

Multimodal infant-directed communication: how caregivers combine tactile and linguistic cues*

RANA ABU-ZHAYA

Purdue University

AMANDA SEIDL

Purdue University

AND

ALEJANDRINA CRISTIA

Laboratoire de Sciences Cognitives et Psycholinguistique, CNRS, DEC-ENS, EHESS

*(Received 7 June 2015 – Revised 31 December 2015 – Accepted 24 July 2016 –
First published online 30 August 2016)*

ABSTRACT

Both touch and speech independently have been shown to play an important role in infant development. However, little is known about how they may be combined in the input to the child. We examined the use of touch and speech together by having mothers read their 5-month-olds books about body parts and animals. Results suggest that speech+touch multimodal events are characterized by more exaggerated touch and speech cues. Further, our results suggest that maternal touches are aligned with speech and that mothers tend to touch their infants in locations that are congruent with names of body parts. Thus, our results suggest that tactile cues could potentially aid both infant word segmentation and word learning.

INTRODUCTION

Although touch is one of the first senses to develop, we are just beginning to scratch the surface in exploring the influence that touch has on our lives. Thus far, we know that touch plays a key role in dyadic interactions and it is a prominent component of the multimodal communication in mother–infant interactions (Anisfeld, Casper, Nozyce & Cunningham, 1990;

[*] Address for correspondence: Rana Abu-Zhaya, Purdue University, Lyles-Porter Hall, West Lafayette, IN 47907. e-mail: rabuzhay@purdue.edu

Feldman, Singer & Zagoory, 2010; Ferber, 2004; Ferber, Feldman & Makhoul, 2008; Franco, Fogel, Messinger & Frazier, 1996; Herrera, Reissland & Shepherd, 2004; Hertenstein, 2002; Jean & Stack, 2009; Jean, Stack & Fogel, 2009; Moszkowski & Stack, 2007; Muir, 2002; Stack & Arnold, 1998). Further, touch has been found to play a role in directing infants' attention, regulating arousal levels, behavioral states, and negative emotions, as well as reducing distress (Hertenstein, 2002; Jean & Stack, 2009, 2012; Stack & Muir, 1990). All of these results could INDIRECTLY contribute to infant language outcomes; however, we do not know whether touch has any DIRECT role in language development, or whether caregivers use touch in a way that could directly contribute to language development. In this paper, we use a micro-genetic approach, coding detailed cues from caregiver touch and speech in mother–infant interactions to examine ways in which touch may be used with speech in the input to the child.

Recent work suggests that touch might be useful to infants' speech perception (Seidl, Tincoff, Baker & Cristia, 2015). Seidl and colleagues familiarized 4-month-olds with a continuous stream of syllables that contained no acoustic or distributional cues to word boundaries in two conditions. In one, infants received a timed tactile stimulation of their elbow or knee that was always synchronous with a specific trisyllabic string (e.g. the syllable sequence *lepoga* was always timed with a touch to the knee). In another condition, infants received similar input, but in the visual modality by watching an experimenter touch her own eyebrow or chin. Infants in both conditions were also touched on, or observed a touch on, another location (e.g. elbow), but this touch/visual cue was not consistently synchronous with a particular syllable sequence (e.g. the syllable sequence *dobita* only occurred one time with a touch on the elbow and all other times without this touch; conversely, touches to the elbow coincided with many other trisyllabic strings). After familiarization, infants in both conditions were tested for their listening preferences to trisyllabic strings using the Head Turn Preference Procedure. The results showed that infants' listening times to trisyllabic sequences that had been paired with consistent touches were shorter than listening times to trisyllabic sequences without consistent touches and sequences that were not in the familiarization string at all. Thus, it appears that providing reliable experimenter touches carefully aligned with units in the speech signal to 4-month-olds can aid in infants' ability to find units in a continuous speech stream. However, although the word segmentation advantage with aligned touches may be found in highly controlled experimental situations, we do not know whether caregivers in the real world actually use touch in synchrony with the speech they direct towards their infants. Therefore, in this study we focus on examining mother–infant interactions and explore the use of touch with speech directed to infants.

Studies have shown that, when examined in isolation, one component of infant-directed communication, Infant-Directed Speech (IDS), may play a role in supporting language acquisition (Falk, 2004; Graf Estes & Hurley, 2013; Hills, 2012; Singh, Nestor, Parikh & Yull, 2009; Thiessen, Hill & Saffran, 2005). Specifically, IDS may aid in early word recognition (Singh *et al.*, 2009) and word learning (Graf Estes & Hurley, 2013). In real-life interactions, however, IDS is not detached from other forms of infant-directed communication. It is part of an intricate multimodal communication system that characterizes caregiver–child interactions. Specifically, mother–infant interactions are replete with different social, emotional, and linguistic cues that are combined together and used in different ways to communicate messages to the infant. Hence, in addition to examining IDS separately, it is vital that we also examine its occurrence with all these other cues and explore how this combination can benefit (or not benefit) the language-learning infant. Examining and analyzing those interactions using micro-genetic approaches and frame-by-frame coding of behaviors yields a wealth of data that can provide us with valuable information on how language is presented to prelinguistic and pre-intentional infants (Tamis-LeMonda, Kuchirko & Tafuro, 2013). Such analyses might help us reach a better understanding and more accurate estimation of the weighting of different cues available to the infant in the process of learning language.

Studies examining multimodal communication in mother–infant interactions reveal patterns of multimodal behavior performed spontaneously by caregivers, often as much as 75–99% of the time (Gogate, Bahrick & Watson, 2000; Nomikou & Rohlfing, 2011). When demonstrating actions to their infants, mothers' speech is well aligned with their actions (Gogate *et al.*, 2000; Meyer, Hard, Brand, McGarvey & Baldwin, 2011; Zukow-Goldring, 1997). For example, Zukow-Goldring (1997) reports a mother producing the phrase 'your head' (in Spanish, *tu cabeza*) while reaching out with her index finger towards her daughter's head, followed by a tap on the head. Moreover, when performing specific actions (e.g. in diaper-changing situations), mothers tend to accompany each step of their action sequences with utterances. Specifically, they pat, tickle, and squeeze their infants in synchrony with their vocal productions, sometimes even lengthening their utterances to fit the length of their actions (Nomikou & Rohlfing, 2011). These behaviors do not only accompany the speech that is directed to infants, they are also distinctively different from those behaviors that occur with speech directed to adults (Brand, Baldwin & Ashburn, 2002; Brand, Shallcross, Sabatos & Massie, 2007). Taken together, these studies point to the natural tendency of infant-directed communication to be multimodal.

Motivated by the findings of Seidl *et al.* (2015), and studies of multimodal infant-directed communication discussed above, we examined whether touch during dyadic interactions may contain cues that could potentially aid the language-learning infant. We had two main questions. First: Is touch in mother–infant dyads used in a way that could help infants pull out words from running speech? This question addresses the temporal alignment between tactile cues and spoken words. Second: Is touch in mother–infant dyads used in a way that could help infants learn the mapping between sound and meaning of some words in their language? This question addresses the congruence between tactile cues and spoken words. To answer our questions, we first examined whether touch is a prevalent component in infant-directed communication, and whether it occurs systematically with speech.

To address these two main questions we used a caregiver–infant book-reading interaction. We chose book-reading because it allowed us to have a high level of control over the linguistic input that infants were receiving compared to free play. Further, book-reading is a very common practice among parents and an important part of early caregiver–infant interactions in Western societies (Sénéchal, LeFevre, Thomas & Daley, 1998). Finally, reading a book is almost always accompanied by social interaction between the adult reader and the child (Mol, Bus, de Jong & Smeets, 2008), and it is one of multiple episodes of physical closeness (Makin, 2006) that characterize early communication. Thus, we predicted that mothers participating in our study would naturally use touch during these interactions with their infants without being told to do so.

We created books to be used in this study; half of the books were about animals and the other half were about body parts. Examining children's books revealed that animals and body parts were popular linguistic categories that appeared in many storybooks for infants. Furthermore, previous studies have shown that 6- to 7-month-olds show some understanding of the meaning of at least some body-part words (Bergelson & Swingley 2012; Tincoff & Jusczyk, 2012). As for animal words, they are also considered to be some of the first words that children learn (Fenson *et al.*, 1993). Using these two sets of word categories enabled us to examine whether the use of touch cues during book-reading interactions was related to a specific set of words – body parts vs. animals. Naturally, tactile cues can be useful for learning body-part words, but may be less useful for animal words, since highlighting the word's referent leads to tactile cues being felt by the child only in the former case. In other words, a caregiver can touch her infant's foot when saying the word *foot*, but cannot touch a congruent location on her infant's body while saying the word *dog*. This touch that accompanies the production of the word *foot* might aid the infant in the early learning of this word. On the other hand,

a touch that accompanies the word *dog* or any other word that does not refer to a body part might help the infant pull out that specific word from the continuous stream of speech, but will not provide the infant with a clue to the word's meaning.

