J. Child Lang. **43** (2016), 284–309. © Cambridge University Press 2015
doi:10.1017/S0305000915000203

Entrainment of prosody in the interaction of mothers with their young children*

EON-SUK KO
Seoul National University

AMANDA SEIDL
Purdue University

ALEJANDRINA CRISTIA
Centre National de la Recherche Scientifique

MELISSA REIMCHEN
University of Manitoba

AND

MELANIE SODERSTROM
University of Manitoba

(*Received* 16 *June* 2014 – *Revised* 28 *January* 2015 – *Accepted* 7 *April* 2015 – *First published online* 3 *June* 2015)

ABSTRACT

Caregiver speech is not a static collection of utterances, but occurs in CONVERSATIONAL EXCHANGES, in which caregiver and child dynamically influence each other's speech. We investigate (a) whether children and caregivers modulate the prosody of their speech as a function of their interlocutor's speech, and (b) the influence of the initiator of the conversation on durational characteristics of the exchange. We analyzed naturalistic conversations from 13 mother–infant/toddler dyads aged

[*] Eon-Suk Ko, Institute for Cognitive Science, Seoul National University; Amanda Seidl, Department of Speech, Language, and Hearing Sciences, Purdue University; Alejandrina Cristia, Laboratoire de Science Cognitives et Psycholinguistiques, Centre National de la Recherche Scientifique; Melissa Reimchen, Department of Psychology, University of Manitoba; Melanie Soderstrom, Department of Psychology, University of Manitoba. This research was supported by the National Research Foundation of Korea Grant NRF-2014S1A5B5A02014474 to EK, and a SSHRC grant 430-2011-0459 to MS. AC acknowledges the institutional support of ANR-10-LABX-0087 and ANR-10-IDEX-0001-02. Address for correspondence: Melanie Soderstrom, Department of Psychology, University of Manitoba, Winnipeg, MB R3 T 2N2. e-mail: M_Soderstrom@umanitoba.ca

12–30 months across full-day recordings of 3–5 days per dyad using LENA and automated analytic tools. We found small, but significant, effects of mothers and their children influencing each other's speech, particularly in pitch measures. We also found longer utterances and shorter response latencies for the initiator of a conversation. While mothers show more mature conversational capabilities (more entrainment, shorter response latencies), our findings converge with prior research to highlight the active role of young children in the conversational exchange.

INTRODUCTION

Many studies suggest that both the amount of caregiver speech (e.g. Hart & Risley, 1995; Hoff & Naigles, 2002; Zimmerman *et al.*, 2009) and the quality of caregiver speech (e.g. Cristia, 2011; Hurtado, Marchman & Fernald, 2008) impact child language acquisition. However, caregiver speech is not a static collection of utterances, passively received by the child. Quite to the contrary, much – if not all – infant-directed speech occurs in the context of CONVERSATIONAL EXCHANGE (Snow, 1977), in which caregiver and child dynamically influence each other's speech. Here, we investigate the possibility that children and caregivers modulate the prosodic characteristics of their speech as a function of their interlocutor's speech patterns, as well as who initiates the conversation.

Examining to what extent individual dyads, as well as children and caregivers in general, are sensitive to one another in running speech is crucial to understanding the nature of these linguistic interactions, and may provide important insights into how language development occurs in both typical and atypical contexts. These insights may emerge on two levels. First, linguistic responsiveness may reflect social sensitivity to other individuals in a global sense, such that more responsive dyads may have higher levels of attachment, which may have a positive impact on language development. In support of this idea, there is evidence that dyads whose speech is optimally correlated in terms of timing may have more secure attachment and the children in such dyads may have better developmental outcomes (Jaffe, Beebe, Feldstein, Crown & Jasnow, 2001). On the other hand, other researchers have found differences between maternal responsiveness in the sense of social contingency and affect communication, and the effects of speech timing per se (Striano, Henning & Stahl, 2006). Second, such responsiveness may have a direct impact on learning mechanisms, such that infants from more responsive dyads may acquire mature linguistic forms more rapidly. This form of learning has been demonstrated experimentally in dyads with infants as young as 9 months of age for characteristics of phonological form (Goldstein & Schwade, 2008),

but has yet to be examined with respect to suprasegmental linguistic characteristics, and/or in naturalistic interactions.

The mechanism responsible for this coordination of speech between mother and child has been called variously 'entrainment' (e.g. Brennan, Galati & Kuhlen, 2010), 'alignment' (e.g. Pickering & Garrod, 2004), or 'imitation' (e.g. Meltzoff & Moore, 1977). These terms have been used in a variety of different ways by researchers, with different implications. Probably the most widely used term in child development is 'imitation'. This term generally has the connotation of a non-reciprocal influence whereby one party (typically the adult) performs an action and another (typically the child) imperfectly performs the same action. The first party's actions therefore influence how the second party's actions are performed, leading to learning over time. In this paper, we take the position that such influences in conversations are, or may be, reciprocal in nature. We therefore use the term ENTRAINMENT (more commonly used in adult–adult interactions) to broadly refer to the phenomenon whereby interlocutors engaged in a conversation adapt their speech patterns in accordance with the interlocutor's speech. This may therefore be an important mechanism leading to learning on the part of the child, but we do not rule out the possibility that the child is also influencing the caregiver's behavior. Our study focuses on two specific and separable aspects of entrainment: SIMILARITY (overall similarity across a sample) and CONVERGENCE (becoming more similar over time). Similarity documents a static relationship that may indicate alikeness due to convergence that has already occurred, or convergence that occurs very quickly or instantaneously. Convergence, on the other hand, documents effects of entrainment as it unfolds within a timeline. We focus our investigation of entrainment on the following acoustic variables: timing (utterance duration and inter-speaker silences), pitch measures (mean, minimum, and maximum pitch and pitch range across the utterance), and speaking rate.

There has been considerable research on the development of conversational timing in young infants. Infants are sensitive to appropriate timing and disprefer interactions in which a significant delay is introduced (Striano *et al*., 2006). At least some aspects of this timing develop in a non-linear fashion, with more overlapping speech occurring in interactions with younger and older infants, but fewer during the period of the emergence of meaningful language, during the first half of the second year of life (Elias & Broerse, 1996). Both mothers and infants appear to influence the duration of overlapping speech and pausing behavior (Feldstein *et al*., 1993), and there may be an important relationship between this ability and later cognitive and language development (e.g. Jaffe *et al*., 2001). One study (Beebe, Alson, Jaffe, Feldstein & Crown, 1988) found a correlation across mother–infant dyads in the duration of

response times at turns, but not of within-speaker pauses. Both this study and another (Shimura & Yamanoucho, 1992) found no correlation in the durations of infant and mother utterances.

Conversational timing is implicated as a key universal feature of human interaction. There are cross-cultural similarities in the qualitative features of turn-taking timing, such as a unimodal distribution of response timing with a peak between 0–200 ms after the interlocutor's offset, and longer response times for disconforming utterances, which are modulated by quantitative differences across languages in the length of mean response times (Stivers *et al.*, 2009). Nonetheless, conversational timing is clearly a domain that requires learning on the part of the infant. It is therefore perhaps unsurprising that durational measures that capture more dynamic, interactional components of the conversation appear to be more sensitive to entrainment effects.

