s) of influence when using only two-event sequences.

1. Introduction

A wealth of evidence indicates the importance of dynamic and contingent vocal interactions in fostering infant vocal communication development. However, much of this research has been conducted in laboratory settings (e.g., Akhtar, Dunham, & Dunham, 1991; Goldstein & Schwade, 2008; Gros-Louis, West, & King, 2014; Peláez, Virues-Ortega, & Gewirtz, 2011; Rollins, 2003), and studies occurring in the home often rely on a researcher being present (e.g., Bornstein, Putnick, Cote, Haynes, & Suwalsky, 2015; Shneidman, Arroyo, Levine, & Goldin-Meadow, 2013) or the use of automated coding (e.g., Warlaumont, Richards, Gilkerson, & Oller, 2014). The present study helps to address these shortcomings by using hand-coding of parent-infant interactions from day-long home audio recordings to examine the bidirectional relationship between infant and adult vocalizations at 12–13 months, a point in development when infants understand and begin to produce meaningful speech.

* Corresponding author at: HopeSparks Children's Developmental Services, 6424 North 9th St, Tacoma, WA 98406, United States.
E-mail address: gpretzer@hopesparks.org (G.M. Pretzer).

https://doi.org/10.1016/j.infbeh.2019.04.007
Received 2 May 2018; Received in revised form 19 April 2019; Accepted 22 April 2019
0163-6383/

1.1. Infant vocal development

The emergence of infant prelinguistic communication is marked by a predictable progression (e.g., Buder, Warlaumont, & Oller, 2013; Koopmans-van Beinum & van der Stelt, 1986; Oller, 1980; 2000; Stoel-Gammon, 1989; Vihman, Macken, Miller, Simmons, & Miller, 1985). Infants begin vocalizing from birth, with primitive vocalizations and cries (Buder et al., 2013; Oller, 1980; Roug, Landberg, & Lundberg, 1989; Stark, 1980). Around 2–3 months of age, infants have gained sufficient control of their vocal tracts to start producing a wider variety of protophones such as full vowels, raspberries, squeals, growls, yells, whispers, and marginal babbling (Buder et al., 2013; Oller, 1980; Roug et al., 1989; Stark, 1980). By about age 6 months, infants begin to demonstrate canonical babbling, which is when a syllable has speech-like timing between consonant and vowel (Buder et al., 2013). From then until at least 18 months, the relative frequency of infant speech-related vocalizations increases steadily (Oller, Eilers, Urbano, & Cobo-Lewis, 1997; Warlaumont & Ramsdell-Hudock, 2016). These vocal milestones provide a foundation for lexical development (Oller, 2000). For example, the phonetic features of canonical babbling are typically similar to the phonetic features of first words (Stoel-Gammon, 1989; Vihman et al., 1985).

1.2. Contingent infant-adult vocal interactions

Importantly, infants acquire these vocal communication abilities in part through interactions with adults. Infant-adult vocal interactions allow parents to demonstrate appropriate conventions for communication, including pragmatic use of vocalizations and rhythms of dialogue (Baldwin & Meyer, 2007; Goldin-Meadow, Goodrich, Sauer, & Iverson, 2007; Golinkoff, 1986; Gros-Louis et al., 2014; Jaffe et al., 2001; Masur, 1987; Olson & Masur, 2011). Contingent adult responses, defined as adult vocalizations that immediately follow infant vocalizations with a likelihood that is greater than what would be expected by chance from the base rates of the infant and adult behaviors, appear to be especially helpful. For example, previous research has shown that infants alter the phonological structure of their vocalizations to match the phonological structure of previous adult vocalizations if those adult vocalizations were contingent responses to infant vocalizations, but not if they were non-contingent on the child's behavior (Goldstein & Schwade, 2008; see also Bloom, Russell, & Wassenberg, 1987, and Bloom, 1988). Further support comes from correlations between sensitive caregiver responses to infant vocalizations (e.g., contingent responses, imitation of the infant's vocalization, or commenting on an object within the infant's visual field) and subsequent rates of canonical syllable production (Gros-Louis et al., 2014). There are also numerous studies linking contingent adult responding with increased frequency of infant speech-related vocalizations (Franklin et al., 2014; Goldstein, Schwade, & Bornstein, 2009; Golinkoff, Can, Soderstrom, & Hirsh-Pasek, 2015; Gros-Louis et al., 2014; Nathani & Stark, 1996; Todd & Palmer, 1968; Warlaumont et al., 2014).

Additionally, prior research has shown that bidirectional infant-caregiver exchanges are important for other aspects of speech, language, and social development (e.g., Golinkoff et al., 2015; Hart & Risley, 1995; Tamis-LeMonda, Bornstein, & Baumwell, 2001). Parents provide models of speech sounds, word labels, and appropriate conversational behaviors, and embedding these models in contingent responses is particularly impactful (although the degree to which children are sensitive to various features of adult language input changes with age; Adamson & Bakeman, 2006). Further, adult responses that label objects promote word learning (both receptively and expressively) in infants even younger than 9 months of age (see Goldstein, Schwade, Briesch, & Syal, 2010; Gogate, Maganti, & Laing, 2013; Matatyaho & Gogate, 2008; Wu & Gros-Louis, 2014).

Adult responsivity to infant vocalizations can also help mediate the relationship between infant prelinguistic vocalizations and infants' later language abilities in children with developmental disabilities (Yoder & Warren, 1999; Harbison et al., in press). Yoder and Warren (2002) found that interventions targeting parent responsivity were associated with growth in language abilities for children with intellectual disabilities, including Down syndrome and William's syndrome. Moreover, infants with higher rates of pretreatment canonical syllable production benefitted most from intervention supporting high parent responsivity. Clearly, sensitive adult responses (i.e., responses that are attuned to the child's social and emotional regulation needs; see Gros-Louis et al., 2014; Field, 1980; Patten, Labban, Casenhiser, & Cotton, 2016) to infant vocalizations can provide important support for communication development.

Infant-adult vocal interactions involve not only adult responses to infant vocalizations but also infant responses to adult vocalizations. Previous studies have demonstrated, across cultures, that infants as young as 4-months-old actively engage in turn taking with adults in their environment, with increased likelihood of infant vocalization following adult utterances (e.g., Bornstein et al., 2015; Van Egeren, Barratt, & Roach, 2001) and that infant and maternal contingencies are significantly correlated (Bornstein et al., 2015). At a larger timescale, Gros-Louis et al. (2014) found that contingent maternal responses predicted maternal-directed infant vocalizations in following months. Considering that adult responses often take the form of speech, and that adult vocalizations (at least those that are non-overlapping with infant vocalizations) often occur in response to and encouragement of further infant speech-like vocalizations, we would predict that infant vocalizations are more likely to be speech-like when they occur in response to adult input. Consistent with this, Warlaumont et al. (2014) found evidence for a social feedback loop wherein infant vocalization is contingent upon adult vocalization and vice versa. Following up on this finding, Gros-Louis and Miller (2018) distinguished between vowel only and consonant-vowel speech-related infant vocalizations. They found that parents were more likely to respond contingently to consonant-vowel than vowel only infant vocalizations. They also found that 10-month-old infants were significantly more likely to produce vowel-like vocalizations following an adult response to the infant's prior vowel-like vocalization and that 12-month-old infants were significantly more likely to produce consonant-vowel vocalizations following an adult response to the infant's prior consonant-vowel vocalization.

1.3. Infant-directed versus other-directed speech

In cultures where infants are regularly provided with infant-directed speech (IDS) from their caregivers, IDS is particularly beneficial for speech and language development compared to adult-directed speech (ADS). IDS exaggerates the phonological features of speech, thereby highlighting important aspects of the speech signal crucial for recognition of vowels and consonants and increasing infant attention to adult vocalizations (1989, Baldwin & Meyer, 2007; Cristia, 2011; Fernald & Kuhl, 1987; Fernald, 1985; Kuhl et al., 1997; Papousek & Papousek, 1989). This slower rate of speech and exaggerated pronunciation permits more effective modeling of words for infants (Song, Demuth, & Morgan, 2010) and is associated with infant word learning (Golinkoff & Alioto, 1995; Ma, Golinkoff, Houston, & Hirsh-Pasek, 2011; Ramírez-Esparza, Garcia-Sierra, & Kuhl, 2014; Shneidman et al., 2013; Weisleder & Fernald, 2013) and appreciation of pragmatics (Golinkoff et al., 2015). Moreover, depressed mothers, who typically have less pitch alteration, were found to have infants with larger productive vocabularies when the mother demonstrated more pitch modification (Porritt, Zinser, Bachorowski, & Kaplan, 2014). Another acoustic feature of speech, intensity, has not been subject to much study. Fernald and Kuhl (1987) investigated whether infants showed listening preferences for intensity patterns of “motherese” vocalizations, but found that infants only showed preference for patterns of pitch.

Although research has shown that IDS promotes language development in various cultures, the role of other-directed speech (ODS), including ADS, is less clear. Many studies, including those sampling from day-long home recordings featuring abundant quantities of ODS, exclude ODS from all analyses (e.g., Ramírez-Esparza et al., 2014). Those studies that do analyze ODS have shown that the quantity of ODS to which a child is exposed seems to have little effect on language learning (Kuhl, Tsao, & Liu, 2003; Oller, 2010; Weisleder & Fernald, 2013). While those studies have not reported beneficial effects of ODS on language learning, other work has demonstrated that toddlers can learn new words from overheard speech (Akhtar, Jipson, & Callanan, 2001). Additionally, Floor and Akhtar (2006)) reported evidence of word learning after overhearing ODS by 18-month-old infants.

