INFANCY

THE OFFICIAL JOURNAL OF THE
INTERNATIONAL CONGRESS
OF INFANT STUDIES

Infancy, **22**(4), 514–539, 2017
Copyright © International Congress of Infant Studies (ICIS)
ISSN: 1525-0008 print / 1532-7078 online
DOI: 10.1111/infa.12165

WILEY Blackwell

Multiple Coordination Patterns in Infant and Adult Vocalizations

Drew H. Abney and Anne S. Warlaumont
Cognitive and Information Sciences
University of California, Merced

D. Kimbrough Oller
School of Communication Sciences and Disorders
University of Memphis and
Institute for Intelligent Systems
University of Memphis and
The Konrad Lorenz Institute for Evolution and Cognition Research

Sebastian Wallot
Interacting Minds Center
Aarhus University

Christopher T. Kello
Cognitive and Information Sciences
University of California, Merced

The study of vocal coordination between infants and adults has led to important insights into the development of social, cognitive, emotional, and linguistic abilities. We used an automatic system to identify vocalizations produced by infants and adults over the course of the day for fifteen infants studied longitudinally during the first 2 years of life. We measured three different types of vocal coordination: coincidence-based, rate-based, and cluster-based. Coincidence-based coordination and rate-based coordination are established measures in the developmental literature. Cluster-based coordination is new and measures the strength of matching in the degree to which vocalization events occur in hierarchically nested clusters. We investigated whether various coordination patterns differ as a function of vocalization type, whether different coordination patterns provide unique information about the dynamics of vocal interaction, and how the various coordination patterns each relate to infant age. All vocal coordination patterns displayed greater coordination for infant speech-related vocalizations, adults adapted the hierarchical clustering of their vocalizations to match that of infants, and each of the three coordination patterns had unique associations with infant age. Altogether, our results indicate that vocal coordination between infants and adults is

---

Correspondence should be sent to Drew H. Abney, Department of Psychological and Brain Sciences,1101 E. 10th St., Indiana University, Bloomington, Indiana, 47405. E-mail: dhabney@indiana.edu

multifaceted, suggesting a complex relationship between vocal coordination and the development of vocal communication.

The progression to speech-like vocalizations is a fundamental component of language learning (Oller, 2000) and is influenced by infant–adult vocal interaction (Bateson, 1975; Bloom, Russell, & Wassenberg, 1987; Goldstein, King, & West, 2003; Goldstein & Schwade, 2008; Gros-Louis, West, Goldstein, & King, 2006; Jaffe et al., 2001; Kokkinaki & Kugiumutzakis, 2000; Nathani & Stark, 1996; Northrup & Iverson, 2015; Papousek & Papousek, 1989; Ramírez-Esparza, García-Sierra, & Kuhl, 2014; Warlaumont, Richards, Gilkerson, & Oller, 2014; Weisleder & Fernald, 2013). Likewise, the quality of these vocal interactions has been shown to predict other social and cognitive behaviors later in development. For example, in the seminal work by Jaffe et al. (2001), the authors found that the degree of vocal rhythmic coordination at 4 months of age predicted levels of attachment and cognitive outcomes at twelve months of age. There are additional studies exemplifying the notion that the degree of vocal interaction, either characterized in terms of temporal coordination (e.g., Feldman & Greenbaum, 1997) or characterized in terms of other properties such as vocalization rate (e.g., Allely et al., 2013), predicts important developmental outcomes. Studies of vocalization properties and vocal coordination patterns are used to build theories of attachment (e.g., Bowlby, 1969) and social learning (e.g., Landry, Smith, & Swank, 2006) in addition to being markers of typical and atypical development (e.g., Oller et al., 2010; Patten et al., 2014; Warlaumont et al., 2014).

Finding new vocal coordination patterns and understanding the relationships between existing vocal coordination patterns might provide new insights into development. Furthermore, to advance our understanding on vocalization properties and vocal coordination, it is important to understand the similarities and differences between different measures of vocal coordination. In this study, we investigate three different types of vocal coordination: coincidence-based, rate-based, and cluster-based. Coincidence-based coordination and rate-based coordination have been previously used in a number of studies to study vocal interactions. Cluster-based coordination is a new measure recently introduced in the study of vocal interaction during adult conversation (Abney, Paxton, Dale, & Kello, 2014).

Coincidence-based coordination is based on the co-occurrence of vocalizations produced by two interlocutors within some minimal period of time. It includes both covocalizations (Harder, Lange, Hansen, Væver, & Køppe, 2015) and turn taking. Jaffe et al. (2001) observed that infant and adult vocalizations were contingent on each other up to a lag of 60 sec. They also observed that the strongest coordination patterns recurred every ~20–30 sec (see also, Feldstein et al., 1993), which they suggested was the optimal interaction "rhythm". The degree of coincidence-based coordination was predictive of various measures of attachment and development. To quantify coincidence-based coordination in this study, we used cross-recurrence quantification analysis to measure the degree to which infant and adult vocalizations occurred close together in time (Coco & Dale, 2014; Cox & van Dijk, 2013; Dale, Warlaumont, & Richardson, 2011; Fusaroli, Konvalinka, & Wallot, 2014; Marwan, Romano, Thiel, & Kurths, 2007; Warlaumont et al., 2014).

We based our measure of coincidence-based coordination on the timing of acoustic onsets of infant and adult vocalizations. Many previous studies have used similar measures of vocalization to study coordination. For example, van Egeren, Barratt, and

Roach (2001) found coordinated interaction within a temporal window of ~3 sec between the onset of a vocalization produced by an infant and the onset of a vocalization response by the mother or vice versa (Harder et al., 2015). Akin to the measure of coincidence-based vocal coordination used in this study, Warlaumont et al. (2014) observed that local coordination in timing of vocalizations across children and their caregivers differed as a function of vocalization type, and whether the infant was typically developing (TD) or diagnosed with autism spectrum disorder (ASD). Child speech-related vocalizations were more likely to receive an adult response relative to non-speech-related vocalizations. Children were also more likely to produce a speech-related vocalization if their previous speech-related vocalization received a response from their caregiver. Furthermore, relative to ASD children, TD children had more frequent vocal interaction with their caregivers and were more likely to lead vocalization interactions.

Rate-based vocal coordination is the degree of matching in the frequency or rate of a particular vocal behavior or property. One example of rate-based coordination is *volubility matching*. Volubility is the quantity or rate of vocalization per unit time, and volubility matching quantifies the similarity between infant and adult volubility across a given recording session. Much work has demonstrated volubility to be an important predictor of vocal development and communication (Franklin et al., 2013; Gilkerson & Richards, 2009; Goldstein & West, 1999; Goldstein, Schwade, & Bornstein, 2009; Hart & Risley, 1995; Hsu, Fogel, & Messinger, 2001; Oller, Eilers, Basinger, Steffens, & Urbano, 1995; Rescorla & Ratner, 1996; Warlaumont et al., 2014), but less work has quantified its coordination across infant and caregiver pairings. In one study, Hart and Risley (1999) found a positive relationship between infant and adult volubility. Other studies have examined effects of adult interactions more generally on infant volubility (Bloom et al., 1987; Franklin et al., 2013; Goldstein et al., 2009) and effects of adult volubility on child language learning (Ramírez-Esparza et al., 2014; Weisleder & Fernald, 2013), and cognitive and perceptual abilities (Greenwood, Thiemann-Bourque, Walker, Buzhardt, & Gilkerson, 2010; Jaffe et al., 2001).

Cluster-based vocal coordination measures the degree to which *temporal events cluster similarly* in infant and adult vocalizations. The acoustic energy in human vocalizations tends to be clustered in time (Abney, Kello, & Warlaumont, 2015; Abney, Paxton et al., 2014; Luque, Luque, & Lacasa, 2015), in that there are frequent starts and stops due to many factors, including breathing, fluctuations in intensity, emotion. Clustering in speech vocalizations also emerges from variations in the sonority of phonetic features, prosody, and pauses due to thought and emphasis. Clustering in acoustic speech energy may also relate to the hierarchical clustering of linguistic units (Grosjean, Grosjean, & Lane, 1979): phonemes cluster into syllables, syllables cluster into words, words cluster into phrases, phrases cluster into sentences, sentences cluster into larger discourse-like structures, etc. (Pickering & Garrod, 2004).

Prelinguistic vocalizations, although not yet bounded by linguistic structure, show precursors to the hierarchical grouping of vocal events of mature speakers. For example, prelinguistic vocalizations produced by infants have been observed to follow a structure of hierarchical clustering at the grouping levels of syllables, utterances, and phrases (Lynch, Oller, Steffens, & Buder, 1995). Here, we aim to extend this work by Lynch et al. by quantifying the degree to which infant vocalizations, and the adult vocalizations to which they are exposed, cluster across the day at timescales ranging from seconds to hours. In this study, we investigate the developmental relationship between hierarchical clustering of temporal events in infant vocalization bouts versus adult vocalizations heard by infants, in addition to other vocal coordination patterns reflecting temporal-based and rate-based vocal coordination. It is generally accepted that the conversational exchange between interlocutors is a dynamic interplay with reciprocal effects (Snow, 1977) and that understanding how infants and adults modulate vocalization properties during conversational exchanges and across development is crucial for understanding typical and atypical communicative development.

In addition to investigating differences in the degree of coordination across the three levels of description (coincidence-based coordination, rate-based coordination, and cluster-based coordination), we can also investigate directions of convergence of these vocalization properties across infants and adults. For example, does volubility rate of caregiver vocalizations adapt to that of the infant? Or vice versa? Additionally, does clustering of caregiver vocalizations adapt to the hierarchical clustering of the infant?