With respect to pitch, Siegel, Cooper, Morgan, and Brenneise-Sarshad (1990) found no effect of interlocutor (mother or father) on measures of the mean pitch (f0) of utterances produced by 9–12-month-old infants. Similarly, McRoberts and Best (1997), examining a single infant longitudinally from 3 to 17 months, found adjustment in the form of higher pitch by the mother and father when speaking with their infant (as would be expected either because of general characteristics of infant-directed speech or because of an entrainment effect) but no difference in the infant's mean pitch depending on whether the infant was alone, with mother, or with father across the age range. Shimura and Yamanoucho (1992) found correlations both between and within mother–infant dyads with respect to mean pitch for 2- to 3-month-olds, but did not attempt to disentangle mother–infant or infant–mother effects. One study, Masataka (1992), found that 3- to 4-month-old infants showed significantly more similarity to their mother in the intonation contours of their vocalizations in response to exaggerated intonation on the part of their mother. However, such response similarity was not seen with less exaggerated contours.

The different results found in the studies reported above are difficult to interpret, as the studies differ on a variety of variables, including age of the infants, language being acquired (Japanese versus English, two languages that differ radically in both linguistic and cultural factors), specific variables being analyzed, and the form of analysis (correlations by participant or by utterance, or more sophisticated statistical analyses). Additionally, many of these studies have necessarily been limited to very small samples of speech, given the labor involved in coding such speech samples. While it has been argued that such small slices produce convergence estimates that are reliable and meaningful for timing (e.g. Jaffe & Feldstein, 1970), the same has not been shown for pitch.

Another limitation of much of this research is that correlations BETWEEN dyads are interpreted as evidence for entrainment effects. However, mothers and infants share much in common going into a conversation, and this prior resemblance unrelated to entrainment may drive correlations across dyads. More fine-grained analysis within a given dyad allows us to more directly examine the extent to which mother and infant influence each other dynamically over the course of a conversational exchange, and to tease apart whether the broad-level correlations found to date across dyads are capturing true entrainment, or are simply the result of pre-existing resemblance between mother and child.

The Language ENvironment Analysis, or LENA, provides us with the unique opportunity to examine mother–child interactions in a large-scale, naturalistic manner (Zimmerman *et al*., 2009). The LENA Research Foundation has developed a small recorder that is capable of storing up to 16 hours of running speech, and which can be worn easily by the child when fitted in the pocket of a custom-made vest, thus capturing the child's production as well as her interlocutors' input over the whole day. Additionally, LENA has developed software that pre-processes the recordings to divide up stretches of recordings into utterances (versus silence or noise), and labels the detected utterances as a function of speaker, based on basic acoustic properties (child versus adult female, among others). In the present study, we used LENA software together with our own scripts to estimate child–caregiver entrainment in terms of both timing (specifically, utterance duration, speech rate, and response time) and pitch characteristics (pitch mean, maximum, minimum, and range).

In our study we collected multiple full-day, naturalistic recordings using LENA, of thirteen monolingual English-learning infants and toddlers aged 12 to 30 months going about their typical day. These recordings were then processed in the laboratory to address the following research questions: First, does entrainment in the form of similarity occur in mother–child conversation? To address this question, we conducted Pearson correlation tests for the speech of mother and child pairs at the level of conversational block (a unit referring to short-term conversational bouts separated by pausing, described in more detail below), and conversational turns (i.e. where a child utterance follows a maternal utterance or vice versa). Similarity in conversational blocks would indicate an adaptation of speech in accordance with the interlocutor's speech patterns within the context of a particular conversation, over and above broad level similarity between a mother–infant/toddler dyad. Turn-level similarity would indicate a more immediate adaptation of speech to match the interlocutor's token-level speech patterns. In adults' speech, there is a report that a tighter similarity exists between speakers at turns compared

to the session level (Levitan & Hirschberg, 2011), and at the conversational block than the session level (Edlund, Heldner & Hirschberg, 2009). A comparison of the correlation coefficients at the turn and block level will shed light on the extent to which the influence of the dyad members on each other varies at different levels of analysis.

Second, we asked whether there were any effects of interlocutor order (i.e. who initiates the conversation and who responds) on the nature and extent of responding. To our knowledge, ours is the first study to examine this potentially important aspect of the conversational dynamic with respect to prosodic influences. We address this question by considering the initiator or the respondent at both the conversational turn and block level. At the conversational turn level, we compare the correlation coefficients in mother-to-child turns with those in child-to-mother turns. Differences between these two analyses would help to tease apart the directionality of mother–child entrainment (i.e. whether mothers adjust more to infants or vice versa). At the level of the conversational block, we examine whether initiating a conversational block has any effects on the temporal pattern, i.e. response time and utterance length, of the speaker as the owner of the block.

Finally, does the child's and the mother's speech converge over the course of a conversational block? In studies of adults' convergence, conversations analyzed typically take place between strangers who need time to get familiarized with each other's speech patterns. Thus convergence usually occurs late in the course of the conversation. Due to the existing familiarity between mother and child, such conversational convergence may be reduced or absent entirely. On the other hand, since infants are immature conversationalists, they may require time to warm up with respect to the dynamics of a particular conversational interchange, and hence may show greater similarity at the end than at the beginning of a conversational block.

METHOD

Participants and recording procedure

Recordings were collected as part of a larger project examining the linguistic environment of infants/toddlers across childcare settings (Soderstrom & Wittebolle, 2013). For the present study, all samples were day-long recordings of infants and toddlers being cared for in the home, primarily by their own mothers.

A total of fourteen mother–child pairs participated in the study, and each pair recorded 3 to 5 days (mean = 3·85, SD = 0·99) of their daily life. However, one of these dyads was excluded from analysis because the father was the primary caregiver, and therefore a much smaller proportion

of the speech segments were maternal speech, raising the baseline error rate of maternal speech segment identification to a much higher proportion of total utterances. The mean duration of the interval between two subsequent recordings was 16 days (SD = 19). The final recording sample consisted of 517·27 hours of recording from a total of 50 days. Due to a technical issue, two adjacent recording days from one participant were treated as a single 'day' in the analysis, so our final recording sample consists of 49 'days'. Recordings ranged from 6·56 to 13·96 hours with a mean of 10·00 hours (SD = 1·84). The total hours of recording from each child ranged from 29 hours to 51 (mean = 40·0, SD = 10·1). Children's age ranged from 12 months to 30 months (mean = 20·4 months, SD = 4·5), with nine male and four female infants/toddlers participating. See Table 1 for participant-specific data.

Participants were recruited from an existing database of families in Winnipeg, Canada, who had expressed an interest in participating in research studies with their child. If verbal interest in the study was expressed by the caregiver over the phone, a research assistant visited the home. During this initial visit, written informed consent was obtained, and the caregivers were given instructions regarding the use of the LENA recording device. After each day of recording, a research assistant would visit the home to collect the recording device in order to download the recording for processing and provide another device for the next recording. Participants were provided a log sheet to note the number of people in the room and the child's activities throughout the day. Participants were compensated at $20/day for the recordings.

