DOI: 10.1111/infa.12532

RESEARCH ARTICLE

INFANCY THE OFFICIAL JOURNAL OF THE INTERNATIONAL CONGRESS OF INFANT STUDIES WILEY

Mothers' use of touch across infants' development and its implications for word learning: Evidence from Korean dyadic interactions

Eon-Suk Ko[1] | **Rana Abu-Zhaya**[2] | **Eun-Sol Kim**[3] | **Taehyeong Kim**[4,5] | **Kyung-Woon On**[6] | **Hyunji Kim**[1] | **Byoung-Tak Zhang**[7] | **Amanda Seidl**[8]

[1]Department of English Language and Literature, Chosun University, Gwangju, Korea

[2]School of Psychology, The University of Plymouth, Plymouth, UK

[3]Department of Computer Science, Hanyang University, Seoul, Korea

[4]AI Lab, CTO Division, LG Electronics, Seoul, Korea

[5]Department of Biosystems Engineering, Seoul National University, Seoul, Korea

[6]Kakao Brain, Seongnam, Korea

[7]Department of Computer Science and Engineering & SNU Artificial Intelligence Institute, Seoul National University, Seoul, Korea

[8]Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, Indiana, USA

Correspondence
Eon-Suk Ko
Email: eonsukko@chosun.ac.kr

Funding information
National Research Foundation of Korea, Grant/Award Numbers: 2021R1I1A2051993, 2021K2A9A1A01102199; NIH, Grant/Award Number: R21 HD108730-01; Chosun University, Grant/Award Number: 2021

Abstract

Caregivers' touches that occur alongside words and utterances could aid in the detection of word/utterance boundaries and the mapping of word forms to word meanings. We examined changes in caregivers' use of touches with their speech directed to infants using a multimodal cross-sectional corpus of 35 Korean mother-child dyads across three age groups of infants (8, 14, and 27 months). We tested the hypothesis that caregivers' frequency and use of touches with speech change with infants' development. Results revealed that the frequency of word/utterance-touch alignment as well as word + touch co-occurrence is highest in speech addressed to the youngest group of infants. Thus, this study provides support for the hypothesis that caregivers' use of touch during dyadic interactions is sensitive to

infants' age in a way similar to caregivers' use of speech alone and could provide cues useful to infants' language learning at critical points in early development.

1 | INTRODUCTION

Child-directed speech has long been considered a core component of caregivers' interactions with their infants (e.g., Saint-Georges et al., 2013; Soderstrom, 2007). During face-to-face interactions, this speech signal is audio-visual and infants are sensitive to this signal and are capable of detecting the correspondences (e.g., Kuhl & Meltzoff, 1982, 1984; Patterson & Werker, 1999) and asynchronies (e.g., Lewkowicz, 2010) between auditory and visual streams. However, audio-visual speech is not the only multimodal component of the communication directed to infants and recent technological and methodological advancements that allow us to collect and analyze video-recorded interactions with infants have made it possible to start characterizing the full range of caregiver-provided cues available to infants during language learning. Thanks to this growing body of work, we now know that caregivers use gestures (e.g., Matatyaho & Gogate, 2008; Matatyaho-Bullaro et al., 2014; Rowe, 2008; Vigliocco et al., 2019), touches (e.g., Anisfeld et al., 1990; Feldman et al., 2002; Ferber et al., 2008; Nomikou & Rohlfing, 2011), and facial expressions (Nomikou & Rohlfing, 2011) when interacting with their infants, in ways that may help bootstrap language-specific learning (e.g., Abu-Zhaya et al., 2017; Gogate et al., 2000; Nomikou & Rohlfing, 2011; O'Neill et al., 2005; Tincoff et al., 2019). For instance, when teaching their infants novel object labels, mothers synchronize their naming events with object motion (Gogate et al., 2000, 2006; Matatyaho & Gogate, 2008), and they do so more frequently for object wholes than object parts, highlighting the whole object in their infants' visual fields (Gogate et al., 2013). While such findings demonstrate that child-directed communication includes an abundance of events that infants are sensitive to (Gogate et al., 2001), some communicative cues that accompany speech have received considerably less attention; one such cue is caregiver touch. In the current paper, we specifically explore how this rich social signal is coupled with speech at different developmental stages and whether such coupling may be beneficial for language learning.

Some of the previously mentioned streams of information appear to be sensitive to infant age and/or development. For example, multimodal labeling of target words, wherein caregivers combine object motion and touch with the production of a novel label, is more prevalent in the input to pre-verbal infants than to infants at more advanced stages of word learning (Gogate et al., 2000, 2015). However, the full extent and nature of these developmentally driven multimodal interactions, how they are used in different contexts, and their significance for language learning are not yet fully understood. The current study aims to fill part of this gap by investigating the distribution and temporal relations between tactile cues and words in child-directed communication in an ecologically valid cross-sectional sample of Korean mother-child dyads. We focus on touch since it is amongst the first senses to develop (by ~7 weeks gestation; Bremner et al., 2012; Bremner & Spence, 2017), is a key cue in social development and emotion regulation (e.g., Feldman et al., 2002; Stack & Muir, 1992), and may play a role in infant language learning (e.g., Seidl et al., 2015; Tincoff et al., 2019).

The mechanisms by which tactile cues might facilitate language learning are not fully understood, but evidence provides support for several possible ways through which touch could play a role in facilitating language learning. First, caregiver touches might aid infants in finding word boundaries by providing direct cross-modal, synchronous cues that may help to bootstrap segmentation of the speech stream. This possibility was explored in a laboratory study of speech segmentation in an artificial

language (Seidl et al., 2015), which showed that presenting infants with consistent touches synchronously aligned with trisyllabic statistical words (three syllable units with high transitional probability between syllables) facilitates infants' ability to extract words from the speech stream. This tight alignment between the boundaries of touches and speech units is not just extant in dummy languages used in the lab, but is also found in the real world. Specifically, using a naturalistic book-reading task, Abu-Zhaya et al. (2017) found that when interacting with infants, caregivers naturally align words and touches creating multimodal events that are temporally tight and non-arbitrary. These findings provide further support for the idea that touch might play a direct role in infants' word segmentation.

Another possible role that touch might play in facilitating language learning is that presentation of caregiver touches with words might aid in word-to-world mapping by boosting the links between words and their referents. This role is supported by a previous body of work showing that caregivers present multimodal cues including auditory-visual-tactile signals during novel word naming (Gogate et al., 2000), and that temporal synchrony between such cues, for example, a spoken word and a visible object/action, facilitates mapping of arbitrary speech signals to objects in pre-verbal infants (Gogate, 2010; Gogate & Bahrick, 1998; Gogate et al., 2006). Similarly, caregivers naturally use auditory-tactile synchrony in real speech when talking about body parts by aligning their production of body-part labels with touches on the location the labels refer to (e.g., touching the belly while saying *belly*; Abu-Zhaya et al., 2017). Given that touch is attention-getting (e.g., Goldin-Meadow & Saltzman, 2000; Jean et al., 2009; Jean & Stack, 2012), this alignment likely increases the allocation of attention to both signals (Gogate & Maganti, 2016; Tincoff et al., 2019). Since allocation of attention and contingency support language learning (e.g., Masek et al., 2021), this could facilitate infants' ability to map body parts to wordforms.

Yet, in order to fully understand the role of the combination of tactile cues with speech during dyadic interactions, it is crucial to explore how they are used across different developmental stages as the infant begins to learn different aspects of their input language (e.g., the child begins to segment words, learn native phonotactics, map wordforms to word meanings, and produce wordforms at different points in development). It has long been assumed that caregivers have a tendency to adapt their speech in accordance with the changing developmental needs of their language learning child (Newport et al., 1977; Snow, 1977; Sokolov, 1993). In particular, the so-called fine-tuning hypothesis of child-directed speech has postulated a noticeable change around the time the child begins to speak their first words, which is a milestone caregivers easily observe. In light of this hypothesis, it is important to consider developmental adaptations when discussing the characteristics of child-directed communication—be it the speech signal or its relationships with any other cue that accompanies it (Jo & Ko, 2018; Odijk & Gillis, 2021).