Furthermore, although modification of speech directed to infants appears to share many properties across cultures, there are cultures in which substantial differences in pragmatic, linguistic and acoustic properties of IDS have been reported. For example, within the Kwara’ae culture, young infants are primarily exposed to a “modified register” of the language that share features of IDS seen in other cultures but has different pragmatic features and emphases (Watson-Gegeo & Gegeo, 1986). Similarly, Kaluli adults have been reported to produce little IDS but have been reported to frequently use a high-pitched register to “translate” for their preverbal infants (Ochs & Schieffelin, 1984). Moreover, pitch modification in IDS has been found to be higher for more highly-educated American mothers (Broesch & Bryant, 2015). Even significant differences in exposure to IDS and in the form that IDS takes do not totally preclude language learning, raising the possibility that ODS might play a substantial role in language learning and making it worth analyzing alongside IDS.

It is also possible that in IDS-heavy cultures, reduced exposure to ODS could be beneficial to language learning under certain circumstances. Specifically, reduction in ODS could reduce background noise, allowing infants to more effectively learn from IDS and their own vocalizations, and as well as signal to infants that adults are paying attention to them. Thus, while infants may or may not gain immediate phonological or lexical information from overhearing ODS, there may be other functions of ODS within these exchanges and *absence* of ODS may itself be valuable at least for some cultures.

Prior research indicates that IDS shows patterns of sequential association with child vocalizations — in fact, the majority of the studies of infant-adult interaction cited above focused exclusively on adult vocalizations that were infant-directed and excluded vocalizations directed toward another adult. It remains to be examined whether ODS also shows patterns of sequential association wherein an infant’s vocalization predicts the subsequent occurrence of overheard adult speech or whether overheard adult speech predicts the subsequent occurrence of an infant vocalization. Better understanding of the temporal contingencies between infant vocalizations, IDS, and ODS may help clarify both the pragmatic properties of IDS and ODS within a given culture, and thereby inform our understanding of the roles each can play in communication development. Study of day-long home audio recording have the advantage of offering ample opportunities to sample both IDS and ODS.

1.4. Studying early vocal interactions in day-long home recordings

Much of the prior research on prelinguistic development has been conducted in laboratories or short sessions in the infant’s home with a researcher present. However, wearable audio recording systems, such as the LENA system (LENA; LENA Research Foundation, Boulder, Colorado, United States), now allow researchers to collect full-day observations of infant-adult interactions in the home. The LENA system is comprised of pocket-sized audio recorders called digital language processors (DLP), specially designed clothing to hold DLPs, and software to process and analyze collected audio data. This methodology allows researchers to study how infants and their caregivers interact across a broad and ecologically valid setting with minimal observer effects. A recent study using first-person and third-person point of view recordings revealed that when using first-person viewpoint cameras in home recordings, there is a higher frequency of “less socially desirable maternal behaviors” (e.g., being distracted or critical) observed than in third-person viewpoint recordings (Lee et al., 2017). Another study comparing day-long audio-only recordings to hour-long video recordings made by parents during the same day found higher rates of language production during the video samples than during the rest of the portions of the day when the infant was awake (Bergelson, Amatuni, Dailey, Koorathota, & Tor, 2018). Thus, the absence of a researcher may provide a more ecologically valid sample than one collected in a laboratory context. Day-long recordings provide a naturalistic means to study vocalization types and contexts beyond those that commonly occur in the lab or when a researcher is present. For example, infants’ distress vocalizations may be more prevalent in day-long home recordings and adults are faced with numerous situations in which they must decide whether to vocalize to the infant or engage with other adult conversational partners or other children.

Previous research by Warlaumont et al. (2014) examined infant and adult vocalization contingencies utilizing the LENA system's automatic labeling of parent and infant vocalizations, which compared to labeling by trained human listeners is less accurate but also less time intensive and less subject to inter and intra-rater variability. The researchers found that infant speech-related vocalizations were more likely to be followed by adult vocalizations than children's non-speech-related vocalizations. Further, children's subsequent rates of production of speech-related vs. not-speech-related sounds were related to these adult responses at both immediate and longer-term timescales. However, analyzing such recordings in detail beyond speaker diarization requires either custom vocalization analysis algorithms that are not yet widely available (e.g., Oller et al., 2010) or human coding/transcription. Even so, hand coding by human listeners remains the gold standard for characterizing many aspects of vocalizations, including whether an utterance contains canonical syllables (Warlaumont & Ramsdell-Hudock, 2016) and to whom adult speech is directed (Schuller et al., 2017). Furthermore, hand-coding is necessary to ensure accurate labeling of the onset and offset of speaker vocalizations, especially during periods of overlap. Thus, utilizing hand coding can allow researchers to examine day-long recordings of infants and their families in more detail and with greater accuracy than that provided by existing automated methods.

1.5. Overview of the present study

This study analyzed hand-coded segments from day-long LENA recordings of 12- to 13-month-old infants focusing on periods of high child vocalization rate (i.e., volubility). We examined the temporal relationships between infant vocalizations and IDS. Additionally, we examined the relationship between infant vocalizations and ODS (i.e., adult speech to other adults or to children other than the target infant), which is often excluded in language development studies. We also distinguished between a range of infant vocalization types, including canonical vocalizations (i.e. babbling or first true words containing at least one adult-like consonant-vowel or vowel-consonant transition; see Oller et al., 1997 and Buder et al., 2013), non-canonical babble (including marginal babbling and other non-canonical, non-reflexive, non-vegetative sounds), and reflexive sounds (laughs and cries). We assessed the degree to which each infant vocalization type predicted subsequent adult vocalizations, and vice versa, using three time windows of analysis, 1 s, 2 s, and 5 s. This design provided a comprehensive analysis of bidirectional contingencies between child vocalizations and adult vocalizations of different types in naturally occurring and unsupervised contexts at a point in development when infants typically begin to produce utterances with recognizable meanings. Finally, we tested for mean pitch, pitch variability, and mean intensity differences between IDS and ODS, to establish a broad sense of the acoustic differences between IDS and ODS in this sample; this is the first analysis of pitch and intensity for IDS and ODS from day-long audio recordings. Pitch and pitch variability of IDS are often explored. Since intensity of IDS (from the child's perspective) might also be a cue to infants and is very easily measured, we decided to include intensity in our analyses.

2. Methods

2.1. Participants

Nineteen infants (12 female) aged 12- to 13-months were selected from a larger, ongoing study of 62 recruited infants. All families were recruited from the San Joaquin Valley in Central California. Families from the larger study were excluded because their recording was shorter than the 10 -h requirement or was not returned (n = 34), the primary language spoken in the home was Spanish (i.e., > 50% of the time, according to the parents' estimates; n = 15), the infant was outside the targeted age range (n = 8), or the recording was split over multiple days (n = 7). The requirement to record 10 -hs within a single day was designed to ensure that the samples were drawn from a large range of different contexts that an infant typically experienced over the course of a day.

Eight infants were from families with both parents having a college degree (including 2 graduate degrees), 5 infants from families with one parent having a college degree (including 1 graduate degree), 5 infants from families with neither parent having a college degree, and one infant's family did not report parental educational experience. Fourteen infants were reported to be of Caucasian descent, 10 infants of Hispanic descent, and 2 infants of Asian descent. No diagnoses of developmental disorders were reported for any of the infants.

2.2. Recording procedure

Each family was mailed a LENA recorder and vest specially designed with a pocket on the chest to hold the recorder. The recorder captured the infant's vocalizations, as well as linguistic and other auditory input from the infant's home environment. Parents were instructed to turn on the audio recorder when the child awoke in the morning and slide it into the pocket of the provided vest. The infant then wore the vest for the entire day, with the exception of baths, naps, or car rides when the vest could be placed nearby while the recording continued. Parents were also allowed to pause the recording for privacy reasons. The recorder automatically shut off after 16 h of audio collection, though some families did not or were unable to record for the entire 16 h.

2.3. Selection of high infant volubility samples

The recordings were first processed using the LENA Pro system's software. The LENA software's automatic labeling system applies a set of mutually exclusive sound source labels to the entirety of each recording. The automated labels are: infant wearing the recorder, male adult, female adult, other child, electronic sounds, noise, and silence, with all labels except silence being further divided into "near" (i.e. relatively loud) and "far" sounds. LENA's Automatic Data Extractor (ADEX) was used to identify the three most voluble (i.e., containing the highest number of "near" infant vocalizations) 5-minute samples for each infant's day-long recording. However, an identified sample was replaced by the next highest-volubility sample if: the highest volubility 5-minute sample was within 30 min of another of the highest-volubility samples; the sample did not have at least 10 infant speech-related vocalizations (i.e., canonical babbling, marginal babbling, or other protophone vocalizations; see Buder et al., 2013 for definitions) as judged by a human listener; or it sounded as if the infant had an object (e.g., a pacifier) obstructing her mouth for a significant portion of the sample. These selection criteria ensured that distinct observational times from the day were included, the automatically assessed high volubility of the sample was valid, and the infant's vocal expressivity was not impeded.