In this study, we ask whether caregivers temporally align touch and speech production during interactions with their infants and whether they produce touch cues that are semantically congruent with their speech. The choice of including two lexical categories in our books allowed us to explore the presentation of animal words in comparison with body-part words. We predicted that mothers would accompany their speech with touch, aligning their touches with words and utterances. Further, we predicted that maternal touch would be more prevalent when reading the body-part books, and that body-part words would be aligned with touches more frequently. Given that infants can benefit from touch as a cue for finding words in continuous speech (Seidl *et al.*, 2015) and since word segmentation is related to later word learning (Newman, Ratner, Jusczyk, Jusczyk & Dow, 2006), we wanted to know whether caregivers use tactile-auditory synchrony during dyadic interactions with their infants in a way that could potentially aid word segmentation and later word learning.

METHODS

Participants

Forty-six dyads were recruited from flyers posted on a university campus and birth announcements published in the local newspaper. Parents were contacted via mail, telephone, e-mail, and Facebook. Out of the total sample of forty-six dyads, twenty were excluded due to fussiness ($n=9$), experimenter error ($n=3$), prematurity/low birth weight ($n=2$), and non-compliance with instructions ($n=6$; e.g. reading the books only once or more than twice). In order to keep our sample as homogenous as possible with respect to acoustic analyses, we excluded one dyad due to the participation of the father, and another due to the dialect of the mother (British English). We did this because previous studies provide evidence for dialect-specific differences in the acoustic features of IDS (Fernald, Taeschner, Dunn, Papousek, de Boysson-Bardies & Fukui, 1989), as well as differences between IDS produced by mothers and fathers (e.g. Shute & Wheldall, 1999). Thus, the final sample included twenty-four dyads. Infants were full-term, had normal birth weight, and were from monolingual English-speaking families. Infants' ages ranged from 0;4.34 to 0;5.82 ($M=0;5.33$, $SD=0;0.414$; 12 female). Mothers' education ranged from twelve to twenty-two years ($M=15.7$ years, $SD=2.56$). All mothers gave informed consent before participation and all infants received a book or a toy for their participation in the study.

Materials

Eight books were created, four on animals (A₁, A₂, A₃, A₄) and four on body parts (B₁, B₂, B₃, B₄). In order to create these books, we used commercial children's books as references for choice of words (see Table A₁, Appendix A). We generated a list of animal and body-part words to be included in our books, to which we added some other body-part words that did not appear in any children's books. These new words (*eyebrow*, *finger*, *chin*, and *heel*) enabled us to avoid any overlap among the target words in our books.

Each book included four target words; one of these was a bisyllabic word with a strong–weak stress pattern, while the other three were monosyllabic words (see Table A₂, Appendix A). We chose to use this combination of prosodic templates because both of these templates are common in young children's books. Each word was accompanied by a picture that was carefully chosen from picture databases. All eight books included the same text in which each target word was repeated four times in sentence-final position (for an example see Appendix B). All of the eight new books included the same number of pages (see Appendix C).

Procedure

Each dyad was randomly assigned a combination of two books from the total of eight books, one on body parts and one on animals (A₁+B₁, A₃+B₃, etc.), such that each combination was assigned to only six dyads and each target word had the opportunity to occur the same number of times in our sample. Prior to the book-reading session, mothers were provided with a brief explanation about the study; specifically, they were told that the study aimed at exploring mother–child book-reading interactions. Researchers did not mention the interest in examining the use of touch during these interactions, and provided all mothers with the following instruction: "We would like you to read each book twice the way you would normally do at home, and please try to feel as comfortable as possible in spite of the new setting and the cameras." The book-reading interactions took place in a quiet room. The infant was seated in a high-chair directly facing the mother, who sat as closely as possible to this high-chair in order to promote as much touching as possible given the constraints of the situation. We chose not to have the infant sit on his/her caregiver's lap because we wanted to elicit touches that were non-accidental in nature. We used two cameras to videotape the interactions; the main camera provided a side view of both the mother and the infant, allowing a good view of the mother's hands, and the other camera was located behind the infant's high-chair and provided a different view of the mother's face and hands, showing part of the infant's body as well. Video-recordings

from this last camera were used in cases where the mother's hands were not visible enough in the video from the main camera. Mothers wore a clip-on microphone that was wirelessly connected to the main camera, allowing us to separate the audio stream from the video stream to allow for separate analyses and coding.

Data analyses

Video coding. Using ELAN (Brugman & Russel, 2004), we coded all the intentional (or non-accidental) maternal touches during the book-reading interactions. Video coding was performed by watching the videos from the main camera without sound. Intentional (or non-accidental) touch was defined as any type of touch that the mother intentionally provided to her infant on any part of the infant's body. Once the infant grabbed or touched his/her mother in any way, coding was ceased (e.g. the mother grabs the baby's hand, but when she is about to release her grip, the baby grabs the mother's finger). Touch that occurred unintentionally or accidentally was not coded (such as when the mother was trying to flip the page in the book and accidentally touched her baby's outstretched arm).

A template was created in ELAN allowing unified coding for all the videos. The template included three tiers for coding the touch event and a fourth tier for coding the type of session ('animal' or 'body-part' book-reading) or a transition between sessions. The three tiers allowed us to annotate three distinct pieces of information for each touch event: touch location, touch type, and number of beats. We coded the start and end times of each touch unit and its location. Possible locations were: head, hair, nose, cheek, eyebrow, eye, ear, chin, mouth, arm, hand, torso (upper body), belly, waist, leg, foot, feet, toe, toes, finger, fingers, knee, and heel. The start and end of each touch were defined using a coding scheme that differentiated touch types (see Table D1, Appendix D). Further, we also coded the number of beats for each touch (e.g. three squeezes, five pokes, ...) which was also defined differently based on the type of touch. Hence, information for each touch event was presented on three separate but connected tiers (see Figure 1 for an example of the coding of a touch event that was aligned with the production of a word). Video coding was performed by two teams of two coders. Each touch unit was agreed upon by the two coders before it was annotated in ELAN. Disagreements were settled through consensus and in some cases through consulting members of the other team. In cases in which a touch unit was not visible from the main camera, and in cases of doubt about the specific features of the touch, coders consulted the video from the other camera. Upon completion of coding, a Praat text-grid file was extracted from ELAN for each video interaction.

MATERNAL TOUCH AND MATERNAL SPEECH

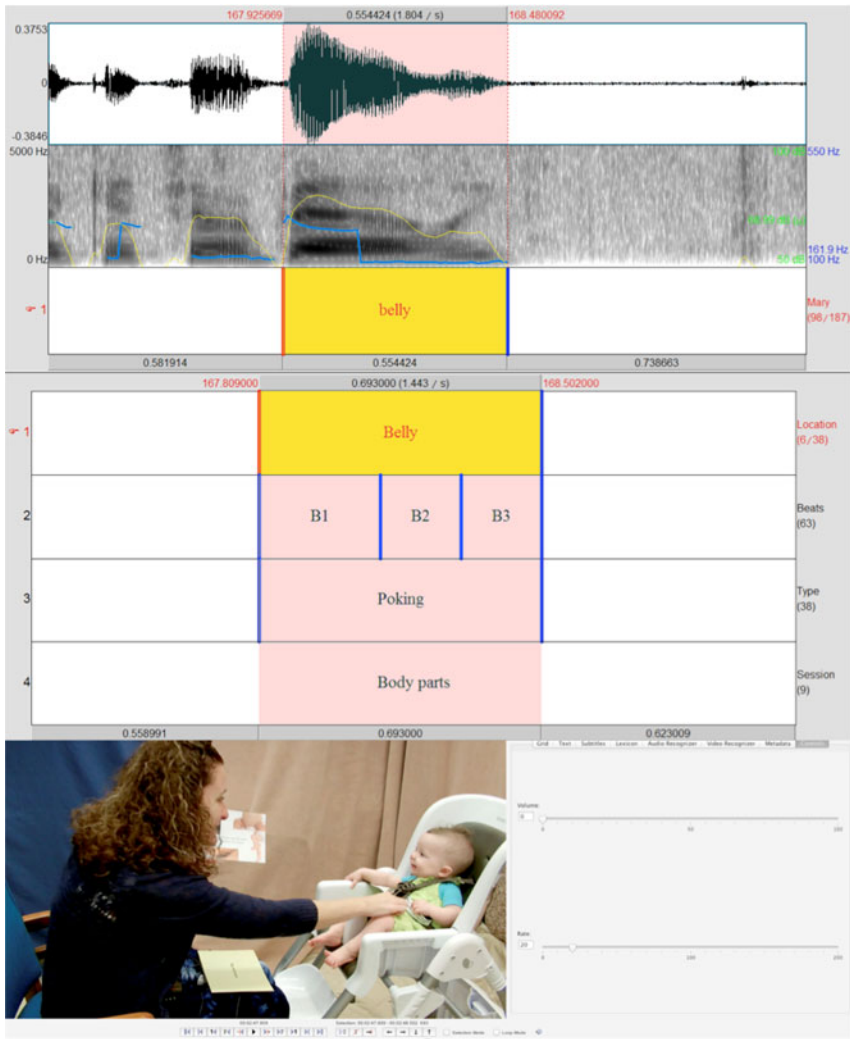


Fig. 1. A time slice from the coding showing the alignment of tactile and auditory information gleaned from the video and audio coding, respectively. Here, the occurrence of a touch unit (poking the belly three times) is congruent with the word produced (*belly*).