Such developmentally driven changes have also been reported for other cues utilized by caregivers during interactions with young infants. Gogate et al. (2000) reported that when mothers teach infants novel words, they produce a higher proportion of synchronous multimodal naming + object-motion events to pre-verbal infants (5- to 8-month-olds) compared to infants in the early-speech (9- to 17-month-olds) or multi-word (21- to 30-month-olds) stages. They also report a similar tendency in the use of tactile cues produced with an object rather than caregivers' hands. Similarly, Ferber et al. (2008) showed that the frequency of maternal touches during dyadic interactions decreases significantly in the second half of the first year of the infant's life. Together, these results show that several features in the input to infants change with development; they also emphasize the necessity of adopting a developmental perspective in analyzing the multimodal features of child-directed communication and how caregivers communicate with their infants during different developmental stages. Crucially, if caregivers are sensitive to infant development in their use of touch coupled with speech, then we can make the prediction that caregivers might unconsciously synchronize or couple touches with speech more frequently to infants in the pre-verbal stage

than to infants in later stages, thereby reflecting a recognition of infants' language skill at each point in development. Understanding caregivers' multimodal input and its potential impact on language learning is possible through observing caregivers' behaviors in ecologically valid environments; such an endeavor offers a close look at the natural cross-modal statistics available for infants while learning language. For example, in her ethnographic naturalistic observations of Mexican mothers' interactions with their infants, Zukow-Goldring (1997) showed that mothers' speech is well aligned with actions and touches (e.g., a mother may say "head" while tapping her daughter's head). Further, using free dyadic play interactions in the lab, Gogate et al. (2013) showed that caregivers use bimodal naming strategies more frequently for labeling whole objects than for labeling object parts. However, these results were obtained during specific points in development, and it is unclear whether any of these behaviors depend on the child's developmental stage. Studies that have examined the developmental adaptations and fine-tuning of mothers' use of multimodal cues (Gogate et al., 2000, 2015) have only obtained data in structured experimental settings. Though findings of maternal attunement in the use of multimodal cues in these contexts contributed to broadening our understanding of the role of multimodal input, the explicit novel word-teaching activity and the intervention of an experimenter do not align with most of infants' experiences in the real world. Thus, in the current study, the data were collected in a setting resembling a typical home environment without any experimental tasks or intervention of an experimenter, and examine caregivers' use of touch with speech across three stages of infant development.

Though much of the cross-modal work on the input to children (discussed above) has focused on Western languages and cultural contexts, we believe that in order to achieve a more comprehensive understanding of child-directed communication, it is essential to expand the contexts we investigate. Thus, in this study, we examine the developmental trajectory of caregivers' use of touch along with child-directed speech in Korean mother-child dyads. The Korean language is typologically distinct from English. For example, verbs come after objects and sentential elements such as subjects and objects are not as frequently realized. Korean parents' interactions with their children may also be different from those observed in Western English-speaking homes. Traditionally, Korean parenting approaches tend to have educational goals (Farver et al., 1995; Kim & Choi, 1994); yet, in recent years, researchers are documenting an increasing recognition of the value of play in children's learning and development (Kim, 2000). In some respects, however, Korean dyads seem to show similar behavioral patterns to those observed in Western cultures; specifically, Korean mothers' touches elicit positive responses from infants (Kwak et al., 2005) in a way that is similar to what was reported in Western cultures (e.g., Peláez-Nogueras et al., 1996; Stack & Muir, 1992). Given the similarities and differences in Korean and Western culture, it would be essential to expand the context in which we examine the touch-speech connection and explore its potential universality.

Our data are sampled from caregivers of three age groups of infants corresponding to three stages of infant language development: pre-verbal (prior to production of first words; 8 months), early-speech (around production of first words; 14 months), and multi-word (production of multi-word phrases; 27 months). Note that these language development stages also correspond to distinct three stages in motor development, that is, sitters, new locomotors, and experienced movers and that these motor development stages may also impact caregiver behaviors with respect to touch and speech (e.g., more touch to infants who are less mobile and more directive/imperative language to children who are more mobile). Thus, we may expect to see differences in caregiver touch-speech behavior if caregivers are sensitive to infants' linguistic development and/or motor development.

Our design is similar to Gogate et al. (2000, 2015) in examining synchrony and its facilitative role in word learning, but we also look into the possibility that touches might aid in segmentation of the speech stream, which is a prerequisite for word learning (e.g., Singh et al., 2012). In order to explore the facilitative role that touch might play in segmentation, we examine the temporal relationship

between touches and words/utterances in caregivers' input, and explore the developmental trajectory of this phenomenon. Our exploration of the temporal relationships between touch and speech includes a differentiation between the two constructs of co-occurrence (roughly, events that co-occur are those that overlap temporally) and alignment (a more specific measure of co-occurrence wherein the edges of events are finely aligned). We first establish that there is a non-random relationship between words and touches (replicating Abu-Zhaya et al., 2017 with a new sample) by testing if the boundaries of words and touches align at a level greater than they would by chance. We then investigate three questions focusing on how caregivers' use of tactile and speech cues throughout development might help bootstrap segmentation (words and utterances) and word learning: (1) Do words co-occur with touch more frequently in early development versus later development? (2) Is there evidence that mothers' alignment of tactile cues with words changes over development? And (3) Does the boundary-marking function of touch for word boundaries extend to the utterance level? We predict that caregivers' production of multimodal touch and speech events will occur at a likelihood higher than chance, in line with previous work (Abu-Zhaya et al., 2017). We also predict, based on the literature reviewed above (e.g., Gogate et al., 2000, 2015), that caregivers will produce a higher frequency of multimodal touch and speech events in the input to younger infants than to older infants, and that these multimodal events will be more likely to be aligned or to co-occur in the input to younger infants.

2 | METHODS

2.1 | Participants

Data from participants in our study were collected as part of a multimodal corpus of 35 Korean mother-child interactions, including 21 boys and 14 girls (Ko et al., 2020). Families were recruited in Seoul through online advertisements and word-of-mouth, and were paid for their participation. The study was conducted in line with the guidelines laid out in the Declaration of Helsinki, with written informed consent obtained from a parent for each child before data collection. All procedures involving participants in this study were approved by the Institutional Review Board of Seoul National University (Approval No. 1602/001-007). Among the 35 mothers, 2 were high school graduates, 22 had a college degree and 11 were pursuing or had obtained an advanced degree. Participants were divided into three age bands according to infants' expected developmental language stage, similar to those tested in Gogate et al. (2000) but with slightly different target ages: pre-verbal (0; 06.2–0; 09.23, M = 0; 08, 12 dyads), early-speech (0; 11.14–1; 04.3, M = 1; 02, 11 dyads), and multi-word (2; 01.10–2; 06.24, M = 2; 03, 12 dyads). Data from an additional dyad from the early-speech group were not included in the analyses due to frequent crying of the child. Note that the age intervals between these groups are not equal: there is a 4-month gap between the pre-verbal and the early-speech groups whereas the gap between the early-speech and multi-word groups is 13 months. These unequal intervals are intentional since our focus was on capturing caregiver behavior at key language development points rather than caregiver differences that may vary with chronological age.

2.2 | Procedure

Each dyad was asked to play freely for about 40 min in an 800 square ft. mock-apartment laboratory. The apartment contained a living room, a bedroom, a kitchen, and a foyer equipped with comfortable furniture and a variety of age-appropriate toys/books for young children. The mother and child each

wore a vest equipped with a lavalier microphone plugged into a small digital recorder in the pocket. Caregiver and child recordings were later merged to form a stereo recording. Dyadic interactions were video-recorded via four wall-mounted cameras, two in the living room, one in the bedroom, and the fourth one in the kitchen. Audio and video data were later synchronized based on a sharp clap of hands recorded by the experimenter on all devices simultaneously at the beginning of the session.

We did not give mothers any particular instructions as to what specific task to perform, but simply asked them to spend time with their children like they would normally do at home. We also did not mention our interest in caregiver touch and its use with child-directed speech. Participants were left alone in the apartment during this free play period and were observed by the researchers in a separate room via wall-mounted cameras. The researchers did not communicate with the participants during the play interaction. The free-play session was followed by an adult-adult communication session, but none of the adult interactions were included in the current analyses (see Ko et al., 2020 for further information about the corpus). We used the video recordings captured by the wall-mounted cameras to analyze touch events, and the audio data captured by the clip-on microphones for forced-alignment and transcription. We selected a *6-min* segment (in line with previous studies that examine linguistic features of caregiver-infant interactions; e.g., Bornstein et al., 2004; Tamis-LeMonda et al., 2013) from each 40-min video-recorded free play session. We ensured that both the mother and child remained visible without interruption, and that the mother's hands and her touching actions were clearly visible during those 6 min. The activities we captured during those 6-min segments mostly involve the dyads playing with toys in the living room in close proximity to each other.[1]

2.3 | Transcription and forced-alignment of text-to-audio

The audio recordings of mother-child interactions were transcribed in the CHAT format (MacWhinney, 2000) based on criteria described in Ko et al. (2020). Since one of the main goals of this study was to investigate the alignment patterns between touch and word boundaries, we needed to segment the speech into words, which we defined as a unit of morpheme(s) separated by space in transcription. We facilitated this task by a forced-alignment of audio and text using a Python script package developed by Gorman et al. (2011), which performed acoustic model training and alignment. The results of the forced-alignment were output in Praat TextGrid format (Boersma & Weenink, 2020). The default output contains a phone and a word tier, in which the speech stream is segmented into the corresponding linguistic units. After forced alignment, word boundaries within the target 6-min segments were manually adjusted by research assistants supervised by the first author to allow for a more accurate analysis of the alignment and overlapping patterns between touch events and words.

