How Speech and Representational Gestures Align in Child-Directed Language: a Corpus-based Study

Yumeng Wang (yumeng.wang.21@ucl.ac.uk)
Experimental Psychology, University College London, 26 Bedford Way, London WC1H 0AP, United Kingdom

Ed Donnellan (ed.donnellan@ucl.ac.uk)
Experimental Psychology, University College London, 26 Bedford Way, London WC1H 0AP, United Kingdom

Gabriella Vigliocco (g.vigliocco@ucl.ac.uk)
Experimental Psychology, University College London, 26 Bedford Way, London WC1H 0AP, United Kingdom

Abstract

Representational gestures are co-speech gestures that carry semantic content related to the content of speech. Previous studies focusing on adult-adult conversation have investigated the temporal alignment of gestures and speech finding that the overwhelming majority of representational gestures are produced right before the lexical content they refer to (their lexical affiliate, LA). However, nothing is yet known about whether caregivers would also time their gestures in the same way in naturalistic interactions. We annotated representational gestures from a large corpus (ECOLANG) of semi-naturalistic conversations between caregivers and their 3-4 year old children (n = 899 gestures from n=36 caregivers). We found that, just as in adult-directed language (ADL), representational gestures in child-directed language (CDL) were more tightly linked to the onset of LAs than the onset of the utterance in which LAs were produced (hence planned when full events are encoded); with the overall majority of the representational gestures starting before their LAs. We further found that age of acquisition (AoA) rating of the LA had a significant effect on the speech-gesture latency. We found that for words acquired earlier, the gesture's stroke (the meaningful part of a gesture) tended to be produced before the LA's onset; for the later acquired word, the stroke tended to be produced at the same time or after the onset of the LA in speech. Our findings suggest that: (1) Regardless of their addressee, speakers always time the production of representational gestures to specific conceptual/linguistic units, rather than the full event/utterance. (2) In contrast to ADL, caregivers' gestures may support addresses' linguistic processing not only by supporting word prediction (of likely better-known words), but also by supporting the learning of conceptual features (of likely less well-known words).

Keywords: multimodal communication; co-speech gestures; representational gestures; child-directed speech; lexical affiliates

Introduction

During conversations, manual gestures always accompany spoken language, (Goldin-Meadow, 2003; Iverson and Goldin-Meadow 1998; McNeill 1992). Representational gestures are co-speech gestures that carry semantic content depicting some characteristic or feature of the referent. These gestures provide semantic information that supplements or complements accompanying speech (McNeill, 1992). This information can support speakers' production at conceptual and/or lexical levels (e.g., Lexical Retrieval Hypothesis, see Krauss, 1998; Rauscher, Krauss & Chen, 1996). It can also support addressees' language comprehension (e.g., Zhang et al., 2021). Developmental studies also suggested the scaffolding role of gestures in children's language development (e.g., Clough & Hilverman, 2018). For gestures to fulfil these different functions, they need to be timed in a precise manner with corresponding units in speech. Here we address two questions concerning the temporal relationship between caregivers' representational gestures and their speech.

What are Gestures Aligned to?

Donnellan et al (2022) carried out a study to address this question in language directed to adults. They contrasted two hypotheses. According to McNeill's growth point theory (McNeill, 1992; McNeill & Duncan, 2000), speech and gestures are systematically organized in relation to one another at a conceptual (event) level. The theory posits that gestures and speech share a computational stage from which both arise, while they are then encoded by either the vocal or gestural systems. This perspective indicates that gestures and speech are typically semantically and pragmatically synchronised so that the speaker's overall communicative intention is implemented by combining verbal and gestural information, but each modality is encoded separately (Bergmann et al, 2011). Specifically, McNeill (1992) suggested that the image (gesture) and the linguistic (speech) elements have always been together in the growth point of the sentences. This growth point, theoretically, should be the utterance's primitive stage and the formation of a growth point is at the utterance level corresponding to an event. Therefore, according to this proposal, the onsets of an utterance and gesture onsets should be tightly linked.

A second possibility suggests that gesture and speech could be more precisely linked at the level of sub-events (lexical). Some studies (Ferré, 2010; Kendon, 1972, 1990; Schegloff, 1984) suggest that gesture tends to align with the semantically most congruent part of the utterance. Other studies show that the gesture is more closely linked to the corresponding word as the asynchrony can be affected by the corresponding lexical affiliate (LA). For example, Morrel-Samuels and Krauss (1992) showed that the gesture onset can

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

be influenced by word familiarity, namely, the more familiar a word, the smaller the asynchrony between the gesture preparation time and the onset of the LA.

In their corpus study, Donnellan et al., (2022) analyzed gesture and speech from a sample of 24 adults in which these adults introduced both familiar and unfamiliar objects to another adult while the objects were either present or absent. They found that gesture phases (both preparation and stroke) onsets were more tightly linked to LA onset than to the beginning of the corresponding utterance, which supported the second hypothesis. However, it is unclear whether the same relationship will hold for the gestures produced by caregivers when they talk to children.

What is the Timing Relationship between Gesture and Speech?

