



Temporal Coordination in Mother–Infant Vocal Interaction: A Cross-Cultural Comparison

Lama K. Farran^{1*}, Hyunjoo Yoo², Chia-Cheng Lee³, Dale D. Bowman⁴ and D. Kimbrough Oller⁴

¹ Communication Sciences and Disorders, University of West Georgia, Carrollton, GA, United States, ² Department of Communicative Disorders, College of Arts and Sciences, University of Alabama, Tuscaloosa, AL, United States,

³ Department of Speech and Hearing Sciences, Portland State University, Portland, OR, United States, ⁴ School of Communication Sciences and Disorders, University of Memphis, Memphis, TN, United States

OPEN ACCESS

Edited by:

Maria Spinelli,
Università degli Studi “G. d’Annunzio”
Chieti – Pescara, Italy

Reviewed by:

Iris Nomikou,
University of Portsmouth,
United Kingdom
Mirco Fasolo,
Università degli Studi “G. d’Annunzio”
Chieti – Pescara, Italy

*Correspondence:

Lama K. Farran
lfarran@westga.edu

Specialty section:

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

Received: 11 May 2019

Accepted: 04 October 2019

Published: 08 November 2019

Citation:

Farran LK, Yoo H, Lee C-C,
Bowman DD and Oller DK (2019)
Temporal Coordination
in Mother–Infant Vocal Interaction:
A Cross-Cultural Comparison.
Front. Psychol. 10:2374.
doi: 10.3389/fpsyg.2019.02374

Temporal coordination of vocal exchanges between mothers and their infants emerges from a developmental process that relies on the ability of communication partners to co-coordinate and predict each other’s turns. Consequently, the partners engage in communicative niche construction that forms a foundation for language in human infancy. While robust universals in vocal turn-taking have been found, differences in the timing of maternal and infant vocalizations have also been reported across cultures. In this study, we examine the temporal structure of vocal interactions in 38 mother–infant dyads in the first two years across two cultures—American and Lebanese—by studying observed and randomized distributions of vocalizations, focusing on both gaps and overlaps in naturalistic 10-min vocal interactions. We conducted a series of simulations using Kolmogorov–Smirnov (K–S) tests to examine whether the observed responsivity patterns differed from randomly generated simulations of responsivity patterns in both Arabic and English for mothers responding to infants and for infants responding to mothers. Results revealed that both mothers and infants engaged in conversational alternation, with mothers acting similarly across cultures. By contrast, significant differences were observed in the timing of infant responses to maternal utterances, with the Lebanese infants’ tendency to cluster their responses in the first half-second after the offset of the Lebanese mothers’ utterances to a greater extent than their American counterparts. We speculate that the results may be due to potential phonotactic differences between Arabic and English and/or to differing child-rearing practices across Lebanese and American cultures. The findings may have implications for early identification of developmental disorders such as autism spectrum disorders within and across cultures.

Keywords: vocal turn-taking, cross-cultural language development, vocal development, vocal coordination, temporal synchrony, mother–infant interaction, niche construction, cross-cultural psycholinguistics

INTRODUCTION

Background on Parent–Infant Interaction

Conversational exchanges between mothers and their young children have been claimed to be ubiquitous across languages and cultures (Bateson, 1975; Stern et al., 1975; Fernald et al., 1989; Papoušek and Papoušek, 1989; Lavelli and Fogel, 2013) and a hallmark of human sociality (Levinson, 2016). These affectively charged interactions are temporally structured and seem to be shared across species (Takahashi et al., 2013), reflecting phylogenetic origins. As such, they constitute the affordances (Trevvarthen, 1979; Loveland, 1991) within which the mother–infant dyad co-coordinates, recruits, and fine-tunes its vocalizations as it engages in communicative niche construction (Odling-Smee, 1988; West et al., 1988), paving the way for the development of language in human infancy.

The temporal structure of early mother–child vocal interactions has been the focus of several recent studies (e.g., Jaffe et al., 2001; Gros-Louis et al., 2006; Kondaurova et al., 2016), suggesting exchanges characterized by precise gaps and overlaps between the utterances of communication partners (Levinson, 2016) across cultures (Keller et al., 1999). The evidence indicates ontogenetic, dyad-specific patterns of mutually co-coordinated, bidirectional attunement from mother and infant (Ko et al., 2015), reflecting primary intersubjectivity (Trevvarthen, 1977), which progressively develops into predictable interactions that form the basis for the development of social communication in infancy (Papoušek and Papoušek, 1989).

Research has suggested that from around 3 to 4 months of age, infants act as communication partners capable of predicting accurately and with remarkable ease when a conversational turn will end and another turn will begin; the infant's predictions are thought to be based on pragmatic, semantic, and syntactic factors (Gratier et al., 2015) as well as suprasegmental (stress, pitch, intonation) aspects of caregiver speech. By around 5–6 months of age, infants have been reported to engage in protoconversations (Bates et al., 1979) whereby they convey their intentions and engage in social communicative exchanges, still long before they become language users. While we recognize that these exchanges are essentially multimodal in nature, we focus on the vocal modality only, as the primary channel for communication during early interactions between mothers and their infants (Van Egeren et al., 2001). Furthermore, we hypothesize that while early conversational turn-taking is very widespread across cultures (Stivers et al., 2009), it may manifest cross-cultural variation as a function of differences in the social use and structure of the languages in which interactions occur (Lupyan and Dale, 2010; Farran et al., 2016). In fact, while quantification has been scanty, some studies have reported major differences in the amount of parent–infant interaction occurring across cultures, with much lower rates in some low-technology societies (Ochs and Schieffelin, 1984; Correa-Chávez and Rogoff, 2009; Cristia et al., 2017). Moreover, there is emerging evidence to suggest reduced amounts of interaction occurring between mothers and their infants and young children who may be at-risk for, or diagnosed with, autism spectrum disorder (ASD) (Warlaumont et al., 2010; Northrup

and Iverson, 2015), which appear associated with derailment of early speech and language development. Importantly, the majority of the studies of parent–infant vocal interaction have been confined to western cultures. This cultural bias in available research suggests a need for comparative empirical studies of mechanisms that underlie cultural differences in typically developing infants and their mothers. Such work may also lay the foundations for examining vocal coordination between at-risk infants and their mothers and designing early intervention studies aimed at understanding the range of young children's vocal—including disordered—development, within and across cultures.

A growing body of research suggests that while both mothers and infants may contribute to early communicative exchanges (Cohn and Tronick, 1988; Feldman, 2007; Sung et al., 2013; Bigelow and Power, 2014; Gros-Louis et al., 2014), mothers seem to play a more influential role than their infants, often driving the conversational flow and ensuring the maintenance of contingent conversational turns (Jaffe et al., 2001; Yoo et al., 2018) and regular conversational turns (Jaffe et al., 2001) that tend to become more refined with experience (Gratier et al., 2015). The amount of interaction is correlated with positive outcomes in attachment, speech, and language development (Goldstein and Schwade, 2008). Recent evidence also suggests that infants, even as young as 8–21 weeks, may play an active role in initiating interactions and maintaining conversational turns (Dominguez et al., 2016). Likewise, Mothers play a critical role in early vocal interaction. Yoo et al. (2018) have shown that mothers tended to alternate with infant speech-like Vocalizations from the beginning of life, suggesting that they intuitively facilitate protoconversation, while they produce overlapping vocalizations that tend to soothe a crying infant.

Examining the temporal intervals between mothers' and infants' vocalizations is a particularly fundamental realm for understanding early typical (Jasnow and Feldstein, 1986; Beebe and Lachmann, 1988; Stivers et al., 2009) and atypical (Warlaumont et al., 2010, 2014; Smith and McMurray, 2018) development within and across cultures. Temporal vocal coordination has been suggested to be crucial in uncovering developmental anomalies characteristic of, for example, ASD in the first two years, thereby potentially informing early detection and intervention (Northrup and Iverson, 2015).

This type of coordination is contingent upon learning by both conversational partners (Goldstein et al., 2003, 2009). The infant must attune to the temporal structure of conversational turns by adapting to the vocal input pattern of the mother. Similarly, the mother must learn her child's vocal cues and adapt her input to suit her child's language and developmental level. Evidently, variation in one conversational partner's vocalization pattern is shaped by variation in the other conversational partner's vocalization pattern, reflecting co-regulation—the coupling of interactions that presumably becomes more coordinated with development (Fogel, 1977; Kaye and Fogel, 1980). Studying the vocalization patterns of both conversational partners is thus necessary to help determine the extent of temporal coordination between mothers and their infants as a precursor to language development.

The back-and-forth temporal coordination between mother and infant has been the focus of numerous studies (e.g., Bateson, 1975; Snow, 1977; Ginsburg and Kilbourne, 1988), and the results have suggested a predictable temporal structure of maternal vocalizations usually occurring within the first second after the termination of infant vocalizations. The pertinent studies have, however, usually not quantified both gaps (indicating alternation between mothers' and infants' utterances) and overlaps (indicating non-alternation between mothers' and infants' utterances). Moreover, many studies have had very small samples, consisting usually of 1–3 children.

In a recent longitudinal study, Hilbrink et al. (2015) presented data that were, to their knowledge, the first to address both gaps and overlaps in vocal responses of infants to mothers and mothers to infants. The study assessed the temporal structure of conversational turns in 12 infants ages 3, 4, 5, 9, 12, and 18 months and their mothers. Based on a permutation test, the authors reported that infants showed systematic vocal responsivity to mother utterances even at 3 months. Accordingly, they reported changes in infant responsivity across time: infants' responses were slower at 9 and 12 months than at the other ages, a finding the authors interpreted as owing to the infants' growing awareness of words in their ambient language.

Subsequent work by Yoo et al. (2018) examining vocal development in 12 infants at 0, 1, and 3 months along with their parents, also addressed gaps and overlaps in maternal vocal responses, but did so in several ways that differed from Hilbrink et al. (2015). First, the effort by Yoo et al. (2018) differentiated maternal responses to speech-like infant vocalizations (termed “protophones” after Oller, 2000) from cry/distress vocalizations, while the Hilbrink et al. (2015) effort had considered maternal responses to all infant sounds (except involuntary [vegetative] sounds), thus including responses to cries, on the grounds that “mothers often treated these sounds as communicative (p. 5).” Yoo et al. (2018) found, however, that mothers' responses tended to overlap with cries, while alternating (showing gaps) in response to protophones. The authors argued that the differentiation was functional, in that mothers seemed to be attempting to soothe infants who were crying, and thus often spoke comfortingly while the infants cried, whereas they treated the protophones as potential material for conversation, and thus tended to wait for infants to complete their conversational turns before speaking themselves. In addition, the two studies differed in that Hilbrink et al. (2015) presented the data on overlap and gap responses separately, while Yoo et al. (2018) presented distributions where all the data were assessed together. The latter approach has the advantage of implying no assumption that either mothers or infants intend to produce overlapping and alternating responses in categorically different ways. Instead the latter approach allows direct portrayal of a continuum of responsivity with degrees of overlap or alternation illustrated in a combined distribution.

The present work follows trends of prior research, with a few methodological adjustments as will be seen below. Furthermore, we are intrigued by the need to examine temporal coordination and timing of vocalizations in mother–infant interaction across different cultures (Keller et al., 1999; Stivers et al., 2009;

Kärtner et al., 2010). It seems likely that there may be important variations across cultures in the timing of interactions, just as it has been claimed that there are important differences in the raw amount of vocal interaction that occurs with infants across cultures (Dixon et al., 1981; Ochs and Schieffelin, 1984; Richman et al., 1992; Lieven, 1994; Whaley et al., 2002; Le Vine, 2004; Cristia et al., 2017). Our effort has been motivated in part by stark differences in the syllable structure and lexical stress patterns across Arabic and English, in addition to cultural differences that we alluded to in the previous section and possible differences in child-rearing practices that characterize the Lebanese and American cultures. We turn to these differences next.