Technical descriptions of the data and pre-processing

Day-long recordings were annotated using the unsupervised algorithms included in the software accompanying LENA, which tags stretches of time as being speech; or else containing speaker overlap, noise, or silence. Thus, a SEGMENT is an individual chunk of time associated with a particular category of speech or acoustic input by the LENA system, such as ADULT FEMALE UTTERANCE or NON-VERBAL NOISE. A detailed description of the LENA system processing may be found in the LENA technical reports (<http://www.lenafoundation.org/customer-resources/technical-reports/>). For our purposes, a speech segment may be considered roughly equivalent to an UTTERANCE, and we will use these terms interchangeably. The speech segments we analyzed within the study were female adult (FAN in LENA's coding system, assumed to be the mother) and the child being recorded (CHN, identified separately from other children in the environment by LENA directly). Speech segments are divided into conversational blocks by LENA, such that a conversational block continues until the gap

TABLE 1. *Demographic and recording details by participant. In the 'age range' row, the age in months of the first and last recording is listed. The bottom row refers to the total number of segments where the speech constitutes more than half of the segment duration (only these segments were included in the analyses).*

Child	A	B	C	D	E	F	G	H	I	J	K	L	M
Gender	F	M	M	F	F	F	M	M	F	M	M	M	M
Age range (months)	12–13	12–13	16–18	20–21	21–21	20–22	22–22	22–22	21–23	22–24	22–25	23–25	29–30
# of days recorded	3	5	5	3	3	3	3	5	5	5	3	3	3
Total N hours	37·47	44·46	58·10	48·70	29·67	28·19	33·42	41·13	52·07	47·51	32·75	39·57	24·23
Total N mother and child segments	10448	13288	26646	12565	9839	8351	17567	10888	9727	26735	12887	21978	6414
Included mother and child segments	3384	4207	11078	4977	4534	2371	8865	4397	1523	14368	5497	11990	1422

between two subsequent speech segments is more than 5 seconds, at which point a new conversational block is declared. Gaps can consist of silence or non-verbal noise. For our purposes, we focused on mother-initiated and child-initiated conversational blocks (AICF and CIC in LENA's coding system), from which the adult female and child segments were selected. The total number of segments spoken by the adult female and the child were similar (collapsing across all days: female adult = 39,745, target child = 38,868). We also analyzed a subset of the data (34,992 segments, 45% of the total) in which it was indicated that ONLY the mother and child were present, based on observational notes taken by the mother at the time of the recording, such that any adult female as labeled by LENA should in fact be the mother. In our initial raw acoustic measures (not reported) we found higher mean pitch in the total dataset compared to the restricted set for both mother and child, but no other basic acoustic differences. Crucially, the overall pattern in the relationship between the two interlocutors was similar. We therefore report only the findings for the larger dataset, and refer to the adult female as the mother for simplicity.

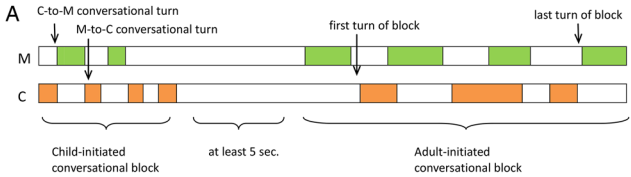
The LENA software provides an estimation of vocalization duration directly from the speaker segment, which may include small within-talker pauses and non-verbal vocalizations such as vegetative noises and crying. LENA makes a first-pass assumption that adult speech will have a minimum length of 1 second, while child speech will have a minimum length of 0·6 s (and 0·8 s for certain other segments). Utterances shorter than these minimums will have their boundaries extended by LENA's segmentation process. Therefore, durations at these minima are artificially inflated. Segments containing less than 50% vocalization (i.e. greater than 50% silence or non-verbal production like crying or laughing) were excluded from our analyses.

Finally, we used custom-written Praat scripts (Boersma & Weenink, 2013) to measure pitch and speaking rate in the dataset gathered using LENA. Pitch was expressed as ERBs (equivalent rectangular band width) since its psycho-acoustic scale provides perceptually relevant quanta (Hermes & van Gestel, 1991). Speaking rate was expressed as number of syllables per second, which was calculated based on the combination of intensity peaks and voicing in each syllable (De Jong & Wempe, 2009). We used a single set of parameters for all participants. Figure 1A schematically demonstrates the definition of segments, conversational turns, and blocks, and Figure 1B shows an example of how acoustic values extracted from the defined turns and blocks are organized into a data frame for a single conversational block combining LENA tagging and Praat measurements.

We processed the data frame further using custom scripts written in R (R Core Team, 2013) to extract the information specific to each of our research questions. For example, a segment was tagged as having constituted a

https://doi.org/10.1017/S0305000915000203 Published online by Cambridge University Press

A



B

	Child ID	date	speaker	begin	end	block number	block type	Duration	speaking rate	pitch mean	pitch min	pitch max	pitch range
1	C003	090708	mother	66.8	67.81	4	Child-init	1.01	1.98	7.16	4.13	9.63	5.51
2	C003	090708	mother	68.61	69.61	4	Child-init	1	1	5.56	3.70	7.18	3.47
3	C003	090708	child	69.61	70.31	4	Child-init	0.7	1.43	6.99	6.37	7.44	1.06
4	C003	090708	mother	70.31	72.03	4	Child-init	1.72	2.33	6.51	5.16	8.56	3.40

Fig. 1A (top). Structure of conversational turns and conversational blocks (M = Mother, C = Child). Note that intervening segments with labels other than mother and child were filtered out. Fig. 1B (bottom). Example of the data frame resulting from the application of custom scripts to the data frame generated by the LENA system.

conversational turn if there was a talker change between that segment and the previous one. We then extracted acoustic values for the previous and following segments at each turn, as well as its linear location within the conversational block. Based on this information, we also extracted the first and last turn of the block to investigate the pattern of convergence or the lack of convergence. More information about these steps of analysis is provided in the 'Results' section, where the investigation of each particular research question is described in detail (see Figure 2).

Reliability test

For reliability purposes, we manually coded 100 segments for each dyad (50 from the mother and 50 from the child). Reliability was performed by volunteer research assistants with undergraduate-level linguistics training. Segments were selected randomly by a custom computer script from a single transcript for each dyad. The research assistant first listened to the segment as given by the LENA system, selected which LENA-generated speaker code best represented the segment (e.g. female adult, silence, overlapping speech), and recorded the number of syllables they heard in the segment. They then listened to the segment a second time with a 1 s context buffer on either side of the segment and again selected a LENA code. After this, the research assistant adjusted the start and stop times for the utterance in Praat until they were satisfied that it accurately reflected

https://doi.org/10.1017/S0305000915000203 Published online by Cambridge University Press

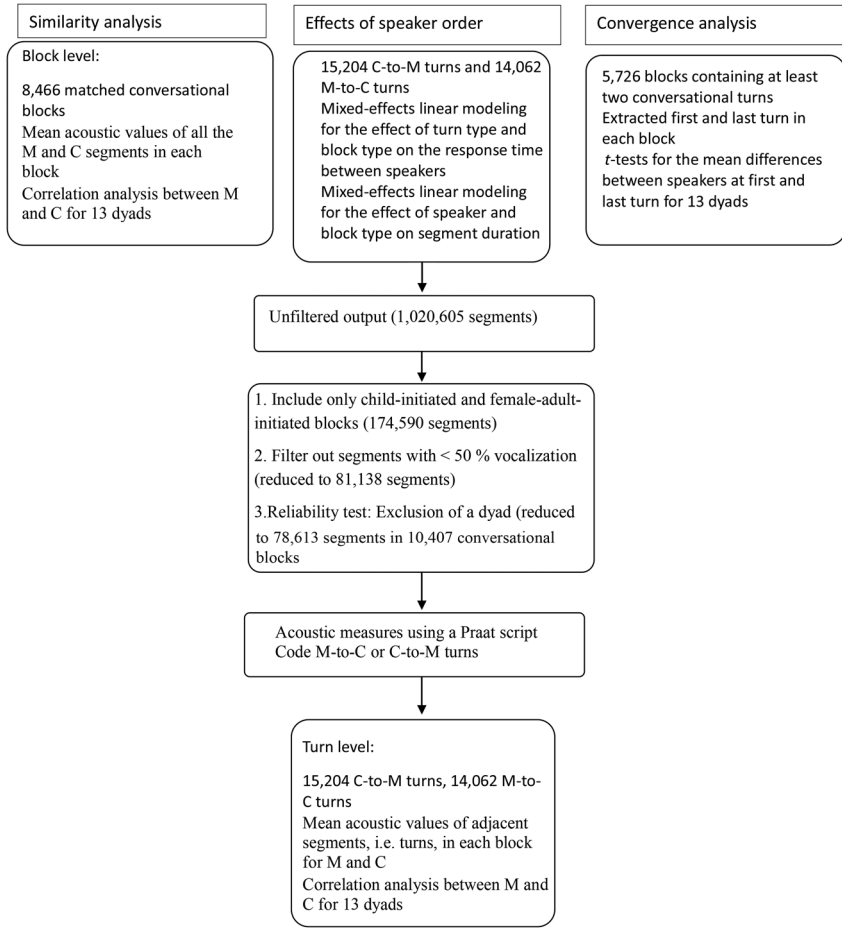


Fig. 2. Flow of data selection and processing.