2.4 | Coding touch events

We coded caregiver touch behaviors during the 6-min segments using version 5.9 of ELAN (2020). Coding was done on ***silent videos***, and was carried out in two phases. Thirteen of the 6-min video segments relied on video data obtained from a single camera for coding the touch events; these

[1]One reviewer asked about the representativeness of our 6-min segments selected compared to the remainder of the recording session. While we did not explicitly explore this, dyads varied in the distribution of their activities, partly conditioned by infants' mobility. For all infants, the touch events we coded usually occurred in a sedentary setting, thus a dyad with a less mobile infant might present with a greater accumulated number of touches than another with a greater mobility. However, our analysis of word-touch alignment is based on the *ratio* rather than raw frequency, so the relation between touches and words captured in the 6-min segments can be regarded to be representative of the dyad's interactions beyond the analyzed segments.

segments were coded by research assistants at Purdue University under the supervision of the second author. For the remaining 22 videos, we had to use videos from more than one camera to accurately capture the events. This second phase of coding was completed by research assistants at Chosun University, supervised by the first author. All coding was done based on the manual in Abu-Zhaya et al. (2017, 2019), and measures were taken to ensure consistency as described below. We only coded non-accidental touch events produced by the mothers either with or without a toy on *any part* on the infant's body. Touch events that were judged as being produced by accident, as in a mother accidently brushing her child while trying to fix the vest equipped with the recording devices or reaching out for an object near the child but hitting the child by mistake, were not coded. We also *did not* code infant touches on the mothers' body, for example, an infant grabbing the mother's finger or hand. We marked the start and end of each touch event, along with information about which hand was used to deliver the touch.[2]

To ensure consistency of coding across different coders in the two phases, we randomly chose 5 out of the 13 dyads coded by assistants at Purdue University, and used them to train the lead graduate student annotator at Chosun University in consultation with the manual in Abu-Zhaya et al. (2017). We then had the lead annotator at the Large Korean University independently code the remaining 8 dyads, which included 139 touch events, to estimate the reliability between the coding schemes in the two sets of measurements. We first compared the accuracy of segmenting the touch events by calculating the differences in the time stamps in demarcating the beginning and the end of the touch events. The absolute difference between the two sets of the coding in marking the beginning and the end of each touch was 19 ms (SD = 81 ms) and 4 ms (SD = 340 ms), respectively. The relatively large SD at the offset of the touch events turned out to be driven by one instance, which had a disagreement of over 3.8 s. After eliminating this particular token, the SD for the offset was reduced to 96 ms. The Pearson correlation coefficient for the touch duration between two sets of coding (from the two institutions) was 0.95, and the results of Cohen's kappa test carried out for the categories of touch events fell within the range of "very good" ($\kappa = 0.892$, $p < 0.001$). The lead annotator at the Large Korean University trained two undergraduate students to complete coding the remaining 22 dyads. These three annotators had established a satisfiable level of agreement via Fleiss's Kappa test ($\kappa = 0.844$, $p < 0.001$) in a separate training session. After completion of the first pass coding, the lead annotator inspected the entire dataset for quality. On a small number of occasions when substantial deviations from her judgment were found, she consulted with the original coder and/or the first author to reconcile or correct them.

2.5 | Merging audio and video data

The alignment of audio to text and the touch coding based on the video, carried out independently from each other, were merged into one Praat TextGrid. For the touch coding, we exported a Praat TextGrid file from each ELAN file. This was then merged with the Praat TextGrid containing the word boundaries, as schematically shown in Figure 1. Thus, the merged TextGrid included a word tier in which word boundaries were demarcated (words tier in Figure 1) and two tiers containing information for touch events imported from ELAN (the right and left hand tiers in Figure 1). These text-grids (available at https://osf.io/v5hsm/ along with other data and analysis scripts) were the basis for our

[2]We also annotated further details about each touch event, for example, their type (e.g., brushing, poking) and location (e.g., belly, nose), and the number of repetitions of each touch type (e.g., 5 taps, 3 brushes) in line with the manual detailed in Abu-Zhaya et al. (2017, 2019), but we do not report on these analyses here.

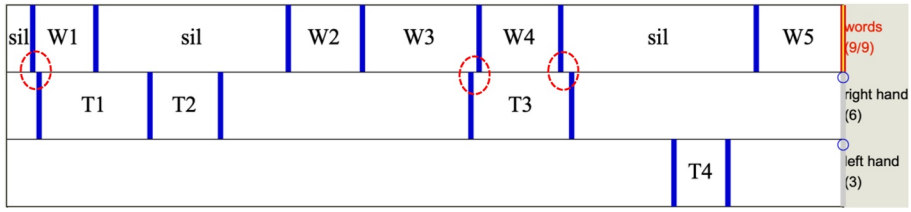

FIGURE 1 A schematic illustration of multimodal coding. The words tier contains word (W) and silence (sil) boundaries, the right and the left-hand tiers represent touch events made with the right and the left hand, respectively. The red dashed circles indicate that the lag between the word and touch boundaries is less than 500 ms, and thus each circled boundary can be classified as being aligned, for example, onset of W1-T1 and W4-T3, offset of W3- onset of T3, and offset of W4-T3. A co-occurrence relation is simply defined by an overlap, for example, W1+T1, W3+T3, W4+T3.

further coding of the word-touch alignment and word + touch co-occurrences, as explained in the next section, and analyses to investigate the use of tactile cues and their temporal relation with words. The frame rate of the video was 30 frames per second, while the sampling rate for the audio recording was 44.1 K/sec. Thus, the audio signal has greater precision than the video. Nevertheless, given the frame length of 33.3 ms in the video, our safety margin of 500 ms should allow us to capture any robust alignment patterns between touches and words.

2.6 | Temporal relationships between words and touches: Co-occurrence and alignment

We define the temporal relation between the occurrences of touch and word in terms of the two closely related, yet separate, notions of *co-occurrence* and *alignment*. The notion of *co-occurrence* focuses on the simultaneous presentation of multi-modal cues in word and touch, regardless of the alignment at the edges. *Alignment* focuses on the onset and offset boundaries of word and touch events, and is a more specific measurement of co-occurrence because for a touch to align with a word it must also co-occur with the word. We consider a touch and a word to co-occur if there is a temporal overlap at any point during the presentation of touches and words, including the 500 ms window at either edge (as was done in Abu-Zhaya et al., 2019). In comparison, we consider auditory and tactile signals to be *aligned* if the edges of a word and a touch both occur within a 500 ms window (again, similar to Abu-Zhaya et al., 2019). Thus, if a short word occurred during the middle part of a longer touch with a greater than 500 ms lag between the word and touch boundaries on *both* ends, it would get counted as co-occurring but not as aligned because there was an overlap but the co-occurrence did not happen at the edges. Crucially, both co-occurrence and alignment might help modulate infants' attention (Tincoff et al., 2019), which could reinforce mothers' behavior. These two measures of caregiver touch are tightly related but could serve somewhat different purposes for infants' learning.

In analyzing the temporal relationships between words and touches, there can be two approaches in setting the anchor between the two modalities. We could view the relationship from the perspective of a word, and examine whether there is a touch event during or in alignment with the production of the word. Alternatively, we could approach it from the perspective of touch, and examine if a word occurs during or in alignment with the touch. In this paper, since our focus is on caregiver cues provided to infants in ways that may impact segmentation and word learning, we mainly take the former approach, that is, we examine whether each word/utterance is temporally aligned with a touch event.

Using a custom-written Python script, we calculated the onset and offset lag between word/utterance and touch edges by subtracting the time of the onset or offset of the touch from the time of the onset or offset of the word/utterance, respectively. We then coded the results on a binary scale. We represented the alignment relation as 1 if the lag duration between the onset or the offset of word/utterance and touch fell within 500 ms of each other; if the edges of word/utterance and touch were more than 500 ms apart, they were judged to be non-aligned and were coded as 0. According to this criteria, W1 in Figure 1 would be coded as 1 both for alignment and co-occurrence while W3 would be 0 for alignment and 1 for co-occurrence. We also recorded the absolute duration of the lag between the word and touch boundaries at either edge in milliseconds.

3 | RESULTS

3.1 | Descriptive statistics of the data

Out of the 9737 word tokens in the corpus, a total of 1596 words (16.4%) occurred with touch events, 1514 with the right-hand, and 82 with the left-hand.[3] The average number of words that were coupled with touch in each dyad was 43.25 (SD = 61.75) and ranged from 1 to 313 (see Table 1 for details about the distribution of words and touches in the three age groups).

The duration of touches (M = 20.6 s, SD = 38.1 s) was much longer than the duration of words (M = 482 ms, SD = 373 ms). As will be shown later, therefore, words often co-occur with touches but the ratio of their boundaries aligning with each other (word-touch alignment) is much smaller.

3.2 | Word-touch co-occurrence and word + touch alignment in natural versus shuffled word orders

The main goal of our study was to examine the use of caregiver-provided touch cues along with Korean child-directed speech during three developmental stages. Before we investigate the differences in the relation between touch and words across developmental stages; however, we must examine whether or not there is an arbitrary relation between touches and words. In previous research, Abu-Zhaya et al. (2017) found that word and touch boundaries are aligned more tightly in mothers' interaction with their infants than would be expected by chance. We, therefore, first investigated if the non-arbitrary alignment of speech and tactile cues found in American mothers' interactions with their infants is also found in our Korean sample.

TABLE 1 Mean number of word tokens, touches, and their overlap within the 6-min segments in three developmental groups (standard deviations are in parentheses).