In addition to the question of whether gestures align with the utterance or LA onset, several previous studies have established that gestures occur slightly before LAs. Based on the existing literature on task-elicited descriptions (de Ruiter, 2000a; Hadar and Butterworth, 1997; Kendon et al, 1980; Morrel-Samuels and Krauss, 1992; Nobe, 2000), it has generally been accepted that the onset of gestures precedes their LAs. More specifically, Butterworth and Beattie (1978) found that gestures tend to have their onsets (the preparation phase) before the onset of the related LAs in relatively fluent phases of speech. Schegloff (1984) also claimed that the gesture onset is typically prior to the LAs.

Recent studies examining naturalistic production have also begun to investigate how much earlier gestures begin with respect to speech. These studies (all investigating adult-adult naturalistic conversations) showed that the onsets of most representational gestures precede their LAs but remain temporally close (Chui, 2005; ter Bekke et al., 2022; Ferré, 2010; Urbanik & Svennevig 2021). For example, in a naturalistic conversation corpus of spontaneous French dialogues, 95% of representational hand gestures (preparation onsets) were found to have begun before LA onset, typically 820ms earlier (Ferré, 2010). Crucially, it was found that 72% of gesture strokes tended to precede LA onsets, on average beginning 450ms earlier. Donnellan et al. (2022) also analysed the latency between representational gesture phases and speech using data from semi-naturalistic conversations between familiar adults. Here, gesture onsets were found to precede LA onset by 814ms and gesture stroke onsets preceded LAs onsets by 370ms. The results are also consistent with previous naturalistic corpus studies (ter Bekker et al., 2022).

However, these studies concerning the timing relationship of co-speech representational gestures have only focused on adult communication with other adults. No research yet has investigated how representational gestures are used by caregivers in child-directed naturalistic conversations.

The Present Study

The present study investigates the timing relationship between caregivers' speech and relevant phases of the gesture (i.e. preparation and stroke) in child-directed language. We further examine whether the patterns observed in adult-adult conversation in previous studies are present in caregiver speech to 3-4 year old children. We focus on 3-4 year old children, an age in which children are still learning at a fast pace and are capable of understanding representational gestures (Tolar, Lederberg, Gokhale & Tomasello, 2008).

We analysed data from the ECOLANG corpus (Vigliocco et al., in prep) where English caregivers talked to their 3-4 year old children about objects that were present or absent, and known or unknown to their children. We annotated representational hand gestures used by caregivers when describing the objects.

With regard to our first question, we predict that just as it was found in the analysis of ADL, gestures will be more tightly related to LA onsets as caregivers may break down events into subparts to make the whole utterance easier for children to understand.

With regard to the second question, if the timing relation between gesture and speech is determined only by production constraints, we should find that gestures precede their LAs in a consistent manner, regardless of the characteristics of the addressee, thus we should replicate here the findings by Donnellan et al (2022). However, there are reasons to believe that the timing relationship may be sensitive to whether the addressee is an adult or a child. In particular, it could be the case that gestures precede their LAs by a longer latency as caregivers may produce the gesture much earlier to allow more time for children to use it to then process the speech information. Alternatively, it may be that the timing relationships dynamically change depending on the knowledge and familiarity of the child toward the concepts/words expressed.

Method

Participants

The data used for this study has been taken from the ECOLANG corpus (Vigliocco et al., in prep), which consisted of 36 caregiver-child dyads based in the UK. Children (18 girls and 18 boys) were aged between 36 and 51 months (Mean = 42.9, SD = 4.25). Caregivers spoke British or American English with their children. Caregivers reported that all children were typically developing without any language disorder. The corpus creation was approved by the Research Ethics Committee, University College London.

Materials

Toys presented in the study were from 4 categories: foods, musical instruments, animals, and tools, which were common for children of this age range. Within each category, there were six toys for each dyad, half of which were unknown to the child and half were known (based on parental reports of the child's knowledge). Toys chosen for each child were from a larger set of approximately 20 toy items per category, each of which was used by a roughly equal number of testing sessions across participants.

Procedure

The corpus collection was carried out in the family home. During the recording, caregiver and child sat 90 degrees from each other at a table. Caregivers were asked to talk to their child in a natural way, but to try to talk about the objects provided. The order of toy categories and whether the toys were present or absent was counterbalanced across participants. When the interaction started with the present condition, the experimenter brought 6 toys from one category (e.g., foods) to the table and then left the room. The caregiver and child talked about these toys for 3-5 minutes, then the experimenter returned to the room, asked the child to assist in tidying up the toys and then left the room for the absent condition. The caregiver and child talked about the objects they had just played with for another 3-5 minutes. The experimenter then reappeared with a new set of toys until all toys in the categories had been used. When the toy absent condition came first, the caregiver was told to begin talking about the upcoming toys that would be brought by one experimenter from another room (caregivers were first taught about and were familiarised with the toys). After 3-5 minutes, the experimenter brought in the set of toys for that category, and left the room, allowing the caregiver and child to talk with the objects present. After 3-5 minutes, the experimenter entered the room to remove the toys and the procedure continued for all 4 toy categories. The full recording session lasted approximately 35-45 minutes.

Coding

Caregivers' speech in the corpus had been annotated in ELAN (Sloetjes & Wittenburg, 2008) and transcribed at both the word and utterance level (Berman & Slobin, 1994). Utterances were defined as a unit that expresses a single situation (such as an activity, event or state). Representational gestures were also coded. For this study, we identified the onset of two gesture phases for each representational gesture: the preparation and the stroke (Seyfeddinipur, 2006; Kita, van Gijn & van der Hulst, 1998). Along with the gesture phases, LAs were marked.