Goals of the current study

The purpose of this study was to examine the development of various types of vocal coordination across infancy and determine whether different patterns are interrelated or independent. We used the LENA™ (Language ENvironment Analysis) system (LENA Foundation, Boulder, CO) to collect naturalistic, daylong audio recordings from fifteen infants. The recordings are from an ongoing study in which infants are followed longitudinally during the first two years of life. The LENA system captures and automatically locates both infant and adult vocalizations. This study seeks to answer three main questions revolving around the general theme of coordination patterns in vocal interaction: (1) Do coincidence-based, rate-based, and cluster-based coordination patterns vary depending on the vocalization type produced by the infant? (2) Do different coordination patterns provide unique information about the dynamics of vocal interaction? (3) How do the various coordination patterns relate to infant age?

METHOD

Participants

Participants were fifteen infants (7 females, 8 males) from an ongoing longitudinal study. Fourteen were exposed primarily to English, and one was exposed primarily to German. The final analysis included 706 recording sessions; thus, the average number of recordings per participant was 47.06 (*SD* = 11.53). The range of earliest recording session age was from 11 days to 162 days. The range of oldest recording session age was from 292 days to 885 days. Thus, the overall span in age range of infants was 11 days to 2 years; 5 months.

Data collection

Recordings of infant and adult vocalizations were made using the LENA™ (Language ENvironment Analysis) system. The LENA system consists of an audio recorder that fits in the front pocket of custom-made clothing and a software system designed to automatically identify various speakers within the recordings (Xu, Yapanel, & Gray, 2009; Xu, Yapanel, Gray, & Baer, 2008). The automated system uses speech recognition technology, trained on human-annotated LENA recordings, to segment and identify onset times for specific vocalization types, taking into account the age of the infant (Richards et al., 2008; Xu, Yapanel, Gray, Gilkerson et al., 2008; Xu, Yapanel & Gray, 2009). The procedure imposes a limit such that the minimal durations of an infant or adult vocalization segment are 600 msec and 1,000 msec, respectively. Accuracy and reliability of the automated system have been tested against human transcribers for over 70 h of American English data (Xu et al., 2009). Segment classification agreement between human transcribers and the automated system was 82% for adult vocalizations and 76% for infant vocalizations. For infant vocalizations, segment classification agreement between human transcribers and the automated system was 75% for speech-related vocalizations and 84% for non-speech-related vocalizations (Xu et al., 2009). Time stamps of classified vocalization segments are reported in the LENA ITS (Interpreted Time Segments) file (Xu, Yapanel, Gray, Gilkerson et al., 2008). Infant and adult vocalization onset times were extracted from this ITS file (Warlaumont et al., 2014).

There are a few noteworthy limitations of this study due to using the LENA system. Segments labeled as having overlap between an infant or adult and any other sound source, a very common label occurring in LENA automated analysis at all ages, were excluded because the system does not indicate the types of sound sources present in those segments. Although the excluded overlapping segments often include infant and/or adult segments, there is no way of knowing based on the automated labels when this is the case. There are also a number of factors that can potentially reduce the accuracy of classification. For example, environmental noise (Soderstrom & Wittebolle, 2013), infant age, speaker variation, and clothing type (Van Dam, 2014) have been observed to influence accuracy (Xu et al., 2009). Our choice to use this system despite these limitations compared to human transcriptions is driven by the fact that the analysis of hierarchical structure of infant vocalization requires long time series in order to incorporate large temporal windows of analysis. The study described here would be impractical to conduct without automatic labeling of event onsets. Many of the same limitations also apply to a number of studies that have also used the LENA system to study language development (e.g., Ambrose, VanDam, & Moeller, 2014; Caskey & Vohr, 2013; Greenwood et al., 2010; Johnson, Caskey, Rand, Tucker, & Vohr, 2014; Oller et al., 2010; VanDam et al., 2015; Warlaumont et al., 2010, 2014; Warren et al., 2010); future studies and technological advances will be necessary to overcome these limitations.

The recorder captured each infant's voice as well as other sounds in the environment including adult vocalizations. In this study, we utilized the automated speaker labeling provided by the software. Only timings of the onsets of each infant's vocalizations and of vocalizations produced by adults in the infant's proximal auditory environment were considered. We treated all recorded adult vocalizations, regardless of which particular individual produced them, as part of the same auditory stream, so when we refer to infant–adult interactions, we are referring to the infant and all adults in the infant's auditory environment. Thus, our analyses do not distinguish between dyadic or triadic interactions where multiple adults are speaking. For the infant, vocalizations included speech-related sounds (e.g., babbling, singing, and gooing), reflexive sounds (e.g., cries and laughs), and vegetative sounds (e.g., burps and grunts). The vocalization onset times were obtained through a program that searched for onset times of CHN (i.e., Child) segments and AN (i.e., Adult) segments within the LENA

ITS file. The program is available at https://github.com/HomeBankCode/lena-its-tools/releases/tag/v1.0 (Warlaumont, 2015).

Caregivers were instructed to begin recording when their infant awoke in the morning and to stop recording when their infant was put to bed at night. Audio recordings could be paused by the parents for privacy reasons throughout the recording sessions. If the caregiver paused and resumed recording in the same day, we treated each segment as a unique session.

A total of 1,322 recordings sessions were collected across all infant–adult interactions. Recording sessions were omitted if the duration was <6 h (505 session; 37.9% of original sample excluded), if the analysis of hierarchical structure could not converge due to low number of onsets (<200 onsets; 105 sessions; 7.9%), or if the estimate of hierarchical structure or volubility was 3.5 *SD*s above or below the mean (16 sessions; 1.2%). This left 706 sessions (approximately 8,492 recording hours) to be used in all the analyses reported below. Average session length was 12.03 h (*SD* = 2.72 h). Sessions omitted due to the 6 h criterion typically reflected the caregiver stopping the recorder at some point in the day and resuming recording at a later point.

For each session, four time series of onset times were created, one for adult vocalizations, and three for the infant: speech-related (speech, nonword babble, and singing), non-speech-related (laughing, crying, burping, coughing, etc.), and the combination of speech-related and non-speech-related. These onset times served as the temporal events used to measure coincidence-based coordination and cluster-based coordination.

Analyses

Coincidence-based coordination

To quantify the coincidence of infant and adult onset events, we used cross-recurrence quantification analysis (CRQA) to obtain a diagonal cross-recurrence profile (DCRP; Coco & Dale, 2014; Dale et al., 2011; Warlaumont et al., 2014). A DCRP uses a 10-sec sliding window to assess overall quantity of coincidence-based coordination at a range of lags. Formal mathematical descriptions of CRQA and DCRP have been documented elsewhere (Coco & Dale, 2014; Dale et al., 2011; Fusaroli et al., 2014; Marwan et al., 2007); therefore, we limit our description to how the analysis relates to quantifying coordination between infant and adult vocalizations.

To obtain DCRPs, vocalization time series were divided into 1 sec bins. Each segment of either infant or adult vocalization was treated as occupying one time bin. This ensured that the interactivity estimated by the DCRPs was not affected by the durations of the segments, but only the timing between infant and adult vocalizations (Warlaumont et al., 2014). Each DCRP returned the rate of co-occurrence of events across the two vocalization time series at 1 sec lags ±10 sec. Note that because overlapping segments between infant and adult vocalizations were excluded from all analyses, there are no lag-0 recurrences reported here. DCRP height was computed by finding the total area under the DCRP profile between lag -10 sec and lag 10 sec. DCRP height measures the quantity of the infant–adult vocal interaction across a 10-sec sliding window. We estimated DCRP height for all three types of vocal interactions (infant speech-related and adult, infant non-speech-related and adult, and infant-combined and adult) for each session.

Rate-based coordination

Vocalization rate was measured in terms of volubility, which was computed as the total amount of vocalization time in each recording sessions, divided by the duration of the recording session. Infant volubility measures were computed separately for speech-related vocalizations, non-speech-related vocalizations and a combination of both types of infant vocalizations. Adult vocalizations were not broken down by type. Volubility matching was measured in terms of the correlations between infant and adult volubilities across sessions and infants.

Cluster-based coordination

The hierarchically nested clustering of vocal onset events was estimated using Allan factor analysis (Allan, 1966). Each time series of acoustic onsets was segmented into M adjacent and nonoverlapping windows of size T; then, the number of events N_j was counted within each window indexed by $j = 1$ to M. The differences in counts between adjacent windows of a given size T were computed as $d_j(T) = N_{j+1}(T) - N_j(T)$. The AF variance $A(T)$ for a given timescale, T, is the mean value of the squared differences, normalized by two times the mean count of events per window (i.e., similar to coefficient of variation, but being constituted from differences between adjacent windows, whereas the typical coefficient of variation ignores temporal relations among elements),

$$A(T) = \frac{\left\langle d(T)^2 \right\rangle}{2\langle N(T)\rangle}.$$

Poisson processes (i.e., random, independent events with exponentially distributed interevent intervals) yield $A(T) \sim 1$ for all T. In contrast, power law clustering yields $A(T) > 1$, specifically with $A(T) \sim (T/T_1)^{\alpha}$, where T_1 is the smallest timescale considered, α the exponent of the scaling relation (Thurner et al., 1997), and $\alpha > 0$. We use the term 'clustering' to refer to the nonequidistributed property observed in vocal onset events. This is a power law with positive exponent α where α provides a metric for the degree to which vocalization events are clustered across timescales. α corresponds to the slope of the plots in panel D of Figure 1, which plots Allan factor, $A(T)$, versus timescale, T, on a log–log plot. The further α is from 0 and the closer it is to 1, the more structured we say the clustering of vocalizations is across scales. AF slope does not necessarily reflect the degree of mature linguistic hierarchical structure although it does reflect the degree of hierarchical structure in the clustering of temporal events.

Ten timescales were used for all event time series. The time bins used were roughly the same across all recordings; there were small differences due to the dependency of the time binning algorithm on the total recording length. The average smallest timescale was ~10 sec, and the average largest timescale was ~88 min. Cluster-based coordination was measured by computing correlations between AF slopes for infant versus adult vocalizations. These correlations measure the extent to which the hierarchical clustering of infant vocalization bouts is similar to that of the adults in their environment across time.