	Pre-verbal	Early-speech	Multi-word
Words	265.92 (87.18)	282.45 (94.73)	290.92 (107.30)
Touches	52.50 (38.68)	17.36 (14.49)	12.25 (9.28)
Words co-occurring with touch	96.0 (89.62)	25.55 (17.22)	14.42 (12.99)
Touches co-occurring with words	45.83 (34.59)	15.64 (13.15)	8.58 (5.82)

[3]Only one of the mothers was left-hand dominant, but we did not concern ourselves with handedness or the hand that touched the infant because we were more interested in the infants' experience of these touches than the caregivers' manual use.

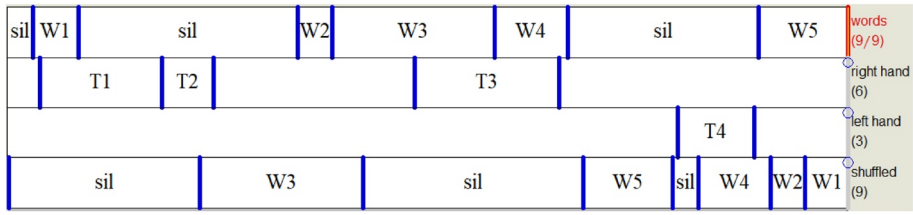

FIGURE 2 Screenshot demonstrating the shuffling procedure. Word (W) and silence (sil) intervals in the words tier are shuffled to generate randomly distributed word strings in the shuffled tier.

The purpose of the analysis in this section is to test if the probability of touch co-occurring or aligning with words is greater than would be expected by chance. There can be various ways of generating unnatural (henceforth shuffled) strings of events whose statistical likelihood is compared against the natural word-touch relations. Here, we implemented two different ways of generating random distributions of word-touch relations. We first adopted the method in previous research (Abu-Zhaya et al., 2017) by shuffling the words and silences within the same dyad to create multiple distributions. We then conducted a second randomization process by shuffling the word or touch tier across dyads.[4] In this approach, the naturally occurring alignment relation between touch and word would be disrupted while the temporally natural structure of communication remains intact.

To replicate the methods in Abu-Zhaya et al. (2017), we used a custom-written Python script and generated random distributions of word order within the 6-min segments by shuffling the words and silence intervals in mothers' utterances. The shuffling procedure leads to a disruption of the association between the word and the touch tiers, providing a novel distribution of the relation between the word and the touch tier while ignoring any underlying connection between the two. An example of this procedure is shown in Figure 2, where the duration of words and silences are intact regardless of the shuffled concatenation. Note that the domain of the shuffling was not at the individual utterance level but the level of the entire duration of the interaction. We applied this shuffling procedure 34 times for each of the 35 files, generating 1190 random distributions of word order from the dataset to be compared with the set of 35 natural alignment patterns between word and touch in the unshuffled files.

We constructed a logistic regression model for the co-occurrence between words and touches using the lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017) package of R (R Core Team, 2019). A logistic regression model was constructed with the CO-OCCURRENCE RELATION (0 = word only, 1 = word + touch co-occurrence) as the dependent variable, WORD ORDER (0 = natural, 1 = shuffled) as a fixed effect, and PARTICIPANT as a random effect (glmer(co-occurrence ~ order + (1|participant), family = "binomial")). We also constructed models to inspect the alignment relation between words and touches for both the onset and offset of each word. The onset and offset of each word were coded in a binary manner as to whether they were aligned with touch or not within the 500 ms window. We entered ALIGNMENT (0 = not aligned, 1 = aligned) as the dependent variable, WORD ORDER (0 = natural, 1 = shuffled) as a fixed effect, and PARTICIPANT as a random effect (glmer(align_onset ~ order + (1|participant), family = "binomial"), glmer(align_offset ~ order + (1|participant), family = "binomial")).

The results of the logistic regression models are shown in Table 2. The log-odds of word + touch co-occurrence was negative for the shuffled word order. This suggests that the log-odds of words to overlap with touch is greater in the *natural* sample than in the shuffled one. In addition, at both edges, the log-odds of word-touch alignment was negative for the shuffled word order. Since the dependent

[4]We thank an anonymous reviewer for suggesting this second type of data randomization.

TABLE 2 Alignment and co-occurrence of words and touches in natural and shuffled samples.

	Estimate	Std. error	*z* value	Pr(>|z|)
Co-occurrence				
(Intercept)	−2.01	0.12	−16.77	<0.001
Order(shuffle)	−0.47	0.10	−4.76	<0.001
Alignment at onset				
(Intercept)	−3.30	0.12	−28.55	<0.001
Order(shuffle)	−0.37	0.13	−2.93	<0.01
Alignment at offset				
(Intercept)	−3.33	0.11	−29.72	<0.001
Order(shuffle)	−0.27	0.12	−2.21	<0.05

TABLE 3 Proportional duration of word + touch overlap in natural and shuffled samples.

	Estimate	Std. error	*df*	*t* value	Pr(>|*t*|)
(Intercept)	17.35	2.63	79.22	6.59	<0.001
Order(shuffle)	−9.84	1.51	8257.8	−6.53	<0.001

variable ALIGNMENT is a dichotomous variable coded 1 if there is alignment and 0 if not, the negative coefficient suggests that the log-odds of word-touch alignment are smaller in the shuffled word order than in the natural one. These results show that the alignment and co-occurrence between words and touches in mothers' interactions with their children were not likely to be by chance.

Further, we tested how the ratio of the word duration that overlaps with touch compared between the natural and the shuffled word orders. We constructed a mixed effects linear model with the percentage of word + touch OVERLAP DURATION as the dependent variable, WORD ORDER (0 = natural, 1 = shuffled) as a fixed effect, and PARTICIPANT as a random effect (lmer(overlap_duration ~ order + (1|participant))). The results (see Table 3) show that the proportion of overlap was significantly lower in the shuffled dataset than in the natural dataset ($p < 0.01$).

We implemented an additional method of generating the random distribution data for the relation between words and touches by replacing the touch tier across dyads. Out of the 35 participants, we set aside one dyad at a time and took the remaining 34 dyads to extract the touch tiers and randomly assign one to each dyad. An iteration of the process for each of the 35 participants yielded 1190 randomly distributed associations between words and touches.

We constructed a logistic regression model for the co-occurrence in exactly the same way as described above for the first shuffling process (glmer(co-occurrence ~ order + (1|participant), family = "binomial")). The results were very similar to the outcome of the first shuffling method, showing a significantly greater ratio of co-occurrence in the natural word-touch relation than in the random distribution ($b = -0.18$, Std. Error = 0.04, $z = -4.84$, $p < 0.001$). The model for the alignment at the onset and offset of words (formula = (glmer(align_onset ~ order + (1|participant), family = "binomial"), glmer(align_offset ~ order + (1|participant), family = "binomial")) also confirmed a significantly greater ratio of word-touch alignment in the data with natural word order than in the data with random word-touch mapping in the onset ($b = -0.24$, Std. Error = 0.05, $z = -4.46$, $p < 0.001$) and the offset ($b = -0.28$, Std. Error = 0.05, $z = -5.13$, $p < 0.001$).

These results show that both the natural alignment between word and touch boundaries and co-occurrence of word and touch in the Korean dataset is not random. Note that the main purpose of the analyses conducted in this section is to test whether or not the relation between word and touch

is random. Having firmly established that these relationships are not random, we proceed to a more systematic investigation of the relation between the two signals across the developmental stages.

3.3 | Word + touch co-occurrence across development

Given the evidence provided above that the temporal relationship between word and touch boundaries is non-arbitrary, we can explore whether the co-occurrence and alignment of touches with words change with infants' age. Specifically, and in line with the fine-tuning hypothesis of the input to children, we predict that the frequency of such word + touch co-occurrence would be higher when the child is younger. To test this prediction, we investigated if the proportion of words co-occurring with touch in the input differs across the groups. We first fitted a linear regression testing the effects of age on the number of words and touches, respectively. The model-estimated number of words within the 6-min segments in the pre-verbal group was 265.92 (SE = 27.95), which was statistically not different from the estimates in the early-speech ($\beta = 16.54$, SE = 40.42, $t = 0.40$, $SD = 0.69$) or the multi-word ($\beta = 25.0$, SE = 39.53, $t = 0.63$, $SD = 0.53$) group. The estimated number of touch events within the 6-min segments in the pre-verbal group was 52.5 (SE = 7.12), and was significantly greater than the estimate of touches in the early-speech ($\beta = -35.1$, SE = 10.3, $t = -3.41$, $p < 0.01$) and the multi-word group ($\beta = -40.3$, SE = 10.1, $t = -3.99$, $p < 0.001$). With the constant number of word tokens ($F(2, 32) = 0.21$, $p > 0.81$) and decreasing number of touch events ($F(2,32) = 8.07$, $SD = 0.001$, adjusted $R^2 = 0.29$) across the three groups as shown in Table 1, it might already seem that the frequency of word + touch co-occurrences declines with children's development. Our goal is to compare the ratio of words that occur with touches beyond the raw frequency of touches.