Gesture Phases. The movement of the hand(s) into a gesture shape indicated the start of the preparation phase. This can be marked where the hand/arm started to rise from a resting position (perhaps on the lap, on a table, or in mid-air). In another case, if the speaker made a gesture right before making the representational gesture, the start of the preparation phase was marked at the moment where the hand relaxed from the previous gesture and moved to make the representational gesture. But this was only true if the hands made a single motion movement towards the representational gesture; in other words, if the hand(s) withdrew from the prior gesture to a resting position before the start of the representational gesture, the preparation onset was noted at the point where the hands began to leave the resting position.

The stroke phase began at the point when the hand(s) started to convey the gesture's meaning by a clearly defined configuration (form) and well-articulated movement that represented some feature, property or action. We recorded the onset time of the first stroke for gestures that include several strokes (but express the same meaning), but if two strokes conveyed different meanings, they were marked as separate gestures.

LA. LAs were defined as the word(s) whose meaning most closely matched a gesture (Holler & Levinson, 2019). Specifically, we considered words that were within 1000ms of the start and end of the gesture as potential LAs for that gesture. The main strategy to determine the corresponding LA for each gesture was to first see what information the gesture carried. We excluded demonstratives before nouns to keep the minimum amount of word(s) which contain the most semantic information.

We took the onset time of LAs for each gesture. If there were multiple LAs within one utterance, we only calculated the onset time of the first LA corresponding with the gesture. For each LA, we obtained an AoA rating (Kuperman et al., 2012) for later analysis.

Utterance. We took the onset time of the utterance from which the LA derived (as marked in the corpus). For each utterance, the topic (i.e., the toy that the utterance was about), object familiarity (known/unknown) and object presence (present/absent of the toy) were coded as part of the corpus.

Reliability Coding. For reliability, 10% of the representational gestures of all 36 participants were double-coded (n = 168 representational gestures) by experienced coders. Both coders coded the preparation and stroke phase of the gesture and marked LAs, with the second coder having no access to the previous version.

Each coder determined whether each potential LA (words appearing within 1000ms of the gesture onset and offset), was a LA of the gesture (n = 2003 potential LAs). 93.61% of these words were agreed upon by two coders (Cohen's $\kappa = .71$, [95% CI = .66 - .76], indicating substantial agreement).

The latencies included in our reliability analysis demonstrated high degree of agreement for gestures that had both stroke and preparation phase, where the initial LA was agreed upon by two coders (n = 133), as measured by correlation (r): (1) gesture preparation to LA onset, $r = .97$, (2) gesture stroke to LA onset, $r = .94$, (3) gesture preparation to utterance onset, $r = .98$, (4) gesture stroke to utterance onset, $r = .96$. Correlations between two coders were still high for all latencies, even when we considered cases with no agreement on the first LA: (1) $r = .75$, (2) $r = .73$, (3) $r = .74$, (4) $r = .73$.

Data Analysis

The present analysis concentrated on the timing relationships between representational gestures and speech in the setting of semi-naturalistic conversations between caregivers and children. For analysis, we only considered representational gestures including both phases (preparation and stroke) and had at least one LA (n = 899 representational gestures). All analyses were performed in R 4.1.2 (R Core Team, 2020), with mixed effects models running with lme4 (Bates et al., 2015), and model summaries generated with p-values calculated using lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017).

What are Gestures Aligned to? We first constructed two mixed effects models with gesture preparation onset predicted by either utterance or LA onset and another two mixed effects with gesture stroke onset time predicted by either utterance or LA onset. For all four models, we included speaker ID as a random effect on the slope. To allow for meaningful comparisons, all variables were mean-centred and scaled (Mean = 0, SD = 1).

Model comparisons using Akaike's Information Criterion (AIC) were conducted. We simply used ΔAIC (the difference between the two AIC values for models being compared, where ΔAIC = 0 as the best fitting model) as a way to compare possible models since lower AIC values indicate a better-fit model (Anderson & Burnham, 2002).

What is the Timing Relationship between Gesture and Speech? Mixed effects models with gesture phase onset predicted by LA onset were conducted to get latencies between them (e.g., the latency between gesture preparation and LA). We entered random effects on the slope for speaker ID and a random intercept for gesture ID. Note that to estimate the precise latencies, all variables were not mean-centred and scaled.

Additionally, as gesture timing may be affected by other variables, we included object presence (object present/absent), object familiarity (known/unknown) and lexical AoA ratings for all LAs in two additional mixed effect models with exact latencies predicted by those factors (i.e., exact latency between preparation and LA predicted by AoA ratings). All continuous variables were mean-centred and scaled (Mean = 0, SD = 1). Note that word frequency and AoA ratings are highly correlated (r = .74), we chose to include AoA ratings considering the developmental characteristics of children.

Results

Gesture Alignment to Utterance and LA

Analysis showed significant results in all four models considering the relative contribution of LA onset and utterance onset to the preparation and stroke onset respectively. Then, model comparison confirmed that LA onset was a stronger predictor than utterance onset of both gesture preparation onset (ΔUtterance onset = 238) and stroke onset (ΔUtterance onset = 256.9). Table 1 and table 2 show the AIC model selection to distinguish among different possible models.

Table 1: AIC model selection to distinguish among two possible models of the relationship between gesture preparation onset, LA onset and utterance onset.