the real start and stop times of the utterances. If necessary, an additional buffer was added, such that the start or stop time could vary further than 1 s from the original LENA parameter.

The median for the differences in segment duration between the LENA-generated and manual annotations was −90 ms, with the inter-quartile range of Q1: −481 ms, and Q3: 112 ms (see Figure 3). Given that the mean duration of manually annotated segments was 1390 ms, the absolute median difference in the LENA-generated and the manual annotation is around the rate of 6%. This appears to us an acceptable level of accuracy for extracting robust patterns of speech in a large dataset.

https://doi.org/10.1017/S0305000915000203 Published online by Cambridge University Press

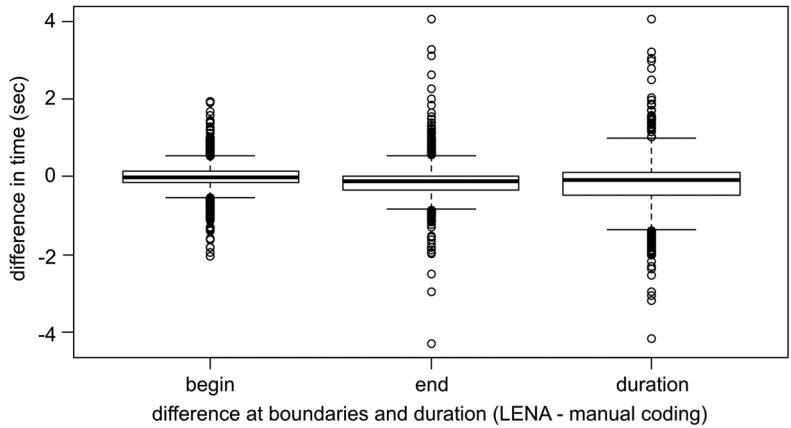


Fig. 3. Differences in LENA and manual annotation at segment boundaries and the duration.

TABLE 2. *Types of LENA's classification errors in labeling speaker ID. For a subset of the segments labeled as CHN (Target Child) or FAN (Female Adult), human annotators manually re-labeled them.*

	Female adult	Unclassified	Male adult	Non-verbal noise	Other child	Target child	TV/ Electronic	Distant speaker	Overlapping
CHN	23	5	3	43	1	616	9	0	0
FAN	556	0	43	17	14	49	13	1	7

We then compared the number of syllables calculated by the Praat script with those generated by the manual coding by applying a Pearson's product-moment correlation test. The result shows that there was a high level of correlation between the two sets of data ($t(1{,}397) = 35{\cdot}9$, $p < {\cdot}001$, $r = 0{\cdot}70$). This level of agreement is more modest than human inter-coder agreement, but within acceptable ranges given the automated nature of the analysis.

Finally, we performed a Pearson's chi-squared test with Yates' continuity correction to evaluate LENA's accuracy of labeling the speaker by comparing the LENA-generated speaker codes with manually annotated ones. The LENA-generated codes are either female adult (FAN) or target child (CHN), whereas manual annotations included as many as nine different categories (see Table 2). The results show that the LENA-generated codes and the manual annotations are not independent from each other ($\chi^2(8,\ N = 1{,}400) = 1{,}045$, $p < {\cdot}001$). The type of classification errors made by the LENA system can be inferred from Table 2, which shows how the segments labeled as female adult and target

https://doi.org/10.1017/S0305000915000203 Published online by Cambridge University Press

child were re-classified by human annotators after listening to the segment in context. The mean proportion of segments whose speaker had been identified correctly was ·84, with a range of ·51 to ·93. After removing the one dyad where the primary caregiver was the father, the mean proportion of correct identification was ·86.

Statistics

Our analyses of entrainment rely on comparing simple acoustic measures of mother and child speech at different levels of analysis – comparisons of overall mean measures across conversational blocks, comparisons of means at turns only across blocks, and comparisons at turns at different time periods within a block. This approach has the advantage that it does not require baseline measures, because we rely on the variance across blocks and turns to control for baseline effects. Additionally, it provides matched or paired datapoints, thus enabling us to use well-described statistical approaches such as correlations and regressions.

We also constructed a linear mixed-effects model where appropriate using the `lme4` package (Bates, Maechler & Bolker, 2013) implemented in R (R Core Team, 2013). As a means to evaluate the statistical significance of the linear model and attain *p*-values, we conducted likelihood ratio tests by comparing nested models. We also report the *t*-values without the degrees of freedom or *p*-values due to the difficulty of determining the right number of degrees of freedom for assessing the *t*-values. Roughly, however, absolute *t*-values of greater than 2 can be interpreted as indicating statistical significance (Baayen, Davidson & Bates, 2008).

RESULTS

Analysis of similarity at the level of conversational turn and conversational block

We first investigated broadly the overall similarity across partners in the measures of segment duration, speaking rate, and pitch by applying Pearson correlation tests. If mother–child dyads have a tendency to coordinate their speech in exchanging conversations, we would expect to see positive correlations in acoustic measures.

The number of conversational blocks contained in each of the 49 days' recording varied between 38 and 623. The number of segments contained within a conversational block ranged from 1 to 254, with a mean of 7·6 segments in a block (SD = 9·5 segments). At the level of conversational block, we took the mean acoustic values for mother and child across the conversational block, including all utterances for mother and child, which reduced the data in each block to two data points, one for each interlocutor. Out of a total of 10,406 blocks, we selected the 8,466 blocks

https://doi.org/10.1017/S0305000915000203 Published online by Cambridge University Press

which contained both the mother's and the child's speech and had valid values in all acoustic variables under investigation. For example, blocks where pitch values could not be measured due to various reasons were eliminated. We then calculated correlation coefficients of the acoustic variables between mother and child across all of their conversational blocks within a single recording day.

We conducted two-tailed one-sample *t*-tests on the correlation coefficients across the session for each dyad, with the null hypothesis that correlations would not be different from zero. This null hypothesis was rejected in all variables except for pitch range (see Table 3, left-hand side). That is, the correlation coefficient was significantly different from zero across the mothers in pitch mean, pitch minimum, and pitch maximum values as well as speaking rate and duration at the block level. However, the mean correlation coefficients were very small, particularly for duration.

For the turn-level analysis, we applied the same statistics to the acoustic values of pairs of segments which constituted conversational turns averaged across the block. This analysis was identical to the preceding one, except that only segments adjacent to turns were included. If several subsequent utterances by the same speaker occurred without the interlocutor taking a turn, these repetitions were excluded. The number of turns in each block ranged from 1 to 76, with a mean of 3·2 turns in a block (SD = 4·0). The rationale for conducting this second analysis was to investigate whether there is a tighter or looser relationship between the two speakers at a more local level in the immediate context of conversational turns, based on the idea that direct feedback between the interlocutors might be masked by including all the utterances in the block.