As described earlier, each word in our data was coded for its overlap with touch dichotomously (1 = word + touch, 0 = word only). The proportion of words that occurred with touches was calculated by dividing the number of words that occurred with touches by the total number of words. Figure 3 shows that the proportion of words occurring with touch was greatest in the pre-verbal group ($M = 0.38$, $SD = 0.29$), followed by the early-speech ($M = 0.10$, $SD = 0.06$) and the multi-word group ($M = 0.05$, $SD = 0.05$). In sum, an average of 38% of words in the speech addressed to children in the pre-verbal stage overlapped with touch, whereas an average of only 5% of the words in child-directed communication occurred with touch in speech addressed to the multi-word group.

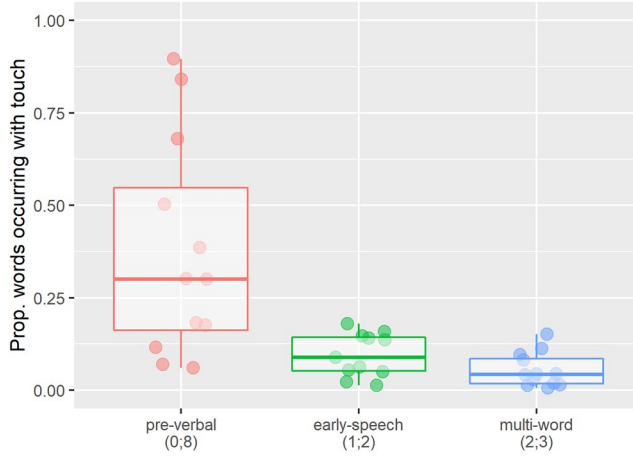

FIGURE 3 The proportion of words that co-occur with touch across development. Individual data points are jittered to avoid overlaps and show details.

We fitted a linear regression model to statistically test the group differences. The distribution of word + touch co-occurrence ratio in the 35 participants was skewed with a long right tail, thus we log-transformed it. The dependent variable of the model was the log-transformed PROPORTION OF WORDS occurring with touch events, and the fixed effect was AGE. A significant regression equation was found ($F(2,32) = 14.31$, $p < 0.001$) with adjusted $R^2 = 0.39$. The results show that the proportion of word + touch co-occurrence in the pre-verbal group ($M = 0.38$, $SD = 0.29$) is significantly higher than the early-speech ($M = 0.10$, $SD = 0.06$; $\beta = 1.29$, $t = 3.4$, $SD = 0.002$) and the multi-word group ($M = 0.05$, $SD = 0.05$; $\beta = 1.99$, $t = 5.3$, $p < 0.001$).

3.4 | Word-touch alignment across development

We additionally asked if word-touch alignment changes across developmental stages. We predicted that word-touch alignment might be more frequent in speech directed to pre-verbal infants, as compared to older children. To test the prediction that age impacts caregivers' behaviors, we analyzed the proportion of words that align with touch in the three age groups. For each participant, we calculated the proportion of word-touch alignment by dividing the number of words coded as 1 (words for which there is a touch boundary within 500 ms) by the total number of words for each of the onset and offset of the word.

To test the difference in word-touch alignment in the three groups, we fitted a multiple linear regression model with the PROPORTION of word-touch alignment as the dependent variable, DEVELOPMENTAL GROUP (pre-verbal, early-speech, and multi-word), EDGE (onset/offset), and the interaction between these two as fixed effects. Our model was statistically significant ($F(5,64) = 13.7$, $p < 0.001$), with adjusted $R^2 = 0.48$. We found a significant main effect of DEVELOPMENTAL GROUP, such that the early-speech ($M = 0.04$, $SD = 0.03$; $\beta = -0.05$, SE $= 0.01$, $t = -3.39$, $p < 0.01$) and the multi-word groups ($M = 0.02$, $SD = 0.01$; $\beta = -0.07$, SE $= 0.01$, $t = -5.29$, $p < 0.001$) were presented with a significantly smaller proportion of word-touch alignment compared to the pre-verbal group ($M = 0.09$, $SD = 0.04$; see Figure 4). There was no effect of EDGE ($t = 0.65$, $SD = 0.52$) nor were there any interactions between EDGE and DEVELOPMENTAL GROUP (t's < -0.35, p's > 0.39).

3.5 | Utterance-touch alignment across developmental stages

We investigated if the demarcative function of touch for words extends to utterances. For the 3473 utterances in the corpus, we binary-coded the alignment relation with touch at the beginning and the end of each utterance. Our definition of utterance in this analysis was based on the transcription guideline as detailed in Ko et al. (2020). Similar to word-touch alignment analyses, we judged an utterance and a touch to be aligned if a same-labeled edge (offset and offset, onset and onset) fell within a 500 ms window from one another (utterance-touch aligned = 1, not aligned = 0). Figure 5 shows a similar developmental trend to word-touch alignment with utterance-touch alignment, with the greatest ratio of alignment occurring in speech to the youngest group. Further, it seems that the ratio of alignment is greater at the end of utterances than the beginning.

To statistically test this observation, we fitted a multiple linear regression model with the PROPORTION of utterance-touch alignment as the dependent variable, DEVELOPMENTAL GROUP (pre-verbal, early-speech, and advanced lexical), EDGE (onset, offset), and the interaction between these two as fixed effects. The model was statistically significant ($F(5,64) = 8.68$, $p < 0.001$) with adjusted $R^2 = 0.36$. There was a main effect of DEVELOPMENTAL GROUP because the proportion of alignment was significantly lower in the early-speech ($\beta = -0.06$, SE $= 0.02$, $t = -2.65$, $SD = 0.01$) and the multi-word ($\beta = -0.07$, SE $= 0.02$, $t = -3.37$, $p < 0.01$) group versus the pre-verbal group. There was also

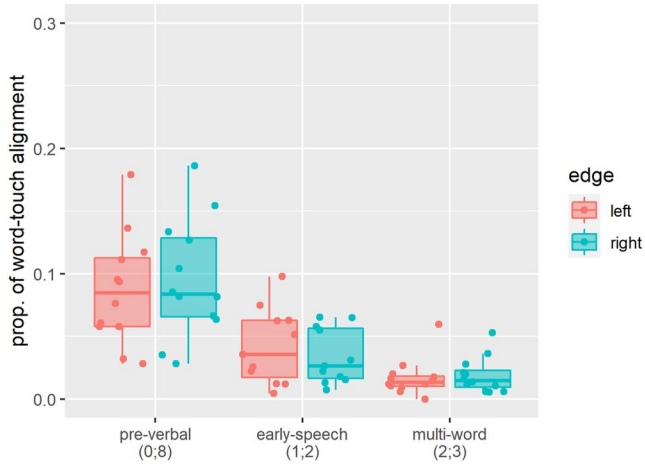

FIGURE 4 Proportion of word-touch alignment at the onset (red) and offset (blue) of words in each developmental group. The proportion of word-touch alignment was calculated by dividing the number of words whose boundaries occur with touch boundaries within 500 ms by the total number of words produced by caregivers in each dyad. Individual data points are jittered to avoid overlaps and show details.

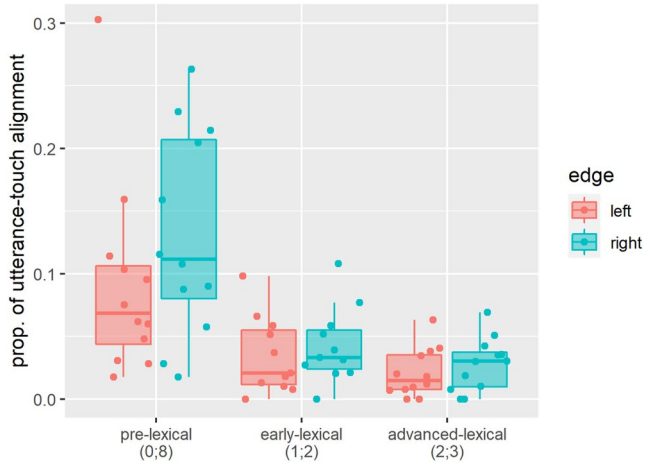

FIGURE 5 Proportion of utterance-touch alignment at the onset (red) and the offset (blue) of utterances in each developmental group. The proportion of utterance-touch alignment was calculated by dividing the number of utterances whose boundaries occur with touch boundaries within 500 ms by the total number of utterances in each dyad. Individual data points are jittered to avoid overlaps and show details.

a marginal effect of EDGE such that the proportion of the utterance-touch alignment was higher at the offset than the onset of utterances ($\beta = 0.04$, SE = 0.02, $t = 1.90$, $SD = 0.061$). There was no significant interaction between the two fixed effects.

4 | DISCUSSION

The last 20 years of research have revealed that language input to children is multimodal (e.g., Gogate et al., 2000, 2015; Nomikou & Rohlfing, 2011; Vigliocco et al., 2019) and that caregivers adapt their use of such cues according to infants' development (e.g., Gogate et al., 2000, 2015). Further,

the alignment and co-occurrence patterns of speech and non-speech cues, for example, tactile cues, within these multimodal events are systematic (e.g., Abu-Zhaya et al., 2017). However, this body of work still leaves many questions unanswered, for example, do tactile cues help infants tackle the feats of segmentation and word learning? does the use of tactile cues with speech change over the child's development? and what do such changes mean for language learning? The current paper aimed to begin addressing these gaps by focusing on the use of tactile cues coupled with speech in caregivers' input to infants at three key developmental timepoints within the first 2 years of life. Since we examine spontaneous interactions in a setting closely resembling the natural home environment, our results allow us to chart the natural history of caregivers' use of multimodal input.