Audio coding. An audio file was extracted from the videos recorded through the main camera. This coding was performed in Praat (Boersma & Weenink, 2013). (Our decision to use Praat for annotating speech stemmed from the fact that it offers good visualization for both spectrograms and waveforms.) We analyzed mothers' speech in two steps.

First, we marked each production of the target words during the reading (see Figure 1). The edges of the target words were marked in Praat based on acoustic features of the phonemes as they appeared in the waveform and the spectrogram. Tags were placed at upward zero crossings to ease extraction of acoustic values for each target word. We tagged words as mothers produced them even if their productions did not correspond precisely with the target words as they appeared in the books (e.g. *kitty* for *cat* or *horsey* for *horse*). Upon completion of this step, and after extracting a text-grid file from the video coding of the same interaction, a third text-grid file was then created in Praat, merging the touch and target word coding. Creating this file allowed us to code other words and sentences (non-target words) that occurred in proximity to each touch event or overlapped with it. (In this coding step, we followed the criteria and steps detailed in Appendix E.) All audio coding was performed by two separate coders who shared notes on the coding process and resolved issues and questions through discussion. Upon completion of all audio coding, we generated another text-grid for each dyad, which included three tiers: words, session, and sentence. In the sentence tier we combined the words that mothers produced into sentences by examining the gaps between the words. If this gap was less than 300 ms, then the words were treated as part of the same sentence; if it was greater, then the words were treated as belonging to two different sentences. This 300 ms criterion was used in line with previous work (e.g. Fernald & Simon, 1984) in which a 300 ms pause was used to separate utterances. Such pauses coincide with sentence boundaries 96–98% of the time in infant- and child-directed speech (e.g. Fernald & Simon, 1984; Fisher & Tokura, 1996).

Extracting the data. A Praat script written specifically for this study allowed us to align and integrate information from the video and audio coding. (This script, as well as other scripts used for this project and descriptions of the design of each script, are available at <<https://osf.io/ybe5g/>>.) The script logged three types of events and generated an output file that included all the events for all the dyads:

1. *Word only*: a word is produced without a concurrent touch between the onset and offset of the word. In this case, the script extracted the start and end times of the word and its duration and logged acoustic information: average, min and max fundamental frequency (fo), and intensity. Further, if this word was part of a sentence, then the script also logged the sentence, its start and end times, and its duration.
2. *Touch only*: a touch is identified, but there is no target word or other non-target words overlapping with it. In this case, the script logged the touch location, start and end times of the touch, and the number of beats.

3. *Word-touch co-occurrence*: there is a touch that overlaps with speech (the target word or any other speech that had been coded because it overlapped with touches). To identify this type, we examined the touch tier at specific points in time, which depended on the word tier. We first looked at whether any touches were ongoing at the word midpoint (the point in time which was halfway between the word onset and offset). If there was no active touch at that point, then we examined the touch tier at the word onset. If no active touch was present at that point, we looked at the time of the word offset. Finally, if no active touches had been identified at any of those three points, we looked for touches that occurred anywhere between the onset and the offset of the word. Once an active touch had been found, we logged both the audio and video information noted in 1 and 2 above. We examined all of these points so that we could guarantee that we were capturing all possible instances of word + touch co-occurrences. For example, based on these criteria, the script will log instances in which the touch overlapped with the word at the word midpoint. In this case, even if the touch actually started before the word midpoint but after the beginning of the word, we would still catch this co-occurrence.

Shuffling the data. Analyses based on words, which have been exhaustively coded, allowed us to assess whether the temporal alignment between speech and touch events is closer than would be expected by chance, by generating a null distribution. To do this, we wrote a Praat script that generated 100 different versions of coded text-grids; this was done by shuffling the temporal position of sentences and the silences separating them within each session separately with respect to the onset of the phase, while keeping words tied with their sentences. (We could have shuffled only the target words, but we decided to work on the sentence level because this is the most extensively coded. Results are certainly exactly the same as if we had shuffled target words, which are a subset of the items shuffled.) [Figure 2](#) shows a sequence extracted from one of the coded files. The top three tiers show the original coding, at the level of words, session, and sentence, one of which is “just like your legs”. The bottom three tiers show the coding arising from one of the shuffles, again including the three tiers: words, session, and sentences. It is evident from the figure that there are exactly the same number of words and sentences across the two versions, because the script only altered the POSITION of sentences and silences. Sometimes, a sentence lands close to its original position, but this occurs only by chance. For example, in [Figure 2](#), the sentence “just like your legs” appears only in its original position in the top three tiers, but it is not visible in this particular shuffling seen in the bottom three tiers; this means that in this shuffling this sentence did not land anywhere near its original position.

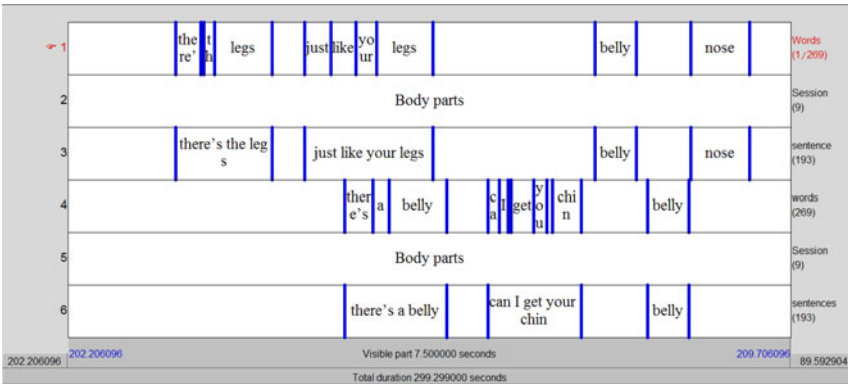


Fig. 2. A sequence extracted from the coding. The first three tiers show the original coding of words, session, and sentence. The bottom three tiers show the coding arising from shuffling the temporal position of sentences.

Since shuffling merely reorders silences and sentences, it does not affect anything about their distributions (duration, frequency of occurrence, etc.). Further, since we shuffled within sessions, we did not perturb the association between certain words and certain sessions. The only effect of shuffling is that it disturbs the actual association between the word tier and the touch tiers, thus providing a distribution of the temporal alignment between the two when there is no real underlying connection, other than the greater or lesser frequency of occurrence of touches in some sessions than others. Since we generated 100 shuffled versions for each dyad, we can compare the actual observed temporal alignment to the temporal alignment that is found by the chance co-occurrence of words and touches. In other words, this constitutes a test via bootstrap resampling (as introduced by, e.g. Fisher, 1935), and the number of samples allows us to establish significance with a precision of two decimals (i.e. up to $p = .01$).

All statistical analyses were performed in R. To answer our research questions, the R scripts tabulated frequencies of occurrence (e.g. how many touches were found, how frequently touches occur with speech, ...), as well as information regarding the characteristics of touch events that occurred with speech and those that did not. Further, the scripts also tabulated information on the specific speech events that occurred with touch, on two levels: sentence and word.

RESULTS AND ANALYSES

We first examined the frequency of touch during book-reading interactions. Results revealed that 3 out of 24 mothers never touched their infants during our observations. Among those who did, we observed between 1 and 66

touches with a mean of 22 touches. Two of the mothers touched their infants only when transitioning between the books (primarily readjusting their infants' position), leaving 19 mothers (79%) who touched their infants during the book-reading sessions. Thus, it appears that while touch is a frequent component during book-reading interactions, it is not a NECESSARY component of book-reading, since it is absent in 5 out of 24 dyads (21%). Moreover, we found that the total number of touches was higher in body-part sessions than in the animal sessions (body-part sessions $M = 14.208$, $SD = 12.254$; animal sessions $M = 3.75$, $SD = 6.954$; Wilcoxon signed rank test $V = 718$, $p = .006$).