At the turn level, the hypothesis of a zero correlation coefficient was rejected for the pitch measures (see Table 3, right-hand side), but not for duration or speaking rate.

Finally, we investigated whether there is a tighter correlation as indicated by the magnitude of correlation coefficients at the turn level compared to the block level, and whether there is an effect of age and gender on the correlation coefficients. To answer this question, we constructed a linear mixed-effects model for the correlation coefficients pooled across the 49 sessions with the acoustic variables that came out with significant correlations at both the block and the turn level, i.e. pitch-related measures. The model included the level (block and turn), age (in days), and gender (male, female) as fixed effects and the dyads (13 groups) and the acoustic measures (4 parameters) as random intercepts. Specifically, the formula used was `lmer(coefficients ~ level + age + gender + (1|dyad) + (1|measure), data = level.data, REML = FALSE)`. The `lmer` function was always run with the same REML specification in this study, so we do not provide this specification in the rest of the paper.

https://doi.org/10.1017/S0305000915000203 Published online by Cambridge University Press

TABLE 3. *Results of one-sample t-tests for the mean correlation coefficients at the block level pooled across the forty-nine days of recording from thirteen mother–child dyads*

		One-sample *t*-tests (turn level)	
	One-sample *t*-tests (block level)	M-to-C turns	C-to-M turns
Duration	$t(12) = 2{\cdot}4$, $p = {\cdot}03$*, mean $r = 0{\cdot}04$	$t(12) = 0{\cdot}27$, $p = {\cdot}79$ (n.s.), mean $r = 0{\cdot}003$	$t(12) = 1{\cdot}5$, $p = {\cdot}16$ (n.s.), mean $r = 0{\cdot}02$
Speaking rate	$t(12) = 7{\cdot}4$, $p < {\cdot}001$***, mean $r = 0{\cdot}09$	$t(12) = 2{\cdot}0$, $p = {\cdot}07$ (n.s.), mean $r = 0{\cdot}03$	$t(12) = 1{\cdot}3$, $p = {\cdot}23$ (n.s.), mean $r = 0{\cdot}02$
Pitch mean	$t(12) = 6{\cdot}9$, $p < {\cdot}001$ ***, mean $r = 0{\cdot}12$	$t(12) = 5{\cdot}8$, $p < {\cdot}001$***, mean $r = 0{\cdot}08$	$t(12) = 10{\cdot}6$, $p < {\cdot}001$***, mean $r = 0{\cdot}14$
Pitch minimum	$t(12) = 5{\cdot}0$, $p < {\cdot}001$***, mean $r = 0{\cdot}09$	$t(12) = 7{\cdot}8$, $p < {\cdot}001$***, mean $r = 0{\cdot}12$	$t(12) = 9{\cdot}4$, $p < {\cdot}001$***, mean $r = 0{\cdot}17$
Pitch maximum	$t(12) = 6{\cdot}2$, $p < {\cdot}001$***, mean $r = 0{\cdot}14$	$t(12) = 8{\cdot}4$, $p < {\cdot}001$***, mean $r = 0{\cdot}11$	$t(12) = 12{\cdot}2$, $p < {\cdot}001$***, mean $r = 0{\cdot}18$
Pitch range	$t(12) = 1{\cdot}4$, $p = {\cdot}20$ (n.s.), mean $r = 0{\cdot}02$	$t(12) = 10{\cdot}2$, $p < {\cdot}001$***, mean $r = 0{\cdot}20$	$t(12) = 10{\cdot}7$, $p < {\cdot}001$***, mean $r = 0{\cdot}22$

Likelihood ratio tests comparing the full model with the one without the fixed effect of level, i.e. turn or block, showed that the discourse level of conversational exchange affected correlation coefficients between mother and child speech ($\chi^2(1) = 15{\cdot}5$, $p < {\cdot}001$). We found that the correlation coefficients were significantly greater at the turn than the block level (Estimate = 0·033, SE = 0·008, $t = 4{\cdot}0$), in line with the findings in previous research. We did not find any significant effect of age or gender on the correlation coefficients between mother and child speech.

In sum, there was a significant correlation between the mother's and child's speech for many acoustic measures both at the conversational block and turn levels, particularly in pitch measures, but these correlations were small. We did not find an effect of age or gender on the correlation coefficients of mother–child speech at either of the analysis levels, although the clustering of our infants in the range between 20–25 months may have limited our ability to find an age effect. We found a higher correlation between mother's and child's speech at the turn level than at the block level in pitch-related measures.

Speaker order effects at the turn and block level

In this section, we explore our second question of whether there is any effect of speaker order, i.e. the initiator/respondent of the turn or the block, on the pattern of mother–child acoustic similarity. Although a tendency of greater correlation coefficients is observed at the child-to-mother than the mother-to-child turns in the previous section, this analysis is based on a substantially reduced number of datapoints from a large amount of raw values by averaging them at each of the blocks and then taking one correlation coefficient across each recording day. In this section, we conduct a more targeted investigation of the effects induced by the linear order of the speakers in the conversation by constructing statistical models based on a more fine-grained dataset. In addition to the effects induced by the speaker order at conversational turns, we also investigate if there is any effect on the conversational patterns in mother–child speech depending on who initiated the series of conversational turns in a conversational block, and any interactions between turn type and the initiator of the block.

We derived the more targeted set of data by calculating the correlation coefficients in mother–child speech by turn type for each block. For this calculation, we selected conversational blocks containing at least 3 conversational turns for mother-to-child (M-to-C) and child-to-mother (C-to-M). We then subjected these data directly to a statistical test to compare the magnitude of correlation coefficients depending on turn types.

To investigate the effects of turn type on the strength of the correlation, we constructed a linear mixed-effects model with the correlation coefficient of

https://doi.org/10.1017/S0305000915000203 Published online by Cambridge University Press

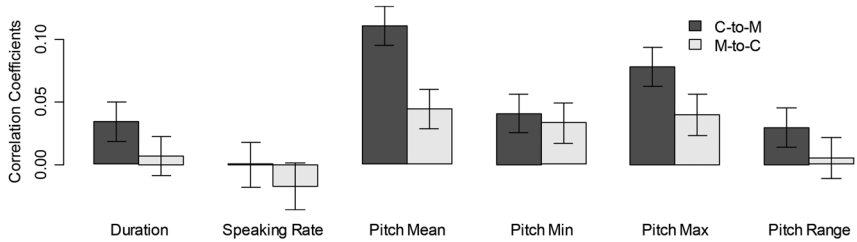


Fig. 4. Effects of turn type (C-to-M and M-to-C turns) on the strength of correlation in conversational turns.

conversational turns in each block as the dependent variable, the turn type (C-to-M and M-to-C) as the fixed effect, and the mother–child dyads (13 groups) and the acoustic measures (6 parameters) as the random factors. The formula used for this modeling was `coefficient ~ turntype+(1|dyad)+(1|measure)`. A likelihood ratio test comparing the full model with the intercepts-only model showed that the turn type has a significant effect on the correlation coefficients ($\chi^2(1) = 11{\cdot}1$, $p < {\cdot}001$), with the turn type C-to-M having a greater correlation coefficient (Estimate = 0·03; SE = 0·009, $t = 3{\cdot}3$) than M-to-C. This result thus indicates that the mother has a tendency to be more accommodating to the speech of her child and increase the similarity of their mean pitch level in their interaction than the child does to the mother (Figure 4).