Our results showed that, similar to previous findings for English-speaking North American mothers (Abu-Zhaya et al., 2017), Korean mothers produce multimodal touch and word events that are more likely to co-occur and be aligned than would be expected by chance. Crucially, this non-arbitrary behavior is more common for caregivers addressing infants in early development as opposed to later development. Specifically, and in line with results by Gogate et al. (2000, 2015) on developmental changes in the co-occurrences of object labels with object motion and touch, our findings also revealed that the proportion of words that occur with touch is significantly higher in the pre-verbal group compared to the later ages. Further, mothers of pre-verbal infants in our sample produce word-touch alignment and word + touch co-occurrence more frequently compared to mothers of infants at the early-speech and multi-word stages. Last, the demarcative function of touches indicating the boundaries of words or phrases was not only found for word boundaries, but was also found at the utterance-level, with the frequency of the utterance-touch alignment being the highest in the pre-verbal group. These results add to a growing body of work focusing on accurately characterizing infants' language environments (e.g., Abu-Zhaya et al., 2017; Nomikou & Rohlfing, 2011; Vigilocco et al., 2019), and how language and language-related input change throughout development (e.g., Gogate et al., 2000, 2015) by demonstrating developmental relationships with caregivers' use of tactile cues aligned and coupled with speech cues.

Though the current results do not allow us to provide evidence for a direct link between multimodal touch-speech cues with speech segmentation and word learning, analyses using the same dataset suggest that mothers who provide more frequent touches also offer additional enhancements at other linguistic levels that may be helpful for speech segmentation and word learning. For example, Ko et al. (2022) report that there is a significant correlation between the ratio of word-touch alignment and the ratio of tag-repetition, a syntactic manipulation thought to help word learning, within the same mothers. Thus, mothers who provide cues beneficial for language learning seem to be consistent in their behavior to provide such cues in several different domains. Such consistent behaviors create multimodal regularities in the input that might ultimately enhance the child's language outcomes. These findings are compatible with the theoretical model in which infants' perceptual detection of invariance (i.e., selectively attending to patterns and regularities in the input) serves as a perceptual gateway to the development of language (Gogate & Hollich, 2010). Crucially, the model suggests that patterns do not need to occur in a 100% stable fashion for them to be detected; it is sufficient for the learner to receive some consistency within particular contexts.

One might question if the relatively small ratio of word-touch alignment, that is, around 9% in the pre-verbal group, might have a beneficial effect on infants' word segmentation. Yet, such low frequency is found in many other input features that have been suggested to be beneficial for infants' language learning. For example, single-word utterances are believed to play an important role in infants' word segmentation and learning (Brent & Siskind, 2001; Lew-Williams et al., 2011) but the ratio of isolated words typically constitutes 7%–10% of caregiver utterances (Brent & Siskind, 2001; Fernald & Morikawa, 1993). Similarly, it is commonly argued that book reading activities are

beneficial for children's word learning but book reading accounts for only a very small proportion, for example, 1%–2% (Soderstrom & Wittebolle, 2013), of children's daily activities. Considering these examples, the relatively small proportion of alignment between words and touches might not be trivial in its role in facilitating infants' word learning, especially if it occurs in particular contexts. Furthermore, despite touch-speech alignment being rarer than speech input alone, it could be a particularly powerful cue for learning as it garners infants' attention (e.g., Jean & Stack, 2012; Jean et al., 2009; Peláez-Nogueras et al., 1996; Tincoff et al., 2019). Further, it is important to note that rarity and/or surprise can make structures more noticeable and learnable, so if word-touch alignment is relatively rare as compared to speech-alone input, then occurrences of word-touch could be particularly helpful to the young infant learner given their ability to garner attention and aid in learning. Specifically, unexpected or rare stimuli show clear differences in brain responses (e.g., the P3 in oddball tasks; Squires et al., 1975; Sakamoto & Love, 2006); and surprising or rare events can induce better learning (e.g., Brod et al., 2018; Vergilova et al., 2022). These explanations are consistent with the long-held idea that invariance detection is enhanced amid variation (Gibson, 1969; Gogate & Hollich, 2010). Since both prosodic and touch cues are strengthened (e.g., longer) when well aligned (Abu-Zhaya et al., 2017), these strengthened cues present during touch-word alignment could be especially helpful in word segmentation even in the absence of a high proportion of touch-word alignment. Finally, it is worth mentioning that the proportion of word + touch co-occurrence is substantially higher (e.g., $M = 38\%$ in the prelexical group, Figure 4) than the proportion of word- or utterance-touch alignment, which may suggest that touch cues could be particularly useful in mapping wordforms to word meanings in the youngest group of infants to whom this frequency is highest.

Our finding of utterance-touch alignment is one of the first reported in the literature (though see similar findings from 3 month olds, wherein caregivers align the edges of their utterances with the onset and offset of facial expressions; Nomikou & Rohlfing, 2011). Though edges of utterances are necessarily edges of words, we examined the subset of those word edges that also demarcate the beginning and the end of an utterance. We found a similar age-dependent developmental change in the ratio of the utterance-word alignment, which suggests that touches might serve a demarcative function for both words and hierarchical linguistic structures above the word level. Interestingly, caregivers' alignment of touch and speech was more frequent at the offset than the onset of an utterance. This suggests either a sort of final strengthening similar to a language-general utterance-final lengthening (Koponen & Lacerda, 2003) or a language-specific relation between touch and verbs in the verb-final Korean language. Further examination of this dataset and datasets from non-verb-final languages will be needed to differentiate these hypotheses.

We now consider the causes behind caregivers' use of touch-speech multimodal behaviors and age-related adaptations of these behaviors. We suggest that the close alignment between touch and word boundaries, and the proportion in which they co-occur early in development, are rooted in caregivers' modulation of multimodal invariant cues for word learning (Gogate & Hollich, 2010) during the interactive alignment and repair processes (Pickering & Garrod, 2004) in the caregiver-infant interactions (Ko et al., 2016). Specifically, we suggest that caregivers may unconsciously use multiple channels in a systematic way to enhance communicative efficiency; they may also monitor their infants' attentional and affective responses in a way that could impact the use of touch and speech given changes in infants' detection of relevant invariant information over time (Bahrick & Lickliter, 2012). For example, if caregivers monitor their infants' attention, affect, and linguistic skills, they may provide aligned word edges and touch edges if such events consistently yield positive attentional (e.g., looks), affective (e.g., smiles), and verbal (e.g., babbles) responses. These behaviors may be reinforced over time and may lead to improved infant learning. Additionally, the change in mothers' use of tactile cues with the child's age might be based on their use of tactile cues for modulating their

children's attention with development. Specifically, if caregivers unconsciously use touch to garner infant attention, we may expect their touch behaviors to change over infants' development with more frequent touches occurring to younger children who have lesser degrees of attentional control (this prediction is supported by previous studies showing a developmental decline in the use of touch cues; Ferber et al., 2008). The use of touch as a multimodal cue relevant for word boundaries and mapping referents to wordforms and its adoption for an attention-modulating function are likely to work in an integrated fashion for infants' word learning. In fact, previous studies have shown that caregivers' greater use of synchronous multimodal labeling events is associated with heightened infant attention (Tincoff et al., 2019) and better learning of word-object relations (Gogate, 2020; Gogate et al., 2006; Matatyaho & Gogate, 2008).

Another, not mutually exclusive, explanation for age-related frequency effects could be that the precise alignment and co-occurrence between touch and speech is driven by infant mobility which increases with development (e.g., Thelen, 1995). Motor milestones can impact the language input children receive and their vocabulary production (Iverson, 2010; Oudgenoeg-Paz et al., 2012) given that caregivers seem to be sensitive to infants development and modify their language with infant motoric development (e.g., Schneider & Iverson, 2022). Further, older infants are more likely to steer their attention from faces to looking at what parents are holding (Fausey et al., 2016; Smith & Yu, 2013), which may impact caregivers' behavior since caregivers attend to where infants are looking or pointing and are likely to label those objects (e.g., Brooks & Meltzoff, 2008). Thus, caregivers may be better able to practice aligning touch and speech when infants are less mobile and their attention is more focused on caregivers' faces rather than objects. With infants' attainment of walking, however, their interest may transition toward the objects being held by caregivers, and caregivers might increasingly use their hands to present objects to children rather than to directly touch their infants' body. This may explain why there is more touch-speech alignment to young infants who are not yet fully mobile. Note, however, that while touch use declines with age, the rate of word-touch alignment also decreases with age as we showed in this paper. In particular, the decrease of touch aligning with words is in the *ratio* of the alignment, rather than *raw numbers*. Thus, the less frequent alignment of touch with words in older children seems to be an effect that exists beyond the general decline in the number of touches with infants' increasing mobility. Therefore, although the greater alignment to younger infants may partially be a result of increased opportunities to practice with less mobile infants, there seems to be a decrease in the multimodal presentation of cues adapted to infants' linguistic development that cannot be attributed to infants' mobility. Adjudicating between these hypotheses with certainty will require a careful control of child mobility within a study context and an examination of within-child mobility variability. Thus, we leave differentiating or weighting these hypotheses as a topic for future study.