Next, we asked how frequently touches co-occurred with speech. Given the naturalistic nature of the interactions we examined, and the fact that we did not provide mothers with any information regarding the nature of the study and the design of the books around specific target words, we did not expect them to stick to the text of the books. Hence, and in order to account for the possible variability in mothers' speech to their infants during the interactions, and in order to cover all possible cases of touch+speech co-occurrences, we treated all speech that occurred in proximity to a touch event as potentially related to that event. Specifically, we coded words in all sentences that overlapped with a touch in any way. We then considered a word (any word) to be temporally aligned, in other words, to coincide with a touch, if the touch and the word occurred within 250 ms of each other. This overlap between word and touch could be complete, partial, or even null. Moreover, this definition does not impose a one-to-one correspondence. For example, a long touch may be classified as co-occurring with five words if it is synchronized with the phrase, "Look at the pretty doggy!" This lax definition of co-occurrence is necessary because we do not know (a) how close speakers synchronize their touch and speech, and (b) how close a tactile event and an auditory event need to be for an infant to perceive them as related. This time period was chosen because it fits with previous research on infants' perception of synchrony for auditory and visual events. Specifically, in order to perceive an auditory event preceding a visual one as asynchronous, 2- to 8-month-olds require a minimum temporal separation of 350 ms between the two events. Similarly, in order to perceive a visual event preceding an auditory one as asynchronous, 2- to 8-month-old infants need the two events to be separated by a minimum of 450 ms (Lewkowicz, 1996). The size of the temporal synchrony window in both cases does not change between 2 and 8 months of age (Lewkowicz, 1996). This means that if an auditory event precedes a visual one by 300 ms, then a 5-month-old is likely to perceive them as temporally synchronous. In the absence of specific evidence regarding tactile cues and speech, we adopted this lax definition to provide a more comprehensive view. We will, however,

investigate finer temporal alignment in touch+speech co-occurrences in subsequent analyses.

Of the 463 touches found across 21 dyads, 340, or 73%, co-occurred with speech (using the previously mentioned lax definition; see Table F1, Appendix F). At the individual level, the percentage of touches co-occurring with speech tended to be between 60% (25th percentile) and 81% (75th percentile), although the full range was 0–100%, with the two instances of 0% touch–speech co-occurrences corresponding to the two mothers who only touched their child during the transitions between books (see Figure 3). Thus, touch very often co-occurs with speech during book-reading. An important finding from these analyses revealed that the total number of touch+speech events (events in which the touch and speech occurred within 250 ms of each other) was higher during body-part sessions than during animal sessions (body-part sessions $M = 11.083$, $SD = 9.573$; animal sessions $M = 2.25$, $SD = 4.425$; Wilcoxon signed rank test $V = 734.5$, $p = .0018$). Hence, during body-parts sessions, mothers produced significantly more touch+speech events as compared to during animal sessions.

In order to examine whether combining speech with touch affects the characteristics of touch in any way, we compared the physical characteristics of the 340 touches that co-occurred with speech with those of the 123 touches that did NOT occur with speech (see Table F1, Appendix F). Our rationale for using any type of speech that occurred with touch was that we wanted to examine the multimodality of the input to the infant from a touch-based approach by examining touch events as a whole. Our coding scheme allowed us to annotate touch events that had varying characteristics in terms of their location, type, number of beats, and duration, hence allowing us to examine the variability in any of these characteristics and whether it can be driven by speech.

We examined the number of beats per touch and the duration of touches. Since these are continuous variables, it was possible to calculate averages within each caregiver, and then use paired *t*-tests to compare the two types. We extracted the mean number of beats and the mean duration of touches for the eighteen mothers who used touch both with and without speech. These analyses revealed that touches accompanied by speech were longer (touch–speech $M = 0.822$ s., $SD = 0.294$ s.; touch-alone $M = 0.487$ s., $SD = 0.399$ s.; Wilcoxon signed rank test $V = 12$, $p = .001$) and had twice as many beats (touch–speech $M = 2.778$, $SD = 1.446$; touch-alone $M = 1.245$, $SD = 0.717$; Wilcoxon signed rank test $V = 13$, $p = .003$) as touches that were not accompanied by speech.

Next, we examined whether the addition of touch to maternal IDS affects the characteristics of that speech in any way. In this set of analyses, we restricted our attention to target words from the text of the books in order to control the content of speech. Specifically, we explored whether speech

MATERNAL TOUCH AND MATERNAL SPEECH

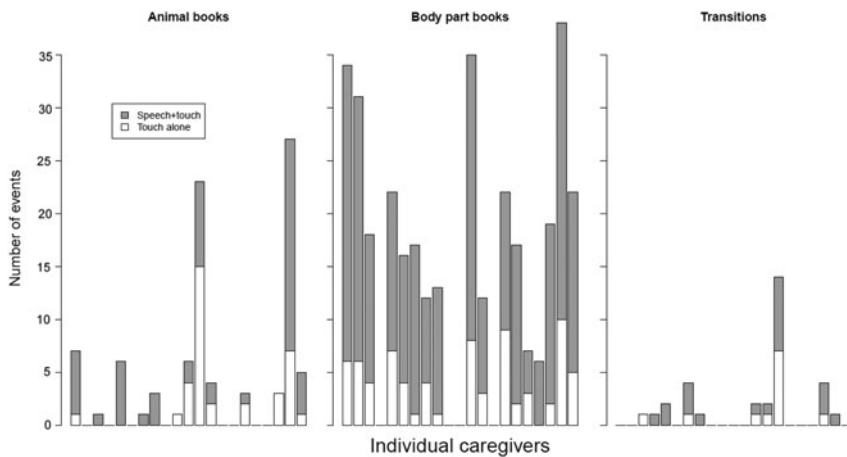


Fig. 3. Prevalence of touch events that occurred with speech and touch events that occurred without speech per dyad.

that occurred with touches differed from speech that occurred without touches by comparing the 1,680 tokens of target words that occurred without any concomitant touches with the 200 tokens of target words that occurred with concomitant touches (see Table F2, Appendix F). We averaged within these two types (words with touch, words without touch) for each speaker separately, and compared these two types along two acoustic dimensions salient to infants, namely, duration and fundamental frequency.

These analyses revealed a trend for shorter duration when words were accompanied by touch than when they were spoken alone (words with touch $M = 0.527$ s., $SD = 0.157$ s.; words alone $M = 0.572$ s., $SD = 0.102$ s.; Wilcoxon signed rank test $V = 115$, $p = .07$, $n = 17$), and significantly higher average f_0 (words with touch $M = 7.518$, $SD = 0.552$; words alone, $M = 6.989$, $SD = 0.499$; Wilcoxon signed rank test $V = 18$, $p = .004$, $n = 17$) and minimum f_0 (words with touch $M = 5.497$, $SD = 0.534$; words alone $M = 5.099$, $SD = 0.279$; Wilcoxon signed rank test $V = 22$, $p = .008$, $n = 17$). However, we did not find a significant difference between words that were accompanied by touch and words that were spoken alone in terms of their maximum f_0 (words with touch $M = 10.042$, $SD = 0.783$; words alone, $M = 9.873$; $SD = 0.777$; Wilcoxon signed rank test $V = 59$, $p = .431$, $n = 17$) or f_0 range (words with touch $M = 4.545$, $SD = 0.664$; words alone $M = 4.773$, $SD = 0.689$; Wilcoxon signed rank test $V = 108$, $p = .145$, $n = 17$). To sum up, the only significant difference between target words spoken with touches and those spoken without a concomitant touch is that the former had higher average and minimum f_0 than the latter.

As noted above, 340 touches co-occurred with any type of speech. Since we coded all the speech that overlapped with touches, sentences co-occurring with touches were not only book extracts. Specifically, we found 326 unique sentences that co-occurred with touch, defining 395 sentence–touch events (each sentence could co-occur with one or more touches, and vice versa). We classified these sentences manually into several categories based on the lexical content. When a classification could not be determined based on the lexical content of the sentence alone, we examined the context in which the sentence was produced. Even after examining the context, we were still unable to classify six cases (4 *yeah, I know, xxx is*). The remaining 389 sentence–touch events were classified into the following categories: verbatim renditions from the books (121), derivations or restatements of phrases from the books (149), phrases aiming at guiding infants’ attention and coding the progress in the task (58), phrases narrating caregiver readjustments of infant position (33), or animal sounds (28). In order to estimate inter-rater reliability in the classification of sentences co-occurring with touches, 53 sentences (out of 326, i.e. 16%) were randomly selected for re-classification by a second coder. One of the sentences was unclassified by both coders; the other 52 sentences yielded an inter-raters’ agreement estimation (unweighted) $\kappa = .606$ (Cohen, 1960; estimated using the package *irr* version 0.84; Gamer, Lemon, Fellows & Singh, 2015), which falls at the boundary between the ‘moderate’ and ‘substantial’ agreement categories proposed by Landis and Koch (1977). Further, our analyses revealed that the majority of the 395 sentence–touch events occurred during body-part book-reading. The average number of sentence–touch events among the 21 caregivers who used touch was $M = 15.65$ events ($SD = 10.97$) compared to $M = 2.95$ ($SD = 5.15$) for animals, and $M = 1.15$ ($SD = 2.01$) for transitions.