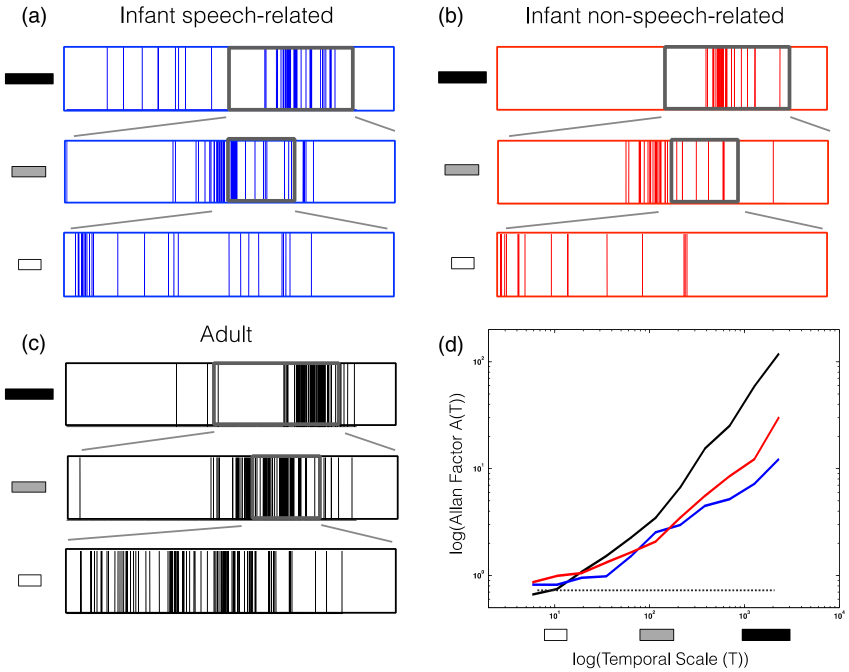

Figure 1 Schematic depiction of procedure of AF analysis at three timescales (~7, ~30, ~60 min). (a–c) Vocalization events are counted within each timescale window. Each vertical line is an acoustic onset for one of the three vocalization types: (a) infant speech-related, (b) infant non-speech-related, and (c) adult. The black, gray, and white rectangles indicate long (~60 min), medium (~30 min), and short timescales (~7 min), respectively. Notice at each of the three timescales, there are clusters of onsets. AF variance is derived from computing the normalized squared difference of onset frequencies between adjacent time windows for the three timescales. AF variance is a measure of the departure from an equidistributed distribution of acoustic onsets. (d) The estimates of hierarchical clustering of vocalization types. The slope, α, of the log(AF) versus log(T) curve estimates the scaling of AF variance across scales. The dotted line indicates a slope of 0, which is evidence for a random (Poisson process) vocalization event series. The other three curves have slopes closer to 1, indicating hierarchical clustering.

RESULTS

Volubility and hierarchical clustering across vocalization types

First, we tested for differences in overall volubility across vocalization types. A one-way ANOVA with volubility as the dependent variable, vocalization type as the predictor variable, and infant as random intercept indicated that volubility differed as a function of vocalization type, $F(3,2806) = 456.02$, $p < .001$. A post hoc Tukey test revealed that volubility for adult vocalizations ($M = .06$, $SE = .002$) was significantly higher than that for infant-combined vocalizations, ($M = .05$, $SE = .0008$), which was significantly higher than infant non-speech-related ($M = .03$, $SE = .0006$), which was significantly higher than infant speech-related ($M = .02$, $SE = .0004$), $ps < .001$.

Second, we tested for differences in hierarchical clustering across vocalization types. A(T) values and timescales were averaged across recordings, and then, A(T) was plotted as a function of T in Figure 2a. See Figure 2b for a scatterplot of each individual recording's values.

The linear trends in Figure 2 suggest that both infant and adult AF functions follow a power law. Flattening at the smallest timescales is expected to occur due to limitations in accuracy of the event onset labeling. To test against the null hypothesis that event time series are random (Poisson distribution), we performed one-sample t-tests for AF slopes against a mean of 0. AF functions for all vocalization types were reliably greater than 0, $ts(705) > 147$, ps < .001. Thus, the positive linear trends in AF functions provide evidence that the onsets for all vocalization types were clustered across multiple timescales.

A one-way ANOVA with AF slope as the dependent variable, vocalization type as the predictor variable, and infant as random intercept indicated that the hierarchical clustering differed as a function of vocalization type, $F(3,2806) = 413.17$, $p < .001$. A post hoc Tukey test showed that AF slopes for the adult vocalizations ($M = .76$, $SE = .004$) were significantly steeper than for the infant-combined vocalizations ($M = .71$, $SE = .004$), which were in turn significantly steeper than infant non-speech-related ($M = .62$, $SE = .004$), which were significantly steeper than infant speech-related ($M = .59$, $SE = .004$), ps < .001. Shallower slopes indicate relatively less nesting of clusters in vocal onset events. (See Appendix for an additional power law analysis).

Finding the same pattern of effects on volubility and AF measures suggests that they may covary. Indeed, correlation analyses showed weak linear relationships between the two measures for infant speech-related ($r[704] = .21$, $p < .001$) and infant-combined ($r[704] = .19$, $p < .001$) vocalizations and moderate relationships for adult ($r[704] = .44$, $p < .001$) and infant non-speech-related ($r[704] = .41$, $p < .001$) vocalizations. Volubility and AF measures appear to reflect one or more common sources of variation, but also exhibit unique effects, as the following analyses show.

To determine whether there was change in volubility and hierarchical clustering over the first year of the infants' lives, we regressed AF slope and volubility on infant age, performing separate analyses for the three types of infant vocalizations and the adult vocalizations. To determine unique effects on each dependent measure, all subsequent analyses were conducted by first computing the correlation between volubility and hierarchical clustering estimates, then obtaining the residual values of either volubility or hierarchical clustering after factoring out their correlation. We then tested for a relationship between the residual values and other variables of interest. For example, if we were interested in the relationship between hierarchical clustering of infant-combined vocalizations and age of infant, we would first compute the residual values of hierarchical clustering after factoring out the (linear) relationship between hierarchical clustering and volubility of infant-combined vocalizations. We then tested whether the residual (unique variance of hierarchical clustering) correlated with age of infant using a first-order correlation, r_{residual}.[1] To control for infant-level variance, we computed the residuals using linear mixed-effects models with infants as random intercepts (Baayen, Davidson, & Bates, 2008). We also present the results of correlation analyses without other variables factored out, to show whether the directions of any effects changed as a result of residualization.

---

[1] A recent paper (Wurm & Fisicaro, 2014) has shown that residualization in regression analyses can be problematic for subsequent interpretations of model coefficients. However, the comparison of two residual variables is less understood. To be sure, we constructed linear mixed-effects models with the original variables and found comparable effects to the results reported in this study.

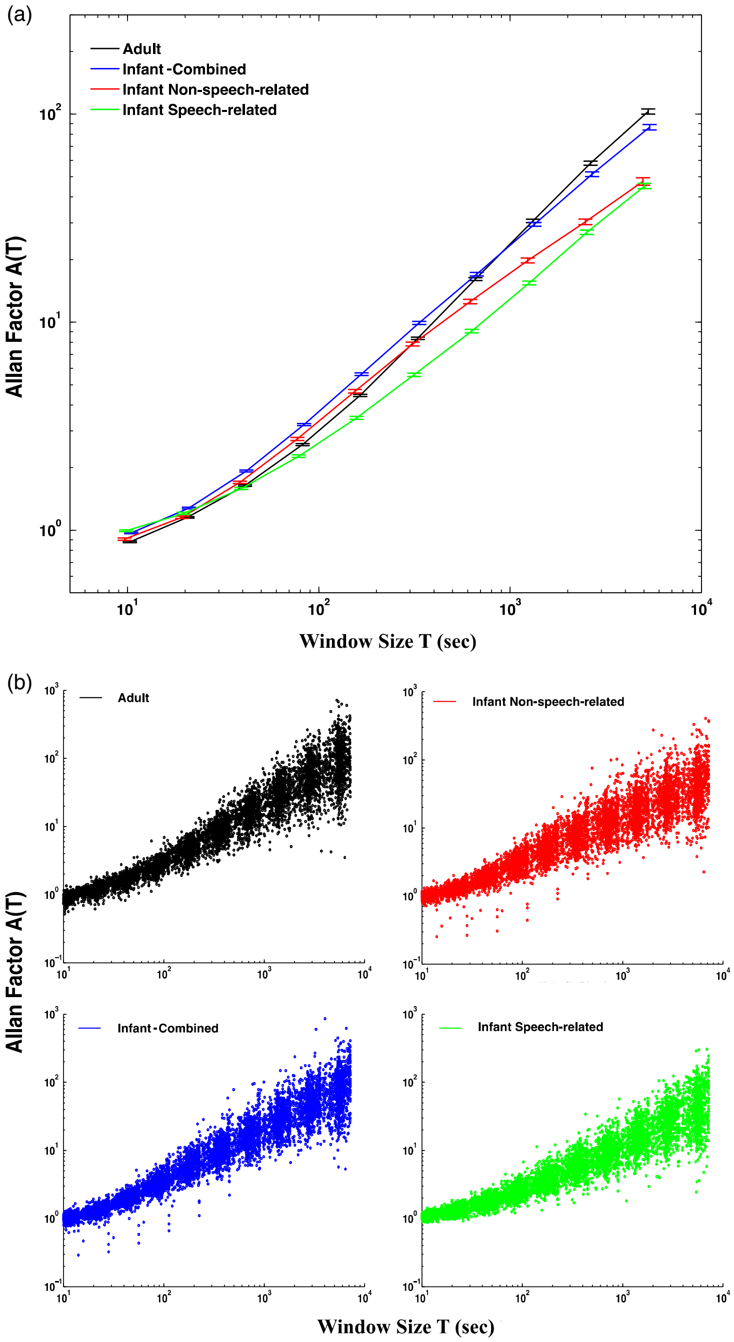

Figure 2 (a) Mean AF functions for adult and infant vocalizations, with standard error bars. (b) Scatterplot of each recording's $A(T)$ values.