Background on the Structure of the Arabic Language

Arabic is a Semitic language, spoken by approximately 300 million people worldwide, and characterized by diglossia, which is the co-existence of two forms of the language that work together to serve different, yet complementary sets of purposes: (1) Modern Standard Arabic (MSA) or “fusha,” which is a derivative of classic Arabic and the variant used for formal education, mainly for reading and writing; and (2) spoken Arabic known as “Amiya” with many varieties (vernaculars or dialects) acquired by native speakers as their first language.

Arabic consists of 28 consonants, with consonant-vowel (CV) as the minimal syllable. Importantly, Arabic stress may vary as a function of the spoken vernacular. It is non-phonemic (Holes, 2004, p. 62), non-disjunctive (Kager, 2009, p. 344) and predictable, though dialect-dependent (Saiegh-Haddad and Henkin-Roitfarb, 2014). To illustrate, Aquil (2011) reviewed Cairene (Egyptian) Arabic and posited that “stress in Cairene Arabic is almost entirely predictable” (p.5). In the case of Lebanese Arabic (one of the target languages in this study), it is presumable that word stress is similarly predictable, although this awaits empirical substantiation.

In addition to the identified differences in language structure, it appears that differences in vocal coordination may be based on different views within the Lebanese and American cultures regarding whether, and degree to which, infants are conceptualized and treated as communicative partners (Keller et al., 2007) and whether they are expected to play an active role in shaping communicative exchanges with their mothers. For research on vocal interaction in the Lebanese society, families or kin groups can reasonably be thought to constitute the appropriate unit of analysis, consistent with the tenets of a collectivist culture. By contrast, the infant or the mother–infant dyad is treated as the appropriate unit of analysis within individualistic cultures such as the US (Bornstein et al., 1991; Keller et al., 2007). These differences in roles, namely focusing on the individual versus the group within societies, may impact the frequency and quality of early vocal interaction in the target cultures. Thus the availability of recorded data on mother–infant interaction in Lebanese and American pairs provides a useful starting point in evaluation of possible cultural effects in mother–infant interaction timing.

Our approach is based on dynamical systems theory (DST; Thelen and Smith, 1994), which conceptualizes early language development as an emergent phenomenon, where the history of interactions influences developmental processes and their constraining functions (Deacon, 2010). Central to DST is the understanding of how patterns relate to the co-regulation between constituents of the system, namely members of the dyad—mother and baby. Relevant to this study, the (vocal) behavior of one member (e.g., the infant) acts to constrain the multiple actions and behaviors of the other member (e.g., the mother), resulting in a complex system that organizes and re-organizes itself into *attractors* (macrolevel, stable patterns) that form the basis for recognizable, dyad-specific patterns of behavior. Thus, change occurs at the *microlevel*, and entails an iterative cycling in infant and mother vocal behavior as both dyad members shift from one co-coordinated communicative exchange to another. Each instance of a given attractor that occurs over time has a similar pattern of vocalizations, albeit different from that of the previous instance of vocalizations pertaining to this same attractor, culminating in within-attractor variability (semi-stable patterns of vocal behaviors). Change also occurs at the *macrolevel* and consists of co-occurring microlevel changes reflecting qualitative, nuanced differences in system organization that emerge from previous microlevel changes within attractors, and shifts are often perceived to be transitions in vocal coordination (for example, from less coordinated to more coordinated). The mother-infant dyads may tend toward tighter coupling of vocalizations as they co-construct the language niche for the two of them and for their communities. Presumably, this pattern would be disturbed in the case of atypical coordination seen in atypical development (e.g., ASD), since disruption in the initial state of back-and-forth vocal exchanges forms the basis upon which change at the microlevel occurs, which then becomes the raw material for macrolevel changes that mother–infant dyads use as they co-construct the language niche in ASD. Our aim in the present study includes examining the structure of mother-infant vocal exchanges, as these hold the potential to inform us regarding the dynamics of the system in both typical and atypical infants across development. We propose that possible developmental trajectories become essentially reliant on contextual factors, with ambient language playing a potentially important constraining role, even before infants produce symbols (Bates et al., 1979). Culture and language thus participate in a ratcheting pattern of niche construction (Deacon, 1997, 2010) with early vocal interaction playing a central role in language emergence.

Purpose of the Current Study

The present study aims at examining the extent to which vocal exchanges between mothers and infants are reciprocally structured across Lebanese-Arabic and American-English dyads. We addressed the following research questions by evaluating combined distributions of both overlaps and alternations of vocal responses by mothers and infants:

- (1) What is the timing of mothers’ vocalizations in response to their infants’ vocalizations within and across cultures?

- (2) What is the timing of infants’ vocalizations in response to their mothers’ vocalizations within and across cultures?
- (3) Are patterns of conversational turns different across the cultures?

MATERIALS AND METHODS

Selection of Participants

The participants were 0-24-month-old infants and their mothers (the same sample as in Farran et al., 2016). The Lebanese sample consisted of 19 Lebanese Arabic-speaking mother-infant dyads (12 male and 7 female) recruited from two private and two public pediatric clinics in Lebanon. They came from low-mid SES backgrounds. The American sample consisted of 19 American English-speaking mother-infant dyads (9 male and 10 female) recruited originally in two University of Memphis longitudinal studies on infant vocal development. They came from low-mid to mid SES backgrounds (see **Table 1**). Infants from both cultures were typically developing and matched on maternal education, age, and gender.

Procedure

Briefly we summarize here the procedure detailed in the prior article (Farran et al., 2016). We recorded 10-min samples of mother–infant interactions. For the Lebanese recordings, we asked the mothers to interact freely with their children as they normally did in the home. The recording equipment was high fidelity and the microphone was placed as near the interacting pair as possible but out of reach of the infant. The American

TABLE 1 | Participant characteristics for infants in the Lebanese and American samples.

Arabic Group		English Group	
Infant ID	Age in Months	Infant ID	Age in Months
A1	0	E4	1
A4	5	E13	5
A5	6	E16	6
A6	6	E19	6
A7	6	E18	6
A11	8	E2	8
A13	9	E5	9
A14	10	E7	10
A17	10	E17	10
A18	11	E14	11
A19	12	E15	12
A2	12	E9	12
A3	13	E6	12
A8	16	E12	16
A9	17	E3	16
A12	18	E1	18
A10	21	E10	20
A15	21	E8	21
A16	24	E11	24

recordings were drawn from an archive of longitudinal research at the University of Memphis. Samples had been digitally recorded for audio using wireless microphones worn on both infant and mother in a laboratory designed to resemble a child’s playroom. To enroll in the study, a written informed consent was obtained from all adult participants and parents of non-adult participants.

The Coding Plan and Software

For the Lebanese recordings, the mothers’ utterances were coded in PRAAT (Boersma and Weenink, 2018), an acoustic analysis system available as on-line freeware that allows coders to view waveform and spectrographic displays in real-time, place cursors on the screen to indicate onset and offset of vocalizations, and determine the locations in time and durations of each parent and infant utterance. For the American recordings, we coded mothers’ and infants’ utterances using AACT (Action Analysis Coding and Training by Delgado et al., 2010) with audio represented in TF32 (Milenkovic, 2015). AACT allowed convenient determination of utterance boundaries and durations. For both Lebanese and American samples, coders identified the onset and offset of each utterance using a breath-group criterion (i.e., one utterance per breath group), as recommended by Lynch et al. (1995). We examined both parent responses to infants and infant responses to parents. All coded PRAAT files were converted into AACT coding files for further analysis.

Following the coding scheme of Farran et al. (2016), maternal utterances were characterized as infant-directed speech (IDS) if they were directed to the infant. A maternal utterance was considered to be a response to a baby utterance only if it was an instance of IDS. Similarly, an infant utterance, which consisted of protophones (speech-like sounds), was considered a response to a parent utterance only if it followed a maternal IDS utterance. Therefore, we conceptualized each data point to consist of a pair of utterances—an infant protophone and a maternal IDS utterance.

The first and third authors completed coding for this study. The first author was the main coder and is a speaker of Arabic as a primary language, English as a second language, and French as a third language. The third author, who served as the reliability coder, is a speaker of Mandarin as a primary language, Southern Min as a second language, and English as a third language. Both coders were trained extensively on the coding scheme and underwent two rounds of training in cursor placement under the supervision of the last author (with PRAAT), each coding a total of 38 sessions (19 sessions in Arabic and 19 sessions in English) independently. Intercoder reliability was computed using Intraclass Correlation (ICC), which resulted in a session-level reliability rating of Optimal-Excellent on both utterances per minute and seconds per minute (see Farran et al., 2016).