Model Specification	*LL*	*AIC*	Δ*AIC*
Preparation onset ~ LA onset	4379.7	-8747.5	0
Preparation onset ~ Utterance onset	4260.7	-8509.5	238

Table 2: AIC model selection to distinguish among two possible models of the relationship between gesture stroke onset, LA onset and utterance onset.

Model Specification	*LL*	*AIC*	Δ*AIC*
Stroke onset ~ LA onset	4383.9	-8755.7	0
Stroke onset ~ Utterance onset	4255.4	-8498.8	256.9

Timing between Gesture Phase and LA

Do Preparation Onsets Precede the LAs? The mixed effects model revealed that overall, gesture onsets significantly preceded LA onsets by around 700ms (β = -700.37, SE = 49.12, t = -14.26, p < .001). According to Figure 1, the overwhelming majority of gesture preparation phases (83%) started before their corresponding LAs.

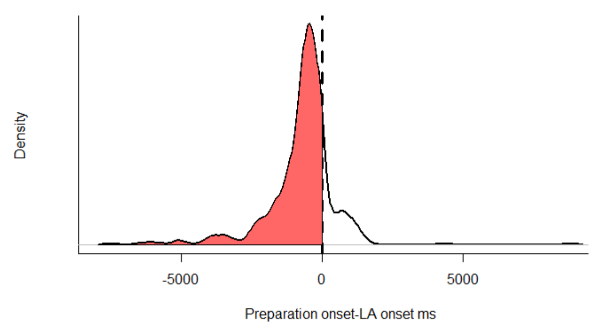

Figure 1. The relationship between gesture preparation and LA. Density plot shows the timing difference between gesture preparation onset and LA onset in ms, where negative values (red) indicate gestures that preceded their LAs. 0ms indicated by the dashed black line.

Do Any Other Factors Affect Time Latency between Preparation Onsets and LA Onsets? Table 3 showed that presence or absence of the items (β = 260.26, SE = 121.91, t = 2.135, p =. 24 033) was a significant predictor of latency between preparation onsets and LA onsets. Figure 2 illustrated the latency differences between gesture preparation onsets and LA onsets across present and absent conditions. Gesture were produced 260.26ms earlier in the absent than in the present condition. The effect of object familiarity (p = .177) and AoA (p = .136) was non-significant, nor was the interaction of familiarity and presence (p =.890).

Table 3: Summary of other factors affecting the time latency between gesture preparation onsets and LA onsets

	β	*SE*	*df*	*t*	*p*
Presence	260.26	121.91	848.52	2.135	.033*
Familiarity	120.39	89.23	849.24	1.349	.177
AoA	27.64	18.53	846.20	1.492	.136
Familiarity *Presence	-36.72	265.85	850.97	-0.138	.890

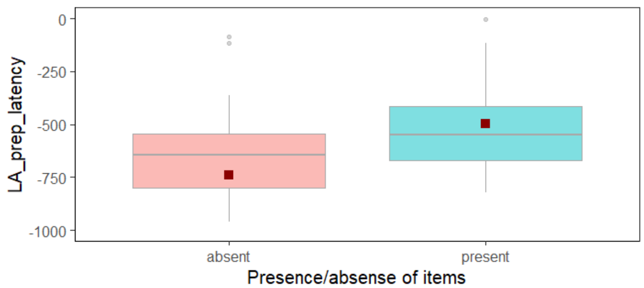

Figure 2. Summary plot showing the latencies between gesture preparation onsets and LA onsets for present and absent conditions, as indicated on the x axis. Red squares highlight the mean values in each box. Confidence intervals are set to 95%.

Do Strokes Onsets Precede LAs? The mixed effects model revealed that gesture stroke onsets significantly preceded LA onsets by around 118 ms (β = -118.92, *SE* = 41.74, *t* = -2.849, *p* =.007). More than half of gesture stroke phases (56%) started before their corresponding LAs (Figure 3).

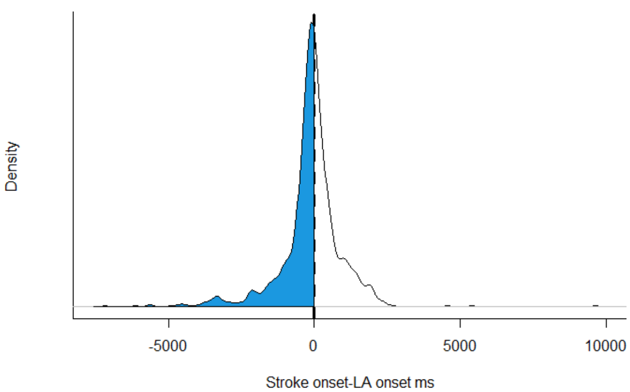

Figure 3. The relationship between gesture stroke and LA onset. The density plot shows the timing difference between gesture stroke onset and LA onset in ms, where negative values (blue) indicate gestures that preceded their LAs. 0ms indicated by the dashed black line.

Do Any Other Factors Affect Time Latency between Stroke Onsets and LA Onsets? Table 4 showed that AoA rating (β = 37.24, *SE* = 18.27, *t* = 2.039, *p* = .042) was a significant predictor of latency between stroke onsets and LA onsets. Figure 4 shows how changes in AoA rating affected the latency between gesture stroke onsets and the LA onsets. Specifically, the estimate of latency between stroke onsets and LA onsets increased in a positive direction as AoA increased. Since the negative value of latency indicated that the stroke came first, what the model showed is that for one standard deviation (2.14) increase in AoA, the stroke was produced 37.24ms later relative to the corresponding LA. The effects of toy presence (*p* = .077), object familiarity (*p* = .279) and their interaction (*p* = .279) were not significant.