2.4. Utterance identification and classification by human listeners

For each segment, infant and adult vocalizations were marked using the EUDICO Linguistic Annotator (ELAN [Computer software], 2018; Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006). Researchers were trained to identify and code infant and adult vocalizations using training procedures created by the authors. Researchers marked vocalizations by listening to the recording and pausing, replaying, and jumping backward and forward as needed to mark the onsets and offsets of infant vocalizations according to the following criteria:

1 *Mark any sound of nontrivial loudness that you think was made by the vocal tract (i.e. larynx, throat, mouth, lips, teeth, nasal passage, etc.).*
2 *Be as accurate as possible when setting utterance boundaries - try to be within tenths of a second.*
3 *Annotations should run from the onset of sound to the offset of sound. [For adult vocalizations, it was also noted that ends of phrases/sentences are often natural ends of vocalizations.]*

Next, vocalizations were coded for the speaker (i.e., infant or adult) and type of utterance. Researchers were trained to classify infant vocalizations within four categories: Reflexive (laugh and cry), Vegetative (e.g., burp, hiccup, cough, yawn, heavy breathing), Canonical (i.e., a syllable that has speech-like timing between consonant and vowel; see Buder et al., 2013), and Non-Canonical/Non-Reflexive (including marginal babbling and other non-canonical protophones; here forward referred to simply as "Non-Canonical"). Vegetative infant vocalizations were subsequently excluded. The adult vocalizations were coded as Infant-Directed, Other-Directed (i.e., speaking to another person who is not the target infant), or unknown, based on the intended addressee. We did not distinguish to whom ODS was directed. At times, ODS was directed to other adults (e.g., one adult discussing dinner plans with another). At other times, ODS was directed to animals (e.g., telling the dog to go outside) or to other children. In the event that adult vocalizations were judged to be directed to other children as well as the target infant, these vocalizations were classified as IDS. Transcribers listened to each vocalization a maximum of three times during the coding process to balance having ample opportunity to hear the vocalization against over-analysis and second-guessing.

Inter-rater reliability for infant vocalization type and adult utterance direction was calculated using percent agreement and Cohen's kappa. Reliability checks were completed on six randomly selected 5-minute samples (one sample each from approximately 30% of the participants). Reliability coders re-coded the originally marked infant vocalizations and the direction of the adult vocalizations, using the same criteria as the original coders. Interrater reliability was adequate for infant vocalization type (percent agreement = 83.40%; Cohen's $k = .58$) and substantial for Adult Utterance Direction (percent agreement = 89.90%; Cohen's $k = .76$) (see Landis & Koch, 1977). The disagreements in infant vocalizations codes were evenly spread across all vocalization types. Specifically, disagreements between canonical and non-canonical syllables represented 55% of the overall discrepancies and disagreements between non-canonical and reflexive vocalizations represented the remaining 45% of disagreements. There were no disagreements between canonical and reflexive infant vocalizations.

2.5. Temporal contingency analyses

Data were converted into a single series of events to test for temporal contingencies between child and adult utterances of various types. The event series allowed for the identification of vocalizations immediately preceding or immediately following any given utterance (Bakeman, 1997; Yoder & Symons, 2010).

Pauses between vocalizations and overlapping vocalizations were first defined in order to convert the original set of utterance boundaries into a single event series (Lloyd, Yoder, Tapp, & Staubitz, 2015). A pause was defined as any duration greater than P between the end time of the current vocal event and the start time of the next vocal event. P was set to either 1 s, 2 s, or 5 s to comprehensively explore different possible timescales at which vocal contingencies might operate (Van Egeren et al., 2001). Fig. 1 illustrates how pauses were inserted for the case where $P = 1$ s.

Previous work has found that adults frequently respond to infants with less than 1 s lag (corresponding to $P = 1$ s in the present analytic approach) and that infants are sensitive to contingent responses operating within this timeframe (e.g., Keller, Lohaus, Völker, Cappenberg, & Chasiotis, 1999; Van Egeren et al., 2001; Warlaumont et al., 2014). However, it is possible that in some cases 1 s might not be long enough to capture temporal contingencies between speakers. Thus, we also conducted all temporal contingency analyses with 2 s maximum lag durations (which, using our analysis approach, corresponded to $P = 2$ s). Research on infant-adult vocalizations has found that 2 s lags capture infant-adult bidirectional vocal interactions and contingent responses across cultures

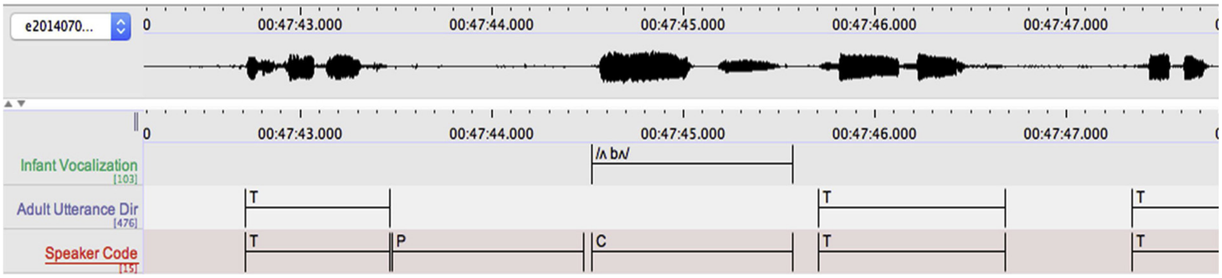

Fig. 1. A coded segment from a participant's recording, with lags between vocalizations of greater than or equal to 2 s coded as a pause (*P*). Lags between vocalizations that are less than 2 s are not coded. /ʌ bʌ/: phonetic transcription of an infant canonical (C) vocalization. T: an infant-directed adult vocalization. Speaker Code: a tier indicating how event series were constructed. The event series for this example was T P C T T.

(Bornstein et al., 2015; Gros-Louis, West, Goldstein, & King, 2006; Hilbrink, Gattis, & Levinson, 2015; Van Egeren et al., 2001). A 2 s pause definition has also been used in research on conversational turn-taking with slightly older children (Yoder, Davies, & Bishop, 1994). Henning and Striano (2011) found that younger infants were sensitive to slightly longer lag times of 3 s. These 3- and 6- month old infants altered their affect (i.e., smiling) based on the lag time before their mothers' responses. Moreover, in addition to including time lags of 1 s and 2 s, we also chose to include a longer time lag of 5 s in our analyses to reduce the likelihood of missing potentially relevant temporal contingencies, especially given research showing that response lags tend to increase around the onset of first words (Hilbrink et al., 2015) and given that the LENA software, which is now being widely used by researchers and interventionists, uses a 5 s lag in defining conversational turns.

For the scope of this study, we operationalize "follow" as meaning an adult utterance occurring within a specified time lag after an infant vocalization with no intervening infant or adult vocalizations and vice versa. For example, when assuming P = 5 s, if the child vocalizes and the parent vocalizes 2 s and 4 s after the child, only the first adult vocalization would be considered to follow the child vocalization and thus end the 5 s time window until the next infant vocalization occurs, creating a new 5 s time window in which a subsequent adult vocalization may or may not follow. Likewise, if an infant produces a vocalization and then vocalizes again 1 s later and the adult vocalizes 1 s after the second infant vocalization, the adult vocalization would only be considered to follow the second infant vocalization, not the first, even though it did also occur within 5 s of the first of the two infant vocalizations.

In contrast to the automatic labeling algorithm used by LENA software that excludes overlapping utterances, overlapping utterances were included in our analyses, with the first speaker's vocalization coded as the first event and the later occurring vocalization coded as the second event (see Figs. 2–4). For instances when the earlier onset vocalization extended greater than or equal to *P* (where *P* is the minimum pause duration in seconds) past the end time of the later onset event, the first event was split into two vocalizations (event 1 and 3). When the later onset overlapping vocalization extended past the end of the initial vocalization for less than *P*, the initial vocalization was coded as the first event and the overlapping vocalization was coded as the second event.

Next, we created pairs of 2-event sequences to analyze the temporal contingencies between vocalization types. The 2-event sequences were created by combining the type of each vocal event (except for the final vocal event within the 5-minute segment) with the type of the following vocal event (e.g., an Infant Reflexive utterance, R, followed by a child-directed, T, would be given a sequence code of RT). These sequence codes were compiled for all three 5-minute segments for all infants.

Finally, we ran generalized linear mixed effects regression models to assess the relations between the first event type and the second event type for each 2-event sequence. All models included participant ID as a random effect and assumed a Binomial distribution, since the dependent variables were all binary. Analyses were programmed in R (R Core Team, 2018) using the lme4 (Bates, Maechler, Bolker, & Walker, 2018), lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017), and car (Fox et al., 2018) packages.

We first analyzed how the first event in each event sequence pair being a Canonical, Non-Canonical Non-Reflexive, or Reflexive infant vocalization uniquely predicted whether the subsequent event was an Infant-Directed adult utterance, coded as 1 if so and 0 if not. The Canonical, Non-Canonical, and Reflexive variables were each dummy coded as 1 if the first event was the relevant infant vocalization type and 0 otherwise. If the first event in the sequence was any other type of event (i.e., if it was a pause or an adult utterance), then all three infant vocal type fixed effect variables were coded as 0. We also ran the analyses with the same predictors but predicting whether the second event in each pair was an Other-Directed adult utterance. The analyses were run separately for

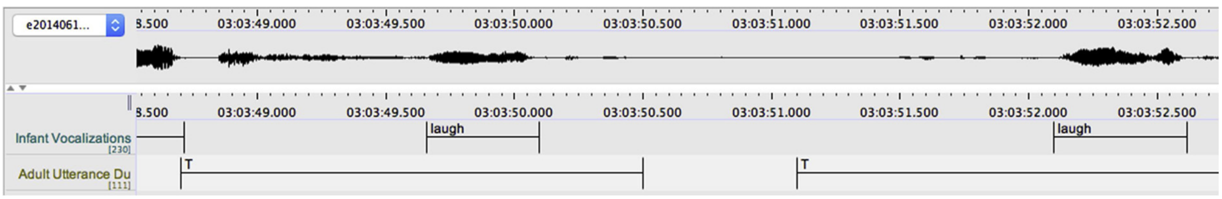

Fig. 2. A coded segment from a participant's recording, depicting examples of infant laughs overlapping infant-directed vocalizations (labeled "T". If the overlapping vocalization was coded as the first event, the later onset overlapping vocalization was coded as the second event. With P = 1, 2, or 5 s, the event series would be T R T R.