Although we present both the first-order correlations (r) and the correlation coefficients from the residuals analyses (r_{residual}), we interpret all results using the magnitude, direction, and statistical significance of the r_{residual} values.

Table 1 shows how AF and volubility vary as a function of age. Volubility increased with infant age for infant speech-related vocalizations and decreased with infant age for infant non-speech-related vocalizations and adult vocalizations. No change in volubility was observed for infant vocalizations when both speech-related and non-speech-related vocalizations were included. AF slopes decreased for infant vocalizations overall as well as for non-speech-related vocalizations, but did not change with age for infant speech-related vocalizations. AF slopes also decreased with age for adult vocalizations. We discuss the implications of this below when presenting results on the relation between infant and adult AF slopes.

Do coincidence-based, rate-based, and cluster-based coordination patterns vary depending on the type of vocalization produced by the infant?

The primary goal of this study was to investigate the different vocal coordination patterns of infant and adult vocalizations. For each of the three coordination patterns, we (1) assessed whether the coordination pattern existed beyond baseline controls, (2) whether such coordination patterns still held after controlling for the other coordination patterns, and (3) whether the degree of the coordination differed as a function of the vocalization type produced by the infant.

To measure coincidence-based vocal coordination between infants and adults, we used DCRP height, derived from CRQA. Higher DCRP heights suggest more coincidence-based vocal coordination. The first step was to set up a baseline measure to compare against empirical pairings of infant and adult vocalization series. Our baseline measure consisted of shuffling the empirical infant and adult time series then submitting them to CRQA to get baseline DCRP height. We chose this baseline measure because it preserves the number of vocalizations and it keeps the shuffled time series the same length as the original time series. We obtained the DCRP height and baseline DCRP height for all three vocalization types. A one-way ANOVA with infant as random intercept indicated that DCRP height for the original time series was on average

TABLE 1
Results of First-Order Correlations, Before and After Residualization, between Vocalization Properties and Infant Age

	ρ	$\rho_{residual}$
All infant vocalizations		
Volubility	.002	.01
AF	−.10**	−.12**
Speech-related		
Volubility	.20***	.26***
AF	.05**	−.03
Non-speech-related		
Volubility	−.15***	−.16***
AF	−.24***	−.25***
Adult		
Volubility	−.24***	−.19***
AF	−.28***	−.18***

Note. $p < .1$, $^{*}p \leq .05$, $^{**}p \leq .01$, $^{***}p \leq .001$. For all analyses, degrees of freedom = 704. AF = Allan factor slope.

higher than shuffled DCRP height across all vocal interaction types, *F* (1,4220) = 221.68, $p < .001$. Because shuffled DCRP height differed as a function of vocal interaction type, we normalized the original DCRP height by subtracting the corresponding shuffled DCRP height from the original DCRP height for each vocal interaction type. A one-way ANOVA with normalized DCRP height as the dependent variable, vocal interaction type as the predictor variable, and infant as random intercept indicated that the degree of coincidence-based coordination differed as a function of vocal interaction type, *F*(2,2101) = 74.81, $p < .001$. A post hoc Tukey test showed that normalized DCRP heights for the infant-combined and adult vocalizations ($M = .001$, $SE = .00009$) were significantly taller than those for the infant speech-related and adult vocalizations ($M = .0009$, $SE = .00005$), which were in turn significantly taller than infant non-speech-related and adult vocalizations ($M = .0002$, $SE = .00004$), *p*s < .001. The same patterns of differences were found when using non-normalized DRCP heights. These results suggest that there was coincidence-based coordination above and beyond a random baseline. Furthermore, coincidence-based coordination was stronger for speech-related relative to non-speech-related interactions. See Figure 3 for DCRPs for the three vocalization types.

To determine the degree of rate-based coordination and cluster-based coordination between infant and adult vocalizations, we correlated volubility and AF slopes measured for adult vocalizations with those for each of the three corresponding infant vocalization types. Correlations were computed between raw infant and adult measures as well as between residuals of the infant and adult measures after taking out any

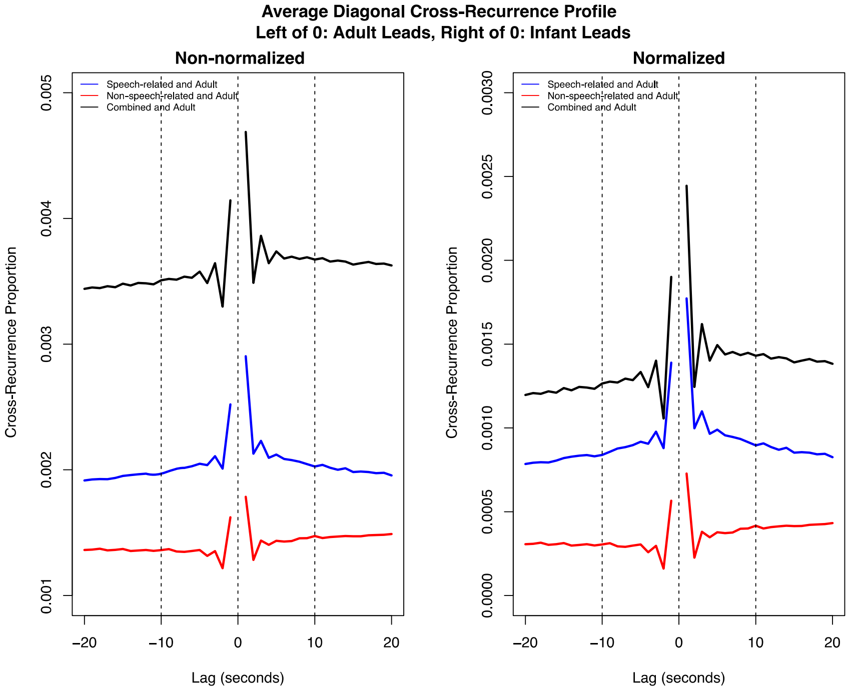

Figure 3 Diagonal cross-recurrence profile (DCRP) averaged across all vocalization types. (Left) Average DCRPs before normalization. (Right) Average DCRPs normalized for shuffled DCRPs.

correlation with age of infant and AF slope, volubility, or DCRP height (whichever two were not the focus of a given comparison). For example, to assess cluster-based coordination, infant and adult AF slopes were each residualized against speech-related volubility of the same speaker type, speech-related DCRP height, and age of infant. As before, to control for infant-level variance, we computed the residuals using linear mixed-effects models with infant as random intercepts. See Table 2 and Figure 4 for results.

For rate-based coordination, infant-combined and infant speech-related vocalization types reliably matched the volubility pattern of adult vocalizations. Using the Fisher *r*-to-*z* transformation to test for differences between correlation strength, infant speech-related volubility matching was marginally stronger than matching between infant non-speech-related vocalization, $z = 1.75$, $p = .08$. For cluster-based coordination, infant-combined and infant speech-related vocalization types reliably matched the structure found in adult vocalizations. Cluster-based vocal coordination between adult vocalizations and infant speech-related vocalizations was significantly stronger than matching between adult vocalizations and infant non-speech-related vocalizations, $z = 4.25$, $p < .001$.

Are adults or infants primarily driving these vocal coordination patterns, and does this change with age?

In the previous section, we observed that different measures of vocal coordination were not statistically reducible to each other. Thus, these measures appear to provide unique information about the relationships between infant and adult vocalization properties. In this section, we explore the question of what information the different vocalization measures provides about whether it is infants or adults who are the primary drivers of vocal coordination during the first 2 years of life.

For coincidence-based coordination, we can measure leader–follower patterns in vocalizations. We computed a leader–follower ratio from the original DCRP for each coincidence-based coordination by taking the ratio of the sum of the right side (infant leading side) to the sum of the left side (adult leading side) of the ±10 sec DCRP profile (Warlaumont et al., 2014). A leader–follower ratio greater than 1.0 indicates that the infant led the adult whereas a ratio <1.0 indicates the adult led the infant.

TABLE 2
Results of First-Order Correlations, Before and After Residualization, between Infant and Adult Vocalization Properties

	ρ	$\rho_{residual}$
Rate-based vocal coordination		
All infant vocalizations	.26***	.10***
Speech-related	.21***	.13***
Non-speech-related	.23***	.06
Cluster-based vocal coordination		
All infant vocalizations	.15***	.20***
Speech-related	.14***	.25***
Non-speech-related	.14***	.04

Note. $p < .1$, $*p \leq .05$, $**p \leq .01$, $***p \leq .001$. For all analyses, degrees of freedom = 704.