Table 4: Summary of other factors affecting the time latency between gesture stroke onsets and LA onsets

	β	*SE*	*df*	*t*	*p*
Presence	212.83	120.39	836.80	1.768	.077
Familiarity	95.29	88.08	826.67	1.082	.279
AoA	37.24	18.27	808.72	2.039	.042*
Familiarity *Presence	-43.80	265.85	848.17	-0.166	.867

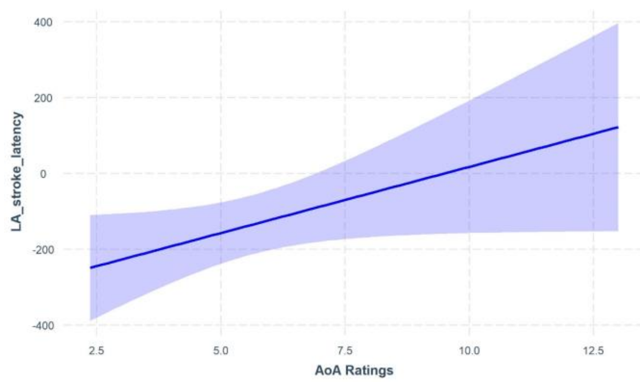

Figure 4. Model-derived predictions of the effect of AoA rating on latency between gesture stroke and LA onsets. Confidence intervals are set to 95%. Negative latencies indicate the stroke comes before the LA, while positive latencies indicate the stroke comes after the LA.

Discussion

In this study, we investigated the timing relationship between caregivers' representational gestures and speech in caregiver-child semi-naturalistic interactions among children aged between 3 to 4 years old. In particular, we focused on the link between gesture phases (gesture preparation and stroke) and speech (utterance onset and LA onset). We also assessed the estimated latency between gesture phase onsets and LA onsets and the influence of factors such as object presence for the latency between gesture preparation and LA, and AoA ratings for the latency between gesture stroke and LA.

We found that gesture phase onsets (both preparation and stroke) were more closely aligned to the corresponding LA than to the onset of the utterance. This finding replicates what has been shown in adult-adult dyads (e.g., Ferré 2010, Donnellan et al., 2022). Therefore, our results do not support the hypothesis based on the growth point theory by McNeill (McNeill, 1992; McNeill & Duncan, 2000), but are consistent with the second possibility, that gestures and speech are linked at a sub-event (lexical) level.

For the estimated latency between gesture phases and LA onsets, we found that gesture phases and LAs were more closely aligned in CDL compared to adults' data (ADL) despite the fact that they both remain within tight temporal proximity. Specifically, the overall estimated latency between preparation and LAs in CDL was approximately 100 ms shorter while the overall estimated latency between gesture stroke and LAs was around 250 ms shorter than ADL.

Our results also showed significant effects of the presence and absence of items on the latency between preparation onsets and LA onsets. In the absent condition, gesture preparation phase was produced significantly earlier relative to the corresponding LA compared to the present condition. We speculate that in the displaced context where the items are absent, caregivers start their gestures earlier to attract children's attention to the upcoming speech and help prepare spatial association between the gesture and upcoming LAs. The finding is in line with the hypothesis that in child-directed language, caregivers may use gestures to support the child's attentional focus on the spoken message and information (McNeill, Cassell, & McCullough, 1994).

Importantly, we found that word AoA ratings were a significant predictor of the latency between gesture stroke phases and LAs. Specifically, for later acquired words, the stroke was produced later relative to the LA, and for early acquired words, the stroke was produced earlier relative to the LA. Particularly, gesture strokes started to be produced after the upcoming of the corresponding LA when the AoA rating was higher than around 10. The finding is in line with the audience design principle that speakers modify their utterances in order to meet the communicative needs of particular conversational partner groups (Clark, 1996; Clark & Murphy, 1982). Going a step further, based on this finding, we speculate that caregivers use gestures in a pedagogical manner: for concepts/words more familiar to the child (acquired earlier), earlier production of representational gestures can support prediction (Donnellan et al., 2022; Zhang et al., 2021) and therefore easier processing. However, as the familiarity of the concept/word decreases, producing gestures at the same time or even after the word may support the activation of semantic properties of the concept/word, or even the acquisition of semantic knowledge associated with uttered words. This possibility is consistent with the fact that we only observed the effect of AoA for the gesture's stroke and not for the preparation phase because the stroke and not the preparation phase marks the meaningful part of the gesture.

Gestures are Communicatively Motivated

Analysis of the temporal alignment between gesture and speech provides insights about whether the production of representational co-speech gestures is driven by production-internal mechanisms (e.g., Kita, 2000) or (or at least to what extent) these gestures are communicatively motivated. While there are already a number of studies showing that in child-directed language, speakers intentionally design both words and gestures to give information that can support the processing of novel information (Campisi and Özyürek, 2013; Clough and Duff, 2020), here we show that this pedagogical function of co-speech gestures extends to their alignment with speech.