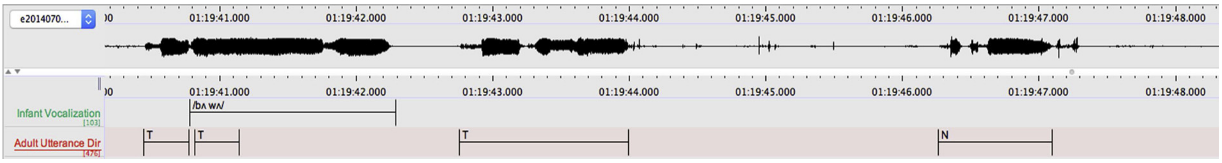

Fig. 3. A coded segment from a participant's recording, depicting an infant-directed vocalization (labeled "T") overlapping a (phonetically transcribed) canonical infant vocalization where a first event begins before and extends beyond the other. If the overlapping vocalization was coded as the first event, the later onset overlapping vocalization was coded as the second event. When P = 1 s, the event series would be: T C T C T P P N. With P = 2 s, the event series would be: T C T T P N. With P = 5 s, the event series would be: T C T T N.

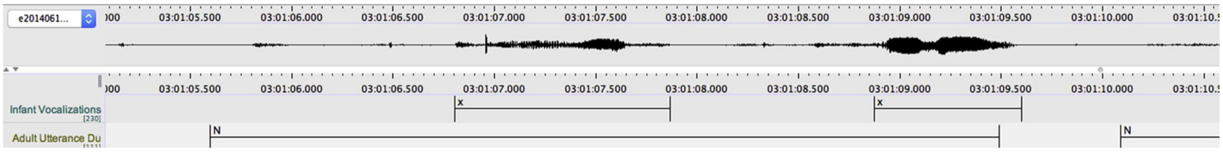

Fig. 4. A coded segment from a participant's recording, depicting non-canonical infant vocalizations (labeled "x") overlapping other-directed speech (labeled "N"). If the overlapping vocalization was coded as the first event, the later onset overlapping vocalization was coded as the second event. With P = 1 s, the event series would be: N X N X N. With P = 2 and 5 s, the event series would be: N X X N.

each of the minimum pause durations (using a 1-, 2-, or 5-second definition of "pause"). For each significant effect, we ran additional regression models to perform post hoc comparisons of the strength of the observed effects; a model was built which directly compared first events of one type to the first events of the contrasting type (ignoring other event types) in their likelihood of being followed by the second event type in question. The same approach was used to test temporal contingencies in the reverse direction; i.e., how well Infant-Directed and Other-Directed adult vocalizations as first events in the sequenced event pairs were uniquely associated with subsequent infant vocalizations of the three types. Although other researchers have used odds ratio analyses (e.g., Van Egeren et al., 2001; Bornstein et al., 2015), we chose to run mixed-effects regression analyses for their ability to easily include multiple predictor variables within the same analysis. Pilot analyses using odds ratio methods were generally consistent with the mixed effects regression results reported below. Separate analyses were run for each adult vocalization type and for each pause duration.

2.6. Acoustic analyses

We used Praat (Boersma & Weenink, 2018) to automatically estimate each adult vocalization's pitch contour. We used the autocorrelation method with a time step of 0.01 s, a pitch floor of 75 Hz, and a pitch ceiling of 1000 Hz. All other parameters were set to default values. The mean and standard deviation of the pitch contour were then calculated. We used mel units rather than Hz for these calculations to account for the nonlinearity pitch perception (and fundamental frequency production). We also used Praat to compute the mean intensity in dB SPL of each adult vocalization. We then ran three mixed effects linear regression models to test for differences between ID and OD vocalizations' mean pitch, pitch variability, and intensity, controlling for infant ID as a random effect. Intensity was included because it is very easy to measure and could conceivably be a very readily accessible cue to infants. Note, however, that intensity is from the perspective of the infant-worn microphone, not from the perspective of the adult speaker. Utterances less than 0.05 s were excluded, as were utterances for which Praat's output was undefined, which can happen in pitch analyses if no voicing is detected within the utterance.

2.7. Code and data availability

All analysis scripts and the 2-event sequence data are available at http://github.com/gpretzer/WWScripts. Additionally, 5-minute audio segments that did not contain last names or especially private episodes and their annotations from infants whose caregivers agreed to share the recordings publicly are available on FigShare (Pretzer, Warlaumont, Lopez, & Walle, 2018). Full LENA recordings from all participants who gave consent are available on HomeBank (VanDam et al., 2016) within the Warlaumont Corpus (Warlaumont, Pretzer, Walle, Mendoza, & Lopez, 2016).

3. Results

This section presents the findings of our current study. In the first subsection, we provide descriptive statistics including the relative frequencies of infant and adult vocal events from the entire data set. The next two subsections detail the relation between infant vocalizations and subsequent adult vocalizations as well as the relations between adult vocalizations and subsequent infant vocalizations. The final subsection presents acoustic analyses that consider both frequency and intensity of IDS and ODS vocalizations.

Table 1
Event counts per vocalization type across all infants, given 1 s, 2 s, or 5 s pause duration.

	Adult IDS	Adult ODS	Infant Canonical	Infant Non-Canonical	Infant Reflexive	Pause
1 s	2127	1541	630	2197	418	7302
2 s	2114	1526	613	2183	399	2919
5 s	2107	1519	604	2179	375	770

3.1. *Frequencies of infant and adult vocalization events*

Table 1 shows the total count for each infant and adult vocalization type for each minimum pause duration. Adult IDS was about 40% more frequent than adult ODS, but both were amply represented in the dataset. Non-canonical non-reflexive vocalizations were the most frequent type of infant vocalization, being about 3.5 times more common than canonical infant vocalizations and about 5.5 times more common than reflexive infant vocalizations. The number of pause events was, unsurprisingly, greatly dependent on pause duration. The numbers of vocalization event types changed slightly as a function of pause duration, due to the way pause duration effected how overlapping events were treated.

3.2. *Infant vocalizations predict subsequent adult vocalizations with various pause durations*

Table 2 shows the results for analyses of how strongly each of the three infant vocalization types uniquely predicted that the next event would be an adult Infant-Directed utterance when pause duration (P) was 1 s. Given the many planned tests, we used a conservative α cutoff ($p =$ .001) to indicate statistical significance. Canonical ($b = 0.99$, $p < .001$), Non-Canonical ($b = 0.70$, $p < .001$), and Reflexive ($b = 1.70$, $p < .001$) infant vocalizations each predicted a subsequent Infant-Directed adult vocalization. Further post-hoc comparisons with a conservative α cutoff ($p =$ .001) revealed that the Reflexive vocalizations more strongly predicted that an Infant-Directed adult utterance would follow than either Canonical or Non-Canonical vocalizations did. Although not included in Table 2, because the comparison between Canonical and Non-Canonical vocalizations here is of particular interest given results reported in prior literature (Gros-Louis & Miller, 2018) it is worth noting that Canonical infant vocalizations were marginally more likely than Non-Canonical to be followed by an Infant-Directed adult vocalization ($b = 0.32$, $p =$ .005) There were no statistically significant relations between Canonical, Non-Canonical, or Reflexive infant vocalizations and subsequent Other-Directed adult vocalizations.

Similarly, analyses of P = 2 s (see Table 2) revealed that Canonical ($b = 0.57$, $p < .001$), Non-Canonical ($b = 0.28$, p < .001), and Reflexive ($b = 1.19$, $p < .001$) infant vocalizations predicted a subsequent Infant-Directed adult vocalization. Post-hoc analyses found that Reflexive infant vocalizations predicted subsequent Infant-Directed adult vocalizations more strongly than did Non-Canonical infant vocalizations. Again, we make an exception to our $p =$ .001 cutoff to note that Canonical infant vocalizations were marginally more likely than Non-Canonical to be followed by an Infant-Directed adult vocalization ($b = 0.30$, $p =$ .007) However, unlike with the 1 s pause duration analyses, we also detected that Non-Canonical infant vocalizations were significantly associated with decreased likelihood of subsequent Other-Directed adult utterances ($b =$ -0.48, $p < .001$).

Finally, when P = 5 s (Table 2), only Reflexive infant vocalizations significantly predicted subsequent Infant-Directed adult

Table 2
Infant Vocalizations Predicting Subsequent IDS or ODS.