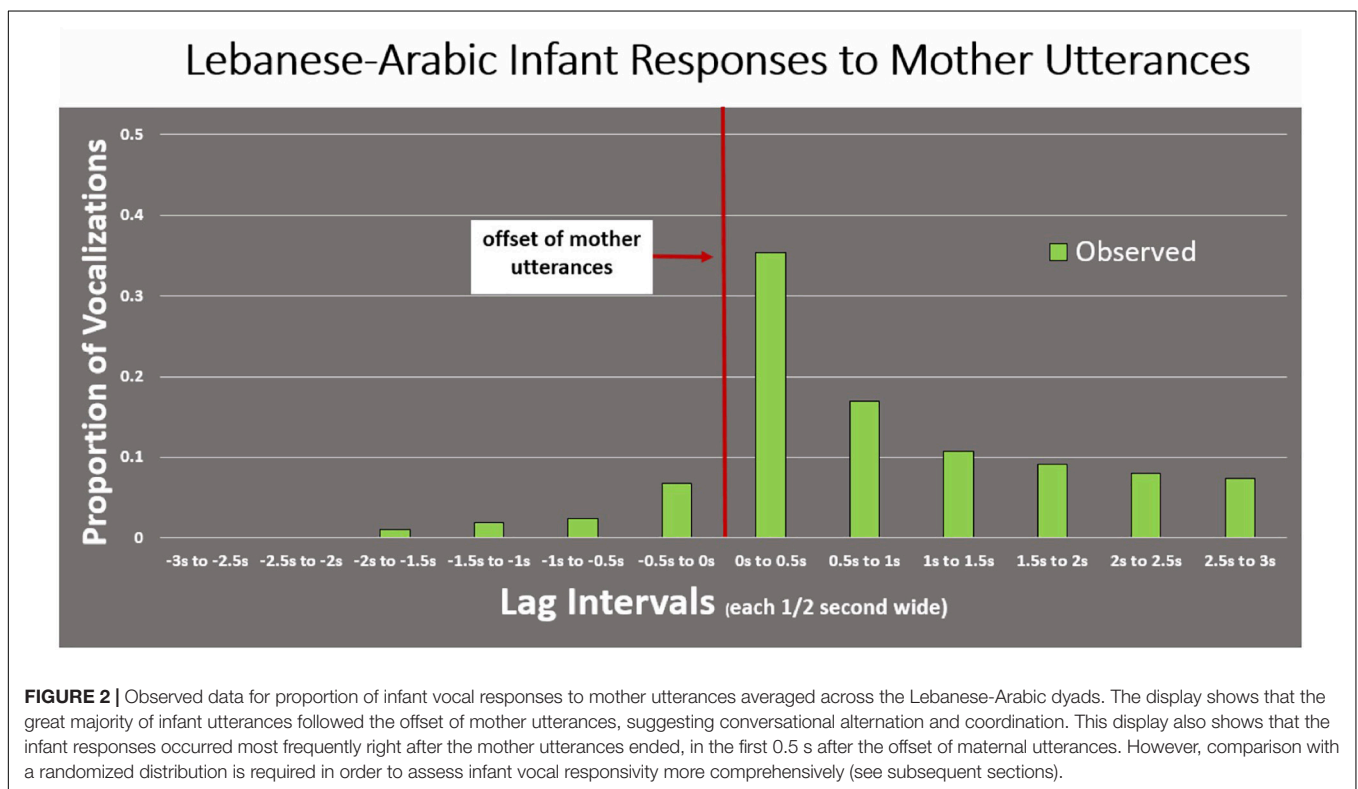
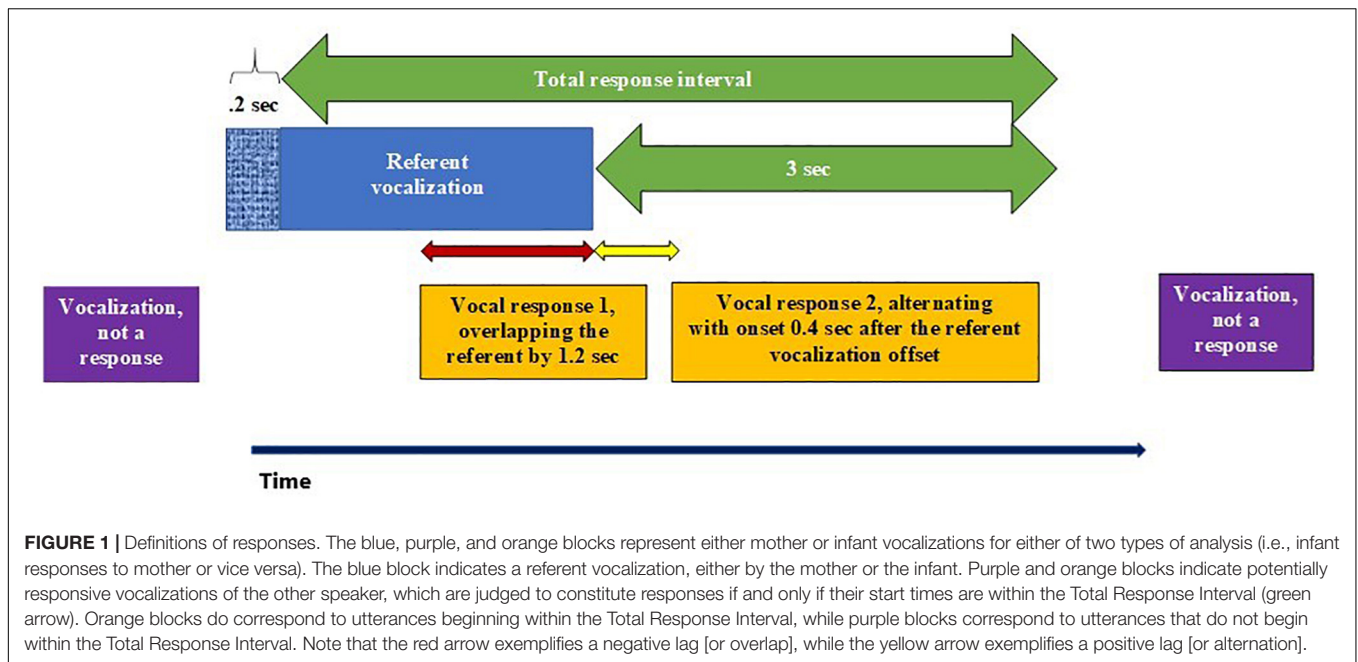
Definition of Response Interval

The results depend upon our definitions of the notions response and response interval, which must be made explicit in order for our effort to be replicable. A Total Response Interval was defined to begin at 0.2 s after the onset of a referent utterance (by the mother or the baby) and to end at 3 s after the offset of the

referent utterance (see **Figure 1**, upper green arrow). In order for an utterance to be deemed a response, it had to *start* within the Total Response Interval. The two purple boxes in **Figure 1**, are not deemed to be responses because their start times do not fall within the Total Response Interval. When defining the responses and calculating lag times, we apply three criteria with regard to the onset and offset of referent and responsive vocalizations. First, as indicated in stippled blue, in order to be treated as a response, a vocalization (e.g., *Vocal response 1* in orange) must begin > 0.2 s after the onset of the referent vocalization (the entire vocalization includes both stippled blue and solid blue). Thus, the first purple vocalization does not qualify as a response because it does not begin after the first 0.2 s of the referent vocalization. *Vocal response 1* and *Vocal response 2*, however, both qualify. Second, a response that follows the offset of the referent must begin within 3 s (bottom green arrow) of that offset. Thus, the second purple vocalization does not qualify as a response because it begins > 3 s after the offset of the referent vocalization. Third, there is a one-to-one mapping of any referent vocalization to a response, and the response must be the first vocalization to qualify. Thus, in **Figure 1**, *Vocal response 2* could only count in the analysis if *Vocal response 1* had not occurred. Any vocalization that qualifies as responsive to more than one referent, must be assigned, in accord with the one-to-one mapping rule, to just one of those potential referents, namely the one whose onset is closest in time to the onset of the potential response. Finally, note that the total response interval duration (upper green arrow) for any referent vocalization, can vary substantially, because it depends on the duration of the referent vocalization itself plus the 3 s following interval.

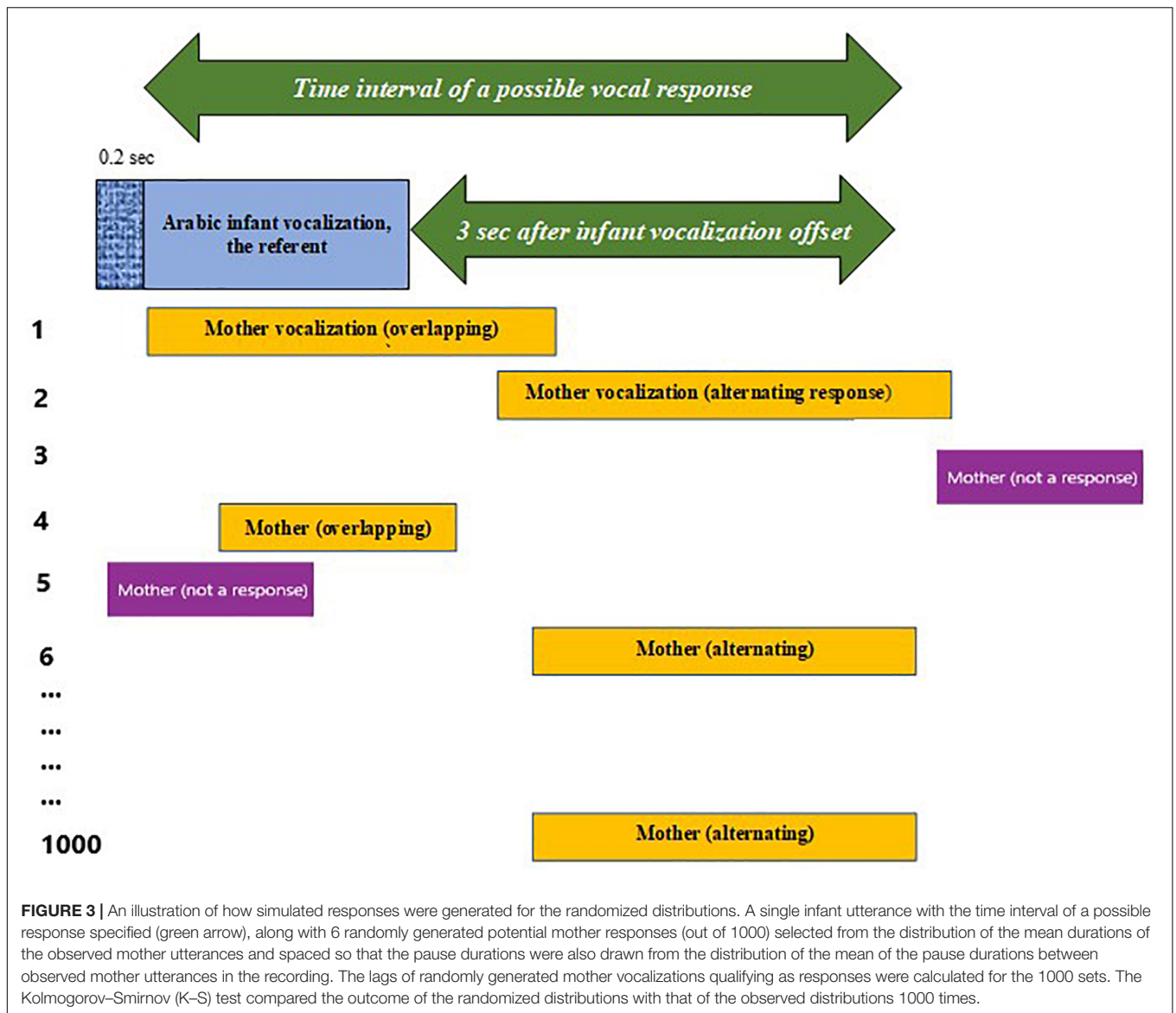
The audio displays allowed ms-level resolution for the visualization of onsets and offsets of utterances. The 0.2 and 3 s time frames are empirically motivated. A vocal response to an auditory stimulus requires about the duration of a short syllable (~ 0.2 s, Jain et al., 2015), and consequently a vocalization occurring in the 0.2 s interval after the onset of another vocalization should not be considered a response. Further, a pause within an adult conversation of more than 3–5 s can often be seen to correspond to a topic change; an utterance occurring after such an interval can reasonably be thought of as a new initiation of the old conversation or the beginning of a brand new one. The 3 s time frame seemed empirically appropriate because, for the data displayed below (illustrated in **Figure 2**), we found that the high responsivity shortly after the offset of referent utterances decreased rapidly thereafter, and if we extended the distributions beyond 3 s, the proportions of “responsive utterances” ceased to decrease further, suggesting that after 3 s, speakers were randomly initiating potentially new conversations, rather than responding to a prior utterance of the other speaker. Thus empirically the 3 s frame after the end of a referent utterance seemed to encompass the real responses to that referent utterance. **Figure 1** illustrates our definitions.

Next we illustrate the response interval idea based on the actual distribution of the Lebanese-Arabic observed data for infants responding to mothers (**Figure 2**). This distribution shows the observed data averaged across the Lebanese-Arabic dyads, where the proportions of infant utterances that met the



requirements to be responses to mother utterances within each time interval are displayed. Thus, the first bar to the right of the vertical red line indicates the proportion of infant vocal responses occurring in the first half-second after the offset of the mother utterance, while the next bar to the right represents the infant responses occurring in the second half-second. Bars to the right of the red line correspond to *positive lag*, that

is proportions of infant utterances that followed the offset of mother utterances. Bars on the left represent the proportion with *negative lag*, that is infant utterances that *preceded* the offset of mother utterances, and thus *overlapped* with them. Clearly, the great majority of infant utterances followed the offset of the mother utterances, suggesting conversational alternation and coordination. It is clear that infant responses occurred most



frequently right after the mother utterances ended, in the first half-second after the mother offset. At the same time, while **Figure 2** suggests that the Arabic infants were alternating with their mothers vocally, the pattern could be misleading unless we compare the observed distributions with randomly distributed possible infant responses, which we address next.