We also investigated whether touches were aligned with word edges, in which case they could potentially provide cues for infants’ segmentation of the speech stream. To this end, we calculated the onset lag as the time of the onset of the touch minus the time of the onset of the word. Smaller numbers indicate better-aligned touches and words. The sign indicates whether the touch started before the word did or vice versa. Thus, a positive onset lag of, for instance, 1 second indicates that the touch started 1 second AFTER the word started. Similarly, we calculated the offset lag as the time of the offset of the touch minus the time of the offset of the word. A positive offset lag indicates that the touch ended AFTER the word ended.¹ These relationships are represented in Table 1.

¹ Inspection of histograms of the distributions of onset and offset lags at the sentence level suggested that sentences and touches might be well aligned, since lags were smaller than

TABLE 1. *The possible temporal alignment relationships between touch events and words; the onset lag is the time of the onset of the touch minus the time of the onset of the word, and the offset lag is the time of the offset of the touch minus the time of the offset of the word. The sign indicates whether the touch precedes the word or vice versa.*

Positive onset and negative offset	Negative onset and negative offset	Positive onset and positive offset	Negative onset and positive offset
---- word ---- ---- touch ----	---- word ---- ---- touch ----	---- word ---- ---- touch ----	---- word ---- ---- touch ----

Given that speech and touch are defined as co-occurring based on temporal information, and considering the lack of data on how caregivers align their touches with speech, we can glean informative data from the temporal alignment analyses incorporating the null distribution generated by shuffling the sentences (see the ‘Shuffling the data’ subsection in the ‘Methods’ section). To this end, we extracted the quartiles of the observed distribution, and each of the 100 shuffled versions, and compared them to each other. These comparisons revealed that the median onset and offset alignments between target words and the touches co-occurring with them were not different than what would occur by chance, but that the VARIANCE in the onset alignment (measured through the inter-quartile range, given the non-normality of the distribution) was significantly lower in the observed distribution than in the null distribution (exact $p = .01$, according to our bootstrap resampling test with $n = 100$). Indeed, as shown in Figure 4, the observed inter-quartile range for onset alignment in the real distribution was 0.62 seconds, whereas the 2.5% and 97.5% bounds of the null distribution were 0.8 and 1.15 seconds, respectively.

To examine speech–touch congruence, the extent of semantic alignment of speech and touch cues was also analyzed. In the context of this study, it is reasonable to wonder whether mothers would touch the body part evoked by their speech. To examine this question, we focused on 185 events in which a body-part target word co-occurred with touch (see Table F2, Appendix F). We classified these events as congruent if mothers touched the body part that was evoked by their speech, and incongruent if the mothers touched some other body part. For nearly all of the speakers, the majority of touches during the production of body-part words were congruent. These proportions were significantly higher than what would

0.5 seconds in 54% of the onsets and 51% of the offsets. However, there is no particular reason to look at 500 ms alignment, and thus we defer to the next set of analyses, based on the target words, which are able to estimate to what extent temporal alignment is precise beyond what would be expected by chance.

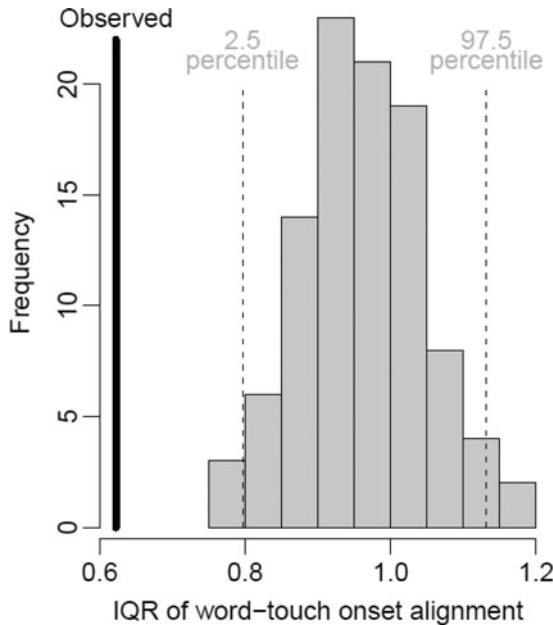


Fig. 4. Variance in the IQR (Inter-Quartile Range) for onset alignment between touches and sentences in the observed distribution (indicated by the black line) and the null distribution (marked by the light bars). The dashed lines indicate that, in the null distribution, the 2.5 and 97.5 percentiles corresponded to 0.8 and 1.15 s. onset alignment, respectively, which is significantly different from that in the observed distribution, i.e. 0.62 s.

be expected by chance ($p < .01$, according to our bootstrap resampling test with $n = 100$), as is evident in Figure 5.

DISCUSSION

In line with previous work (Herrera *et al.*, 2004; Weiss, Wilson, Hertenstein & Campos, 2000), these data revealed that, though there is considerable individual variation in the frequency of use of touch, it appears to be a frequent component in mother–infant book-reading interactions; specifically, touch has been found to be more frequent during body-part sessions than animal sessions. Further, our findings show that the majority of touch events co-occur with speech, pointing again to the multimodal nature of infant-directed communication.

These touch+speech events were found to be significantly different from corresponding unimodal events in terms of their physical characteristics. First, touches that were accompanied by speech were longer and had twice as many beats as touches that were produced without any accompanying

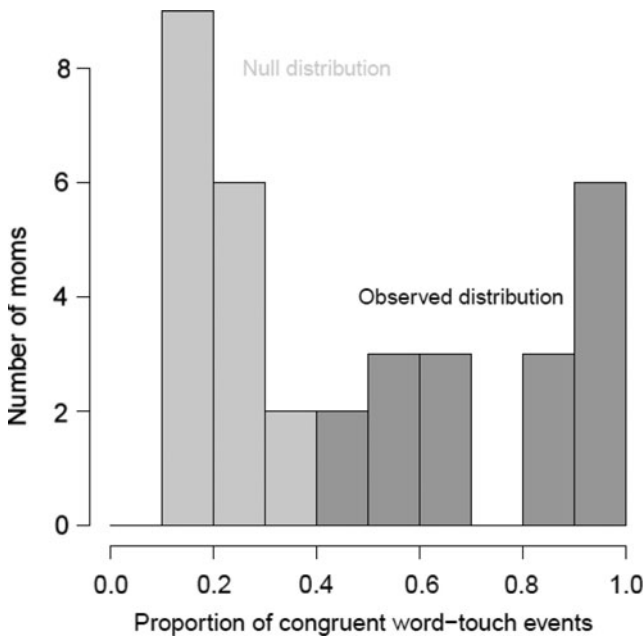


Fig. 5. Frequency of different proportions of congruent word+touch events as seen in the observed distribution (dark grey bars) and the null distribution (light grey bars).

speech. Second, a close examination of the speech that overlapped with touches (target words that appeared in the books) revealed that words that were produced with touches had both a higher average *f*₀ and a higher minimum *f*₀ than words that were spoken without any accompanying touch. These touch+speech events were not only unique in their physical characteristics (exaggeration of touch duration and the higher *f*₀ in speech), they were also well aligned. Specifically, 73% of touch events occurred with speech within a time period of 250 ms. This time-window has been previously shown to be sufficient for 2- to 8-month-olds to perceive auditory and visual events as synchronous (Lewkowicz, 1996). However, given that visual and auditory events could potentially differ from tactile and auditory events, and given the lack of data on the nature of perceptual synchrony in the latter types of events, we can only speculate that such tactile–auditory events could be perceived as synchronous if separated by less than 250 ms. Minimally, we can conclude that infants are not only receiving multimodal touch+speech communication in the form of touch and speech that are separately exaggerated, but they are also receiving touch+speech events that are temporally aligned in a way that could be POTENTIALLY perceived by infants as synchronous.

Hence, maternal touch, when temporally aligned with speech, has the potential for rendering this multimodal input particularly useful to the infant struggling to segment the speech stream into units. It seems plausible that touch might be useful for this purpose, given recent experimental data in Seidl *et al.* (2015) showing that experimenter-provided touch can aid word segmentation in infants. These recent findings by Seidl and colleagues are relevant to child linguistic development because speech segmentation can be viewed as an important early benchmark that infants need to reach in the process of learning words and building a lexicon (Graf Estes, Evans, Alibali & Saffran, 2007; Junge, Kooijman, Hagoort & Cutler, 2012; Jusczyk, 1999; Kooijman, Junge, Johnson, Hagoort & Cutler, 2013). Further, it appears that word segmentation is also related to early expressive language skills (Junge *et al.*, 2012; Kooijman *et al.*, 2013; Newman *et al.*, 2006; Singh, Reznick & Xuehua, 2012), comprehension skills (Kooijman *et al.*, 2013), and later syntactic and semantic language profiles (Newman *et al.*, 2006). While our findings shed some light on the multimodality of infant-directed communication, we do not know how and whether the alignment of touch and speech in the real world might serve infants in their segmentation of speech. Hence, further research is still needed before we can pinpoint the specific contributions of different cues to word segmentation in a multitude of situations both in and outside of the lab.