We next analyzed the effects of turn type on each speaker's response time to the other speaker. Time stamps at each conversational turn were extracted from the LENA output. Out of the total of 78,613 segments, there were 15,204 C-to-M and 14,062 M-to-C conversational turns. As stated in the 'Introduction', we constructed a subset of the master data by selecting only the segments labeled as spoken by either the mother or the child from the original LENA-generated dataset to be used for our study. Therefore, many of the M-to-C or C-to-M turns in our data contained intervening segments of other types, for example other child speech (CXN) or overlapping speech (OLN). Nevertheless, a robust pattern of longer response time at the M-to-C turn compared to the C-to-M turn was found (see Figure 5). The response time at M-to-C turns was longer (M = 1·92 s, SD = 2·43 s) compared to the response time at C-to-M turns (M = 1·46 s, SD = 2·10 s). This finding is consistent with previous research suggesting longer response times for children than their caregivers (Tice, Bobb & Clark, 2011). There is relatively little research on the response time in mother–child interaction, but response times in child-to-child exchanges have been measured at 1·5 to 2·1 seconds (Garvey & Berninger, 1981; Lieberman & Garvey, 1977).

https://doi.org/10.1017/S0305000915000203 Published online by Cambridge University Press

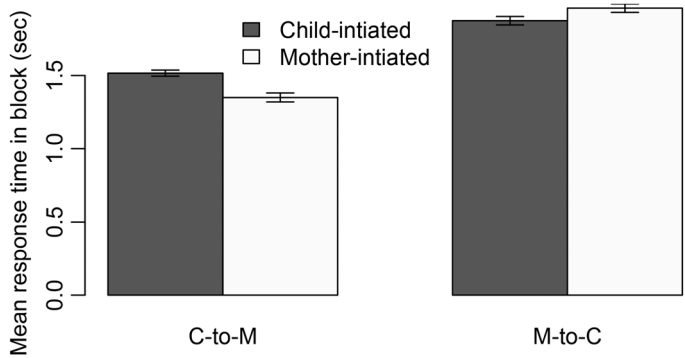


Fig. 5. Mean response time as a function of turn type and block type. C-to-M: response time between a child utterance and a maternal response. M-to-C: response time between a maternal utterance and a child response.

To statistically test the effect of turn type on response time, we again constructed a mixed-effects model. The turn type (M-to-C and C-to-M) and block type (Child-initiated and Mother-initiated) were entered as fixed effects with an interaction term between them. We included a random intercept for the mother–child dyads. The R formula used for this model was `coefficients ~ turntype * blocktype + (1 | dyad)`. Likelihood ratio tests showed that the full model accounts for the data significantly better than the one without an interaction term ($\chi^2(1) = 18{\cdot}0$, $p < {\cdot}001$) or other reduced models, i.e. the one with only the turn type ($\chi^2(2) = 19{\cdot}3$, $p < {\cdot}001$) or the block type ($\chi^2(2) = 312{\cdot}8$, $p < {\cdot}001$). The effect of turn type reflects the significantly longer response time at the M-to-C than the C-to-M turn type (Estimate = 0·6, SE = 0·04, $t = 14{\cdot}4$). The effect of block type reflects a significantly longer response time in the child-initiated than the mother-initiated blocks (Estimate = 0·15, SE = 0·04, $t = 3{\cdot}8$). The significant interaction between the turn type and the block type reflects the shorter mean response time at C-to-M turns in mother-initiated blocks (1·35 s) than in child-initiated blocks (1·52 s), and the shorter mean response time at M-to-C turns in child-initiated blocks (1·87 s) than in mother-initiated blocks (1·96 s). This interaction suggests that speakers tend to respond more quickly in blocks that they themselves initiated.

We next investigated the effect of initiator on duration of speech segments produced by each speaker, and found that both types of speakers produced longer segment durations in conversational blocks that they initiated (see Figure 6). That is, the mean duration of maternal utterances is longer in the mother-initiated (M = 1·53 s, SD = 0·88) than in the child-initiated (M = 1·35 s, SD = 0·66) blocks, whereas the mean duration of child

https://doi.org/10.1017/S0305000915000203 Published online by Cambridge University Press

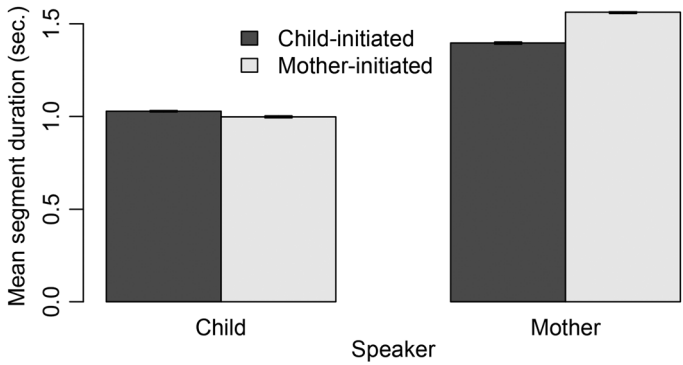


Fig. 6. Mean segment duration as a function of speaker and block type.

utterances was longer in the child-initiated (M = 1·01 s, SD = 0·79) than in the mother-initiated (M = 0·98 s, SD = 0·7) blocks.

We constructed a mixed-effects model including the speaker and block type (child-initiated and mother-initiated) as fixed factors with an interaction term between the two, and the mother–child dyad as a random factor. The R formula used for this analysis is `duration ~ speaker * blocktype + (1|dyad)`. A likelihood ratio test showed that the full model with the interaction term fits the data significantly better than the model with only the speaker ($\chi^2(2) = 382{\cdot}2$, $p < {\cdot}001$), with only the block type ($\chi^2(2) = 5{,}950$, $p < {\cdot}001$), or the one without the interaction ($\chi^2(1) = 291{\cdot}2$, $p < {\cdot}001$). The significant interaction between speaker and block type (Estimate = –0·2, SE = 0·01, $t = -17{\cdot}1$) reflects the tendency for both mothers and children to produce a longer segment in blocks that they initiated than in blocks initiated by their conversational partner reported above.

In sum, we found evidence that there is an effect of turn type on the correlation of mean pitch and response time in mother–child conversation. The greater correlation coefficient in C-to-M turns found in mean and maximum pitch suggests that the mother adapts her speech to the child more actively than vice versa. In addition, our results found an initiator-effect and its interaction with turn type in response time, and with the speaker in the segment duration. In other words, both children and mothers spoke for longer durations and responded more quickly in conversational blocks that they initiated themselves.

Testing convergence over a conversational block

In this section, we investigate whether the speech of mother and child become more similar to each other over the course of a conversation with

https://doi.org/10.1017/S0305000915000203 Published online by Cambridge University Press

focus on the change in the trend at the level of conversational block. Much previous research reporting entrainment in adult speech is based on datasets recorded over 30-minute or 1-hour sessions. Often, the mean of the entire session is taken to compare with the conversational partner to check similarity, or the data are divided into half, and the mean acoustic values in the earlier part of the session are compared with the ones in the latter part to examine convergence (e.g. Levitan & Hirschberg, 2011). When such a method is applied to a day-long recording like ours, however, information that could have been provided by a more temporally fine-grained reduction of the data is likely to be lost. In this regard, the availability of the unit conversational block in our data provides an efficient way of dividing the long stretch of recording into ecologically valid units of smaller time scale without causing radical data reduction by collapsing data across the entire day.