Pause Duration	First Event	Second Event	*b*	95% CI Lower Bound	95% CI Upper Bound	*p*	Statistically Significant Post-Hoc Comparisons
1 s	Canonical	IDS	0.57	0.66	1.32	< .001	
1 s	Non-Canonical	IDS	0.28	0.48	0.92	< .001	
1 s	Reflexive	IDS	1.19	1.32	2.09	< .001	Reflexive > Canonical; Reflexive > Non-Canonical
1 s	Canonical	ODS	−0.25	−0.15	0.69	0.03	
1 s	Non-Canonical	ODS	−0.48	−0.32	0.25	0.73	
1 s	Reflexive	ODS	−0.38	−0.44	0.58	0.53	
2 s	Canonical	IDS	0.57	0.23	0.90	< .001	
2 s	Non-Canonical	IDS	0.28	0.06	0.50	< .001	
2 s	Reflexive	IDS	1.19	0.80	1.58	< .001	Reflexive > Non-Canonical
2 s	Canonical	ODS	−0.25	−0.70	0.16	0.05	
2 s	Non-Canonical	ODS	−0.48	−0.77	−0.21	< .001	
2 s	Reflexive	ODS	−0.38	−0.92	0.10	0.01	
5 s	Canonical	IDS	0.13	−0.21	0.47	0.19	
5 s	Non-Canonical	IDS	−0.09	−0.31	0.13	0.19	
5 s	Reflexive	IDS	0.70	0.29	1.11	< .001	Reflexive > Non-Canonical
5 s	Canonical	ODS	−0.67	−1.11	−0.26	< .001	
5 s	Non-Canonical	ODS	−0.88	−1.14	−0.60	< .001	
5 s	Reflexive	ODS	−0.83	−1.38	−0.34	< .001	

Table 3
IDS or ODS Preceding Infant Vocalizations.

Pause Duration	First Event	Second Event	b	95% CI Lower Bound	95% CI Upper Bound	p	Statistically Significant Post-Hoc Comparisons
1 s	IDS	Canonical	0.51	0.15	0.86	< .001	
1 s	IDS	Non-Canonical	0.39	0.16	0.61	< .001	
1 s	IDS	Reflexive	1.52	1.13	1.90	< .001	Reflexive > Canonical; Reflexive > Non-Canonical
1 s	ODS	Canonical	0.35	−0.07	0.75	0.01	
1 s	ODS	Non-Canonical	0.12	−0.15	0.39	0.12	
1 s	ODS	Reflexive	0.15	−0.39	0.63	0.33	
2 s	IDS	Canonical	0.21	−0.15	0.73	0.05	
2 s	IDS	Non-Canonical	0.11	−0.11	0.34	0.09	
2 s	IDS	Reflexive	0.98	0.58	1.38	< .001	Reflexive > Canonical; Reflexive > Non-Canonical
2 s	ODS	Canonical	−0.15	−0.59	0.25	0.23	
2 s	ODS	Non-Canonical	−0.39	−0.67	−0.13	< .001	
2 s	ODS	Reflexive	−0.27	−0.79	0.21	0.08	
5 s	IDS	Canonical	−0.13	−0.48	0.21	0.23	
5 s	IDS	Non-Canonical	−0.16	−0.38	0.06	0.02	
5 s	IDS	Reflexive	0.53	0.11	0.94	< .001	Reflexive > Non-Canonical
5 s	ODS	Canonical	−0.50	−0.92	−0.10	< .001	
5 s	ODS	Non-Canonical	−0.76	−1.03	−0.49	< .001	
5 s	ODS	Reflexive	−0.66	−1.18	−0.17	< .001	

utterances ($b = 0.70$, $p < .001$). Post-hoc analyses revealed that Reflexive vocalizations were significantly stronger predictors of subsequent Infant-Directed speech than Non-Canonical vocalizations were ($p < .001$). Conversely, analyses of the 5 s minimum pause duration revealed that Canonical ($b = -0.67$, $p < .001$), Non-Canonical ($b = -0.88$, $p < .001$), and Reflexive ($b = -0.83$, $p < .001$) infant vocalizations all were associated with decreased likelihood of the following event being an Other-Directed adult utterance.

3.3. Adult vocalizations predict subsequent infant vocalizations with various pause durations

Table 3 shows the results for analyses of how each of the three infant vocalization types predicted that the previous event would have been an adult Infant-Directed utterance. When the maximum lag was $P = 1$ s, Infant-Directed adult utterances were significantly associated with subsequent Canonical ($b = 0.51$, $p < .001$), Non-Canonical ($b = 0.39$, $p < .001$), and Reflexive ($b = 1.52$, $p < .001$) infant vocalizations. Post-hoc comparisons determined that the Reflexive vocalizations were a stronger predictor of a preceding infant directed utterance than Canonical ($p < .001$) and Non-Canonical ($p < .001$) infant vocalizations. There were no significant relations between Other-Directed adult vocalizations and subsequent Canonical, Non-Canonical, and Reflexive infant vocalizations.

Analyses with P = 2 s (see Table 3) revealed that preceding Infant-Directed adult utterances were significantly associated with subsequent Reflexive infant vocalizations, and that this was a stronger association ($b = 0.98$, $p < .001$) than with Canonical and Non-Canonical infant vocalizations ($p < .001$). Further, we found that preceding Other-Directed adult vocalizations were negatively associated with subsequent Non-Canonical infant vocalizations ($b = -0.39$, $p < .001$).

Finally, we ran analyses with P = 5 s (see Table 3). Reflexive infant vocalizations ($b = 0.53$, $p < .001$) succeeded Infant-Directed adult utterances with a stronger association than Non-Canonical infant vocalizations. Second event Canonical ($b = -0.50$, $p < .001$), Non-Canonical ($b = -0.76$, $p < .001$), and Reflexive ($b = -0.66$, $p < .001$) infant vocalizations were negatively associated with first event Other-Directed adult utterances.

3.4. Acoustic analyses of IDS and ODS

Table 4 shows results of mixed effects regressions predicting acoustic features as a function of whether an adult vocalization was Infant-Directed or Other-Directed. Positive b values indicate that the measure was greater for ID vocalizations. Mean pitch (in mel) was significantly higher for ID vocalizations ($b = 23.28$, $p < .001$), as was vocalization intensity (in dB; $b = 5.57$, $p < .001$). Pitch

Table 4
Acoustic Differences for Infant-Directed vs. Other-Directed Adult Vocalizations.

Acoustic feature	IDS mean (SD)	ODS mean (SD)	b	p
Mean pitch (mel)	238.20 (73.53)	210.35 (81.27)	23.28	< .001
Pitch standard deviation (mel)	45.77 (36.65)	53.78 (48.16)	−3.09	.06
Mean intensity (dB)	70.23 (7.41)	64.46 (8.09)	5.57	< .001

variability, as measured by standard deviation in mel of a vocalization's pitch contour, did not significantly differ for ID vocalizations compared to OD vocalizations (b = -3.09, p = .06).

4. Discussion

4.1. Infant vocalizations predict subsequent infant-directed adult vocalizations, and vice versa

All three types of infant vocalizations (Reflexive, Canonical, and Non-Canonical Non-Reflexive) predicted that the following event would be an Infant-Directed adult vocalization, as long as the maximum lag between events was relatively short (1 or 2 s). When the allowed time lag was 5 s, a significant tendency for Infant-Directed adult vocalizations to follow infant vocalizations was detected only for reflexive infant vocalizations. This result is consistent with prior research (Bornstein et al., 2015; Van Egeren et al., 2001) showing that adults' infant-directed vocalizations toward younger infants are more likely to occur following infant vocalizations (both distress and non-distress), and that such responses typically occur within 1 or 2 s.

When looking at first-order sequential contingencies, these sequential relations between infant vocalizations and infant-directed adult vocalizations appear to be bidirectional. All three types of infant vocalizations were more likely to occur following infant-directed adult vocalizations than at other times. This finding is also consistent with prior research, which has typically found fairly symmetrical patterns of temporal association (Jaffe et al., 2001; Van Egeren et al., 2001).

4.2. Differences between infant vocalization types

Reflexive vocalizations were stronger predictors of subsequent IDS than non-canonical vocalizations. This is consistent with prior work utilizing hand-coding of younger infants' vocal interactions with their caregivers during shorter observation windows (Van Egeren et al., 2001), which has also found strong relationships between infant distress vocalizations of younger, 4-month-old infants, and infant-directed adult utterances. Although it is possible that reflexive vocalizations in 4-month-olds are different than in 12-month-olds (e.g., younger infants may produce a higher proportion of cries to laughs than older infants; younger infants may produce more reflexive vocalizations relative to speech-related vocalizations than older infants; adults may respond more sympathetically to cries from a 4-month-old than to cries from a 12-month-old etc.), it is still interesting to note that non-speech-related vocalizations are such strong predictors of adult responsivity. Given our treatment of overlapping vocalizations, these results are also consistent with Yoo, Bowman, and Oller (2018) who found that IDS is more likely to overlap with cries.

On the other hand, this finding seems to differ from Warlaumont et al. (2014) finding that adult responses were more likely to follow speech-related than reflexive infant vocalizations. Two methodological differences could account for this discrepancy. First, Warlaumont et al. (2014) relied on the automatic labels provided by the LENA system, which may have resulted in the speech-related vs. non-speech-related distinction not mapping as faithfully onto protophones vs. reflexive sounds as the present study's human listener based labels. Moreover, the use of automatic labeling necessitated that Warlaumont and colleagues ignore vocalizations overlapping with other sound sources, which might be especially likely when adults are responding to infant cries. Second, Warlaumont et al. (2014) analyzed the entire day-long audio recordings, whereas the present study analyzed only high-infant-volubility samples from within the day-long recordings. We would speculate that the difference in how overlapping vocalizations were treated across the studies is the main factor, but this and other suppositions should be subject to testing in future research.