The differences we observed in the characteristics of touch+speech events compared with touch-alone and speech-alone events are intriguing. Why is it the case that certain cues were exaggerated when they were multimodal? One possibility is that, while speaking to their infants and touching them, mothers were unconsciously trying to align these cues, causing them to lengthen their touches to fit with the length of their utterances. In the diaper-changing interactions examined in Nomikou and Rohlfing (2011), the authors speculated that some mothers accentuated the duration of their actions by lengthening their speech, thus creating a synchronous ‘hands and language’ event. However, it is just as possible that mothers might lengthen their speech to align with the length of their actions. Regardless, this possibility of unconscious alignment points towards mothers’ unconscious awareness of the benefits of a synchronous multimodal presentation of speech for their infants’ developing language skills. In an examination of the occurrence by chance of the previously described events, we created a null distribution by shuffling the sentences to manipulate the distribution of temporal alignment between touch and speech events. These analyses revealed that the variance of onset alignment with respect to the target words was significantly lower in the observed distribution than in the null distribution. A possible explanation for the small variation in the observed distribution compared to the null

distribution is that perhaps, in our observed sample, mothers are unconsciously trying to align their speech with touches every time they produce such multimodal events. Such a hypothesis, however, does not explain the change in the speech (raised *fo*) that occurred with touches. However, perhaps there are alternative explanations for this effect. First, it is plausible that touching their infants causes mothers to use more affective language, which would in turn result in higher *fo*, as we observed in our sample. Alternatively, it is possible that IDS with particularly high *fo* means that the mother is being particularly affectionate with her child and a side effect of this affect is simply an increase in touch. Either way, while our results cannot adjudicate between these alternatives, they do clearly show that touch and speech are both exaggerated when they occur together.

If we attempt to explain why mothers exaggerate their productions of touch and speech when those occur together, one possibility might follow the suggestion offered by Brand and Shallcross (2008) that infant-directed modifications seem to offer the benefit of enhancing infants' attention. Hence, it is reasonable to assume that the infant-directed modifications we found – the exaggeration of touches that occur with speech and the higher *fo* of the speech that occurs with touch – actually reflect caregivers' attempts at garnering infants' attention. Similarly, it is possible that mothers' use of tactile cues is related to their attempts at regulating their infants' arousal, as previous studies have found that touch regulates arousal levels and reduces distress (Hertenstein, 2002; Jean & Stack, 2009, 2012; Stack & Muir, 1990). In fact, some of the sentences that occurred with touch in our sample were classified as serving the function of guiding the infants' attention (e.g. “look at the book”, “look at mommy”, or producing the infant's name). Thus, touch (referential or not) might heighten or dampen the infant's arousal, and caregivers might exploit it for this reason. If this were the case, temporal alignment and congruence might simply be a secondary cue or side effect of the caregiver's main goal of arousal regulation. Nonetheless, this cue could help the infant to pay more attention to whatever occurs in synchrony with the touch. Specifically, the use of touch might allow infants to be more attentive to the speech stream and to the accompanying cues simply because he or she might be more aroused. Our data may partially allow us to address this arousal hypothesis. If touch is used primarily by caregivers in this language-rich setting to regulate arousal, then we might predict that caregivers would trade-off touch with IDS cues (such as pitch), since IDS has also been reported to heighten arousal (Nakata & Trehub, 2004), so that the infant is not overly aroused due to excessive use of multiple arousing cues. However, as mentioned earlier, we found that words that were produced with touches actually had higher *fo* than words that were spoken without any accompanying touch. This might lead to the

conclusion that touch is used as an accompanying cue to speech rather than a main arousal cue which might trade-off with cues from other modalities.

Another significant finding of this work is that a semantic examination of word+touch events revealed that the proportion of touch locations that were congruent with the meaning of the word that accompanied them was higher than would be expected by chance. This highlights another important function that caregiver touch could come to have. Specifically, our findings suggest that touches could potentially aid in word learning by highlighting word–referent relationships to the language-learning infant. Nearly all mothers in our sample produced word–touch events that were congruent. This means that most infants in our sample heard words referring to body parts while they were being touched on those same locations on their own bodies (e.g. a mother reading the section about belly in the book produced the word *belly* in temporal alignment with a touch on her infant’s belly). It is intriguing that mothers produced these multimodal temporally aligned and congruent cues specifically for body-part words following a simple request to merely read the books to their infants; books that included images of other infants’ body parts. Yet, even with such symbolic representation of body parts, most mothers seized this opportunity for teaching their infants about body parts in a more body-oriented way. This specific and timed production of such events could be extremely helpful for the infant to map words onto their referents in the face of noisy environments in which neither speech is parsed out into its components, nor objects are presented in isolation. Such presentation of body-part words to those prelinguistic infants might provide some explanation for their early acquisition and the ease at which they enter infants’ proto-lexica.

Given the way this study was designed and conducted, our data are limited and do not allow us to make definite conclusions regarding the benefits of using touch in any manner with all kinds and topics involving speech. Nonetheless, we believe that our data might be generally useful to infant word segmentation, even if touch is only provided in this aligned and exaggerated way when discussing body parts with infants. We conclude this for two key reasons. First, we suspect that caregivers naturally (outside of the body-part book-reading context) discuss body parts with their infants in diapering and feeding situations. Second, even if such aligned and exaggerated touches are only found in body part discussions, once the infant segments these words she could use them as a toehold from which to acquire the rest of her proto-lexicon. For example, if infants segment the word *foot*, then when *foot* occurs in a sentence flanked by other words (e.g. “Look at your tiny foot baby!”), the infant could then use the word *foot* as an anchor to segment the novel forms *tiny* and *baby*. Such ideas are discussed in Bortfeld, Morgan, Golinkoff, and Rathbun, (2005), in which the authors suggest that the infant’s own name and the

word baby (other early segmented and learned words) could function as similar anchors to build more general segmentation skills.

Another possibility that might explain the occurrence of speech with touch cues in a referential and temporally synchronized manner might be related to the nature of human communication patterns. It is possible that our spoken language system evolved from a gestural or tactile system, or that the two systems evolved together. Thus, these two systems might still operate in a dependent manner (McNeill, 2012). If this is the case, then it is not surprising that mothers use this feature when communicating with their infants. Once again, only future work further exploring the dependence of these two communication systems (touch and spoken language) will contribute to support this hypothesis.

Given that mothers and infants communicate in different ways depending on the context in which they are interacting (Tamis-LeMonda, Song, Leavell, Kahana-Kalman & Yoshikawa, 2012), the frequency, types, and functions of touch could be affected by the type of interaction and the degree of physical closeness that is observed (feeding, floor play, face-to-face interactions, ...; Jean *et al.*, 2009), and by the cultural and ethnic background of caregivers (Franco *et al.*, 1996). Thus, it is not surprising that some mothers in our sample ($n = 5$) did not touch their infants at all during the book-reading interactions. However, the current data do not allow us to provide an explanation for the lack of touch in these dyads, and we can only speculate on the reasons behind it based on previous work. For instance, maternal responsiveness, including touch, is related to mothers' years of education and socioeconomic status (SES; Richman, Miller & Levine, 1992). However, and fitting with conclusions from previous work (Weiss, Wilson & Morrison, 2004), we did not find a correlation between SES and touch frequency (correlation coefficient $r(22) = 0.089$, $p = .67$). This lack of effect of SES might be due to the fact that our sample was very homogeneous; hence, examining more heterogeneous samples might yield different effects. Further, it is worth noting that the lack of touch in some dyads could be somewhat due to our design; since the nature of a book-reading interaction leaves only one free hand for the mother to use, this could potentially make it more difficult to interact with the infant using touch.

In sum, we found clear relationships between caregiver touch and speech, suggesting that most caregivers produce touch aligned with, and congruent to, spoken language. Thus, touch cues appear to have another function in early interactions that is distinct from the previously reported functions of touch; i.e. touch could serve a referential and aligning function highlighting words in the speech stream that could aid the infant in the task of speech segmentation and later word learning.