5 | CONCLUSION

To our knowledge, this is the first study to investigate word-touch and utterance-touch alignment in a cross-sectional sample. The temporal alignment of multimodal touch and speech events in the input of Korean mothers to their infants is non-arbitrary, and its frequency decreases with infants' age. No other studies seem to have demonstrated the developmental effects on caregivers' use of tactile cues based on data collected in a home-like environment with no tasks or experimental intervention involved. These results extend findings with middle-class Midwestern American mothers in the context of book-reading interactions (Abu-Zhaya et al., 2017) as well as Caucasian- and Hispanic-American mothers (Gogate et al., 2000) and Asian-Indian mothers (Gogate et al., 2015) in experimental lab settings. This extension shows that these previous findings are not limited in their scope to specific populations and contexts, but can rather be extended to at least one more cultural group (Korean

mothers) and context (naturalistic play interactions). We showed that mothers' use of tactile cues in alignment or in co-occurrence with speech cues change over children's development, with greater word-touch alignment and word + touch co-occurrences presented in the youngest group. There is a possibility that such changes reflect mothers' adaptation of the multimodal cues in accordance with the developmental needs and or feedback from their language learning children, but we cannot rule out the possibility that it could be a reflection of mothers' response to other aspects of the child's development such as attentional focus or mobility. Future research could more narrowly focus on child's individual linguistic ability (e.g., babbling behaviors) or other aspects of development (e.g., mobility) to help clarify the motivation behind such changes in caregiver behaviors. Finally, the constant high proportion of touches co-occurring with words throughout infants' development suggests that mothers are more likely to use touch in combination with words than alone, which highlights the potential importance of multimodal cues to infants' word learning.

ACKNOWLEDGMENTS

We thank the families who participated in this study, and Jihyo Kim, Taeseong Lee, and Suhan Kim for their help with coding the data. We also thank the anonymous reviewers for their careful reading of our paper and many insightful comments and suggestions. This research was supported by the National Research Foundation of Korea (2021R1I1A2051993; 2021K2A9A1A01102199) and research fund from Chosun University (2021) to EK, and R21 HD108730-01 to AS. The authors declare no conflicts of interest with regard to the funding source for this study.

ORCID

Eon-Suk Ko https://orcid.org/0000-0003-3963-4492
Rana Abu-Zhaya https://orcid.org/0000-0001-9532-7935

REFERENCES

Abu-Zhaya, R., Kondaurova, M. V., Houston, D., & Seidl, A. (2019). Vocal and tactile input to children who are deaf or hard of hearing. *Journal of Speech, Language, and Hearing Research*, *62*(7), 2372–2385. https://doi.org/10.1044/2019_JSLHR-L-18-0185

Abu-Zhaya, R., Seidl, A., & Cristia, A. (2017). Multimodal infant-directed communication: How caregivers combine tactile and linguistic cues. *Journal of Child Language*, *44*(5), 1088–1116. https://doi.org/10.1017/s0305000916000416

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), 1617–1627. https://doi.org/10.2307/1130769

Bahrick, L. E., & Lickliter, R. (2012). The role of intersensory redundancy in early perceptual, cognitive, and social development. In A. J. Bremner, D. J. Lewkowicz, & C. Spence (Eds.), *Multisensory development* (pp. 183–206). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199586059.003.0008

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Package lme4. *Journal of Statistical Software*, *67*(1), 1–91.

Boersma, P., & Weenink, D. (2020). Praat: Doing phonetics by computer [Computer program]. Version 6.1.20 beta. Retrieved September 10, 2020 from http://www.praat.org/

Bornstein, M., Hahn, C.-S., & Haynes, O. (2004). Specific and general language performance across early childhood: Stability and gender considerations. *First Language*, *24*(3), 267–304. https://doi.org/10.1177/0142723704045681

Bremner, A. J., & Spence, C. (2017). The development of tactile perception. *Advances in Child Development and Behavior*, *52*, 227–268. https://doi.org/10.1016/bs.acdb.2016.12.002

Bremner, A. J., Lewkowicz, D. J., & Spence, C. (2012). The multisensory approach to development. In A. J. Bremner, D. J. Lewkowicz, & C. Spence (Eds.), *Multisensory development* (pp. 1–26). Oxford University Press.

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, *81*(2), B33–B44. https://doi.org/10.1016/s0010-0277(01)00122-6

Brod, G., Hasselhorn, M., & Bunge, S. A. (2018). When generating a prediction boosts learning: The element of surprise. *Learning and Instruction*, *55*, 22–31. https://doi.org/10.1016/j.learninstruc.2018.01.013

Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language*, *35*(1), 207–220. https://doi.org/10.1017/s030500090700829x

ELAN. (2020). Nijmegen: Max Planck Institute for Psycholinguistics, the language archive (Version 5.9) [Computer software]. https://archive.mpi.nl/tla/elan

Farver, J. A., Kim, Y. K., & Lee, Y. (1995). Cultural differences in Korean- and Anglo-American preschoolers' social interaction and play behaviors. *Child Development*, *66*(4), 1088–1099. https://doi.org/10.2307/1131800

Fausey, C. M., Jayaraman, S., & Smith, L. B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition*, *152*, 101–107. https://doi.org/10.1016/j.cognition.2016.03.005

Feldman, R., Eidelman, A. I., Sirota, L., & Weller, A. (2002). Comparison of skin-to-skin (kangaroo) and traditional care: Parenting outcomes and preterm infant development. *Pediatrics*, *110*(1 Pt 1), 16–26. https://doi.org/10.1542/peds.110.1.16

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), 363–370. https://doi.org/10.1016/j.earlhumdev.2007.09.019

Fernald, A., & Morikawa, H. (1993). Common themes and cultural variations in Japanese and American mothers' speech to infants. *Child Development*, *64*(3), 637–656. https://doi.org/10.2307/1131208

Gibson, E. J. (1969). *Principles of perceptual learning and development*. Appleton–Century–Crofts.

Gogate, L. (2020). Maternal object naming is less adapted to preterm infants' than to term infants' word mapping. *Journal of child psychology and psychiatry, and allied disciplines*, *61*(4), 447–458. https://doi.org/10.1111/jcpp.13128

Gogate, L. J. (2010). Learning of syllable–object relations by preverbal infants: The role of temporal synchrony and syllable distinctiveness. *Journal of Experimental Child Psychology*, *105*(3), 178–197. https://doi.org/10.1016/j.jecp.2009.10.007

Gogate, L. J., & Bahrick, L. (1998). Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in seven-month-old infants. *Journal of Experimental Child Psychology*, *69*(2), 133–149. https://doi.org/10.1006/jecp.1998.2438

Gogate, L. J., & Hollich, G. (2010). Invariance detection within an interactive system: A perceptual gateway to language development. *Psychological Review*, *117*(2), 496–516. https://doi.org/10.1037/a0019049

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), 878–894. https://doi.org/10.1111/1467-8624.00197

Gogate, L. J., Bolzani, L. H., & Betancourt, E. A. (2006). Attention to maternal multimodal naming by 6- to 8-month-old infants and learning of word-object relations. *Infancy*, *9*(3), 259–288. https://doi.org/10.1207/s15327078in0903_1

Gogate, L. J., Maganti, M., & Laing, K. B. (2013). Maternal naming of object wholes versus parts to preverbal infants: A fine-grained analysis of scaffolding at 6–8 months. *Infant Behavior and Development*, *36*(3), 470–479. https://doi.org/10.1016/j.infbeh.2013.03.012

Gogate, L. J., Walker-Andrews, A. S., & Bahrick, L. E. (2001). The intersensory origins of word comprehension: An ecological–dynamic systems view. *Developmental Science*, *4*(1), 1–18. https://doi.org/10.1111/1467-7687.00143

Gogate, L., & Maganti, M. (2016). The dynamics of infant attention: Implications for crossmodal perception and word-mapping research. *Child Development*, *87*(2), 345–364. https://doi.org/10.1111/cdev.12509

Gogate, L., Maganti, M., & Bahrick, L. E. (2015). Cross-cultural evidence for multimodal motherese: Asian Indian mothers' adaptive use of synchronous words and gestures. *Journal of Experimental Child Psychology*, *129*, 110–126. https://doi.org/10.1016/j.jecp.2014.09.002

Goldin-Meadow, S., & Saltzman, J. (2000). The cultural bounds of maternal accommodation: How Chinese and American mothers communicate with deaf and hearing children. *Psychological Science*, *11*(4), 307–314. https://doi.org/10.1111/1467-9280.00261

Gorman, K., Howell, J., & Wagner, M. (2011). Prosodylab-aligner: A tool for forced alignment of laboratory speech. *Canadian Acoustics*, *39*(3), 192–193.