Canonical vocalizations were also more likely than non-canonical vocalizations to be followed by IDS (p's > .001 but < .01), when considering 1 s and 2 s lags. This corresponds with previous work by Gros-Louis et al. (2006) finding that mothers in a laboratory setting were more likely to respond to infant consonant-vowel sounds (i.e., canonical sounds) than to vowel-like sounds (which we include in our definition of non-canonical vocalizations). These results also align with those of Gros-Louis and Miller (2018), who found that adults were more likely to provide vocal responses following consonant-vowel vocalizations (equivalent to our canonical vocalizations) than following vowel only vocalizations (included within our non-canonical vocalization).

Our results, particularly the bidirectionality of the findings, also fit with work by Goldstein, King, and West (2003) indicating that contingent maternal responses increased subsequent infant vocalizations that were more mature, namely canonical syllables, and of Papousek and Papousek (1989) demonstrating that infants alter the phonological structure of their vocalizations to better match their mothers' vocalizations. In light of such prior research, the bidirectionality observed in the present study may indeed reflect a true contingency of infant vocal type on preceding adult vocalization, though we cannot be certain because the same methodological issue with inferring causality discussed later also applies here.

4.3. Other-directed adult vocalizations are negatively associated with infant vocalizations

Our examination of first-order sequential contingencies with 5 s maximum lag between vocal utterances (i.e. a 5 s minimum pause duration) found that all three infant vocalization types were associated with a decreased likelihood that an Other-Directed adult vocalization would follow. This relationship was again bidirectional, with Other-Directed adult vocalizations predicting decreased likelihood of any type of infant vocalization as the next event.

This is a novel and interesting finding, suggesting that other-directed adult utterances, such as adult-directed adult vocalizations, are not unrelated to infant behavior. Prior work (Kuhl et al., 2003; Oller, 2010; Shneidman et al., 2013; Weisleder & Fernald, 2013) has found that unlike infant-directed adult speech, other-directed adult speech is not predictive of children's language learning. However, the present results suggest that OD speech should not be totally discounted. One plausible explanation for our results is that a reduction in adults' other-directed vocalizations may cue the infant that adults are paying attention to them. Infants might also vocalize to prompt adults in the environment to focus attention on the infant, which would in turn reduce other-directed adult speech as adults. Alternatively, infants might be more likely to vocalize at times when they are better able to hear their own vocalizations above the background noise of other conversations. Yet another possible explanation for the negative association between ODS and infant vocalization is that it is a byproduct of the positive temporal association between infant vocalizations and IDS (and potentially between successive infant vocalizations, not analyzed here). It is also possible that the human coders used in the present study were better able to hear infant vocalizations (especially short and quiet sounds that would have been categorized as Non-Canonical Non-Reflexive) without other-directed adult conversation taking place in the background. Regardless, the present findings make a case for the importance of considering not only infant-directed adult vocalizations but also other-directed adult vocalizations when considering the full range of contexts that infants experience over the course of a day. This may be particularly relevant when studying cultures that vary in the forms and/or frequency of infant-centered speech (see next section). Incorporation of video or other data could also help provide a clearer picture of the physical locations of people and other aspects of infants' physical and social contexts (see Fausey, Jayaraman, & Smith, 2016).

It is also noteworthy that the detection of the relationship (albeit a negative one) between other-directed adult vocalizations and infant vocalizations necessitated the use of a longer, 5 s criteria for a pause. The 2 s maximum lag revealed only a negative association for one of the infant vocalization types (Non-Canonical Non-Reflexive), and no negative associations with infant vocalization types were found for the 1 s maximum lag. Thus, while focusing on shorter timescales appears appropriate when examining infant-directed adult behavior occurring during typical parent-infant dyadic play sessions, it is important to consider longer timescales of possible associations between infant and adult behavior when examining naturalistic observations in which infants may compete with other individuals for adults' attention (and vice versa).

4.4. Acoustic differences between infant-directed and other-directed adult speech

In addition to finding qualitative differences in how IDS and ODS related temporally to infant vocalizations, we also observed differences in the acoustic features of IDS and ODS. On average, the mean pitch of infant-directed adult vocalizations was higher than that of other-directed adult vocalizations, consistent with the large body of prior work analyzing pitch of infant- versus adult-directed speech in more constrained settings (Broesch & Bryant, 2015; Fernald, Taeschner, Dunn, & Papousek, 1989).

We also found that infant-directed adult vocalizations had greater intensity (a.k.a. amplitude and the acoustic correlate of perceived loudness) as recorded by the infant-worn microphone compared to other-directed adult vocalizations; this is not something that has been found in prior research on ID speech acoustics. One possible explanation could be that observed ID speech was produced in closer proximity to the infant than overheard OD speech. Regardless of the cause, the intensity of adult vocalization could provide both an additional benefit for children's communication development and an additional cue to infants about which speech is intended for their benefit. The focus here on child-centered naturalistic recordings thus permitted the identification of a new acoustic feature distinguishing infant-directed speech in the real world, from the child's perspective.

Finally, unlike prior work (Fernald et al., 1989), we did not find greater pitch variability for infant-directed vocalizations. Our pitch measurements were based on automated analyses, which may be less accurate, and our data were from a wider range of contexts. Either of these factors could account for the lack of pitch variability effects seen here. Nevertheless, it suggests the possibility that caregivers' pitch variability may be less of a cue and aid to infant communication development than increased pitch and intensity overall.

The acoustic differences observed between infant-directed and other-directed adult utterances could help explain why infant-directed adult vocalizations show positive temporal association with infant vocalizations. The louder, infant-directed vocalizations may be more salient to infants. Moreover, their higher pitch, which more closely matches infants' own voices, may be beneficial for infants' learning to map their own vocalizations onto adult targets. Thus, the acoustic features that distinguish ID speech may serve to stimulate infant vocalization. In turn, infant vocalizations, especially those indicating distress and those that are more mature, may cue adults to modify their vocalization acoustics in order to distract, teach, or soothe the infant. It also seems plausible that different types of infant vocalization may elicit different types of acoustic modifications by adults, both in their infant-directed and in their other-directed speech, and vice versa. Other research automatically analyzing pitch of infant and adult vocalizations within day-long audio recordings (but not distinguishing child-directed from other-directed adult speech) has found evidence for pitch matching between 12–30 month old children and their caregivers (Ko, Seidl, Cristia, Reimchen, & Soderstrom, 2016).

4.5. The difficulty of inferring causality from first-order temporal contingencies

Our findings that patterns of temporal association were predominantly bidirectional, including the associations between reflexive vocalizations (where cries greatly outnumbered laughs) and infant-directed adult vocalizations, highlights issues of inferring causality from first-order temporal associations. Though possible, it seems unlikely that the high tendency for infant-directed adult vocalizations to be followed by infant cries is due to infant cries being *caused* by adult vocalizations. A more likely explanation is that infant cries tended to occur in bouts, with adults responding to infant cries within such bouts. More formally, where $P(B \mid A)$ = the probability of vocal event B occurring given that vocal event A has just occurred, it is unclear if the pattern of infant vocalizations tending to follow adult vocalizations (i.e., $P(\text{infant} \mid \text{adult}) > P(\text{infant} \mid \text{not adult})$), is due to $P(\text{infant} \mid \text{infant} \rightarrow \text{adult}) > P(\text{infant} \mid \text{infant} \rightarrow \text{not adult})$ or whether it is due to $P(\text{infant} \mid \text{infant} \rightarrow \text{any event}) > P(\text{infant} \mid \text{not infant} \rightarrow \text{any event})$ and $P(\text{adult} \mid \text{infant}) > P(\text{adult} \mid \text{not infant})$. This latter scenario would be the case if infants tend to produce their vocalizations over a protracted period of time or in bouts while adults have a tendency to respond to the earlier infant vocalizations within those series. A symmetrical set of indistinguishable possibilities also applies for explaining the P (adult | infant) > P (adult | not infant) pattern.

4.6. Additional future directions

Although we have referred to the present study as a “comprehensive” study of first-order temporal contingencies between infant and adult vocalizations (because we included three main categories of infant vocalizations, both infant-directed and non-infant-directed input from adults and analyzed three different possible timescales at which such contingencies could operate), there exist many possible useful extensions. For instance, Golinkoff et al. (2015) have discussed the idea of quality vs. quantity in regard to IDS. The present study identified the *quantity* of vocalizations in temporal relation to infant behavior. However, future studies could examine certain *qualitative* aspects of the IDS utterances, such as the pragmatics (e.g., imperative, interrogative, prohibition, etc.) or function (i.e., is the utterance labeling an object, expanding upon an infant utterance, recasting an infant utterance, etc.) of the IDS utterances within day-long naturalistic recordings. Although previous work has examined some qualitative aspects of IDS, including part vs. whole within word labels (Gogate et al., 2013), adjusting levels of sensitivity (Zukow-Goldring, 1996), or directives (Reddy, Liebal, Hicks, Jonnalagadda, & Chintalapuri, 2013), such research has not used data from daylong home recordings. Similarly, infant vocalizations could also be broken down into a much more fine-grained set of types, such as distinguishing between different types of non-canonical babble.

We also encourage future work examining three-event sequences and/or use other analytic approaches or experimental designs to investigate the causal underpinnings of the observed bidirectional associations. Some prior studies have addressed three-event sequential patterns (Gros-Louis & Miller, 2018; Warlaumont et al., 2014; Harbison et al., submited, though their methods would need to be adapted in order to distinguish the specific alternatives posed above. A related approach would be to apply Granger causality analysis to consider a larger range of timescales and possible patterns of temporal association across speakers and vocalization types (see Xu, Abney, & Yu, 2017, for an example of this approach applied to multi-modal parent-infant interactions).