REFERENCES

- Anisfeld, E., Casper, V., Nozyce, M. & Cunningham, N. (1990). Does infant carrying promote attachment? An experimental study of the effects of increased physical contact on the development of attachment. *Child Development* **61**(5), 1617–27.
- Bergelson, E. & Swingle, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the United States of America* **109**(9), 3253–8.
- Boersma, P. & Weenink, D. (2013). Praat: doing phonetics by computer [Computer program], Version 5.3.45, online: <<http://www.praat.org/>> (retrieved 30 December 2013).
- Bortfeld, H., Morgan, J., Golinkoff, R. M. & Rathbun, K. (2005). Mommy and Me: familiar names help launch babies into speech-stream segmentation. *Psychological Science* **16**(4), 298–304.
- Brand, R. J., Baldwin, D. A. & Ashburn, L. A. (2002). Evidence for ‘motionese’: modifications in mothers’ infant-directed action. *Developmental Science* **5**(1), 72–83.
- Brand, R. J. & Shallcross, W. L. (2008). Infants prefer motionese to adult-directed action. *Developmental Science* **11**(6), 853–61.
- Brand, R. J., Shallcross, W. L., Sabatos, M. G. & Massie, K. P. (2007). Fine-grained analysis of motionese: eye gaze, object exchanges, and action units in infant- versus adult-directed action. *Infancy* **11**(2), 203–14.
- Brugman, H. & Russel, A. (2004). Annotating multimedia/multi-modal resources with ELAN. In *Proceedings of LREC 2004, Fourth International Conference on Language Resources and Evaluation*. Nijmegen: The Language Archive, Max Planck Institute for Psycholinguistics Online: <<http://tla.mpi.nl/tools/tla-tools/elan/>>.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* **20**(1), 37–46.
- Falk, D. (2004). Prelinguistic evolution in early hominins: Whence motherese? *Behavioral and Brain Sciences* **27**(4), 491–541.
- Feldman, R., Singer, M. & Zagoory, O. (2010). Touch attenuates infants’ physiological reactivity to stress. *Developmental Science* **13**(2), 271–8.
- Fenson, L., Dale, P., Reznick, S., Thal, D., Bates, E., Hartung, J., ... Reilly, J. (1993). *The MacArthur Communicative Development Inventories*. San Diego, CA: Singular Publishing Group.
- Ferber, S. G. (2004). The nature of touch in mothers experiencing maternity blues: the contribution of parity. *Early Human Development* **79**(1), 65–75.
- Ferber, S. G., Feldman, R. & Makhoul, I. R. (2008). The development of maternal touch across the first year of life. *Early Human Development* **84**(6), 363–70.
- Fernald, A. & Simon, T. (1984). Expanded intonation contours in mothers’ speech to newborns. *Developmental Psychology* **20**(1), 104–13.
- Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B. & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers’ and fathers’ speech to preverbal infants. *Journal of Child Language* **16**(03), 477–501.
- Fisher, C. & Tokura, H. (1996). Acoustic cues to grammatical structure in infant-directed speech: cross-linguistic evidence. *Child Development* **67**(6), 3192–218.
- Fisher, R. A. (1935). *The design of experiments*. Edinburgh: Oliver and Boyd.
- Franco, F., Fogel, A., Messinger, D. S. & Frazier, C. A. (1996). Cultural differences in physical contact between Hispanic and Anglo mother–infant dyads living in the United States. *Early Development and Parenting* **5**(3), 119–27.
- Gamer, M., Lemon, J., Fellows, I. & Singh, P. (2015). irr: Various Coefficients of Interrater Reliability and Agreement. Online: <<http://CRAN.R-project.org/package=irr>> (retrieved 3 June 2015).
- Gogate, L. J., Bahrick, L. E. & Watson, J. D. (2000). A study of multimodal motherese: the role of temporal synchrony between verbal labels and gestures. *Child Development* **71**(4), 878–94.
- Graf Estes, K., Evans, J. L., Alibali, M. W. & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? *Psychological Science* **18**(3), 254–9.

- Graf Estes, K. & Hurley, K. (2013). Infant-directed prosody helps infants map sounds to meanings. *Infancy* **18**(5), 797–824.
- Herrera, E., Reissland, N. & Shepherd, J. (2004). Maternal touch and maternal child-directed speech: effects of depressed mood in the postnatal period. *Journal of Affective Disorders* **81**(1), 29–39.
- Hertenstein, M. J. (2002). Touch: its communicative functions in infancy. *Human Development* **45**(2), 70–94.
- Hills, T. (2012). The company that words keep: comparing the statistical structure of child-versus adult-directed language. *Journal of Child Language* **40**(3), 1–19.
- Jean, A. D. L. & Stack, D. M. (2009). Functions of maternal touch and infants' affect during face-to-face interactions: new directions for the still-face. *Infant Behavior & Development* **32**(1), 123–8.
- Jean, A. D. L. & Stack, D. M. (2012). Full-term and very-low-birth-weight preterm infants' self-regulating behaviors during a still-face interaction: influences of maternal touch. *Infant Behavior & Development* **35**(4), 779–91.
- Jean, A. D. L., Stack, D. M. & Fogel, A. (2009). A longitudinal investigation of maternal touching across the first 6 months of life: age and context effects. *Infant Behavior & Development* **32**(3), 344–9.
- Junge, C., Kooijman, V., Hagoort, P. & Cutler, A. (2012). Rapid recognition at 10 months as a predictor of language development. *Developmental Science* **15**(4), 463–73.
- Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends in Cognitive Sciences* **3**(9), 323–8.
- Kooijman, V., Junge, C., Johnson, E. K., Hagoort, P. & Cutler, A. (2013). Predictive brain signals of linguistic development. *Frontiers in Psychology* **4**(25), 1–13.
- Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* **33**(1), 159–74.
- Lewkowicz, D. J. (1996). Perception of auditory-visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception and Performance* **22**(5), 1094–106.
- Makin, L. (2006). Literacy 8–12 months: What are babies learning? *Early Years* **26**(3), 267–77.
- McNeill, D. (2012). *How language began*. New York: Cambridge University Press.
- Meyer, M., Hard, B., Brand, R. J., McGarvey, M. & Baldwin, D. A. (2011). Acoustic packaging: maternal speech and action synchrony. *IEEE Transactions on Autonomous Mental Development* **3**(2), 154–62.
- Mol, S. E., Bus, A. G., de Jong, M. T. & Smeets, D. J. H. (2008). Added value of dialogic parent–child book readings: a meta-analysis. *Early Education and Development* **19**(1), 7–26.
- Moszkowski, R. J. & Stack, D. M. (2007). Infant touching behavior during mother–infant face-to-face interactions. *Infant and Child Development*, **16**(3) 307–19.
- Muir, D. W. (2002). Adult communications with infants through touch: the forgotten sense. *Human Development* **45**(2), 95–9.
- Nakata, T. & Trehub, S. E. (2004). Infants' responsiveness to maternal speech and singing. *Infant Behavior and Development* **27**(4), 455–64.
- Newman, R., Ratner, N. B., Jusczyk, A. M., Jusczyk, P. W. & Dow, K. A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: a retrospective analysis. *Developmental Psychology* **42**(4), 643–55.
- Nomikou, I. & Rohlfing, K. J. (2011). Language does something: body action and language in maternal input to three-month-olds. *IEEE Transactions on Autonomous Mental Development* **3**(2), 113–28.
- Richman, A. L., Miller, P. M. & LeVine, R. A. (1992). Cultural and educational variations in maternal responsiveness. *Developmental Psychology* **28**(4), 614–21.
- Seidl, A., Tincoff, R., Baker, C. & Cristia, A. (2015). Why the body comes first: effects of experimenter touch on infants' word finding. *Developmental Science* **18**(1), 155–64.
- Sénéchal, M., LeFevre, J. A., Thomas, E. M. & Daley, K. E. (1998). Differential effects of home literacy experiences on the development of oral and written language. *Reading Research Quarterly* **33**(1), 96–116.