We compared the acoustic values of mother and child speech at the first and last conversational turn of each conversational block within each of the thirteen dyads. For this analysis, we selected the conversational blocks containing at least two conversational turns, which totaled 5,726 blocks out of the original 9,170 blocks that contained at least one turn. We then extracted the first and last conversational turn of each block, took the mean differences between the two interlocutors in the first and last turn of the block across the entire session for each dyad, and compared them using two-tailed paired *t*-tests. If mother–child speech converges, we would expect the difference between the mother and child speech to be smaller towards the end of a conversational block compared to the beginning.

We found that the mean differences of the acoustic values in the first and last turn of the conversational block between mother and child differed significantly in mean and minimum pitch (see Table 4), with mother and child becoming more similar over the course of a conversational block. This convergence in pitch measures was small, and driven by a decrement in pitch on the part of the child over the course of the conversation. No other measures showed significant convergence.

DISCUSSION

With respect to our first question, we found evidence of similarity in mother–child utterances across a variety of different analyses of our data, particularly in pitch measures. This effect of similarity cannot be accounted for by overall similarity of the mother–child dyad, since our analyses captured variance across blocks, turns and utterances WITHIN a given dyad. Our findings therefore provide some evidence of an entrainment effect whereby mother and child influence each other toward more similar speech patterns across conversational blocks. However, given

TABLE 4. *Testing convergence by comparin g the difference in acoustic values in the first and last conversational turns between mother and child in conversational blocks*

		First turn of block		Last turn of block	
	t-test	Child's mean value	Mother's mean value	Child's mean value	Mother's mean value
Duration	$t(12) = -0{\cdot}1$, $p = {\cdot}93$ (n.s.)	1·0 (s)	1·35 (s)	0·98 (s)	1·37 (s)
Speaking rate	$t(12) = -1{\cdot}5$, $p = {\cdot}16$ (n.s.)	2·29 (syl/s)	3·06 (syl/sec)	2·28 (syl/s)	3·08 (syl/s)
Pitch mean	$t(12) = -2{\cdot}9$, $p = {\cdot}01$*	8·0 (ERB)	6·54 (ERB)	7·93 (ERB)	6·52 (ERB)
Pitch minimum	$t(12) = 2{\cdot}7$, $p = {\cdot}02$*	6·12 (ERB)	4·80 (ERB)	6·02 (ERB)	4·76 (ERB)
Pitch maximum	$t(12) = -0{\cdot}2$, $p = {\cdot}84$ (n.s.)	9·53 (ERB)	8·89 (ERB)	9·45 (ERB)	8·85 (ERB)
Pitch range	$t(12) = -1{\cdot}12$, $p = {\cdot}29$ (n.s.)	3·41 (ERB)	4·09 (ERB)	3·43 (ERB)	4·08 (ERB)

that the size of the correlations is quite small, the implications of these effects must not be overstated. It is important to remember that mother and child will start out very similar in these global speech measures due to shared environmental and (often) genetic influences. Therefore, detecting this subtle variance ACROSS time between mother and child may be very difficult over and above this baseline similarity. Nonetheless, these findings build on prior findings analyzing across dyads to suggest that even with our more rigorous analysis, there is evidence for entrainment effects.

Our second question tackled the extent to which initiator and respondent effects influenced mother–child speech patterns. We found stronger correlations in the mean and the maximum pitch at the turn level when mothers responded to their child than vice versa, suggesting that mothers are adapting their speech more to their child than the reverse. Nonetheless, in some measures, correlations of child responses to maternal utterances reached significance, suggesting at least some adaptation on the part of the child. Additionally, children (and mothers) produce more mature speech forms in conversations that they themselves initiated (i.e. shorter response latencies, and longer segment duration). To our knowledge, this is the first study to identify such an effect, and it has strong implications for the role of locus of control in the development of language. Our findings suggest that providing infants/toddlers with opportunities to initiate dialog may drive learning. This possibility suggests that the notion of language environment quality for infants needs to be expanded to include consideration of agency on the part of the infant, and converges with research finding that infants learn better when mothers are responsive to their requests for linguistic input (e.g. Begus, Gliga & Southgate, 2014).

Our third and final question looked for evidence of convergence of acoustic measures between mother and child, rather than simply similarity, across a conversational block. We found a significant effect of convergence in mean and minimum pitch (i.e. mother and child became more similar to each other in mean and minimum pitch over the course of a conversation), and no other significant changes in the acoustic similarity across a conversational block. It is noteworthy that the convergence in pitch was driven by a decrease in the child's mean pitch and not that of the mother. While this may indicate an entrainment effect, it is also possible that this reflects a more general pitch decrement over the course of a conversation, and is not driven by the mother's pitch.

The finding that child-to-mother turns show a stronger correlation than mother-to-child turns bears additional consideration. This systematic pattern lends credence to the notion that while our correlations are small, we are tapping into a real phenomenon in mother–child interactions, in that it is consistent with the findings in previous research. In McRoberts and Best (1997), a similar direction of adjustment is found in that both the mother and father increased their mean pitch when interacting with their child compared to their adult-directed speech, but the child's pitch was not significantly different compared to their baseline pitch, e.g. pitch of vocalization when alone. Our results replicate this by showing that mothers entrain more than infants/toddlers do, but extends it by suggesting that finer-grained analyses can reveal small but consistent alterations in the child's responses to their mother (cf. Table 3 and further discussion below). Thus, we should not conclude that children are entirely passive, a notion that is also inappropriate given the initiator effects found. Indeed, the initiator influences on utterance duration and response time for both mother and child indicate that children are active contributors to the paralinguistic interactions in these conversations.

Our study diverged from previous research in that we found entrainment of pitch from child-to-mother, while some previous studies, such as McRoberts and Best (1997) and Siegel *et al*. (1990), did not. One possible reason for their differing results could be the coarse timescale of analysis in their study and/or the smaller sample size. In both the McRoberts and Best and Seigel *et al.,* studies, comparisons were made of mean pitch across different contexts (i.e. comparing interactions with mother and interactions with father), rather than more direct comparisons of adjacent utterances. Although Seigel *et al.,* did perform an analysis examining adjacent turns, and McRoberts and Best also report an attempted finer-grained timescale analysis, in both cases this was done with a relative small sample and might not have had enough power to find the subtle effects found in our analysis (although see below for another possibility related to developmental changes). Previous literature of adult speech

indicates that the correlation between the interlocutors increases at a local level. For example, Levitan and Hirschberg (2011) found significant correlations in all acoustic variables related to pitch, intensity, and vocal quality at the turn level but only some of them displayed correlation at the session level. Likewise, Gorisch, Wells, and Brown (2012) found that the pitch contour of insertions (short utterances) were significantly more similar to the immediately preceding turn, suggesting a local management of pitch contour. Thus our finding of higher correlation coefficients at the turn level than the block level for pitch-related measures may reflect a similar effect of tighter correlation at a local level. The tighter correlation observed over a short time span of conversation in various studies suggests that the entrainment in mother–child speech is essentially a process of continuously coordinating the speech in response to the interlocutor at a local level of exchange.