Generating Randomized Distributions and Testing for Responsivity

We conducted Kolmogorov–Smirnov (K–S) tests to examine whether the observed responsivity patterns differed from randomly generated simulations of responsivity patterns in both Arabic and English for mothers responding to infants and for infants responding to mothers. The following describes the generation of the randomized distributions, which was based on principles in common usage in the literature on randomized

designs (Kempthorne, 1955; Edgington, 1969; Romano, 1990; Gadbury, 2001). This kind of randomized approach has been widely specified and used in a variety of fields, although to our knowledge has not been used before in parent–infant interaction research. Our intent is to specify our procedure sufficiently for potential replication. **Supplementary Appendices A and B** provide both a technical description of the procedure and the R code that implemented it. In the following we describe the procedure conceptually.

A “pair” of utterances was defined here as consisting of a sequence of a mother utterance (the referent) followed by an infant utterance (the response), or an infant utterance (the referent) followed by a mother utterance (the response), meeting the timing requirements of our response interval definition as in **Figure 1**. Thus, a response always started at least 0.2 s after the onset of the referent. Further, a response could start up to 3 s after the offset of the referent.

For each of the four data sets (Arabic mothers responding to Arabic infants, American mothers responding to American infants, Arabic infants responding to Arabic mothers, American infants responding to American mothers), the mean duration of the responses and the mean duration of the pauses between the responses was computed for each recording. Using these means as parameters of exponential distributions, a sequence of random response durations and pause durations was generated for each recording. This randomized sequence of (possible) responses was compared with the fixed set of referents within each recording to calculate the lags (either positive or negative) corresponding to the qualifying responses (in accord with the response definition) of the randomized distribution for that recording. One thousand such random sets were generated for each recording, and the corresponding lags were computed for each set of infant–mother or mother–infant pairs across each set of recordings from each of the four data sets. Randomized distributions for each of the four data sets represented data collapsed across all the recordings within each data set.

For example, in a K–S test for mother responses to infant utterances in Arabic, we started by fixing in time the infant utterances (in this case, the referents) for an Arabic recording. **Figure 3** shows an example infant utterance placed in time as it was observed, with the time interval of a possible response specified. Also displayed are six possible randomly generated potential mother responses (out of the 1000 generated for each recording), the durations of which were all selected from the distribution of the mean durations of the observed mother utterances in the Arabic recording in question and all spaced within each of the set of 1000 so that the pause durations were also drawn from the distribution of the mean of the pause durations between mother utterances. Having generated the randomized sets, the lags of randomly generated mother vocalizations qualifying as responses were calculated for the 1000 sets, and the outcome responsivity for the randomized distributions was compared with that of the observed distributions 1000 times by the K–S test.

RESULTS

Illustration of Comparisons Between Observed and Randomized Distributions of Vocal Responses: The Case of Lebanese Arabic-Learning Infants Vocally Responding to Their Mothers' Vocalizations

Understanding the results requires appreciation of an asymmetry in the shapes of the randomized distributions obtained through our procedure. One might expect random events to be distributed roughly equivalently across the time frame evaluated for responses. However, actually randomized events show much larger numbers of positive than negative lag values, and are thus more similar to the observed distributions of responses than one might imagine. This asymmetry in the randomized distributions is the result of asymmetry in the durations of referent utterances

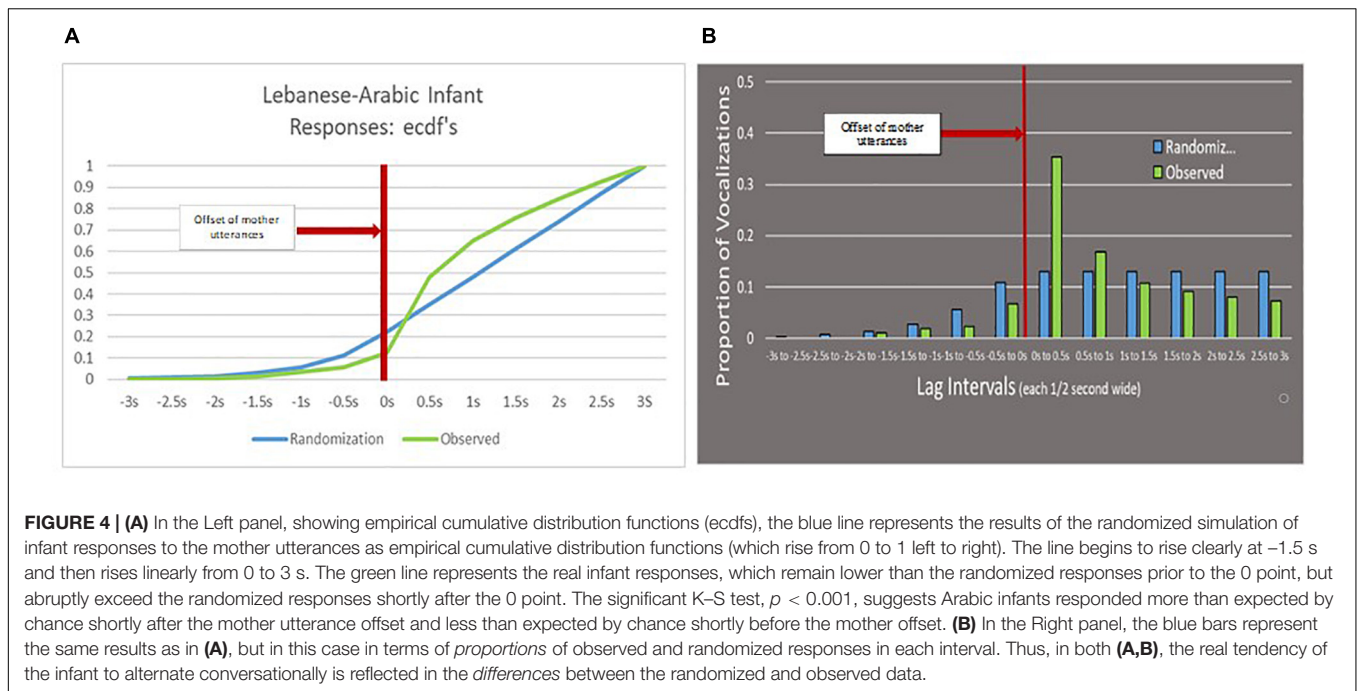
and the following response intervals. The former were variable (typically from 0.4 to 1 s for infant utterances, and 0.5 to 1.5 s for mother utterances) and much shorter than the fixed following response intervals (3 s).

We conducted the K–S test to examine whether our observed data (either mother or infant vocal responses) showed random or systematic responses with respect to the referent utterances. If the distribution of the observed data significantly differs from the distribution of the randomized data, it corroborates that the observed data indeed show systematic behaviors. **Figures 4A,B** provide examples of the K–S test and we provide details of the analysis as follows.

Figure 4A (empirical cumulative distribution functions of responses, ecdfs) and **Figure 4B** (proportional occurrence of responses) illustrate the outcome for Lebanese infants responding to Lebanese mothers in two ways. In **Figure 4A** the outcome is displayed as two empirical cumulative distribution functions (ecdfs) for the observed and randomized lags, while in **Figure 4B** the same data are displayed in bar-graph form, with each bar representing proportions of responses in the observed and randomized data (in this illustration, the observed data are the same as in **Figure 2**). In both cases, the data are binned at 0.5 s starting 3 s before the 0 point (the offset of the referent utterance, by the mother in this case) to 3 s after the 0 point. Because the referent utterances were of variable duration, and were typically much shorter than 3 s, there are very few responses in the Figures for either the observed or the randomized distributions prior to a lag of -1.5 s.

The slow rise in the negative lag range for both **Figures 4A,B** and the much more rapid rise in the positive lag range are of course attributable to the much lower likelihood of a responsive utterance occurring *during* a mother referent utterance (which was typically much less than 3 s in duration) than during the following pause (which was typically three or more times longer). In **Figure 4A** the randomized distribution rises essentially linearly in the positive lag range, and in **Figure 4B** the same distribution is essentially flat in the positive lag range, because the positive lag range is fixed in duration. It is the relative positioning of the observed and randomized data in these figures that supply the most meaningful portrayal of infant vocal responsivity—in the absence of the randomized comparison, the responsivity patterns of the infants in **Figures 4A,B** would appear misleadingly large, because we would be tempted to base our judgment only on the relative amounts of response before and after the 0 point.

The K–S test for whether the observed Lebanese Arabic-learning infant responses were random with respect to the fixed Lebanese Arabic-speaking mother speech patterns rejected the null hypothesis at $p < 0.001$ (test statistic = 0.168), indicating that Lebanese infants did indeed respond systematically. In **Figure 4A**, it can be seen that the two distributions begin to separate in the 1.5 s before the offset of the referent mother utterance (the 0 point) indicating the tendency of the real infants (in green) to respond less often with negative lag or overlap than in the case of randomly distributed simulated responses (in blue). Then immediately after the offset of the mother referent utterance, the real infant responses abruptly rise above



the randomized responses, indicating infants tended to respond non-randomly, especially in the first 0.5 s after the mother utterance offset. In **Figure 4B**, the observed infant responses also show lower proportions prior to the mother utterance offset than randomly occurring simulated responses, with a sharply higher proportion for the observed infant responses in the 0.5 s after the mother utterance offset. Clearly, as seen in **Figure 4B**, the Lebanese infants seemed to cluster their responses shortly after the offset of the mother utterances, suggesting a pattern of conversational alternation.

Summary of Various Comparisons of Observed and Randomized Distributions of Vocal Responses

The relation between response patterns as seen in the ecdfs and in the proportion displays is similar for all the comparisons to follow, and K–S tests produced significant rejection of the null hypothesis that the observed and randomized distributions were the same at $p < 0.001$ for all the four key cases discussed below (Lebanese infants responding to Lebanese mothers [as indicated above], Lebanese mothers responding to Lebanese infants, American infants responding to American mothers, and American mothers responding to American infants). Moreover, the data show a similar pattern in all four cases, in that the responders showed fewer vocalization responses in the half-second interval before the referent speaker's offset than in the randomized set, and more responses in the half-second interval after the referent speaker's offset than in the randomized set.