Iverson, J. M. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of Child Language*, *37*(2), 229–261. https://doi.org/10.1017/S0305000909990432

Jean, A. D., & 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 and Development*, *35*(4), 779–791. https://doi.org/10.1016/j.infbeh.2012.07.023

Jean, A. D., 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 and Development*, *32*(3), 344–349. https://doi.org/10.1016/j.infbeh.2009.04.005

Jo, J., & Ko, E.-S. (2018). Korean mothers attune the frequency and acoustic saliency of sound symbolic words to the linguistic maturity of their children. *Frontiers in Psychology*, *9*, 2225. https://doi.org/10.3389/fpsyg.2018.02225

Kim, S. J. (2000). *Contextual differences in Korean mother–child interactions during pretend play, puzzle games, and story retelling*. University of Texas at Austin, Ph. D. Dissertation.

Kim, U., & Choi, S.-H. (1994). Individualism, collectivism, and child development: A Korean perspective. In P. M. Greenfield & R. R. Cocking (Eds.), *Cross-cultural roots of minority child development* (pp. 227–257). Lawrence Erlbaum Associates, Inc.

Ko, E.-S., Jo, J., & Chai, J. H. (2022). Word-order adaptation and lexical repetition in Korean child-directed speech: Implications for infants' learning nouns in a verb-dominant language. psyarxiv.com/dxbhm

Ko, E.-S., Jo, J., On, K.-W., & Zhang, B.-T. (2020). Introducing the Ko Corpus of Korean mother–child interaction. *Frontiers in Psychology*, *11*, 3698. https://doi.org/10.3389/fpsyg.2020.602623

Ko, E.-S., Seidl, M., Cristia, A., Reimchen, M., & Soderstrom, M. (2016). Entrainment of prosody in the interaction of mothers with their young children. *Journal of Child Language*, *43*(2), 284–309. https://doi.org/10.1017/S0305000915000203

Koponen, E., & Lacerda, F. (2003). Final lengthening in infant directed speech may function as a cue to phrase constituents. *PHONUM*, *9*, 9–12.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, *218*(4577), 1138–1141. https://doi.org/10.1126/science.7146899

Kuhl, P. K., & Meltzoff, A. N. (1984). The intermodal representation of speech in infants. *Infant Behavior and Development*, *7*(3), 361–381. https://doi.org/10.1016/S0163-6383(84)80050-8

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Kwak, K., Kim, S., & Jeong, Y. (2005). Infants' sensitivity on the changes of mothers' touch. *Korean Journal of Child Studies*, *26*(5), 123–137.

Lewkowicz, D. J. (2010). Infant perception of audio-visual speech synchrony. *Developmental Psychology*, *46*(1), 66–77. https://doi.org/10.1037/a0015579

Lew-Williams, C., Pelucchi, B., & Saffran, J. R. (2011). Isolated words enhance statistical language learning in infancy. *Developmental Science*, *14*(6), 1323–1329. https://doi.org/10.1111/j.1467-7687.2011.01079.x

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (3rd ed.). Lawrence Erlbaum Associates.

Masek, L. R., McMillan, B. T. M., Paterson, S. J., Tamis-LeMonda, C. S., Golinkoff, R. M., & Hirsh-Pasek, K. (2021). Where language meets attention: How contingent interactions promote learning. *Developmental Review*, *60*, 100961. https://doi.org/10.1016/j.dr.2021.100961

Matatyaho, D. J., & Gogate, L. J. (2008). Type of maternal object motion during synchronous naming predicts preverbal infants' learning of word-object relations. *Infancy*, *13*(2), 172–184. https://doi.org/10.1080/15250000701795655

Matatyaho-Bullaro, D., Gogate, L., Mason, Z., Cadavid, S., & Abdel-Mottaleb, M. (2014). Type of object motion facilitates word mapping by preverbal infants. *Journal of Experimental Child Psychology*, *118*, 27–40. https://doi.org/10.1016/j.jecp.2013.09.010

Newport, E., Gleitman, H., & Gleitman, L. (1977). Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children* (pp. 109–149). Cambridge University Press.

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), 113–128. https://doi.org/10.1109/tamd.2011.2140113

O'Neill, M., Bard, K. A., Linnell, M., & Fluck, M. (2005). Maternal gestures with 20-month-old infants in two contexts. *Developmental Science*, *8*(4), 352–359. https://doi.org/10.1111/j.1467-7687.2005.00423.x

Odijk, L., & Gillis, S. (2021). Fine lexical tuning in infant directed speech to typically developing children. *Journal of Child Language*, *48*(3), 591–604. https://doi.org/10.1017/S0305000920000379

Oudgenoeg-Paz, O., Volman, M. C., & Leseman, P. P. (2012). Attainment of sitting and walking predicts development of productive vocabulary between ages 16 and 28 months. *Infant Behavior and Development*, *35*(4), 733–736. https://doi.org/10.1016/j.infbeh.2012.07.010

Patterson, M. L., & Werker, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior and Development*, *22*(2), 237–247. https://doi.org/10.1016/S0163-6383(99)00003-X

Peláez-Nogueras, M., Field, T. M., Hossain, Z., & Pickens, J. (1996). Depressed mothers' touching increases infants' positive affect and attention in still-face interactions. *Child Development*, *67*(4), 1780–1792. https://doi.org/10.2307/1131731

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–190. https://doi.org/10.1017/s0140525x04000056

R Core Team. (2019). R: A language and environment for statistical computing, Vienna, Austria. https://www.R-project.org/

Rowe, M. L. (2008). Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language*, *35*(1), 185–205. https://doi.org/10.1017/s0305000907008343

Saint-Georges, C., Chetouani, M., Cassel, R., Apicella, F., Mahdhaoui, A., Muratori, F., Laznik, M. C., & Cohen, D. (2013). Motherese in interaction: At the cross-road of emotion and cognition? (A systematic review). *PLoS One*, *8*(10), e78103. https://doi.org/10.1371/journal.pone.0078103

Sakamoto, Y., & Love, B. C. (2006). Vancouver, Toronto, Montreal, Austin: Enhanced oddball memory through differentiation, not isolation. *Psychonomic Bulletin & Review*, *13*(3), 474–479. https://doi.org/10.3758/bf03193872

Schneider, J. L., & Iverson, J. M. (2022). Cascades in action: How the transition to walking shapes caregiver communication during everyday interactions. *Developmental Psychology*, *58*(1), 1–16. https://doi.org/10.1037/dev0001280

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), 155–164. https://doi.org/10.1111/desc.12182

Singh, L., Steven Reznick, J., & Xuehua, L. (2012). Infant word segmentation and childhood vocabulary development: A longitudinal analysis. *Developmental Science*, *15*(4), 482–495. https://doi.org/10.1111/j.1467-7687.2012.01141.x

Smith, L. B., & Yu, C. (2013). Visual attention is not enough: Individual differences in statistical word-referent learning in infants. *Language Learning and Development*, *9*(1), 25–49. https://doi.org/10.1080/15475441.2012.707104

Snow, C. E. (1977). Mothers' speech research: From input to interaction. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children* (pp. 31–49). Cambridge University Press.

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, *27*(4), 501–532. https://doi.org/10.1016/j.dr.2007.06.002

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. https://doi.org/10.1371/journal.pone.0080646

Sokolov, J. (1993). A local contingency analysis of the fine-tuning hypothesis. *Developmental Psychology*, *29*(6), 1008–1023. https://doi.org/10.1037/0012-1649.29.6.1008

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, *38*(4), 387–401. https://doi.org/10.1016/0013-4694(75)90263-1

Stack, D. M., & Muir, D. W. (1992). Adult tactile stimulation during face-to-face interactions modulates five-month-olds' affect and attention. *Child Development*, *63*(6), 1509–1525. https://doi.org/10.2307/1131572

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), 202–209. https://doi.org/10.1109/TAMD.2013.2269905

Thelen, E. (1995). Motor development: A new synthesis. *American Psychologist*, *50*(2), 79–95. https://doi.org/10.1037/0003-066X.50.2.79

Tincoff, R., Seidl, A., Buckley, L., Wojcik, C., & Cristia, A. (2019). Feeling the way to words: Parents' speech and touch cues highlight word-to-world mappings of body parts. *Language Learning and Development*, *15*(2), 103–125. https://doi.org/10.1080/15475441.2018.1533472

Vergilova, Y., Jachmann, T. K., Mani, N., & Kray, J. (2022). Age-related differences in expectation-based novel word learning. *Psychophysiology*, *59*(8), e14030. https://doi.org/10.1111/psyp.14030

Vigliocco, G., Motamedi, Y., Murgiano, M., Wonnacott, E., Marshall, C., Milán-Maillo, I., & Perniss, P. (2019). Onomatopoeia, gestures, actions and words: How do caregivers use multimodal cues in their communication to

children? In *Proceedings of the 41st annual conference of the cognitive science society*. Cognitive Science Society. https://doi.org/10.31234/osf.io/v263k

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). American Psychological Association. https://doi.org/10.1037/10265-006

How to cite this article: Ko, E.-S., Abu-Zhaya, R., Kim, E.-S., Kim, T., On, K.-W., Kim, H., Zhang, B.-T., & Seidl, A. (2023). Mothers' use of touch across infants' development and its implications for word learning: Evidence from Korean dyadic interactions. *Infancy*, 1–22. https://doi.org/10.1111/infa.12532