One important difference between our study and that of many of the studies reported in the 'Introduction' is the age of our sample. Most of the studies to date have been performed on relatively young infants (e.g. 3-month-olds) with the oldest ages being 9–12 months (Siegel *et al*., 1990) and up to 17 months (McRoberts & Best, 1997). The one exception was the Elias and Broerse (1996) study, which examined infants from ages 0;3 up to 2;0 and found developmental changes in overlapping speech across the ages studied. The youngest child in our sample was 13 months old, with the oldest being 2;6. It is therefore possible that the differences in pitch entrainment between our sample and prior research might be driven by developmental changes. Our analyses did not find strong support for age effects, but the clustering of our infants' ages around 20–25 months may have obscured developmental effects. Similarly, McRoberts and Best did not find evidence of developmental differences in their longitudinal case study. Nevertheless, given the small amount of data in that study, it is still possible that developmental changes are taking place between the first and second year of life that they were unable to detect that account for these differences.

Our findings suggest an important role for robust automated analyses in examining questions regarding the acoustic properties of mother–child interactions. This is an exciting time for this kind of research, as there is a convergence of methodological, analytical, and statistical innovations. Recent work by Buder, Warlaumont, Oller, and Chorna (2010), for example, demonstrates a unique analytic approach to the automated analysis of pitch dynamics between mother and child that may allow for the detection of more complex relationships than simple similarity or convergence, and may be generalizable to multiple other acoustic features.

CONCLUSION

Our study examined the dynamic relationship between the acoustic properties of mother and infant/toddler speech, by examining correlations WITHIN mother–child dyads, in order to reduce the influence of prior resemblance on measures of entrainment. We found small, but significant effects of entrainment in pitch, and less robust effects in utterance duration and speaking rate. We also found evidence of convergence in pitch measures across a single conversational block between mother and child, but these effects were small and driven by a general decrease in pitch on the part of the child. In general, maternal entrainment toward the child was stronger than vice versa. However, child entrainment toward mother was also found. In addition, we found effects of the initiator of a conversation, with longer utterances and shorter response latencies associated with the initiator of a conversational block. While mothers show more mature conversational capabilities (more entrainment, shorter response latencies, longer utterances), our findings converge with prior research to highlight the active role of young children in the conversational exchange.

REFERENCES

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.

Bates, D., Maechler, M. & Bolker, B. (2013). lme4: linear mixed-effects models using S4 classes, R package version 1.0-4. Online: <http://cran.r-project.org/web/packages/lme4/lme4.pdf>.

Beebe, B., Alson, D., Jaffe, J., Feldstein, S. & Crown, C. L. (1988). Vocal congruence in mother–infant play. *Journal of Psycholinguistic Research* **17**, 245–59.

Begus, K., Gliga, T. & Southgage, V. (2014). Infants learn what they want to learn: responding to infant pointing leads to superior learning, *PLoS ONE* **9**(10), online: <e108817.doi: 10·1371/journal.pone.0108817>.

Boersma, P. & Weenink, D. (2013). Praat: doing phonetics by computer [Computer program]. Version 5.3.39. Online: <http://www.fon.hum.uva.nl/praat/>.

Brennan, S. E., Galati, A. & Kuhlen, A. K. (2010). Two minds, one dialog: coordinating speaking and understanding. *Psychology of Learning and Motivation* **53**, 301–44.

Buder, E. H., Warlaumont, A. S., Oller, D. K. & Chorna, L. B. (2010). Dynamic indicators of mother–infant prosodic and illocutionary coordination. *Proceedings of the 5th International Conference on Speech Prosody*. Online: <https://umdrive.memphis.edu/awarlmnt/www/Buder_Warlaumont_Oller_Chorna_2010.pdf>.

Cristia, A. (2011). Fine-grained variation in caregivers' /s/ predicts their infants' /s/ category. *Journal of the Acoustical Society of America* **129**, 3271–80.

De Jong, N. H. & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods* **2**, 385–90.

Edlund, J., Heldner, M. & Hirschberg, J. (2009). Pause and gap length in face-to-face interaction. *Proceedings of Interspeech 2009, 10th Annual Conference of the International Speech Communication Association, Brighton, UK*. Online: <http://academiccommons.columbia.edu/download/fedora_content/download/ac:159987/CONTENT/edlund_al_09.pdf>.

Elias, G. & Broerse, J. (1996). Developmental changes in the incidence and likelihood of simultaneous talk during the first two years: a question of function. *Journal of Child Language* **23**(1), 201–17.
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), 455–70.
Garvey, C. & Berninger, G. (1981). Timing and turn-taking in children's conversations. *Discourse Processes* **4**, 27–57.
Goldstein, M. H. & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science* **19**(5), 515–23.
Gorisch, J., Wells, B. & Brown, G. J. (2012). Pitch contour matching and interactional alignment across turns: an acoustic investigation. *Language and Speech* **55**(1), 57–76.
Hart, B. & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Paul H Brookes Publishing.
Hermes, D. & van Gestel, J. C. (1991). The frequency scale of speech intonation. *Journal of the Acoustical Society of America* **90**, 97–102.
Hoff, E. & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development* **73**(2), 418–33.
Hurtado, N., Marchman, V. A. & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science* **11**(6), F31–9.
Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L. & Jasnow, M. D. (2001). Rhythms of dialogue in infancy: coordinated timing in development. *Monographs of the Society for Research in Child Development* Serial **66**(2). Wiley. Article Stable URL: <http://www.jstor.org/stable/3181589>.
Jaffe, J. & Feldstein, S. (1970). *Rhythms of dialogue*. New York: Academic Press.
Levitan, R. & Hirschberg, J. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. *Proceedings of Interspeech, 12th Annual Conference of the International Speech Communication Association*, Florence, Italy. Online: <http://www.cs.columbia.edu/~julia/papers/levitan&hirschberg11.pdf>.
Lieberman, A. F. & Garvey, C. (1977). Interpersonal pauses in preschoolers' verbal exchanges. Paper presented at the biennial meeting of the Society for Research in Child Development, New Orleans, LA.
Masataka, N. (1992). Pitch characteristics of Japanese maternal speech to infants. *Journal of Child Language* **19**, 213–23.
McRoberts, G. W. & Best, C. T. (1997). Accommodation in mean fo during mother–infant and father–infant vocal interactions: a longitudinal case study. *Journal of Child Language* **24**, 719–36.
Meltzoff, A. N. & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science* **198**(4312), 75–8.
Pickering, M. J. & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* **27**, 169–225.
R Core Team (2013). A language and environment for statistical computing, Version 3·0, R Foundation for Statistical Computing. Online: <http://www.R-project.org>.
Shimura, Y. & Yamanoucho, I. (1992). Sound spectrographic studies on the relation between Motherese and pleasure vocalization in early infancy. *Pediatrics International* **34**(3), 259–66.
Siegel, G. M., Cooper, M., Morgan, J. L. & Brenneise-Sarshad, R. (1990). Imitation of intonation by infants. *Journal of Speech, Language, and Hearing Research* **33**(1), 9–15.
Snow, C. E. (1977). The development of conversation between mothers and babies. *Journal of Child Language* **4**(1), 1–22.
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.
Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K.-E. & Levinson, S. C. (2009). Universals and

cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, 106(26), 10587–10592.
Striano, T., Henning, A. & Stahl, D. (2006). Sensitivity to interpersonal timing at 3 and 6 months of age. *Interaction Studies* **7**(2), 251–71.
Tice (Casillas), M., Bobb, S. & Clark, E. (2011). Timing in turn-taking: children's responses to their parents' questions. *Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue*, Los Angeles, California, 202–203. Online: <http://projects.ict.usc.edu/nld/semdial2011/proceedings/semdial2011_tice_1.pdf>.
Zimmerman, F. J., Gilkerson, J., Richards, J. A., Christakis, D. A., Xu, D., Gray, S. & Yapanel, U. (2009). Teaching by listening: the importance of adult–child conversations to language development. *Pediatrics* **124**(1), 342–9.