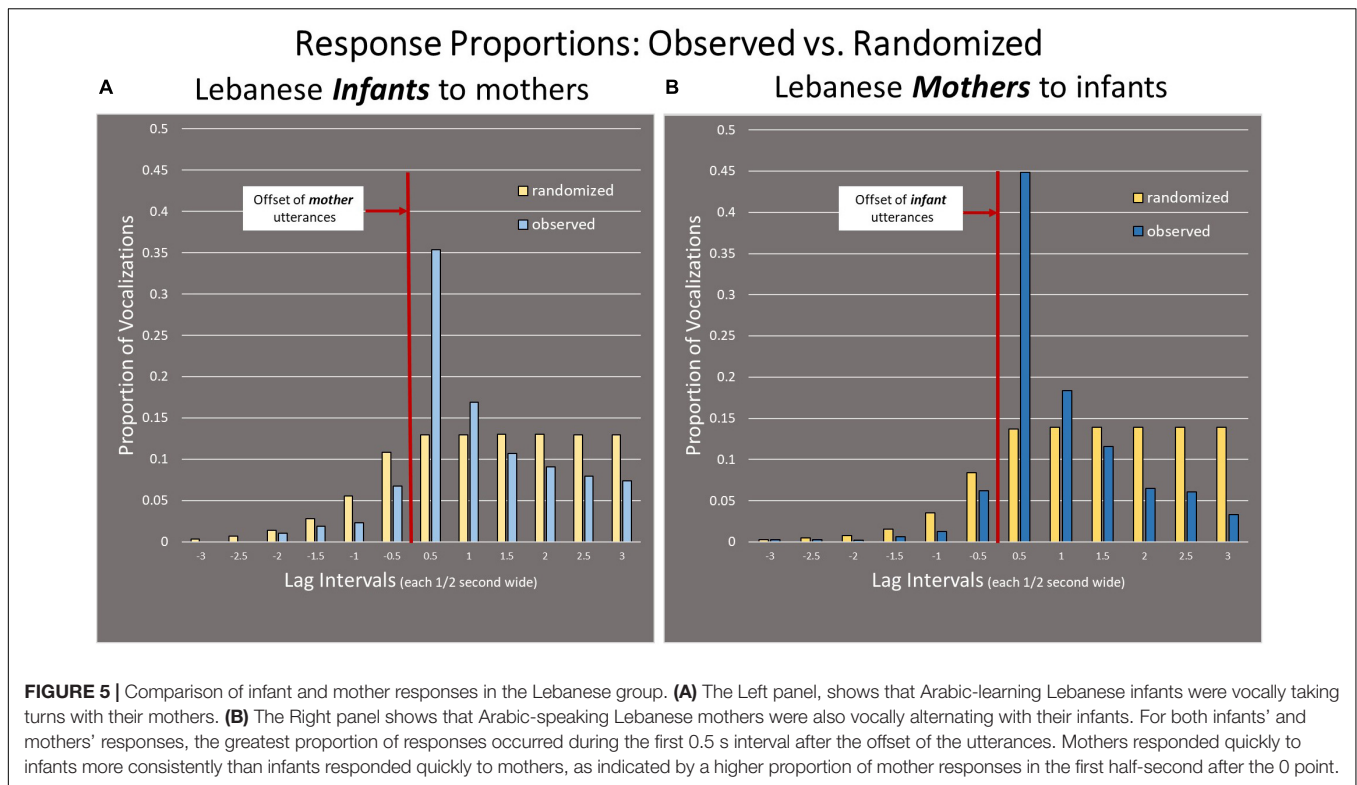
Although not the focus of the study, we conducted exploratory analyses to examine whether the temporal structure of mother-child vocalizations varied as a function of children's age. The infants followed the same pattern of vocal responsivity at both

the younger (<10 mo) and older ages (>10 mo) in both cultural groups with all four comparisons of observed and randomized distributions, in all cases showing $p < 0.001$. As expected, mothers showed more distinct responsivity than infants in both cultural groups ($p < 0.001$), with higher proportions of responses in the first half-second after the infant referent vocalizations than the infants showed with respect to the mother referent vocalizations.

For the remaining displays in the main text of the article, we shall present the data in terms of observed and randomized proportions only. Additional ecdfs are provided in **Supplementary Appendix C**. It might be speculated that with our relatively small N, outlier infants might have strongly affected the patterns of results. We have supplied ecdfs in **Supplementary Appendix C** that show observed lag distributions for the individual infants, revealing that outliers played a small role, with most infants conforming to the patterns reported below.

Arabic-Learning Infants Compared to Arabic-Speaking Mothers

In **Figure 5**, the Arabic-learning infants as responders are shown on the Left panel and the Arabic-speaking mothers as responders are shown on the Right panel. Here again, the intervals on the x-axis represent half-second time frames, from -3 s to $+3$ s with respect to offset of the referent utterances. The results indicate that the mothers alternated vocalizations with infants in a similar way to that of the infants alternating with the mothers. The greatest proportion of responses of mothers also occurred during the first half-second interval after offset of the infant utterances, and there was a tendency for fewer responses to occur in the intervals before the infant offset than in the case of randomly distributed responses. Moreover,



the mothers were more consistent in responding quickly than the infants were ($p < 0.001$)—45% of responses of mothers versus 35% of responses of infants were in the first half-second after the offset of the referent utterance, showing that Lebanese mothers were more precisely able to time their vocally responsive utterances than infants.

English-Learning American Infants Compared to English-Speaking American Mothers

Figure 6 illustrates the data for the American dyads with infants on the Left panel and mothers on the Right panel. As in the Lebanese Arabic case, the K–S test showed there was a statistically significant difference from chance for both infants and mothers ($p < 0.001$). Similar to the Lebanese Arabic data, the greatest proportion of responses of infants and mothers occurred during the first half-second after the referent utterance offset. It is thus evident that both mothers and infants in both cultures showed significant vocal alternation. As in the case of the Lebanese dyads, the American infants' preference for responding in the first half-second was much less strong than in the case of the American mothers ($p < 0.001$), with 42% of maternal responses occurring in the first half-second compared to only 23% of infant responses.

American Infants Compared to Lebanese Infants

Based on data in Figure 7, we conducted the K–S statistical significance test for possible differences between American and Lebanese Arabic infants and found that Lebanese Arabic infants

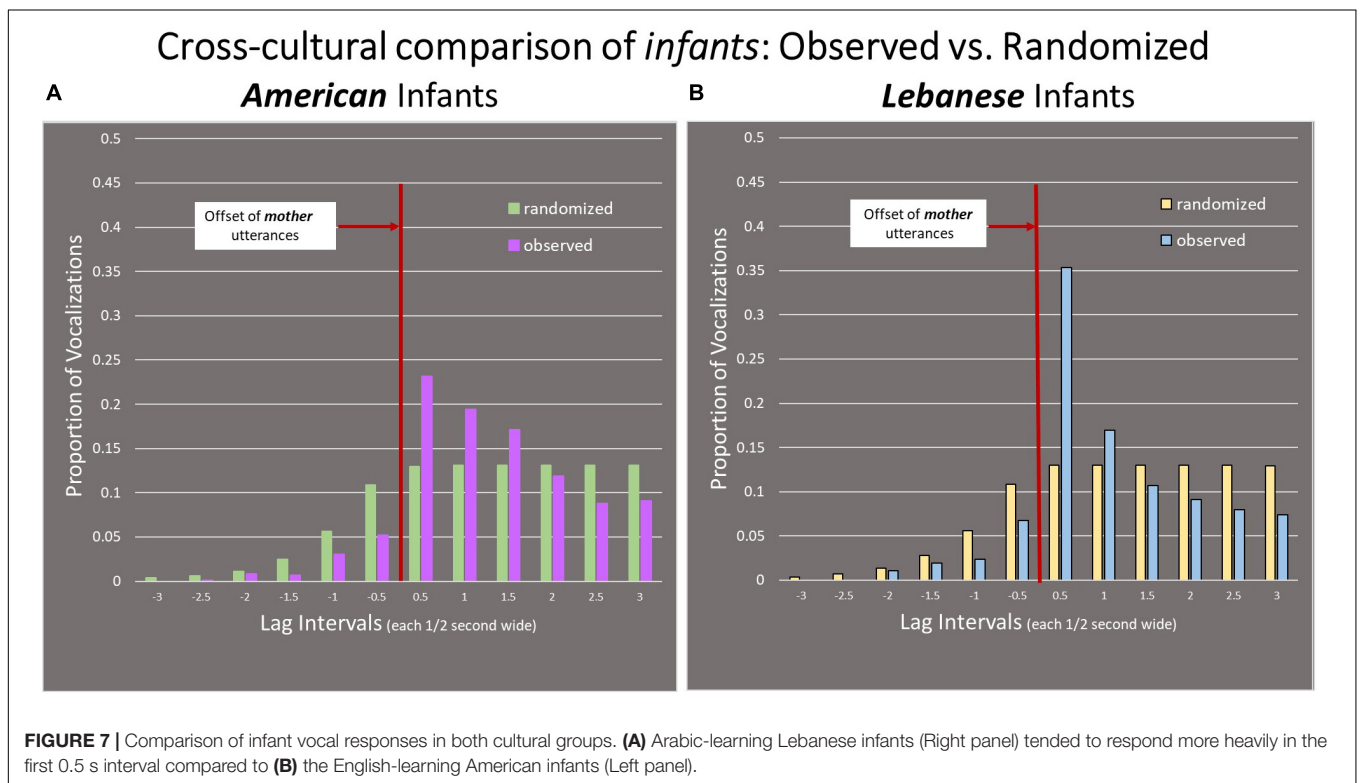
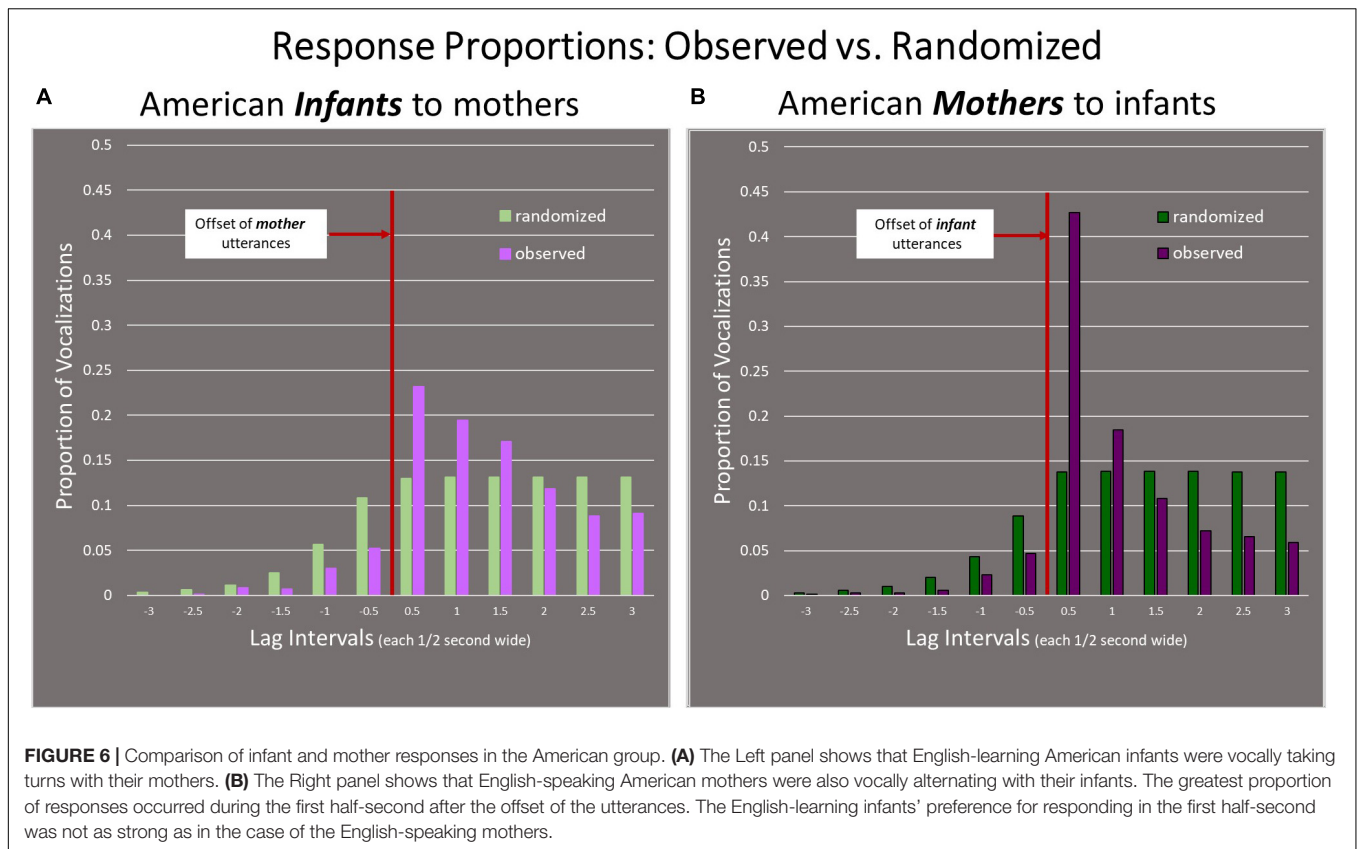
tended to respond more heavily in the first half-second interval after mother utterance offset compared to the American infants (the lag distributions of observed responses differed across cultural groups at $p < 0.001$; the K–S test statistic was 0.14 with the critical value for $p < 0.001$ being 0.09). The pattern shows 35% of Arabic infants responses occurred in the first half-second compared to 23% for the American infants. In fact, the American infants continued to show response levels exceeding chance even in the third interval, out to a second and a half beyond the offset of the mother utterances. In contrast, the Lebanese Arabic infants' responsivity beyond the chance level did not continue beyond the one second period after the offset of the mother utterances.