Additionally, the methods used in the present study could also be applied to recordings of children at other ages (e.g., Jaffe et al., 2001; Tamis-LeMonda et al., 2001), cultures (Lee, Jhang, Relyea, Chen, & Oller, 2018), or populations, such as children with clinical or medical diagnoses, or who are at-risk for language disorders (see Yoder & Warren, 2002). For example, different cultures have different norms regarding the frequency and nature of child-directed speech and speech directed to other children and adults. As mentioned above, it has been reported that in Kaluli culture, infant-directed speech is rare and is not typically used to elicit an infant response, whereas these infants instead hear a good deal of high-pitched other-directed but infant-centered speech in which caregivers speak “for” younger (under 6-months-old) prelinguistic infants (Ochs & Schieffelin, 1984). It would be interesting to separately code infant-directed, infant-centered other-directed (both by the adults speaking-for and others’ speech to the adults doing the speaking-for), and not-infant-centered other-directed adult Kaluli speech to assess the temporal contingencies between each adult vocalization type and infant vocal behavior, and the relation of these contingencies with specific acoustic features. It has been asserted that Kaluli caregivers don’t usually speak for a prelinguistic infant as an interpretation of infant vocalizations, but rather in response to other contextual factors and non-vocal infant actions. Thus, one might predict that infant-centered other-directed speech would not have the same temporal associations with infant vocalization as was observed for infant-directed speech in the present study.

Even within industrialized Western cultures there exists substantial variability in the frequency of adult input directed to the child and in how much adults facilitate infant-adult conversational and proto-conversational exchanges (Hart & Risley, 1995). The temporal contingency analyses controlled for base rates of adult and infant vocalizations and individual or subgroup differences were not examined. It is possible that base rates of adult vocalization may correlate with the strength of infant-adult temporal contingencies, but it is also possible that contingencies can be strong despite relatively low base rates. This would be an interesting topic for future work comparing both within and across cultures. Additionally, as discussed in the section above, this present study does not allow us to make any conclusions about causality. In order to make definitive claims about causality would require tightly controlled experiments manipulating IDS and ODS and testing for differences in infant vocalizations. Finally and relatedly, future work could relate patterns of contingency, base rates of behaviors, and acoustic modifications in infant-adult vocal interactions from within day-long audio recordings to measurable language, social, emotional and other outcomes.

5. Conclusion

Sampling day-long home audio recordings provides a set of vocalizations that are more representative of infants’ typical experiences than those used in many prior studies. Manually annotating segments of day-long home audio recordings enabled us to assess the relationship between various types of infant vocalizations and *infant*-directed and *other*-directed adult speech. Our findings support prior findings of temporal relationships between infant-directed adult speech and infant vocalizations of all types, indicating that such patterns generalize to natural home environments. Moreover, we identified additional patterns of parent-infant vocal interactions, particularly that other-directed adult vocalization and infant vocalization are negatively associated with each other. Finally, that all the detected contingencies were found to be bidirectional highlights some of the issues with inferring causality from first-order temporal contingencies.

Acknowledgments

We would like to thank Iman Mosleh, Juan Duran, Alisa Hajou, and Vanessa Romo Ortega for help with data collection, Evelyn Alvarez, Paola Flores, and Jill Mead for annotating and transcribing utterances, Saraching Chao for assistance with the reliability analyses, and Ale Fontana and Brian MacWhinney for assistance preparing the files for public sharing. We would also like to thank Stephanie Shih and Teenie Matlock for comments on an earlier draft. This work was funded by the National Science Foundation through grant numbers BCS-1529127 and SMA-1539129/1827744, by a James S. McDonnell Foundation Fellowship in Understanding Human Cognition, and by a Faculty Research Grant from the University of California, Merced Graduate and Research Council.

References

Adamson, L. B., & Bakeman, R. (2006). Development of displaced speech in early mother–child conversations. *Child Development, 77*(1), 186–200.
Akhtar, N., Dunham, F., & Dunham, P. J. (1991). Directive interactions and early vocabulary development: The role of joint attentional focus. *Journal of Child Language, 18*(1), 41–49.
Akhtar, N., Jipson, J., & Callanan, M. A. (2001). Learning words through overhearing. *Child Development, 72*(2), 416–430.
Bakeman, R. (1997). *Observing interaction: An introduction to sequential analysis* (2nd ed.). New York: Cambridge University Press.
Baldwin, D., & Meyer, M. (2007). *How inherently social is language? Blackwell Handbook of Language Development*87–106.
Bates, D., Maechler, M., Bolker, B., & Walker, S. (2018). *lme4: Linear mixed-effects models using 'Eigen' and S4 [R Package]*.
Bergelson, E., Amatuni, A., Dailey, S., Koorathota, S., & Tor, S. (2018). Day by day, hour by hour: Naturalistic language input to infants. *Developmental Science*e12715.
Bloom, K. (1988). Quality of adult vocalizations affects the quality of infant vocalizations. *Journal of Child Language, 15*(03), 469–480.
Bloom, K., Russell, A., & Wassenberg, K. (1987). Turn taking affects the quality of infant vocalizations. *Journal of Child Language, 14*(02), 211–227. https://doi.org/10.1017/S0305000900012897.
Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer [Computer program]. Version 6.0.43.* retrieved 11 November 2018 fromhttp://www.praat.org/.
Bornstein, M. H., Putnick, D. L., Cote, L. R., Haynes, O. M., & Suwalsky, J. T. D. (2015). Mother-infant contingent vocalizations in 11 countries. *Psychological Science, 26*(8), 1272–1284.
Broesch, T. L., & Bryant, G. A. (2015). Prosody in infant-directed speech is similar across Western and traditional cultures. *Journal of Cognition and Development, 1*, 31–43.
Buder, E. H., Warlaumont, A. S., & Oller, D. K. (2013). *An acoustic phonetic catalog of prespeech vocalizations from a developmental perspective. Comprehensive perspectives on child speech development and disorders: Pathways from linguistic theory to clinical practice.* 103–134.
Cristia, A. (2011). Fine-grained variation in caregivers' /s/ predicts their infants' /s/ category. *The Journal of the Acoustical Society of America, 129*, 3271–3280.
ELAN [Computer software] (2018). *Nijmegen: Max planck institute for psycholinguistics.* Retrieved fromhttps://tla.mpi.nl/tools/tla-tools/elan/.
Fausey, C. M., Jayaraman, S., & Smith, L. B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition, 152*, 101–107.
Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior & Development, 8*(2), 181–195.
Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development,* 1497–1510.
Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior & Development, 10*(3), 279–293.
Fernald, A., Taeschner, T., Dunn, J., & Papousek, M. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language, 16*(3), 477–501.
Field, T. (1980). Supplemental stimulation of preterm neonates. *Early Human Development, 4*(3), 301–314.
Floor, P., & Akhtar, N. (2006). Can 18-month-old infants learn words by listening in on conversations? *Infancy, 9*(3), 327–339.
Fox, J., Weisberg, S., Price, B., Adler, D., Bates, D., Baud-Bovy, G., ... R-Core (2018). *Car: Companion to applied regression [R package]*.
Franklin, B., Warlaumont, A. S., Messinger, D., Bene, E., Nathani Iyer, S., Lee, C.-C., ... Oller, D. K. (2014). *Language Learning and Development, 10*(3), 279–296.
Gogate, L. J., Maganti, M., & Laing, K. B. (2013). Maternal naming of object wholes versus parts to preverbal infants: A fine-grained analysis of scaffolding at 6–8 months. *Infant Behavior & Development, 36*(3), 470–479.
Goldin-Meadow, S., Goodrich, W., Sauer, E., & Iverson, J. (2007). Young children use their hands to tell their mothers what to say. *Developmental Science, 10*(6), 778–785.
Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science, 19*, 515–522.
Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences, 100*(13), 8030–8035.
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.
Goldstein, M. H., Schwade, J., Briesch, J., & Syal, S. (2010). Learning while babbling: Prelinguistic object-directed vocalizations indicate a readiness to learn. *Infancy, 15*(4), 362–391.
Golinkoff, R. M. (1986). 'I beg your pardon?': The preverbal negotiation of failed messages. *Journal of Child Language, 13*(3), 455–476.
Golinkoff, R. M., & Alioto, A. (1995). Infant-directed speech facilitates lexical learning in adults hearing Chinese: Implications for language acquisition. *Journal of Child Language, 22*, 703–726.
Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby)Talk to me: The social contexts of IDS and its effects on early language acquisition. *Current Directions in Psychological Science, 24*(5), 339–344.
Gros-Louis, J., & Miller, J. L. (2018). From 'ah'to 'bah': social feedback loops for speech sounds at key points of developmental transition. *Journal of Child Language, 45*(3), 807–825.
Gros-Louis, J., West, M. J., Goldstein, M. H., & King, A. P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of Behavioral Development, 30*(6), 509–516.
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.
Harbison, A. L., Woynaroski, T. G., Wade, J. W., Warlaumont, A. S., & Yoder, P. J. (2019). A new measure of child vocal reciprocity in children with autism spectrum disorder (in press) *Autism Research*. https://doi.org/10.1002/aur.1942.
Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experiences of young American children.* Baltimore, MD: Brookes Publishing.
Henning, A., & Striano, T. (2011). Infant and maternal sensitivity to interpersonal timing. *Child Development, 82*(3), 916–931.
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, 1492*. https://doi.org/10.3389/fpsyg.2015.01492.
Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., Jasnow, M. D., Rochat, P., ... Stern, D. N. (2001). Rhythms of dialogue in infancy: Coordinated timing in development. *Monographs of the Society for Research in Child Development, 66*(2), 1–149.
Keller, H., Lohaus, A., Völker, S., Cappenberg, M., & Chasiotis, A. (1999). Temporal contingency as an independent component of parenting behavior. *Child Development, 70*(2), 474–485.
Ko, E. S., Seidl, A., Cristia, A., Reimchen, M., & Soderstrom, M. (2016). Entrainment of prosody in the interaction of mothers with their young children. *Journal of Child Language, 43*(2), 284–309.
Koopmans-van Beinum, F. J., & van der Stelt, J. M. (1986). *Early stages in the development of speech movements. Precursors of early speech.* UK: Palgrave Macmillian37–50.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., ... Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science, 277*(5326), 684–686.