Conclusion

In sum, the current research is the first study investigating how caregivers time their representational gestures in naturalistic interactions in CDL. Our results align with previous co-speech gesture timing studies that look at natural conversations by adults even with tighter gesture speech synchrony, thus extending the finding of the timing relationship between gestures and speech to the child-directed and natural caregiver-children conversations. More importantly, our findings further suggest that caregivers' gestures may support addresses' linguistic processing not only by supporting word prediction for well-known words, but also by supporting the learning of conceptual features of unfamiliar words by adjusting the timing of speech and gesture in a pedagogical and efficient way.

Acknowledgements

The work reported here was funded by a European Research Council Advanced Grant (ECOLANG, 743035) and Royal Society Wolfson Research Merit Award (WRM\R3\170016) to GV.

References

Anderson, D. R., & Burnham, K. P. (2002). Avoiding pitfalls when using information-theoretic methods. *The Journal of wildlife management*, 912-918.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.

Bergmann, K., Aksu, V., & Kopp, S. (2011). The relation of speech and gestures: Temporal synchrony follows semantic synchrony. *In Proceedings of the 2nd Workshop on Gesture and Speech in Interaction* (GeSpIn 2011).

Berman, R. A., & Slobin, D. I. (1994). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Erlbaum.

Campisi, E., & Özyürek, A. (2013). Iconicity as a communicative strategy: Recipient design in multimodal demonstrations for adults and children. Journal of Pragmatics, 47(1), 14-27.

Cartmill, E. A., Armstrong III, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences*, 110(28), 11278-11283.

Chui, K. (2005). Temporal patterning of speech and iconic gestures in conversational discourse. *Journal of Pragmatics*, 37(6), 871-887.

Church, R. B., Kelly, S., & Holcombe, D. (2014). Temporal synchrony between speech, action and gesture during language production. *Language, Cognition and Neuroscience*, 29(3), 345-354.

Clark, H. H. (1996). Using language. Cambridge, England: Cambridge University Press

Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. In J.-F. Le Ny & W. Kintsch (Eds.), Language and comprehension (pp. 287–299). Amsterdam: North-Holland

Clough, S., & Duff, M. C. (2020). The role of gesture in communication and cognition: Implications for understanding and treating neurogenic communication disorders. *Frontiers in Human Neuroscience*, *14*, 323.

Clough, S., & Hilverman, C. (2018). Hand gestures and how they help children learn. *Frontiers for young minds*, *6*.

Cristia, A. (2013). Input to Language: The Phonetics and Perception of Infant-Directed *Speech. Language and Linguistics Compass*, 7(3), 157-170. doi: 10.1111/lnc3.12015

Cristia, A., Dupoux, E., Ratner, N., & Soderstrom, M. (2019). Segmentability Differences between Child-Directed and Adult-Directed Speech: A Systematic Test with an Ecologically Valid Corpus. *Open Mind*, 3, 13-22. doi: 10.1162/opmi_a_00022

de Ruiter, J.P., 2000a. *The production of gesture and speech.* In: McNeill, David (Ed.), Language and Gesture. Cambridge University Press, pp. 248–311

de Villiers Rader, N., & Zukow-Goldring, P. (2012). Caregivers' gestures direct infant attention during early word learning: the importance of dynamic synchrony. *Language Sciences*, 34(5), 559-568.

Dick, A., Mok, E., Beharelle, A., Goldin-Meadow, S., & Small, S. (2012). Frontal and temporal contributions to understanding the iconic co-speech gestures that accompany speech. *Human Brain Mapping*, 35(3), 900-917. doi: 10.1002/hbm.22222

Donnellan, E., Özder, L., Man, H., Grzyb, B., Gu, Y., & Vigliocco, G. (2022). Timing relationships between representational gestures and speech: A corpus based investigation. *Proceedings of the Annual Meeting of the Cognitive Science Society, 44.*

Goldin-Meadow, S., & Alibali, M. (2013). Gesture's Role in Speaking, Learning, and Creating Language. *Annual Review of Psychology*, 64(1), 257-283. doi: 10.1146/annurev-psych-113011-143802

Farwell, C., Snow, C., & Ferguson, C. (1979). Talking to Children: Language Input and Acquisition. *Language*, 55(2), 449. doi: 10.2307/412603

Fernald, A. (2000). Speech to Infants as Hyperspeech: Knowledge-Driven Processes in Early Word Recognition. *Phonetica*, 57(2–4), 242–254.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology*, 27(2), 209–221.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A crosslanguage study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, 16(3), 477–501. Cambridge Core.

Ferré, G. (2010). Timing relationships between speech and co-verbal gestures in spontaneous French. In Language Resources and Evaluation, *Workshop on Multimodal Corpora* (Vol. 6, pp. 86-91).

Freedman, N. (1977). *Hands, words, and mind: On the structuralization of body movements during discourse and the capacity for verbal representation.* In Communicative structures and psychic structures (pp. 109-132). Springer, Boston, MA.

Gogate, L., Bolzani, L., & Betancourt, E. (2006). Attention to Maternal Multimodal Naming by 6- to 8-Month-Old Infants and Learning of Word-Object Relations. *Infancy*, 9(3), 259-288. doi: 10.1207/s15327078in0903_1

Hadar, U., & Butterworth, B. (1997). Iconic gestures, imagery, and word retrieval in speech. *Semiotica*, 115(1-2). doi: 10.1515/semi.1997.115.1-2.147

Hall, C., Wakefield, E., & Goldin-Meadow, S. (2022). *Gesture can facilitate children's learning and generalization of verbs.* In A. Morgenstern & S. Goldin-Meadow (Eds.), Gesture in language: Development across the lifespan (pp. 185–202). De Gruyter Mouton; American Psychological Association.