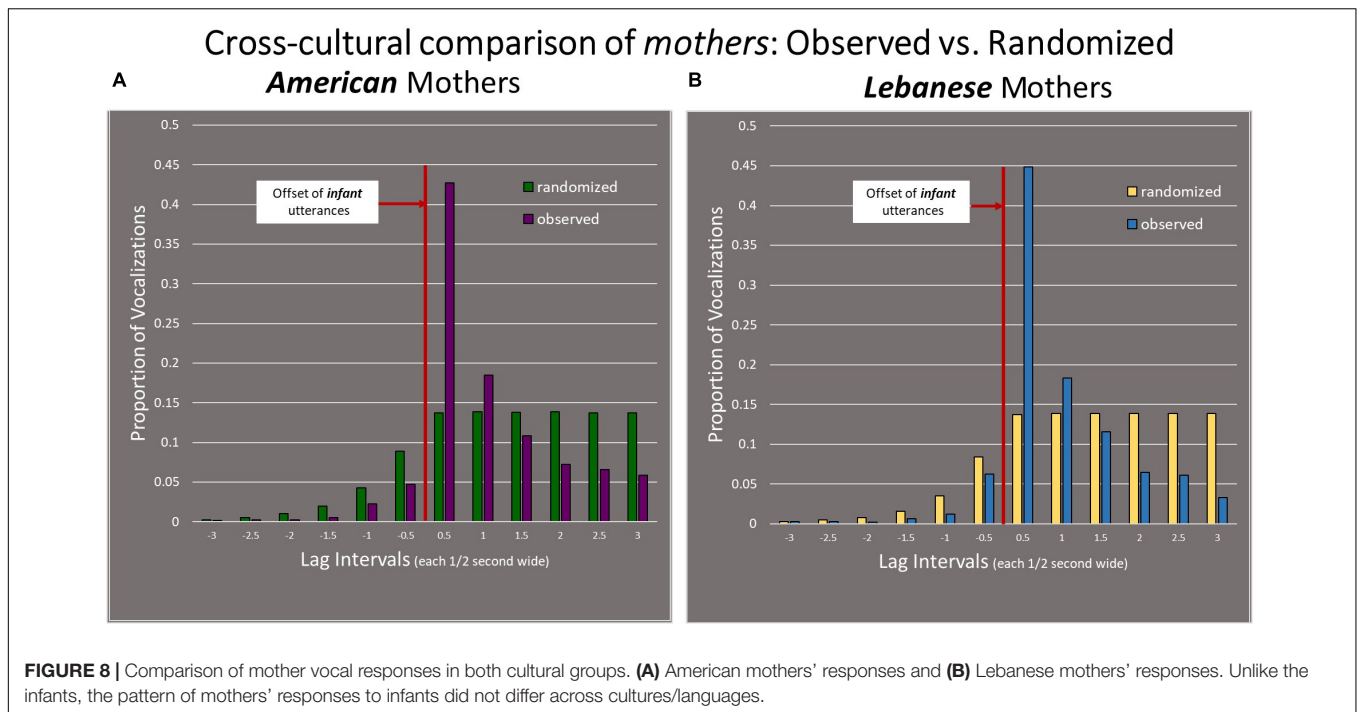
American Mothers Compared to Lebanese Mothers

The mothers did not differ in the way the infants did across cultures (Figure 8). In both cultures, the responses of mothers were concentrated within the first half-second after the offset of the infant utterances, and the proportions of responses within that interval were similar (42 and 45%, $p > 0.05$; the test statistic was 0.03808, while the critical value for $p < 0.05$ was 0.0624).

DISCUSSION

The current study sought to investigate temporal coordination in vocal exchanges between mothers and infants across two cultural groups, Lebanese-Arabic and American-English dyads in the first 2 years of life. Our method addressed both gaps and overlaps





in both mother–infant and infant–mother responses. The study aimed at answering three research questions:

- (1) *What is the timing of mothers' vocalizations in response to their infants' vocalizations within and across cultures?*

We found that the mothers appeared to engage in vocal alternation and behave similarly in terms of vocal response timing with their infants within and across Lebanese and American cultures and across the first 2 years of life. This suggests that mothers may be responsive to their infants' vocalizations in both cultures and may perceive their infants as capable communication partners, thus treating their vocalizations as intentional communication bids (Schieffelin and Ochs, 1986).

- (2) *What is the timing of infants' vocalizations in response to their mothers' vocalizations within and across cultures?*

Similarly, our findings reveal within each culture that infants seem to respond contingently to mothers' vocalizations: Lebanese infants followed Lebanese mothers vocally, and American infants followed American mothers vocally. These patterns support the view that temporal interactional dynamics is a universal phenomenon (Stivers et al., 2009). That this pattern applied both to the Lebanese and American dyads and to infants across the first and the second year suggests that contingent vocal interaction with alternation is a fundamental characteristic of humanity, supporting the supposition that such interaction forms a foundation for language (Trevathan, 1979, 2001).

While both mothers and infants showed systematic patterns of vocal response within the first half-second after the offset of vocalizations by the other speaker, the tendency to favor that first half-second was notably stronger in the mothers than

the infants in both cultures. The pattern suggests that mothers across the two cultures may have been guiding their infants, whether consciously or unconsciously, toward conversational competence. Other recent studies suggest that mothers, even in the first month, engage in such conversational guidance (Dominguez et al., 2016; Yoo et al., 2018). In fact, it is not entirely clear in the first few months of age that infants actively respond to vocalizations of mothers at all. We cannot rule out the possibility that mothers are able to predict infant vocal patterns and that they tend to curtail their own vocalizations when they sense their babies are about to speak. If this is true, their predictive curtailments would surely yield the impression that their infants are more actively involved in conversation than they actually are.

By 5–6 months of age, there exists other evidence that infants do indeed actively play a role in conversation—they tend to double their rate of vocalization in the still-face paradigm during the still-face episode, indicating presumably that they are trying to repair the “broken” conversation when the mother goes silent (Yale et al., 2003; Goldstein et al., 2009; Franklin et al., 2013).

These thoughts are important in the present context because they highlight the fact that even our own pattern of results—comparing observed and randomized distributions—does not rule out a role for the mother in guiding conversations with her infant, even at the older ages of the study, producing patterns that may suggest a more systematic and active role for the infant than is veridical. Surely, infants are learning to predict mothers' vocal patterns across the first years of life, and they might also be learning to guide conversations, as the mothers clearly do. It is important to note that while our method shows unambiguously that there are systematic relations in time between mother and infant vocalizations, and that those timing relations suggest conversation, the results cannot show by themselves that infants

are responding actively, given that mothers clearly have at least some ability to predict infant vocalization timing. We do, however, believe it is safe to conclude that the results suggest active entrainment by the mothers, encouraging infants to learn the patterns of contingency that conversation requires.

(3) *Are the patterns of conversational turns different across cultures?*

Despite the observed universal pattern of vocal responsivity in mothers and infants, some cultural variation emerged. While mothers responded to infant vocalizations in a similar time frame across the two cultures, we found cross-cultural differences in the infants' responsivity. These differences were observed in the speed with which Lebanese infants responded to their mothers, with more than one third of the Lebanese infant responsive utterances tending to occur within a half-second after the offset of maternal utterances, compared to less than one quarter in the American infant sample.

While the mechanisms that might drive infants to show different patterns of temporal responses across cultures are unknown, here we provide our favored speculation, which is based on the possibility that the Arabic language may provide more predictable indicators of utterance termination than the English language. We are not asserting this difference to exist—both languages have complex stress systems (Kenstowicz and Abdul-Karim, 1980; McCarthy, 1980; Selkirk, 1980; Osterling and Dawson, 1994; Kiparsky, 2003; Watson, 2011)—but if it does, it could help explain the infant responsivity pattern, where Lebanese Arabic infants seemed more able than American English infants to time their responses with the ending of mother utterances.

Let us exemplify: We speculate that English may not be as predictable in terms of location of stress in words as Arabic (Watson, 2011). Word stress is well known in some languages to be a strong predictor of boundaries. For example, French tends to have final word stress, Czech initial stress, and both Polish and Spanish penultimate. Thus, any content word pronounced in any of these languages provides a clue about when it may end as soon as the stress is identified by the listener in real time. Both English and Arabic have very complicated stress patterns. In fact, recent evidence suggests that stress is highly predictable in Modern standard Arabic (Saiegh-Haddad and Henkin-Roitfarb, 2014) and the Cairene (Egyptian) dialect of Arabic (Aquil, 2011). Accordingly, our suggestion is that perhaps the English stress is the less predictable and consequently, the American infants may have tended to lag a little longer in their responses, because they did not have such a strong basis as the Lebanese infants to predict when the mothers' utterances would end.

An alternative interpretation for the differences in infant temporal coordination across the cultures could be based on differences in maternal utterance duration (Farran et al., 2016). IDS has been thought to encourage rapid infant vocal responsivity, and short IDS utterances may be particularly effective in facilitating rapid responsivity. It could be the case that the Lebanese infants' ability to respond quicker to their mothers' utterances was brought about by Lebanese mothers'

shorter utterance duration (and Lebanese infants' awareness of it) compared to that of American mothers. Similarly, intonation patterns could have differed between the languages, cuing the Lebanese infants more fully than American infants about maternal utterance termination.

Our aim at offering a culturally motivated explanation leads us to a related speculation. The fact that infants across the two cultural groups differed in their ability to time their turns to match their mothers' communicative bids could be driven by the differing mothers' expectations of their children, owing to potential cross-cultural differences in child rearing practices and the degree to which infants are conceptualized as communicative beings (Goodnow et al., 1984; Keller et al., 2007). The Lebanese mothers may have thus been more likely to converse using significantly shorter, simpler utterances, which in turn facilitated infants' ability to co-coordinate the timing of their vocalizations to match that of their mothers. At the same time, it is important to emphasize that mothers in both cultural groups appeared to be seamlessly and effortlessly capable of tailoring the timing of their vocalizations to adjust to the timing of their infants' vocalizations, presumably enabling both groups of mothers to recruit their infants' rudimentary turn-taking abilities toward increasingly sophisticated language competence (Fernald et al., 1989).

The fact that both mothers' and infants' vocalizations were temporally structured, corroborates previous research (Jaffe et al., 2001; Keller et al., 2007; Bornstein et al., 2015) pointing to sensitivity on the part of the dyad to adapt to the cultural environment (for example to patterns of interaction of the other party to conversation or to special characteristics of the language structure) (Thelen and Smith, 1994). The results suggest it may be important to consider not only universal variables in child rearing practices, but also cultural and language-specific variables which may shape social interaction uniquely, thus creating special cultural niches (Odling-Smee, 1988) between mothers and their infants that form the basis for the development of language in infancy and early childhood.