Kuhl, P. K., Tsao, F.-M., & Liu, H.-M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences of the United States of America, 100*(15), 9096–9101.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). *lmerTest: Tests in linear mixed effects models [R package]*.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310.

Lee, C. C., Jhang, Y., Relyea, G., Chen, L. M., & Oller, D. K. (2018). Babbling development as seen in canonical babbling ratios: A naturalistic evaluation of all-day recordings. *Infant Behavior & Development, 50*, 140–153.

Lee, R., Skinner, A., Bornstein, M. H., Radford, A. N., Campbell, A., Graham, K., ... Pearson, R. M. (2017). Through babies' eyes: Practical and theoretical considerations of using wearable technology to measure parent-infant behavior from the mothers' and infants' view points. *Infant Behavior & Development, 47*, 62–71.

Lloyd, B. P., Yoder, P. J., Tapp, J., & Staubitz, J. L. (2015). The relative accuracy and interpretability of five sequential analysis methods: A simulation study. *Behavior Research Methods, 48*, 1482–1491.

Ma, W., Golinkoff, R. M., Houston, D., & Hirsh-Pasek, K. (2011). Word learning in infant- and adult- directed speech. *Language Learning and Development, 7*, 209–225.

Masur, E. F. (1987). Imitative interchanges in a social context: Mother-infant matching behavior at the beginning of the second year. *Merrill-Palmer Quarterly, 4*(33), 453–472.

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.

Nathani, S., & Stark, R. E. (1996). Can conditioning procedures yield representative infant vocalizations in the laboratory? *First Language, 16*(48), 365–387.

Ochs, E., & Schieffelin, B. B. (1984). Language acquisition and socialization: Three developmental stories and their implications. In R. A. Shweder, & R. A. LeVine (Eds.). *Culture theory: Essays on mind, self, and emotion* (pp. 276–320). Cambridge University Press.

Oller, D. K. (1980). The emergence of speech sounds in infancy. *Child phonology, 1*, 93–112.

Oller, D. K. (2000). *The emergence of the speech capacity.* Mahwah, NJ: Erlbaum.

Oller, D. K. (2010). All-day recordings to investigate vocabulary development: A case study of a trilingual toddler. *Communication Disorders Quarterly, 31*(4), 213–222.

Oller, D. K., Eilers, R. E., Urbano, R., & Cobo-Lewis, A. B. (1997). Development of precursors to speech in infants exposed to two languages. *Journal of Child Language, 24*(02), 407–425.

Oller, D. K., Niyogi, P., Gray, S., Richards, J. A., Gikerson, J., Xu, D., Yapanel, U., & Warren, S. F. (2010). Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. *Proceedings of the National Academy of Sciences of the United States of America, 107*(30), 13354–13359.

Olson, J., & Masur, E. F. (2011). Infants' gestures influence mothers' provision of object, action and internal state labels. *Journal of Child Language, 38*(5), 1028–1054.

Papousek, M., & Papousek, H. (1989). Forms and functions of vocal matching in interactions between mothers and their precanonical infants. *First Language, 9*(6), 137–157.

Patten, E., Labban, J. D., Casenhiser, D. M., & Cotton, C. L. (2016). Synchrony detection of linguistic stimuli in the presence of faces: Neuropsychological implications for language development in ASD. *Developmental Neuropsychology, 41*(5-8), 362–374.

Pelaéz, M., Virues-Ortega, J., & Gewirtz, J. L. (2011). Reinforcement of vocalizations through contingent vocal imitation. *Journal of Applied Behavior Analysis, 44*(1), 33–40.

Porritt, L. L., Zinser, M. C., Bachorowski, J. A., & Kaplan, P. S. (2014). Depression diagnoses and fundamental frequency-based acoustic cues in maternal infant-directed speech. *Language Learning and Development, 10*(1), 51–67.

Pretzer, G., Warlaumont, A., Lopez, L., & Walle, E. (2018). *Infant and adult vocalizations in home audio recordings.* https://doi.org/10.6084/m9.figshare.6108392.v1.

R Core Team (2018). *R: A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing.* Retrieved fromhttps://www.R-project.org.

Ramírez-Esparza, N., Garcia-Sierra, A., & Kuhl, P. K. (2014). Look who's talking: Speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science, 17*(6), 880–891.

Reddy, V., Liebal, K., Hicks, K., Jonnalagadda, S., & Chintalapuri, B. (2013). The emergent practice of infant compliance: An exploration in two cultures. *Developmental Psychology, 49*(9), 1754.

Rollins, P. R. (2003). Caregivers' contingent comments to 9-month-old infants: Relationships with later language. *Applied Psycholinguistics, 24*(2), 221–234.

Roug, L., Landberg, I., & Lundberg, L.-J. (1989). Phonetic development in early infancy: A study of four Swedish children during the first eighteen months of life. *Journal of Child Language, 16*(1), 19–40. https://doi.org/10.1017/S0305000900013416.

Schuller, B., Steidl, S., Batliner, A., Bergelson, E., Kajewski, J., Janott, C., ... Zafeiriou, S. (2017). The INTERSPEECH 2017 computational paralinguistics challenge: Addressee, cold, & snoring. *Proceedings of INTERSPEECH 2017,* 3442–3446. https://doi.org/10.21437/Interspeech.2017-43.

Shneidman, L. A., Arroyo, M. E., Levine, S., & Goldin-Meadow, S. (2013). What counts as effective input for word learning? *Journal of Child Language, 40*, 672–686.

Song, J. Y., Demuth, K., & Morgan, J. L. (2010). Effects of the acoustic properties of infant-directed speech on infant word recognition. *The Journal of the Acoustical Society of America, 128*, 389–400.

Stark, R. E. (1980). Stages of speech development in the first year of life. *Child phonology, 1*, 73–90.

Stoel-Gammon, C. (1989). Prespeech and early speech development of two late talkers. *First Language, 9*, 207–224.

Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development, 72*(3), 748–767.

Todd, G. A., & Palmer, B. (1968). Social reinforcement of infant babbling. *Child Development,* 591–596.

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.

VanDam, M., Warlaumont, A. S., Bergelson, E., Cristia, A., Soderstrom, M., De Palma, P., ... MacWhinney, B. (2016). HomeBank: An Online Repository of Daylong Child-Centered Audio Recordings. *Seminars in Speech and Language, 37*(02), 128–142.

Vihman, M. M., Macken, M. A., Miller, M., Simmons, H., & Miller, J. (1985). From babbling to speech: A re-assessment of the continuity issue. *Language, 61*(2), 297–445.

Warlaumont, A. S., & Ramsdell-Hudock, H. L. (2016). Detection of total syllables and canonical syllables in infant vocalizations. *Proceedings of INTERSPEECH 2016,* 2676–2680. https://doi.org/10.21437/Interspeech.2016-1518.

Warlaumont, A. S., Pretzer, G. M., Walle, E. A., Mendoza, S., & Lopez, L. D. (2016). *Warlaumont HomeBank Corpus.* https://doi.org/10.21415/T54S3C.

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.

Watson-Gegeo, K. A., & Gegeo, D. W. (1986). Calling-out and repeating routines in Kwara'ae children's language socialization. *Language socialization across cultures, 3*, 17–50.

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experiences strengthens processing and builds vocabulary. *Psychological Science, 24*(11), 2143–2152.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). *ELAN: A professional framework for multimodality research. Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*1556–1559.

Wu, Z., & Gros-Louis, J. (2014). Infants' prelinguistic communicative acts and maternal responses: Relations to linguistic development. *First Language, 34*(1), 72–90.

Xu, T. L., Abney, D. H., & Yu, C. (2017). Discovering multicausality in the development of coordinated behavior. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.). *Proceedings of the 39th annual meeting of the cognitive science society* (pp. 1369–1374). .

Yoder, P. J., & Symons, F. J. (2010). *Observational measurement of behavior.* New York: Springer Pub.

Yoder, P. J., & Warren, S. F. (1999). Maternal responsivity mediates the relationship between prelinguistic intentional communication and later language. *Journal of*

Early Intervention, 22(2), 126–136.
Yoder, P. J., & Warren, S. F. (2002). Effects of prelinguistic milieu teaching and parent responsivity education on dyads involving children with intellectual disabilities. *Journal of Speech Language and Hearing Research, 45*(6), 1158–1174.
Yoder, P. J., Davies, B., & Bishop, K. (1994). Reciprocal sequential relations in conversations between parents and children with developmental delays. *Journal of Early Intervention, 18*(4), 362–379.
Yoo, H., Bowman, D. A., & Oller, D. K. (2018). The origin of protoconversation: An examination of caregiver responses to cry and speech-like vocalizations. *Frontiers in Psychology, 9*, 1510.
Zukow-Goldring, P. (1996). Sensitive caregiving fosters the comprehension of speech: When gestures speak louder than words. *Early Development and Parenting: An International Journal of Research and Practice, 5*(4), 195–211.