Harling, P., Edwards, A., 1997. *Hand tension as a gesture segmentation cue*. In: Harling, P., Edwards, A. (Eds.), Progress in Gestural Interaction: Proceedings of the Gesture Workshop 1996. Springer, Berlin, Heidelberg, New York, pp. 75–87.

Hockett, C, D, (1960). *The origin of speech*. Scientific American. 203(3), 88-96.

Holler, J., & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in Cognitive Sciences*, 23(8), 639-652.

Holler, J., Shovelton, H., & Beattie, G. (2009). Do Iconic Hand Gestures Really Contribute to the Communication of Semantic Information in a Face-to-Face Context?. *Journal of Nonverbal Behavior*, 33(2), 73-88. doi: 10.1007/s10919-008-0063-9

Hostetter, A. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin*, 137(2), 297-315. doi: 10.1037/a0022128

Hostetter, A., Alibali, M., & Kita, S. (2007). I see it in my hands' eye: Representational gestures reflect conceptual demands. *Language and Cognitive Processes*, 22(3), 313-336. doi: 10.1080/01690960600632812

Ianì, F., Cutica, I., & Bucciarelli, M. (2016). Timing of Gestures: Gestures Anticipating or Simultaneous With Speech as Indexes of Text Comprehension in Children and Adults. *Cognitive Science*, 41, 1549-1566. doi: 10.1111/cogs.12381

Ianì, F., Cutica, I., & Bucciarelli, M. (2016). Timing of Gestures: Gestures Anticipating or Simultaneous With Speech as Indexes of Text Comprehension in Children and Adults. *Cognitive Science*, 41, 1549-1566. doi: 10.1111/cogs.12381

Iverson, J. M., & Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature*, *396*(6708), 228-228.

Lindell, M. K., & Perry, R. W. (1992). *Behavioral foundations of community emergency planning*. Hemisphere Publishing Corp.

Long, J. A., (2022). jtools: Analysis and Presentation of Social Scientific Data. R package version 2.2.0, https://cran.r-project.org/package=jtools.

Iverson, J., Capirci, O., & Caselli, M. (1994). From communication to language in two modalities. *Cognitive Development*, 9(1), 23-43. doi: 10.1016/0885-2014(94)90018-3

Iverson, J., Capirci, O., Longobardi, E., & Cristina Caselli, M. (1999). Gesturing in mother-child interactions. *Cognitive Development*, 14(1), 57-75. doi: 10.1016/s0885-2014(99)80018-5

Iverson, J. M., & Thelen, E. (1999). Hand, mouth and brain. The dynamic emergence of speech and gesture. J*ournal of Consciousness studies*, 6(11-12), 19-40.

Kelly, S., & Church, R. (1998). A Comparison between Children's and Adults' Ability to Detect Conceptual Information Conveyed through Representational Gestures. *Child Development*, 69(1), 85-93. doi: 10.1111/j.1467-8624.1998.tb06135.x

Kelly, S., & Church, R. (1997). Can Children Detect Conceptual Information Conveyed Through Other Children's Nonverbal Behaviors?. *Cognition and Instruction*, 15(1), 107-134. doi: 10.1207/s1532690xci1501_4

Kelly, S., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language*, 89(1), 253-260. doi: 10.1016/s0093-934x(03)00335-3

Kendon, A. (1994). *Gesture and understanding in social interaction.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. *The relationship of verbal and nonverbal communication*, 25(1980), 207-227.

Kisa, Y. D., Achen, C., Goldin-Meadow, S., & Casasanto, D. (2022). Why do people gesture more during disfluent speech? A pragmatic account. *In Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 44, No. 44).

Kita, S. (2000). How representational gestures help speaking. *Language and gesture*, 1, 162-185.

Kita, S., Gijn, I. V., & Hulst, H. V. D. (1997, September). Movement phases in signs and co-speech gestures, and their transcription by human coders. *In International Gesture Workshop* (pp. 23-35). Springer, Berlin, Heidelberg.

Kipp, M. (2005). *Gesture generation by imitation: From human behavior to computer character animation*. Universal-Publishers.

Krauss, R., Dushay, R., Chen, Y., & Rauscher, F. (1995). The Communicative Value of Conversational Hand Gesture. *Journal of Experimental Social Psychology*, 31(6), 533-552. doi: 10.1006/jesp.1995.1024

Krauss, R. M. (1998). Why do we gesture when we speak?. *Current directions in psychological science*, 7(2), 54-54.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82.

Kuperberg, G., & Jaeger, T. (2015). What do we mean by prediction in language comprehension?. *Language, Cognition and Neuroscience*, 31(1), 32-59. doi: 10.1080/23273798.2015.1102299

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior research methods*, *44*, 978-990.

Leonard, T. and Cummins, F. (2009). Temporal Alignment of Gesture and Speech. *In Proceedings of Gespin*, Poznan, Pologne (24-26 September).

Levy, E. T., & McNeill, D. (1992). Speech, gesture, and discourse. *Discourse processes*, 15(3), 277-301.

Loehr, D. P. (2004). *Gesture and intonation* (p. 205). Washington, DC: Georgetown University.

McNeill, D. (1992). *Hand and mind: what gestures reveal about thought.* The University of Chicago Press.