Several efforts have aimed at gaining a fuller understanding of vocal development across cultures (Bornstein et al., 2015) and socioeconomic strata (Hoff, 2003). While the evidence to date corroborates IDS and vocal coordination between infants and their mothers across a variety of cultures, it has also been speculated that vocal coordination may be derailed, if, for example, endogenous infant vocalization or maternal response to infant vocalizations is very low (Oller et al., 2019). The current study supports the need to investigate language structure and cultural practices as important influences on of vocal coordination.

Understanding how vocal interaction unfolds between typically developing infants and their mothers also has the potential to inform our understanding of the emergence of developmental anomalies (Leezenbaum et al., 2014) across cultures. This may in turn be crucial in contributing to the assessment of and intervention with infants and young children who have developmental vulnerabilities (e.g., ASD, see Northrup and Iverson, 2015) and shaping developmental trajectories of turn-taking for optimal language outcomes across cultures.

Future research attention should be directed to understanding the genesis of turn-taking between mothers and their infants with developmental disorders (e.g., ASD) across cultures and languages, characterizing the contribution of each member of the dyad in the co-coordination of communicative exchanges. Moreover, future studies could help uncover the mechanisms that undergird maternal input and interactivity such as varying arousal levels of infants associated with different speech registers (Farran et al., 2016), level of infant volubility (Oller et al., 2019) and lexical diversity, and varying cultural views of infants thought to be associated with the emergence of language in human infancy (Kärtner et al., 2010). A challenge for future work will be to tease apart the input cues (e.g., acoustic, visual, physical) that likely interact with infant characteristics (e.g., presence of a developmental disorder, communicative fitness, attachment) to shape the trajectories of turn-taking in infants at-risk for or diagnosed with developmental disorders (Loveland, 1991). Current research points to the importance of multimodal sampling of vocal interaction—a limitation of the present study. This is particularly crucial given evidence underscoring the importance of the visual modality in infant volubility (blind infants have been reported to show low volubility, Fraiberg, 1979) and the potential protective role multimodal cues may exert in cases where maternal input has been found to be less-than-optimal (e.g., low SES, low technology societies) (Cristia et al., 2017). Such an understanding will contribute immensely to the design and implementation of better intervention studies that may ultimately improve future language and overall developmental outcomes of at-risk children. Future research could also benefit from employing more nuanced approaches to data collection and analysis by introducing random sampling of home recordings of face-to-face interactions, multimodal data collection, and targeting languages that come from different language families as well as various dialects within specific languages such as Arabic and English. Such expanded research should enable direct evaluation of the contribution of various language and dialect structures to early vocal coordination and ultimately language emergence across cultures.

DATA AVAILABILITY STATEMENT

The datasets generated for this study will not be made publicly available. The recordings cannot be made publicly available because the conditions of IRB approval for making the recordings

REFERENCES

- Aquil, R. (2011). *Empirical Evidence: Stress As a Perceptual Unit in Cairene Spoken Arabic. Perspectives on Arabic Linguistics*. Amsterdam: John benjamins.
- Bates, E., Benigni, L., Bretherton, I., Camaioni, L., and Volterra, V. (1979). *The Emergence of Symbols: Cognition and Communication in Infancy*. New York, NY: Academic Press.
- Bateson, M. C. (1975). Mother-infant exchanges: the epigenesis of conversational inter-action. *Ann. N. Y. Acad. Sci.* 263, 101–113. doi: 10.1111/j.1749-6632.1975.tb41575x

do not include permission from the parents to make the recordings available publicly. To provide the raw recordings would violate the confidentiality requirements of the data collection. However, we can supply the data sheets including durations obtained from both coders. From these sheets, all the calculations have been made that are included in the paper. Requests to access the datasets should be directed to the corresponding author.

ETHICS STATEMENT

The University of West Georgia Institutional Review Board and the University of Memphis Institutional Review Board approved this study. This study was carried out in accordance with the recommendations of the University of West Georgia and University of Memphis guidelines and IRB Committees' approval with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the IRB committee at the University of West Georgia and the University of Memphis.

AUTHOR CONTRIBUTIONS

LF and DO contributed to conception and design of the study. LF, HY, and C-CL organized the database. DB performed the statistical analysis. LF wrote the first draft of the manuscript. LF, DB, and DO wrote sections of the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

FUNDING

This work was funded in part by Grants R01 DC006099, DC011027, and DC015108 from the National Institute on Deafness and Other Communication Disorders and by the Plough Foundation, which supports Oller's Chair of Excellence.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02374/full#supplementary-material>

- Beebe, B., and Lachmann, F. (1988). The contribution of mother-infant mutual influence to the origins of self and object representations. *Psychoanal. Psychol.* 5, 304–333.
- Bigelow, A. E., and Power, M. (2014). Effects of maternal responsiveness on infant responsiveness and behavior in the still-face task. *Infancy* 19, 558–584. doi: 10.1111/inf.12059
- Boersma, P., and Weenink, D. (2018). *Praat: Doing Phonetics by Computer [Computer Program]. Version 6.0.43*. Available at: <http://www.fon.hum.uva.nl/praat/> (accessed June 20, 2015).
- Bornstein, M., Putnick, D., Cote, L., Haynes, M., and Suwalsky, J. (2015). Mother-infant contingent vocalizations in elevn

- countries. *Psychol. Sci.* 26, 1272–1284. doi: 10.1177/0956797615586796
- Bornstein, M., Tal, J., and Tamis-LeMonda, C. (1991). “Parenting in cross-cultural perspective. The United States, France, and Japan,” in *Cultural Approaches to Parenting*, ed. M. H. Bornstein, (Hillsdale, NJ: Lawrence Erlbaum Associates), 69–90.
- Cohn, J. F., and Tronick, E. Z. (1988). Mother–infant face-to-face interaction: influence is bidirectional and unrelated to periodic cycles in either partner’s behavior. *Dev. Psychol.* 24, 386–392. doi: 10.1037/0012-1649.24.3.386
- Correa-Chávez, M., and Rogoff, B. (2009). Children’s attention to interactions directed to others: guatemalan mayan and European American patterns. *Dev. Psychol.* 45, 630–641. doi: 10.1037/a0014144
- Cristia, A., Dupoux, E., Gurven, M., and Stieglitz, J. (2017). Child-directed speech is infrequent in a forager–farmer population: a time allocation study. *Child Dev.* 90, 759–773. doi: 10.1111/cdev.12974
- Deacon, T. (1997). *The Symbolic Species: The Co-evolution of Language and the Brain*. New York, NY: Norton & Co.
- Deacon, T. (2010). A role for relaxed selection in the evolution of the language capacity. *Proc. Natl. Acad. Sci. U.S.A.* 2010, 9000–9006. doi: 10.1073/pnas.0914624107
- Delgado, R. E., Buder, E. H., and Oller, D. K. (2010). *AACT (Action Analysis Coding and Training)*. Miami, FL: Intelligent Hearing Systems.
- Dixon, S., Tronick, E., Keefer, C., and Brazelton, T. B. (1981). “Mother–infant interaction among the Gusii of Kenya,” in *Culture and Early Interactions*, eds T. Field, A. M. Sostek, P. Vietze, and P. H. Leiderman, (Hillsdale, NJ: Lawrence Erlbaum).
- Dominguez, S., Devouche, E., Apter, G., and Gratier, M. (2016). The roots of turn-taking in the neonatal period. *Infant Child Dev.* 25, 240–255. doi: 10.1002/icd.1976
- Edgington, E. (1969). Randomization tests. *J. Psychol.* 57, 445–449.
- Farran, L. K., Lee, C. C., Yoo, H., and Oller, D. K. (2016). Cross-cultural register differences in infant-directed speech: an initial study. *PLoS One* 11:e0151518. doi: 10.1371/journal.pone.0151518
- Feldman, R. (2007). Parent–infant synchrony: biological foundations and developmental outcomes. *Curr. Dir. Psychol. Sci.* 16, 340–345. doi: 10.1111/j.1467-8721.2007.00532.x
- Fernald, A., Taeschner, T., Dunn, J., Papoušek, M., De Boysson-Bardies, B., and Fukui, I. (1989). A cross-language study of prosodic modifications in mothers’ and fathers’ speech to preverbal infants. *J. Child Lang.* 16, 477–501. doi: 10.1017/s0305000900010679
- Fogel, A. (1977). “Temporal organization in mother–infant, face-to-face interaction,” in *Studies in Mother–Infant Interaction*, ed. H. R. Schaffer, (London: Academic Press), 119–151.
- Fraiberg, S. (1979). “Blind infants and their mothers: an examination of the sign system,” in *Before Speech: The Beginning of Interpersonal Communication*, ed. M. Bullowa, (Cambridge: Cambridge University Press), 147–169.
- Franklin, B., Warlaumont, A. S., Messinger, D., Bene, E., Iyer, S. N., Lee, C. C., et al. (2013). Effects of parental interaction on infant vocalization rate, variability and vocal type. *Lang. Learn. Dev.* 10, 279–296. doi: 10.1080/15475441.2013.849176
- Gadbury, G. (2001). Randomization inference and bias of standard errors. *Am. Stat.* 55, 310–313. doi: 10.1198/000313001753272268
- Ginsburg, G. P., and Kilbourne, B. K. (1988). Emergence of vocal alternation in mother–infant interchanges. *J. Child Lang.* 15, 221–235. doi: 10.1017/s0305000900012344
- Goldstein, M. H., King, A. P., and West, M. J. (2003). Social interaction shapes babbling: testing parallels between birdsong and speech. *Proc. Natl. Acad. Sci. U.S.A.* 100, 8030–8035. doi: 10.1073/pnas.1332441100
- Goldstein, M. H., and Schwade, J. A. (2008). Social feedback to infants’ babblings facilitates rapid phonological learning. *Psychol. Sci.* 19, 515–523. doi: 10.1111/j.1467-9280.2008.02117.x
- Goldstein, M. H., Schwade, J. A., and Bornstein, M. H. (2009). The value of vocalizing: five-month-old infants associate their own noncry vocalizations with responses from caregivers. *Child Dev.* 80, 636–644. doi: 10.1111/j.1467-8624.2009.01287.x
- Goodnow, J., Cashmore, J., Cotton, S., and Knight, R. (1984). Mothers’ developmental timetables in two cultural groups. *Int. J. Psychol.* 19, 193–205. doi: 10.1080/00207598408247526
- Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., and Parlato-Oliveira, E. (2015). Early development of turn-taking in vocal interaction between mothers, and infants. *Front. Psychol.* 6:1167. doi: 10.3389/fpsyg.2015.01167
- Gros-Louis, J., West, M. J., Goldstein, M. H., and King, A. P. (2006). Mothers provide differential feedback to infants’ prelinguistic sounds. *Int. J. Behav. Dev.* 30, 509–516. doi: 10.1177/0165025406071914
- Gros-Louis, J., West, M. J., and King, A. P. (2014). Maternal responsiveness and the development of directed vocalizing in social interactions. *Infancy* 19, 385–408. doi: 10.1111/desc.12641
- Hilbrink, E. E., Gattis, M., and Levinson, S. C. (2015). Early developmental changes in the timing of turn-taking: a longitudinal study of mother–infant interaction. *Front. Psychol.* 6:1492. doi: 10.3389/fpsyg.2015.01492
- Hoff, E. (2003). The specificity of environmental influence: socioeconomic status affects early vocabulary development via maternal speech. *Child Dev.* 74, 1368–1378. doi: 10.1111/1467-8624.00612
- Holes, C. (2004). *Modern Arabic: Structures, Functions, and Varieties*. Washington, DC: Georgetown University Press.
- Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., and Jasnow, M. D. (2001). Rhythms of dialogue in infancy: coordinated timing in development. *Monogr. Soc. Res. Child Dev.* 66, 1–132.
- Jain, A., Bansal, R., Kumar, A., and Singh, K. D. (2015). A comparative study of visual and auditory reaction times on the basis of gender and physical activity levels of medical first year students. *Int. J. Appl. Basic Med. Res.* 5, 124–127. doi: 10.4103/2229-516X.157168
- Jasnow, M., and Feldstein, S. (1986). Adult-like temporal characteristics of mother–infant vocal interactions. *Child Dev.* 57, 754–761. doi: 10.2307/1130352
- Kager, R. (2009). “Stress,” in *The Encyclopedia of Arabic Language and Linguistics*, ed. K. Versteegh, (Leiden: E. J. Brill), 344–353.
- Kärtner, J., Keller, H., and Yovsi, R. D. (2010). Mother–infant interaction during the first 3 months: the emergence of culture-specific contingency patterns. *Child Dev.* 81, 540–554. doi: 10.1111/j.1467-8624.2009.01414.x
- Kaye, K., and Fogel, A. (1980). The temporal structure of face-to-face communication between mothers and infants. *Dev. Psychol.* 16, 454–464. doi: 10.1037//0012-1649.16.5.454
- Keller, H., Abels, M., Borke, J., Lamm, B., Su, Y., Wang, Y., et al. (2007). Socialization environments of Chinese and Euro-American middle-class babies: parenting behaviors, verbal discourses and ethnotheories. *Int. J. Behav. Dev.* 3, 210–217. doi: 10.1177/0165025407074633
- Keller, H., Lohaus, A., Volker, S., Cappenberg, M., and Chasiotis, A. (1999). Temporal contingency as an independent component of parenting behavior. *Child Dev.* 70, 474–485. doi: 10.1111/1467-8624.00034
- Kemphorne, O. (1955). The randomization theory of experimental inference. *J. Am. Stat. Assoc.* 50, 946–967. doi: 10.1080/01621459.1955.10501979
- Kenstowicz, M., and Abdul-Karim, K. (1980). Cyclic stress in levantine arabic. *Stud. Linguist. Sci.* 10, 55–76.
- Kiparsky, P. (2003). “Syllables and moras in Arabic,” in *The Syllable in Optimality Theory*, eds C. Féry, and R. V. D. Vijver, (Cambridge: Cambridge University Press), 147–182. doi: 10.1017/cbo9780511497926.007
- Ko, E.-S., Seidl, A., Cristia, A., Reimchen, M., and Soderstrom, M. (2015). Entrainment of prosody in the interaction of mothers with their young children. *J. Child Lang.* 43, 284–309. doi: 10.1017/S0305000915000203
- Kondaurova, M., Fagan, M., Dille, L., and Bergeson-Dana, T. (2016). Vocal matching in interaction between mothers and their normal-hearing and hearing-impaired twins. *Proc. Meet. Acoust.* 26:060007. doi: 10.1121/2.000373
- Lavelli, M., and Fogel, A. (2013). Interdyad differences in early mother–infant face-to-face communication: real-time dynamics and developmental pathways. *Dev. Psychol.* 49, 2257–2271. doi: 10.1037/a0032268
- Le Vine, E. V. M. (2004). “Challenging expert knowledge: findings from an African study of infant care and development,” in *Advances in Applied Developmental Psychology Childhood and Adolescence: Cross-Cultural Perspectives and Applications*, eds U. P. Gielen, and J. Roopnarine (Westport, CT: Greenwood Publishing Group), 149–165.
- Leezenbaum, N. B., Campbell, S. B., Butler, D. M., and Iverson, J. M. (2014). Maternal verbal responses to communication of infants at low and heightened risk for autism. *Autism* 18, 694–703. doi: 10.1177/1362361313491327