- Shute, B. & Wheldall, K. (1999). Fundamental frequency and temporal modifications in the speech of British fathers to their children. *Educational Psychology* **19**(2), 221–33.
- Singh, L., Nestor, S., Parikh, C. & Yull, A. (2009). Influences of infant-directed speech on early word recognition. *Infancy* **14**(6), 654–66.
- Singh, L., Reznick, J. S. & Xuehua, L. (2012). Infant word segmentation and childhood vocabulary development: a longitudinal analysis. *Developmental Science* **15**(4), 482–95.
- Stack, D. M. & Arnold, S. L. (1998). Changes in mothers' touch and hand gestures influence infant behavior during face-to-face interchanges. *Infant Behavior and Development* **21**(3), 451–68.
- Stack, D. M. & Muir, D. W. (1990). Tactile stimulation as a component of social interchange: new interpretations for the still-face effect. *British Journal of Developmental Psychology* **8**, 131–45.
- Tamis-LeMonda, C. S., Kuchirko, Y. & Tafuro, L. (2013). From action to interaction: infant object exploration and mothers' contingent responsiveness. *IEEE Transactions on Autonomous Mental Development* **5**(3), 202–9.
- Tamis-LeMonda, C. S., Song, L., Leavell, A. S., Kahana-Kalman, R. & Yoshikawa, H. (2012). Ethnic differences in mother–infant language and gestural communications are associated with specific skills in infants. *Developmental Science* **15**(3), 384–97.
- Thiessen, E. D., Hill, E. A. & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy* **7**(1), 53–71.
- Tincoff, R. & Jusczyk, P. W. (2012). Six-month-olds comprehend words that refer to parts of the body. *Infancy* **17**(4), 432–44.
- Weiss, S. J., Wilson, P., Hertenstein, M. J. & Campos, R. (2000). The tactile context of a mother's caregiving: implications for attachment of low birth weight infants. *Infant Behavior and Development* **23**(1), 91–111.
- Weiss, S. J., Wilson, P. & Morrison, D. (2004). Maternal tactile stimulation and the neurodevelopment of low birth weight infants. *Infancy* **5**(1), 85–107.
- Zukow-Goldring, P. (1997). A social ecological realist approach to the emergence of the lexicon: educating attention to amodal invariants in gesture and speech. In C. Dent-Read & P. Zukow-Goldring (eds), *Evolving explanations of development: ecological approaches to organism-environment systems*, (pp. 199–250). Washington, DC: American Psychological Association.

Appendix A: Other resources

TABLE A1. *Reference books*

1	“Where is Baby’s belly button?” (Katz, K. (2000). <i>Where is Baby’s belly button?</i> New York, NY: Little Simon.)
2	“The ME Book” (Tymms, J. (2005). <i>The ME book</i> . New York, NY: Golden Books.)
3	“All of baby nose to toes” (Adler, V. (2009). <i>All of baby nose to toes</i> . New York: Penguin Group Inc.)
4	“Dear zoo” (Campbell, R. (1982). <i>Dear zoo</i> . London: Macmillan.)
5	“Where’s Spot?” (Hill, E. (1980). <i>Where’s Spot?</i> New York, NY: G. P. Putnam’s Sons.)
6	“I went walking” (Williams, S. (1989). <i>I went walking</i> . New York, NY: Red Wagon Books.)
7	“Peek-a-moo!” (Cimarusti, M. T. (1998). <i>Peek-a-moo</i> . New York, NY: Dutton Children’s Books.)

TABLE A2. *Target words in each book*

Animal books		Body-parts books	
A1	A2	B1	B2
Camel (SW)	Puppy (SW)	belly (SW)	tummy (SW)
Bear (S)	Bird (S)	nose (S)	eye (S)
Cat (S)	Horse (S)	chin (S)	waist (S)
Sheep (S)	Cow (S)	leg (S)	feet (S)
A3	A4	B3	B4
Lion (SW)	Hippo (SW)	finger (SW)	eyebrow (SW)
Frog (S)	Snake (S)	mouth (S)	ear (S)
Mouse (S)	Dog (S)	knee (S)	heel (S)
Duck (S)	Pig (S)	toe (S)	hand (S)

NOTES: The letters in parenthesis refer to the stress pattern of the word: SW = Strong-Weak; S = Strong

Appendix B: Sample book text

Example – Book B1

Do you see the belly? Where's the belly?

Here's the belly.

Do you see the nose? Where's the nose?

Here's the nose.

Do you see the chin? Where's the chin?

Here's the chin.

Do you see the leg? Where's the leg?

Here's the leg.

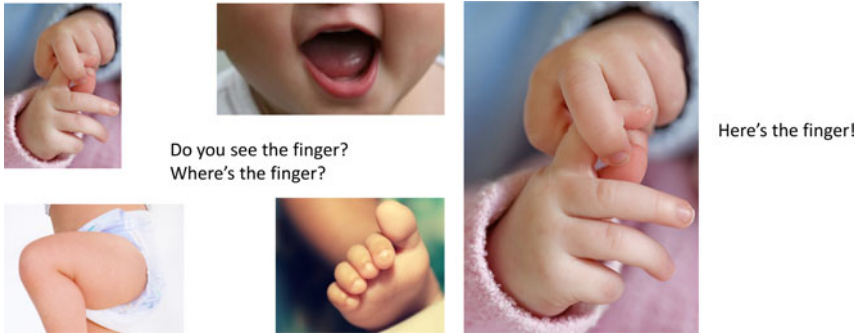
Here's the belly.

Here's the nose.

Here's the chin.

And here's the leg.

Appendix C: Sample pages from the books



SOURCES: baby (2007), online: <<https://plus.google.com/104016225791303606375/posts?pid=5829343309702875042&oid=104016225791303606375>> (retrieved 21 September 2013); baby (2013a), online: <<http://clinicajuniorsaude.pt/en/ms/ms/-paediatric-specialities-8200-083-albufeira/ms-90063238-p-2/>> (retrieved 21 September 2013); baby (2013b), online: <http://thetir.cafemom.com/pregnancy/159693/autism_risk_linked_to_induction> (retrieved 21 September 2013); Devine, R. (2008), brother and sister, online: <<https://www.flickr.com/photos/sesameellis/2967885621/in/set-72157606194417543>> (retrieved 21 September 2013).



SOURCES: Arabian camel (2010), online: <<http://www.todayifoundout.com/index.php/2010/07/camels-humps-are-not-filled-with-water/>> (retrieved 10 June 2013); Eastcott, J. & Momatiuk, Y. (2006), grizzly bear, online: <http://animals.nationalgeographic.com/animals/photos/bears/#/grizzly-bear_566_600x450.jpg> (retrieved 10 June 2013); Llanwenog sheep (2013), online: <<http://www.fermaime.com/?id=19&l=2049>> (retrieved 10 June 2013); McClellan, L. (2006), online: Barnyard the cat, online: <http://animals.nationalgeographic.com/animals/wallpaper/domestic-cat_image.html> (retrieved 10 June 2013).

Appendix D: The different types of touch

TABLE D I. *The different types of touch*

Brushing	A motion that begins in one location and ends in another, performed either with one finger or the whole hand. Each movement in one direction was coded as one beat; going back on the opposite direction was coded as another beat.
Grabbing	Coded only when noticeable as a separate touch. No beats.
Moving	Mother moves infant's body part in any way (shaking, moving towards the book). Beats are counted based on the direction of the movement; once the direction changes, a new beat is coded.
Pinching	A squeezing motion with two fingers only. Coding starts when the fingers are stretched before the pinch, and ends with the fingers stretched again as in the initial position.
Poking	Coding starts with the actual touch on the body part and ends when the finger is pulled back, either to start a new poke or to end the whole touch. Each poke is coded as a separate annotation unless mothers do not pull their fingers off the infant's body.
Readjustment	Mother adjusts infant's position; location varies. No beats.
Resting	Mother is resting her hand on any of the infant's body parts. No beats.
Squeezing	A squeezing motion with the whole hand. Coding starts with the hand stretched before the squeeze, and ends with the hand stretched again as in the initial position.
Tapping	Touch with the whole hand. Similar to poking, coding starts with the actual touch on the body part and ends when the hand is pulled back, either to start a new tap or to end the whole touch. Each tap is coded as a separate annotation unless mothers do not pull their fingers off the infant's body.
Tickling	No beats.
Unspecified	Other types of touch that do not apply to any specific category. No beats.
Wiping	Mother wipes baby's drool. No beats.

Appendix E: Criteria for the second step in audio coding

1. First, we examined if the touch overlapped either completely or partially with one word, and if that was the case then only that word was coded.
2. We also examined if a word occurred within a 0.5 s. window before or after the touch; if such a word was detected, then it was also tagged using the same method as described in the first step.
3. If we detected a longer sequence of speech occurring with a touch event or in proximity to the touch event, then we examined whether that sequence was an utterance. Utterances were defined based on a temporal criterion: as long as the space between the words was no more than 0.5 s., they were regarded as part of the same utterance. If the sequence fulfilled our criterion and was indeed an utterance, and it either began or ended within 0.5 s. from a touch event, or had a complete or partial overlap with a touch event, then the whole utterance was coded in the form of separate words.

4. If we detected a touch that occurred between utterances where the last word of one utterance and the first word of the next utterance were separated out by more than 0.5 s., but the touch itself occurred less than 0.5 s. after the end of one utterance and less than 0.5 s. prior to the beginning of the other utterance, then they were both coded and treated as having a temporal relationship with the touch event.

Appendix F: Data tables

TABLE F1. *The number of touches in total and the frequency of touch + speech events compared to speech alone events*

Touch events (<i>n</i> = 463)	
Touches that co-occurred with speech	340
Touches that did not co-occur with speech	123

TABLE F2. *The number of target words per category and the frequency with which each occurred with or without touch*

	Target words (<i>n</i> = 1880)		Total target words
	Animals (<i>n</i> = 899)	Body-parts (<i>n</i> = 981)	
With touch	15	185	200
Without touch	884	796	1680