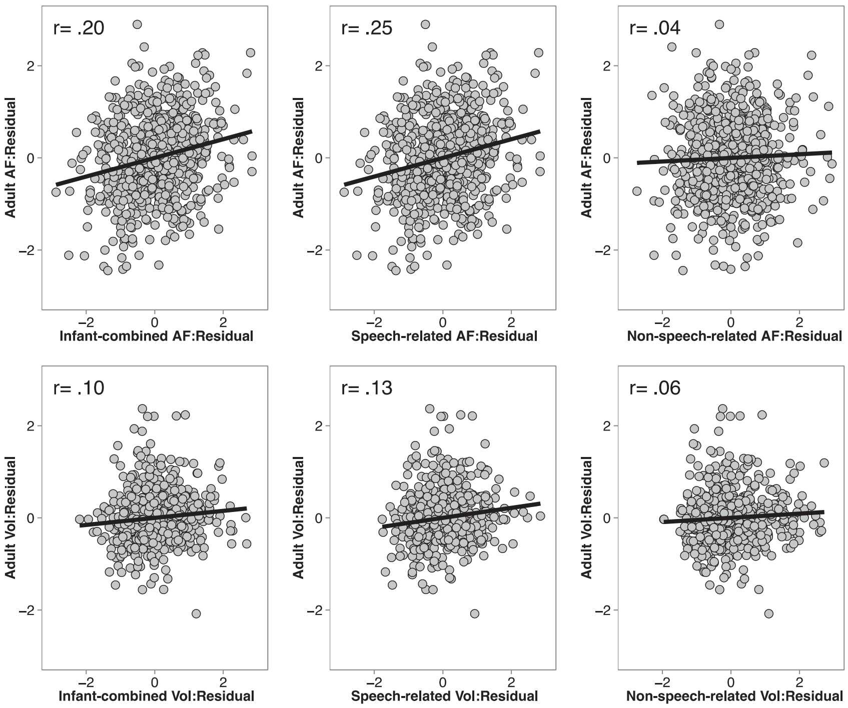

Figure 4 (Top row) Cluster-based vocal coordination results for adult and (left to right) infant-combined, infant speech-related, and infant non-speech-related. (Bottom row) Rate-based vocal coordination results. All variables are standardized. Each circle represents an individual recording.

A one-way ANOVA with leader–follower ratio as the dependent variable, vocal interaction type as the predictor variable, and infant as random intercept indicated that infant leading differed as a function of vocalization type, $F(2,2101) = 14.85$, $p < .001$. A post hoc Tukey test showed that leader–follower ratios for the infant-combined and adult vocalizations ($M = 1.049$, $SE = .002$) were higher than the ratios for infant non-speech-related and adult vocalizations ($M = 1.041$, $SE = .002$, $p = .006$), and infant speech-related and adult vocalizations ($M = 1.035$ $SE = .002$, $p < .001$). Leader–follower ratios for infant non-speech-related and adult vocalizations were higher relative to ratios for infant speech-related and adult vocalizations, $p = .048$.

To determine whether leader–follower ratios changed across infant age, we tested for correlations between ratios for each vocalization type and infant age. We observed no reliable association between infant speech-related ($r[704] = -.05$, $p = .19$) or infant-combined ($r[704] = -.05$, $p = .19$) leader–follower ratios and age. We did observe a reliable negative association between infant non-speech-related leader–follower ratios and age ($r[704] = -.08$, $p = .04$), suggesting that as infants grew older, there was a decrease in the tendency for infant non-speech-related vocalizations to precede adult vocalizations rather than vice versa.

For volubility and hierarchical clustering, we computed absolute similarity scores and then tested for correlations between the difference scores and infant age. For the similarity score (SS), we computed an absolute similarity score by subtracting infant vocalization property (AF or volubility) from the adult vocalization property, taking the absolute value, and subtracting the value from 1, for example,

$$SS_{AF} = 1 - ABS(\text{Adult AF slope} - \text{Infant AF slope}).$$

A similarity score of 1.0 suggests the vocalization properties across infant and adult were identical. A positive correlation between SS and age indicates greater matching between infant and adult on that property as age increased. Figure 5 provides a graphical depiction of these results.

Adults and infants showed statistically significant increases in coincidence-based vocal coordination for all infant vocalization types (speech-related: $r[704] = .27$, $p < .001$; non-speech-related: $r[704] = .21$, $p < .001$, all: $r[704] = .21$, $p < .001$) as well as in cluster-based vocal coordination for infant speech-related vocalizations ($r[704] = .18$, $p < .001$) but not in cluster-based vocal coordination for infant all vocalizations ($r([704] = .04$, $p = .25$), or infant non-speech-related vocalizations, $r(704) = .05$, $p = .18$.

Using the Fisher r-to-z transformation to test for differences between correlation strength, we observed stronger convergence for speech-related vocalizations relative to non-speech-related vocalizations for both volubility ($z = 9.04$, $p < .001$) and hierarchical clustering, $z = 7.31$ $p < .001$.

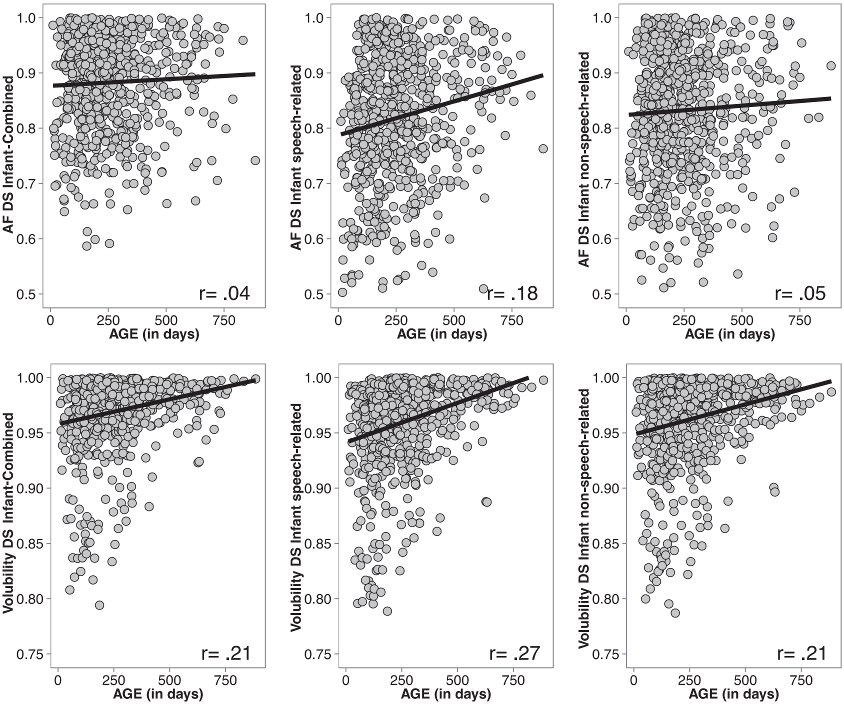

Figure 5 (Top row) Similarity score (SS) results for (left to right) infant-combined hierarchical clustering estimates, speech-related hierarchical clustering estimates, and non-speech-related hierarchical clustering estimates as a function of infant age. (Bottom row) SS results for (left to right) infant-combined volubility, speech-related volubility, and non-speech-related volubility as a function of infant age. *Note*. AF and volubility SS axes have different ranges.

For infant speech-related hierarchical clustering, combining the observation that infants and adults converge with age with the result that infant hierarchical clustering does not change with age and the result that adult hierarchical clustering decreases with age, we can infer that the adult vocalization environment is adapting its hierarchical clustering to that of the infant over the course of the first 2 years of life. Because infant speech-related volubility increases whereas adult volubility decreases over infant age, the results from the difference score analyses suggest bidirectional convergence: Both infants and adults adjust volubility rates toward each other over infant age.

Do the different coordination measures have unique developmental trends?

In the previous sections, we established that the three vocal coordination patterns are not reducible to each other and provide different perspectives on the interpersonal dynamics of infant–adult vocal coordination. In this final section, we investigate whether the various coordination patterns are independently associated with infant age.

In addition to the three vocal coordination patterns that have been the foci of this study, for this section, we also included a conversational turn-taking measure computed by the LENA system. The conversational turn-taking measure computed by LENA is frequently used in the literature (Caskey, Stephens, Tucker, & Vohr, 2011; Gilkerson & Richards, 2009; Gilkerson, Richards, & Topping, 2015; Greenwood et al., 2010; Suskind et al., 2015; Warren et al., 2010) and is therefore an important measure to include when assessing independent associations with infant age. A conversational turn is identified when a sequence of speech-related sound segments from an adult then an infant, or vice versa, occurs within 5 sec without an intervening non-speech-related segment or speech-related segment from another adult or infant. Conversational turn count can be considered a measure of infant–adult interaction (Warren et al., 2010). Because recording sessions in our sample greatly varied in length, we computed *turn-taking rate* by dividing conversational turn count by the length of the recording session.

Because the turn-taking rate is computed using only speech-related segments, we limited our analyses in this section to speech-related coordination patterns. Table 3 reports first-order correlations and also correlations with residualized variables. Coincidence-based coordination ($r_{\text{residual}} = .07$, $p = .05$), rate-based coordination ($r_{\text{residual}} = .31$, $p < .001$), and conversational turn-taking rate ($r_{\text{residual}} = .15$, $p < .005$) were all independently positively associated with infant age. Cluster-based coordination was not independently associated with infant age, $r_{\text{residual}} = .02$, $p = .61$.

DISCUSSION

This study examined coordination patterns that arise from different measures of infant and adult vocalizations. We aimed to answer three specific questions: (1) Do coincidence-based, rate-based, and cluster-based coordination patterns vary depending on the vocalization type produced by the infant? (2) Do different coordination patterns provide unique information about the interpersonal dynamics of vocal interaction? (3) How do the various coordination patterns relate to infant age?

TABLE 3
Results of First-Order Correlations, Before and After Residualization, between Each Coordination Type and Infant Age

Coordination pattern	ρ	$\rho_{residual}$
Coincidence-based	<.001	.07*
Rate-based	.27***	.31***
Cluster-based	.18**	.02
Turn-taking rate	−.03	.15***

Note. $p < .1$, $^{*}p \le .05$, $^{**}p \le .01$, $^{***}p \le .001$. For all analyses, degrees of freedom = 704. Rate-based coordination and cluster-based coordination reflect similarity scores.

We observed that all three coordination patterns displayed higher rates of coordination for infant speech-related vocalizations relative to infant non-speech-related vocalizations. These results point to a difference in coordination as a function of speech relatedness and could perhaps be due to speech-related vocalization holding more social value to caregivers. Properties derived from the coordination patterns provided new insights into unidirectional and bidirectional adaptation between infants and their caregivers. Finally, we observed unique trajectories between the coordination patterns and infant age.