McNeill, D. (2014). Gesture–speech unity: Phylogenesis, ontogenesis, and microgenesis. *Language, Interaction and Acquisition*, 5, 137–184.

McNeill, D., Cassell, J., & McCullough, K. E. (1994). Communicative effects of speech-mismatched gestures. *Research on language and social interaction*, 27(3), 223-237.

Morrel-Samuels, Palmer, Krauss, Robert M., 1992. Word familiarity predicts temporal asynchrony of hand gestures and speech. Journal of Experimental Psychology: *Human Learning and Memory* 18 (3), 615– 622.

Namy, L., Campbell, A., & Tomasello, M. (2004). The Changing Role of Iconicity in Non-Verbal Symbol Learning: A U-Shaped Trajectory in the Acquisition of Arbitrary Gestures. *Journal of Cognition and Development*, 5(1), 37-57. doi: 10.1207/s15327647jcd0501_3

Nobe, Shuichi, 2000. Where to most spontaneous representational gestures actually occur with respect to speech? In: McNeill, David (Ed.), *Language and Gesture.* Cambridge University Press, pp. 186– 198.

Özçalışkan, Ş., & Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition*, 96(3), B101-B113. doi: 10.1016/j.cognition.2005.01.001

R Core Team. (2020). R: A Language and Environment for Statistical Computing. https://www.r-project.org/

Sanjeevan, T., Mainela-Arnold, E., Alibali, M., & Evans, J. (2016). The temporal relationship between speech and manual communicative gesture in children with specific

language impairment. *Gesture*, 15(3), 321-339. doi: 10.1075/gest.15.3.03san
Schegloff, E. A. (1984). *On Some Gestures' Relation to Talk.* In J. M. Atkinson and J. Heritage (Eds.), Structures of Social Action. Cambridge: CUP, pp. 266- -298.
Schmidt, C. L., & Lawson, K. R. (2002). Caregiver attention-focusing and children's attention-sharing behaviours as predictors of later verbal IQ in very low birthweight children. *Journal of Child Language*, 29(1), 3-22.
Seyfeddinipur, M. (2006). *Disfluency: Interrupting speech and gesture*. [Sl]:[Sn].
Shneidman, L., & Goldin-Meadow, S. (2012). Language input and acquisition in a Mayan village: how important is directed speech?. *Developmental Science*, 15(5), 659-673. doi: 10.1111/j.1467-7687.2012.01168.x
Shneidman, L., & Goldin-Meadow, S. (2012). Language input and acquisition in a Mayan village: how important is directed speech?. *Developmental Science*, 15(5), 659-673. doi: 10.1111/j.1467-7687.2012.01168.x
Showalter, D. (2003). Thinking about Military Revolution. *Historically Speaking*, 4(4), 9-10. doi: 10.1353/hsp.2003.0061
Singmann, H., Bolker, b., Westfall, J., & Aust, F. (2016). afex: Analysis of Factorial Experiments. R package version 0.16-1. https://CRAN.R-project.org/package=afex
Sloetjes, H., & Wittenburg, P. (2008). Annotation by category-ELAN and ISO DCR. *In 6th international Conference on Language Resources and Evaluation* (LREC 2008).
Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, 27(4), 501–532
Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological science*, 18(7), 587-592.
Sowden, H., Perkins, M., & Clegg, J. (2008). The co-development of speech and gesture in children with autism. *Clinical Linguistics &Amp; Phonetics*, 22(10-11), 804-813. doi: 10.1080/02699200801919745
Ter Bekke, M., Drijvers, L., & Holler, J. (2020). The predictive potential of hand gestures during conversation: An investigation of the timing of gestures in relation to speech.
ter Bekke, M., Özyürek, A., & Ünal, E. (2022). Speaking but not gesturing predicts event memory: A cross-linguistic comparison. *Language and Cognition*, 1-23.
Tolar, T., Lederberg, A., Gokhale, S., & Tomasello, M. (2007). The Development of the Ability to Recognize the Meaning of Iconic Signs. *Journal of Deaf Studies and Deaf Education*, 13(2), 225-240. doi: 10.1093/deafed/enm045
Vigliocco, G., Gu, Y., Motamedi, Y., Grzyb, G., Brekelmans, M., Murgiano, M., Brieke, R., & Perniss, P. (unpublished). The ECOLANG corpus of dyadic interactions between caregivers and their 2–4 year-old child and between two adults.
Wickham, H., Chang, W., & Wickham, M. H. (2016). Package 'ggplot2'. Create elegant data visualisations using the grammar of graphics. *Version*, 2(1), 1-189.
Yap, D., So, W., Melvin Yap, J., Tan, Y., & Teoh, R. (2010). Iconic Gestures Prime Words. *Cognitive Science*, 35(1), 171-183. doi: 10.1111/j.1551-6709.2010.01141.x
Yap, M. B. H., & Jorm, A. F. (2011). The influence of stigma on first aid actions taken by young people for mental health problems in a close friend or family member: findings from an Australian national survey of youth. *Journal of affective disorders*, 134(1-3), 473-477.
Zhang, Y., Frassinelli, D., Tuomainen, J., Skipper, J. I., & Vigliocco, G. (2021). More than words: word predictability, prosody, gesture and mouth movements in natural language comprehension. *Proceedings of the Royal Society B: Biological Sciences*, 288, 20210500.