- Levinson, S. C. (2016). Turn-taking in human communication, origins, and implications for language processing. *Trends Cogn. Sci.* 20, 6–14. doi: 10.1016/j.tics.2015.10.010
- Lieven, E. V. M. (1994). “Cross-linguistic and crosscultural aspects of language addressed to children,” in *Input and Interaction in Language Acquisition*, eds C. Gallaway, and B. J. Richards, (Cambridge: Cambridge University Press), 56–73. doi: 10.1017/cbo9780511620690.005
- Loveland, K. (1991). Social affordances and interaction II: autism and the affordances of the human environment. *Ecol. Psychol.* 3, 99–119. doi: 10.1207/s15326969eco0302_3
- Lupyan, G., and Dale, R. (2010). Language structure is partly determined by social structure. *PLoS One* 5:e8559. doi: 10.1371/journal.pone.0008559
- Lynch, M., Oller, D., Steffens, M., and Buder, E. (1995). Phrasing in prelinguistic vocalizations. *Dev. Psychobiol.* 28, 3–25. doi: 10.1002/dev.420280103
- McCarthy, J. J. (1980). A note on the accentuation of Damascene Arabic. *Stud. Linguist. Sci.* 10, 77–98.
- Milenkovic, P. (2015). *TF32 (Computer Software)*. Madison, WI: University of Wisconsin-Madison.
- Northrup, J. B., and 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, 523–547. doi: 10.1111/inf.12090
- Ochs, E., and Schieffelin, B. (1984). “Language acquisition and socialization: three developmental stories and their implications,” in *Culture Theory: Essays on Mind, Self and Emotion*, eds R. Shweder, and R. LeVine, (Cambridge, MA: Cambridge University Press), 276–320.
- Odling-Smee, F. J. (1988). “Niche constructing phenotypes,” in *The Role of Behavior in Evolution*, ed. H. C. Plotkin, (Cambridge, MA: MIT Press), 73–132.
- Oller, D. K. (2000). *The Emergence of the Speech Capacity*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Oller, D. K., Griebel, U., Iyer, S. N., Jhang, Y., Warlaumont, A. S., Dale, R., and Call, J. (2019). Language origins viewed in spontaneous and interactive vocal rates of human and bonobo infants. *Front. Psychol.* 10:729. doi: 10.3389/fpsyg.2019.00729
- Osterling, J., and Dawson, G. (1994). Early recognition of children with autism: a study of first birthday home videotapes. *J. Autism Dev. Disord.* 24, 247–257. doi: 10.1007/bf02172225
- Papoušek, M., and Papoušek, H. (1989). Forms and functions of vocal matching in interactions between mothers and their precanonical infants. *First Lang.* 9, 137–157. doi: 10.1177/014272378900900603
- Richman, A. L., Miller, P. M., and LeVine, R. A. (1992). Cultural and educational variations in maternal responsiveness. *Dev. Psychol.* 28, 614–621 doi: 10.1037/0012-1649.28.4.614
- Romano, J. (1990). On the behavior of randomization tests without a group invariance assumption. *J. Am. Stat. Assoc.* 85, 686–692. doi: 10.1080/01621459.1990.10474928
- Saiegh-Haddad, E., and Henkin-Roitfarb, R. (2014). “The structure of Arabic language and orthography,” in *Handbook of Arabic Literacy: Insights and Perspectives*, eds E. Saiegh-Haddad, and M. Joshi, (Dordrecht: Springer), 3–28. doi: 10.1007/978-94-017-8545-7_1
- Schieffelin, B. B., and Ochs, E. (eds) (1986). *Language Socialization Across Cultures*. New York, NY: Cambridge University Press.
- Selkirk, E. (1980). The role of prosodic categories in English word stress. *Linguist. Inq.* 11, 563–605.
- Smith, N., and McMurray, B. (2018). Temporal responsiveness in mother–child dialogue: a longitudinal analysis of children with normal hearing and hearing loss. *Infancy* 23, 410–431. doi: 10.1111/inf.12227
- Snow, R. E. (1977). Individual differences and instructional theory. *Educ. Res.* 6, 11–15. doi: 10.3102/0013189x006010011
- Stern, D. N., Jaffe, J., Beebe, B., and Bennett, S. L. (1975). Vocalizing in unison and in alternation: two modes of communication within the mother–infant dyad. *Annu. N. Y. Acad. Sci.* 263, 89–100. doi: 10.1111/j.1749-6632.1975.tb41574.x
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10587–10592. doi: 10.1073/pnas.0903616106
- Sung, J., Fausto-Sterling, A., Garcia Coll, C., and Seifer, R. (2013). The dynamics of age and sex in the development of mother–infant vocal communication between 3 and 11 months. *Infancy* 18, 1135–1158. doi: 10.1111/inf.12019
- Takahashi, D. Y., Narayanan, D. Z., and Ghazanfar, A. A. (2013). Coupled oscillator dynamics of vocal turn-taking in monkeys. *Curr. Biol.* 23, 2162–2168. doi: 10.1016/j.cub.2013.09.005
- Thelen, E., and Smith, L. B. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: Bradford Books/MIT Press.
- Trevarthen, C. (1977). “Descriptive analyses of infant communicative behavior,” in *Studies in Mother-Infant Interaction*, ed. H. R. Schaffer, (London: Academic Press.), 227–270.
- Trevarthen, C. (1979). “Communication and cooperation in early infancy. A description of primary intersubjectivity,” in *Before Speech: The Beginnings of Interpersonal Communication*, ed. M. Bullowa, (Cambridge: Cambridge University Press), 321–347.
- Trevarthen, C. (2001). Infant intersubjectivity: research, theory, and clinical applications. *J. Child Psychol. Psychiatry* 42, 3–48.
- Van Egeren, L., Barratt, M., and Roach, M. (2001). Mother–infant responsiveness: timing, mutual regulation, and interactional context. *Dev. Psychol.* 37, 684–697.
- Warlaumont, A. S., Oller, D. K., Dale, R., Richards, J. A., Gilkerson, J., and Xu, D. (2010). “Vocal interaction dynamics of children with and without autism,” in *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, eds S. Ohlsson, and R. Catrambone, (Austin, TX: Cognitive Science Society).
- Warlaumont, A. S., Richards, J. A., Gilkerson, J., and Oller, D. K. (2014). A social feedback loop for speech development and its reduction in autism. *Psychol. Sci.* 25, 1314–1324. doi: 10.1177/0956797614531023
- Watson, J. C. E. (2011). “Word Stress in Arabic,” in *Blackwell Companion to Phonology*, eds M. V. Oostendorp, C. Ewen, E. Hume, and K. Rice, (Hoboken, NJ: Wiley), 2990–3019.
- West, M. J., King, A. P., and Arberg, A. A. (1988). “The inheritance of niches: the role of ecological legacies in ontogeny,” in *Developmental Psychobiology and Behavioral Ecology*, ed. E. M. Blass, (New York, NY: Plenum Press), 41–62.
- Whaley, S. E., Sigman, M., Beckwith, L., Cohen, S. E., and Espinosa, M. P. (2002). Infant–caregiver interaction in Kenya and the United States: the importance of multiple caregivers and adequate comparison samples. *J. Cross Cult. Psychol.* 33, 236–247.
- Yale, M. E., Messinger, D., and Delgado, C. F. (2003). The temporal coordination off early infant communication. *Dev. Psychol.* 39, 815–824.
- Yoo, H., Bowman, D., and Oller, D. K. (2018). The origin of protoconversation: an examination of caregiver responses to cry and speech-like vocalizations. *Front. Psychol.* 9:1510. doi: 10.3389/fpsyg.2018.01510

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

Copyright © 2019 Farran, Yoo, Lee, Bowman and Oller. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.