Hierarchical vocalization patterns and volubility

To answer the three research questions provided at the outset of this paper, estimations of vocalization properties such as hierarchical clustering and volubility were required. An important finding from this study was that the onsets of infant vocalization bouts have hierarchical structure at timescales ranging from seconds to hours. This result expands upon previous work using subjective ratings to assess hierarchical structure or phrasing of infant vocal productions at shorter timescales (Lynch et al., 1995) and also converge with evidence of hierarchical structure in speech based on other algorithms evaluating other vocalization patterns (Abney, Warlaumont, Haussman, Ross, & Wallot, 2014; Abney, Paxton et al., 2014; Luque et al., 2015). Lynch et al. identified the hierarchical organization of syllables, utterances, and prelinguistic phrases and identified hierarchical structure spanning the typical duration of syllables (<500 msec) to less than several seconds in duration. Because of the temporal resolution of the automated vocalization segmentation used in our study, the shortest timescale included in our estimate of hierarchical clustering was approximately 10 sec. The hierarchical structure we identified spanned from ~10 sec to ~1.5 h. Therefore, the hierarchical clustering observed in the present study is at the level of bouts of vocalization and does not reflect the structure within utterances. Future work is required to better understand the hierarchical structure of infant vocalizations at shorter timescales, for example, spanning milliseconds to seconds. These results also suggest that infant prelinguistic vocalizations are not equidistributed and are power-law-distributed. Subsequent investigation demonstrated that the interevent intervals of the vocalization events were power-law-distributed with a slope approximating −2 (see Appendix). Our results therefore provide evidence for fractal properties of prelinguistic vocalizations.

Evidence for hierarchical clustering of vocalizations was found at even the youngest session, recorded from an infant who was aged 11 days. Although estimates of hierarchical clustering for infant speech-related vocalizations were not observed to change with age, we observed a reliable decrease in hierarchical clustering (more random) for infant non-speech-related vocalizations. The results presented here suggest that infant vocalization bouts exhibit nonrandom temporal patterning from shortly after birth and that, for speech-related vocalizations, this hierarchical nature of vocalization bouts is fairly stable across the period of prelinguistic and early linguistic development.

We also investigated patterns of infant volubility. Previous work has suggested that by about 3–5 months of age, infants learn that vocalizations have social value, with more communicative types of vocalizations influencing parental engagement (Goldstein et al., 2009). Previous work has also found that adult responsiveness to infant vocalizations increases during the second year (Bloom, Margulis, Tinker, & Fujita, 1996). In the present study, volubility for infant speech-like vocalizations increased with infant age, replicating prior findings that also used the LENA system (Greenwood et al., 2011) and strengthening the idea that, over time, infants learn that vocalizations hold social value and serve a communicative function. We also found that volubility for infant non-speech-related vocalizations decreased with infant age (similar to Warlaumont et al., 2014).

It is important to point out a few possible limitations to the observed results. It is always possible that the increases in volubility are influenced by decreasing sleeping time relative to neonates. Although this is a possibility, naps are a component of an infant's daily routine and among the many factors of the complex interaction between infant vocalization bouts and adult vocalization bouts. Additionally, the ability of the LENA system to discern infant vocalizations may improve with age. Therefore, it is possible that changes in volubility across age are at least partially due to differences in the ability of the LENA system to discern between infant vocalizations across age. Future work combining automatic and manual coding procedures is important to establish the reliability of increased volubility across age.

We found that changes in hierarchical clustering and in volubility across age held even when other variables were factored out through residualization. These results, combined with the different developmental patterns observed for volubility versus hierarchical clustering, suggest that volubility and hierarchical clustering provide at least partially independent information about infant prespeech and early speech development. The estimation of hierarchical clustering of vocalizations may provide additional measures that can help predict later infant behaviors and abilities. For example, the hierarchical clustering of infant behavior may reflect the daily routines of a family and/or daycare environment, and the predictability of these routines may be reflected in the consistency of AF slopes across recordings. Future work is required to test whether or not hierarchical clustering is a vocalization property with predictive value for important developmental outcomes.

Vocal coordination patterns vary by vocalization type and provide unique information based on level of description

We introduced a typology of coordination patterns that spans across levels of description and timescale: coincidence-based coordination, rate-based coordination, and cluster-based vocal coordination. Using CRQA, we observed that coincidence-based coordination was greater than a random baseline based on shuffled time series. One potential issue with the data collection technique used in this study is that we are not directly aware of specific bouts of interaction relative to incidental vocalizations made by infants and adults in the infants' auditory environments. Showing that empirical DRCP heights were greater than surrogate-based DRCP heights provides evidence for the nonincidental, vocal interaction of infants and adults in close proximity to the infant.

Across the different vocal coordination patterns, we found that coordination patterns based on infant speech-related vocalizations were stronger and more frequent relative to coordination patterns based on infant non-speech-related vocalizations. These results point to the sensitivity of the coordination patterns based on child vocalization type.

Different coordination patterns provide unique information about the interpersonal dynamics of vocal interaction

For coincidence-based vocal coordination, we computed leader–follower dynamics across vocalization type and across temporal lag. We found that within a 10 sec window, infant vocalizations precede adult vocalizations and more so for non-speech-related vocalizations. Rate-based and cluster-based vocal coordination patterns offered a different perspective on leading and following in vocal dynamics. Focusing on rate-based patterns, we found bidirectional convergence of volubility across infant age: infants and adults both adjusted volubility rates toward each other across age. Focusing on cluster-based vocal coordination, we found that adults adapted the hierarchical clustering of their vocalizations to that of their infants' vocalizations as infant age increased.

Also studying daylong home audio recordings, Ko, Seidl, Cristia, Reimchen, and Soderstrom (2015) investigated the relationship between acoustic properties of mother and infant/toddler vocalizations. Ko et al. observed that mothers and infants/toddlers converged across various vocalization properties such as pitch. Specifically, mothers adapted their speech to the infant/toddler more than vice versa. The results of the current paper extend what Ko et al. observed by pointing to another vocalization property, hierarchical clustering, that shows similar convergence patterns. Notably, there was adult-to-infant convergence of both hierarchical clustering and volubility. Our results diverge from Ko et al. in the timescales of convergence: Ko et al. found convergence of pitch at the local level of conversational exchange, whereas the results in this study found convergence of hierarchical clustering and volubility across the entire span of daylong recording session (e.g., ≥6 h).

The observation that adults adapted the hierarchical clustering of their vocalizations to that of their infants' vocalizations adds additional support for the fine-tuning hypothesis (Snow, 1989, 1995), suggesting that adults adapt the complexity of their child-directed language in response to properties of child-produced language. Most of the support for the fine-tuning hypothesis focused on measures of linguistic complexity (Kunert, Fernández, & Zuidema, 2011; Snow, 1995; Sokolov, 1993; Yurovsky, Doyle, & Frank, 2016). Our results support the fine-tuning hypothesis, but use a metric focused on the hierarchical organization (a hallmark of 'complex systems') of vocal clustering instead of linguistic complexity. Future work testing the fine-tuning hypothesis should consider multiple measures of ‘complexity’ spanning various levels of linguistic and vocal alignment.

Coordination patterns and infant age

Since Bateson (1975) and Stern, Jaffe, Beebe, and Bennett (1975) first proposed that an important property of interpersonal exchange and communicative function was the development of turn-taking dynamics, several studies have illuminated developmental patterns of vocal turn-taking rate and timing (Caskey et al., 2011; Harder et al., 2015; Hilbrink, Gattis, & Levinson, 2015). These studies provide important information about the timing of turn taking (e.g., Hilbrink et al., 2015) or the transition from covocalizations to turn taking across development (e.g., Harder et al., 2015; Rutter & Durkin, 1987). But turn taking is only one type of vocal coordination. Our investigation of multiple vocal coordination patterns across development adds to prior research by showing the relationships between vocal coordination patterns focusing on different levels of analysis and infant age. We found that different coordination patterns had different associations with infant age. Rate-based vocal coordination had the strongest association with infant age: speech-related rate-based vocal coordination increased with infant age. Turn-taking rate and coincidence-based coordination both increased with infant age as well. When controlling for all other coordination patterns, cluster-based coordination was not associated with infant age. Although cluster-based speech-related vocal coordination did not change significantly with increasing infant age once other coordination patterns were controlled for, cluster-based coordination may nevertheless reflect an aspect of coordination between infant and caregivers that have developmental significance, for example, by facilitating information transfer between infant and caregiver across the first year.

We found that infants’ vocal timing became more similar to their caregivers’ vocal timing across the first 2 years of life. In other words, within a 10-sec temporal window, infant and caregiver vocalizations occurred more frequently across infant age. This finding, in conjunction with the results of increased turn-taking rate and increased rate-based matching across age, suggests a dynamic trajectory of vocal development. Throughout the first few years of life, infant and caregiver vocalizations become more temporally coordinated (coincidence-based vocal coordination), vocalize at similar rates across the day (rate-based vocal coordination) and increase the rate of structured turn-taking patterns (turn-taking rate).

Future directions

An important potential application of infant–adult vocal coordination patterns is to the study of language development and atypical development. Jaffe et al.’s (2001) contribution is an example of the utility of using coordination patterns to predict developmental outcomes. Future work should incorporate a pluralistic approach to coordination patterns to determine the predictive value of different coordination patterns for important developmental outcomes. To that end, it is important to understand what information different coordination patterns provide.

Coincidence-based vocal coordination provides information about the similarities and differences in vocal timing. Rate-based vocal coordination provides information about the similarities and differences in overall volubility rates across a recording session. Cluster-based vocal coordination provides information about the similarities and differences in the production of hierarchical clustering across a recording session.

Although all three coordination patterns provide important information about vocal interaction, cluster-based vocal coordination is motivated by a theory in statistical mechanics investigating the outcomes of interacting complex networks. More specifically, work in statistical mechanics has shown that when two complex systems interact, information transfer between them is enhanced and may even become optimal when their multiscale dynamics are matched (West, Geneston, & Grigolini, 2008), a term called *complexity matching*. Previous research studying adult conversations has shown that the degree of cluster-based vocal coordination or *complexity matching* differed depending on specific conservational contexts (Abney, Paxton et al., 2014). Perhaps, a function of cluster-based vocal coordination is increased communication? Indeed, the question of *function* for any coordination pattern or collection of coordination patterns should be the focus of future research.

This information transfer hypothesis requires much more empirical attention before any substantive conclusions can be made. For example, recent work on infant language development has utilized the LENA system along with various standardized measures of language and communication development (e.g., MacArthur-Bates, Communicative Development Inventory; Fenson et al., 2007) to investigate language learning in naturalistic environments (Ramírez-Esparza et al., 2014; Weisleder & Fernald, 2013; Walle & Warlaumont, 2015). Future studies should investigate the role of the production and convergence of specific vocalization properties like volubility and hierarchical clustering on vocabulary or other aspects of language development (see Northrup & Iverson, 2015).

CONCLUSION

Our results support the proposal that various vocal coordination patterns spanning multiple levels of description provide unique information about infant–adult vocal interactions. We found increased coincidence-based, rate-based, and cluster-based vocal coordination for infant speech-related vocalizations relative to non-speech-related vocalizations. We also found different infant–adult convergence patterns depending on the measure used. For instance, leader–follower dynamics derived from coincidence-based coordination measurements suggest that infants lead vocal exchanges, whereas adults adapt their hierarchical clustering to that of the infant over time. Finally, we found divergent associations between infant age and the various vocal coordination patterns. In particular, higher degrees of speech-related coincidence-based, rate-based, and conversational turn taking were independently associated with increased rates of turn taking. Future work should address the question of how the various coordination patterns relate to the different contexts and event types the infant experiences over the course of the day and should attempt to discover the unique functions the different coordination patterns serve (if any). Future work should focus on utilizing multiple vocal coordination patterns in combination to test whether multiple levels of description increase the predictive value for identifying important developmental milestones or diagnosing various clinical disorders.

ACKNOWLEDGMENTS

This work was supported by the National Institute on Deafness and other Communication Disorders (R01 DC011027) and the Hatano Cognitive Development Fellowship.

REFERENCES

Abney, D. H., Kello, C. T., & Warlaumont, A. S. (2015). Production and convergence of multiscale clustering in speech. *Ecological Psychology*, *27* (3), 222–235.

Abney, D. H., Paxton, A., Dale, R., & Kello, C. T. (2014). Complexity matching in dyadic conversation. *Journal of Experimental Psychology: General*, *143*(6), 2304–2315.

Abney, D. H., Warlaumont, A. S., Haussman, A., Ross, J. M., & Wallot, S. (2014). Using nonlinear methods to quantify changes in infant limb movements and vocalizations. *Frontiers in Psychology*, *5*(771), 1–15.

Allan, D. W. (1966). Statistics of atomic frequency standards. *IEEE Proceedings*, *54*, 221–230.

Allely, C. S., Purves, D., McConnachie, A., Marwick, H., Johnson, P., Doolin, O., . . .Wilson, P. (2013). Parent-infant vocalisations at 12 months predict psychopathology at 7 years. *Research in Developmental Disabilities*, *34*(3), 985–993.

Ambrose, S. E., Van Dam, M., & Moeller, M. P. (2014). Linguistic input, electronic media, and communication outcomes of toddlers with hearing loss. *Ear and Hearing*, *35*, 139–147.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Bateson, M. C. (1975). Mother-Infant exchanges: The epigenesis of conversational interaction. *Annals of the New York Academy of Sciences*, *263*(1), 101–113.

Bloom, L., Margulis, C., Tinker, E., & Fujita, N. (1996). Early conversations and word learning: Contributions from child and adult. *Child Development*, *67*(6), 3154–3175.

Bloom, K., Russell, A., & Wassenberg, K. (1987). Turn taking affects the quality of infant vocalizations. *Journal of Child Language*, *14*(02), 211.

Bowlby, J. (1969). *Attachment, vol. 1 of attachment and loss*. New York, NY: Basic Books.

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

Caskey, M., Stephens, B., Tucker, R., & Vohr, B. (2014). Adult talk in the NICU with preterm infants and developmental outcomes. *Pediatrics*, *133*(3), e578–e584.

Caskey, M., & Vohr, B. (2013). Assessing language and language environment of high-risk infants and children: A new approach. *Acta Paediatrica*, *102*, 451–461.

Coco, M. I., & Dale, R. (2014). Cross-recurrence quantification analysis of categorical and continuous time series: An R package. *Frontiers in Psychology*, *5*, 510.

Cox, R. F., & van Dijk, M. (2013). Microdevelopment in parent-child conversations: From global changes to flexibility. *Ecological Psychology*, *25*(3), 304–315.

Dale, R., Warlaumont, A. S., & Richardson, D. C. (2011). Nominal cross recurrence as a generalized lag sequential analysis for behavioral streams. *International Journal of Bifurcation and Chaos*, *21*(04), 1153–1161.

Feldman, R., & Greenbaum, C. W. (1997). Affect regulation and synchrony in mother—infant play as precursors to the development of symbolic competence. *Infant Mental Health Journal*, *18*(1), 4–23.

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*(4), 455–470.

Fenson, L., Marchman, V. A., Thal, D., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *The MacArthur-Bates communicative development inventories user's guide and technical manual* (2nd edn). Baltimore, MD: P.H. Brookes.

Franklin, B., Warlaumont, A. S., Messinger, D., Bene, E., Nathani Iyer, S., Lee, C.-C., . . . Oller, D. K. (2013). Effects of parental interaction on infant vocalization rate, variability and vocal type. *Language Learning and Development*, *10*, 279–296.

Fusaroli, R., Konvalinka, I., & Wallot, S. (2014). Analyzing social interactions: Promises and challenges of cross recurrence quantification analysis. Springer Proceedings in Mathematics & Statistics, *103*, 137–155).

Gilkerson, J., & Richards, J. A. (2009). *The LENA natural language study*. Boulder, CO: LENA Foundation.

Gilkerson, J., Richards, J. A., & Topping, K. J. (2015). The impact of book reading in the early years on parent–child language interaction. *Journal of Early Childhood Literacy*. Advance online publication. doi: 10.1177/1468798415608907.

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 of the United States of America*, *100*(13), 8030–8035.

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, *19*(5), 515–523.

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., & West, M. J. (1999). Consistent responses of human mothers to prelinguistic infants: The effect of prelinguistic repertoire size. *Journal of Comparative Psychology*, *113*(1), 52.

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

Grosjean, F., Grosjean, L., & Lane, H. (1979). The patterns of silence: Performance structures in sentence production. *Cognitive Psychology*, *11*(1), 58–81.

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*, 509–516.

Harder, S., Lange, T., Hansen, G. F., Væver, M., & Køppe, S. (2015). A longitudinal study of coordination in mother–infant vocal interaction from age 4 to 10 months. *Developmental Psychology*, *51*(12), 1778.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: P.H. Brookes.

Hart, B., & Risley, T. R. (1999). *The social world of children: Learning to talk*. Baltimore, MD: Paul H. Brookes Publishing Co.

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*, 10–3389.

Hsu, H. C., Fogel, A., & Messinger, D. S. (2001). Infant non-distress vocalization during mother-infant face-to-face interaction: Factors associated with quantitative and qualitative differences☆. *Infant Behavior and Development*, *24*(1), 107–128.

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*, i–149.

Johnson, K., Caskey, M., Rand, K., Tucker, R., & Vohr, B. (2014). Gender differences in adult-infant communication in the first months of life. *Pediatrics*, *134*(6), e1603–e1610.

Ko, E. S., Seidl, A., Cristia, A., Reimchen, M., & Soderstrom, M. (2015). Entrainment of prosody in the interaction of mothers with their young children. *Journal of Child Language*, *43*, 1–26.

Kokkinaki, T., & Kugiumutzakis, G. (2000). Basic aspects of vocal imitation in infant-parent interaction during the first 6 months. *Journal of Reproductive and Infant Psychology*, *18*(3), 173–187.

Kunert, R., Fernández, R., & Zuidema, W. (2011). Adaptation in Child Directed Speech: Evidence from Corpora. In *Proc. SemDial* (pp. 112–119).

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.

Luque, J., Luque, B., & Lacasa, L. (2015). Scaling and universality in the human voice. *Journal of the Royal Society, Interface/the Royal Society*, *12*(105), 20141344.

Lynch, M. P., Oller, D. K., Steffens, M. L., & Buder, E. H. (1995). Phrasing in prelinguistic vocalizations. *Developmental Psychobiology*, *28*(1), 3–25.

Marwan, N., Romano, M. C., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports*, *438*(5), 237–329.

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

Northrup, J. B., & Iverson, J. M. (2015). Vocal coordination during early parent-infant interactions predicts language outcome in infant siblings of children with autism spectrum disorder. *Infancy*, *20*(5), 523–547.

Oller, D. K. (2000). *The emergence of the speech capacity*. Lawrence Erlbaum and Associates, Mahwah, N.J.: Psychology Press.

Oller, D. K., Eilers, R. E., Basinger, D., Steffens, M. L., & Urbano, R. (1995). Extreme poverty and the development of precursors to the speech capacity. *First Language*, *15*(44), 167–187.

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

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., Belardi, K., Baranek, G. T., Watson, L. R., Labban, J. D., & Oller, D. K. (2014). Vocal patterns in infants with autism spectrum disorder: Canonical babbling status and vocalization frequency. *Journal of Autism and Developmental Disorders*, *44*(10), 2413–2428. doi:10.1007/s10803-014-2047-4.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *The Behavioral and Brain Sciences*, *27*, 169–190; discussion 190–226.

Ramírez-Esparza, N., García-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.

Rescorla, L., & Ratner, N. B. (1996). Phonetic profiles of toddlers with specific expressive language impairment (SLI-E). *Journal of Speech, Language, and Hearing Research*, *39*(1), 153–165.

Richards, J. A., Gilkerson, J., Paul, T., & Xu, D. (2008). The LENA automatic vocalization assessment. Technical Report LTR-08–1. Available at: http:// www.lenafoundation.org/wp-content/uploads/ 2014/10/ LTR-08-1_Automatic_Vocalizatio- n_Assessment.pdf.

Rutter, D. R., & Durkin, K. (1987). Turn-taking in mother–infant interaction: An examination of vocalizations and gaze. *Developmental Psychology*, *23*(1), 54.

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). Cambridge, England: Cambridge University Press.

Snow, C. E. (1989). Understanding social interaction and language acquisition; sentences are not enough. In M. H. Bornstein, & J. S. Bruner (Eds.), *Interaction in human development* (pp. 83–103). Hillsdale, NJ: Lawrence Erlbaum Associates.

Snow, C. E. (1995). Issues in the study of input: Fine-tuning, universality, individual and developmental differences, and necessary causes. In P. Fletcher, & B. MacWhinney (Eds.), *The handbook of child language* (pp. 180–193). Oxford, England: Blackwell.

Soderstrom, M., & Wittebolle, K. (2013). When do caregivers talk? The influences of activity and time of day on caregiver speech and child vocalizations in two childcare environments. *PLoS ONE*, *8*(11), e80646.

Sokolov, J. L. (1993). A local contingency analysis of the fine-tuning hypothesis. *Developmental Psychology*, *29*, 1008–1023.

Stern, D. N., Jaffe, J., Beebe, B., & Bennett, S. L. (1975). Vocalizing in unison and in alternation: Two modes of communication within the mother-infant dyad. *Annals of the New York Academy of Sciences*, *263*, 89–100.

Suskind, D. L., Leffel, K. R., Graf, E., Hernandez, M. W., Gunderson, E. A., Sapolich, S. G., … Levine, S. C. (2016). A parent-directed language intervention for children of low socioeconomic status: a randomized controlled pilot study. *Journal of Child Language*, *43*(02), 366–406.

Thurner, S., Lowen, S. B., Feurstein, M. C., Heneghan, C., Feichtinger, H. G., & Teich, M. C. (1997). Analysis, synthesis, and estimation of fractal-rate stochastic point processes. *Fractals*, *5*, 565–595. doi: 10.1142/ S0218348X97000462

Van Dam, M. (2014). Acoustic characteristics of the clothes used for a wearable recording device. *The Journal of the Acoustical Society of America*, *136*(4), EL263–EL267.

VanDam, M., Oller, D. K., Ambrose, S. E., Gray, S., Richards, J. A., Xu, D., … Moeller, M. P. (2015). Automated vocal analysis of children with hearing loss and their typical and atypical peers. *Ear and Hearing*, *36*(4), e146–e152.

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.

Warlaumont, A. S. (2015). HomeBankCode: lena-its-tools (Version 1.0) [Computer software]. Available from https://github.com/HomeBankCode/lena-its-tools/releases/tag/v1.0

Warlaumont, A. S., Oller, D. K., Dale, R., Richards, J. A., Gilkerson, J., & Xu, D. (2010). Vocal interaction dynamics of children with and without autism. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, pp. 121–126. Austin, TX: Cognitive Science Society.

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. doi:10.1177/0956797614531023.

Warren, S. F., Gilkerson, J., Richards, J. A., Oller, D. K., Xu, D., Yapanel, U., & Gray, S. (2010). What automated vocal analysis reveals about the vocal production and language learning environment of young children with autism. *Journal of Autism and Developmental Disorders*, *40*(5), 555–569.

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

West, B. J., Geneston, E. L., & Grigolini, P. (2008). Maximizing information exchange between complex networks. *Physics Reports*, *468*(1), 1–99.

Wurm, L. H., & Fisicaro, S. A. (2014). What residualizing predictors in regression analyses does (and what it does not do). *Journal of Memory and Language*, *72*, 37–48.

Xu, D., Yapanel, U., & Gray, S. (2009). *Reliability of the LENA Language Environment Analysis System in Young Children's Natural Home Environment*. Boulder, CO. doi:xu

Xu, D., Yapanel, U., Gray, S., & Baer, C. (2008). *The LENA Language Environment Analysis System: The interpretative time segments (ITS) file*.

Xu, D., Yapanel, U., Gray, S., Gilkerson, J., Richards, J., & Hansen, J. (2008). Signal processing for young child speech language development. In *Proceedings of the 1st Workshop on Child, Computer, and Interaction*. Chania, Crete, Greece.

Yurovsky, D., Doyle, G., & Frank, M. C. (2016). Linguistic input is tuned to children's developmental level. *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*. 2093–2098.

APPENDIX

Hierarchical or multiscale clustering as estimated using AF is a power law defining temporal clustering across timescales. West et al. (2008) suggested that complex systems exhibit complexity matching when their interevent intervals (IEIs) are power-law-distributed with an exponent near two, $P(\text{IEI}) \sim 1/\text{IEI}^{\gamma}$, where $\gamma \sim 2$. An IEI is the temporal duration between two events in an event series. Previous work observed that IEIs of adult vocalizations during conversations exhibited a power law, $\gamma \sim 2$ (Abney, Paxton, Dale, & Kello, 2014). To corroborate the AF results showing power law scaling of temporal clustering, we tested whether IEIs of each vocalization type followed a power law across timescales. A histogram of IEIs was computed for the time series from each infant and adult, across all recording sessions and vocalization types. The smallest bin of the histogram was set relative to the shortest IEI value in each time series. Eight subsequent bins were logarithmically spaced to capture IEIs of a range of lengths for each time series. Figure A1 shows the resulting histograms for all time series of each vocalization type, plotted together in a single graph. For each vocalization type, the figure shows a clear trend of a negatively sloped line in logarithmic coordinates. The slopes of the trend are about −2 for all vocalization types. Therefore, the data meet the theoretically derived precondition for complexity matching (Abney et al., 2014; West et al., 2008) and corroborate the observation of a power law for temporal clustering across multiple timescales.

The slopes of IEIs for adult vocalizations ($M = -2.09$, $SE = .02$) were steeper than slopes for infant speech-related IEIs ($M = -2.03$, $SE = .02$, $p = .01$) and infant non-speech-related IEIs ($M = -1.92$, $SE = .02$, $p < .001$), but not infant all IEIs ($M = -2.09$, $SE = .02$, $p = .67$). Overall the general pattern of IEI slopes was consistent with the patterns of AF slopes across vocalization types: Adult vocalizations have stronger hierarchical clustering and steeper IEI slopes relative to all infant vocalization types. One notable difference across the hierarchical clustering and IEI slope results is that the IEI slopes for infant speech-related vocalizations were steeper than IEI slopes for infant non-speech-related vocalizations, $p = .003$. For hierarchical clustering, we observed that AF slopes were steeper for infant non-speech-related vocalizations relative to infant speech-related vocalizations. It is important to point out that the two analyses are not identical and provide subtly different information about the timing of vocalizations: The AF slopes provide information about the clustering of onset events across timescale, and the IEI slopes provide information about the distribution of vocalization interevent intervals.

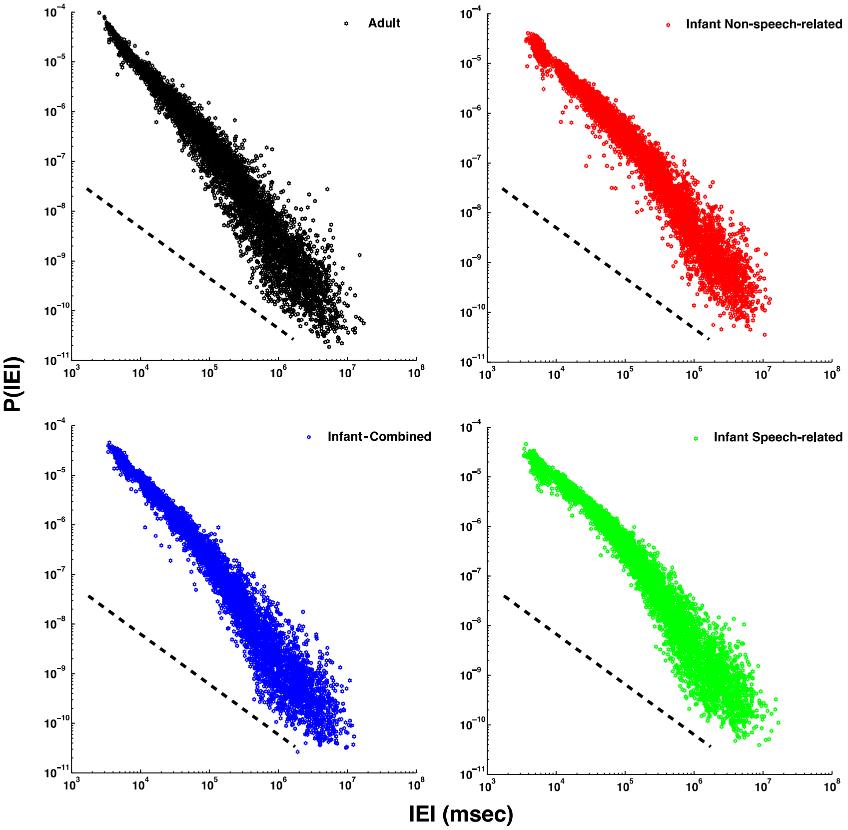

Figure A1 Interevent interval (IEI) probability density functions for all recordings for each vocalization type, plotted in logarithmic coordinates using logarithmic binning. Dashed lines show idealized slope of −2.