

BUCLD 38 Proceedings

To be published in 2014 by Cascadilla Press

Rights forms signed by all authors

Children's Comprehension of Cohesive Use of Space by Gestures Accompanying Spoken Discourse

Kazuki Sekine* and Sotaro Kita*

We rarely hear speech without any other visual information, such as gesture, in our everyday life. Studies of young adults have shown that we cannot help but take into account information from both speech and gesture (e.g., Kelly, Özyürek, & Maris, 2010). Children typically learn their language in a multimodal environment. Given that gestures often convey information that is not conveyed in the accompanying speech (McNeill, 1992), children need to integrate information from gesture and speech, not only at word and sentence levels, but also at the discourse level in order to fully understand a speaker's intended message. This study focused on children's abilities to integrate information from gesture and speech at the discourse level. Discourse is defined in this study as a structure in communication signals that span over multiple sentences and multiple gestures.

Previous gesture research has showed that during a narrative, an adult speaker creates coherent discourse by using linguistic devices and idiosyncratic speech-accompanying gestures (Gullberg 2006; McNeill 1992, 2005; McNeill & Levy 1993; Yoshioka, 2005). A speaker sometimes produces gestures to assign particular referents to a specific area in front of the speaker's body. These gestures can contribute to discourse cohesion (McNeill, 1992). For example, when introducing a new protagonist in a narrative, adult speakers often assign them to a specific area by a pointing gesture or an iconic gesture (iconic gestures are gestures that depict objects, actions and movements on the basis of similarity). When the same referent is mentioned again later, the same location is gesturally indicated (Gullberg 2006; So, Kita & Goldin-Meadow, 2009). In other words, once such a space is assigned a particular meaning, it is often maintained throughout the discourse, not unlike the use of space for co-reference in sign language (e.g., Klima & Bellugi, 1982). Such systematic use of space in gestures enhances the

* Department of Psychology, University of Warwick, UK.

All correspondence should be directed to K. Sekine, Email: K.Sekine@warwick.ac.uk

cohesiveness of the discourse (Gullberg, 2006). It has been observed that children start using gestures that co-occur with spoken referential expressions and locate the referents in abstract space around 8 year old (McNeill, 1992), and then use them frequently from 10 or 11 years old (Cassell, 1991; Sekine & Furuyama, 2010).

In comparison to studies about production of gestures that use space cohesively, comprehension of such gestures is understudied. Two studies have shown that adult listeners take up information from cohesive gestures (Cassell, McNeill, & McCullough, 1998; McNeill, Cassell, & McCullough, 1994). They presented a short video clip where a person tells a short story with gesture and speech to participants. Participants were instructed to retell the story to a listener. In the stimulus, the speaker set up two referents in his gesture space with deictic gestures and then linguistically referred back to one of the referents and pointed to the wrong space (the space for the other referent) at the same time. When retelling the narrative, participants attempted to incorporate information from speech and gesture even when they were incongruent with each other. This indicates that participants gleaned information from cohesive gestures. However, it is not clear whether participants integrated information from cohesive gesture and speech during comprehension or during production (retelling).

To our knowledge, no study in the literature has investigated how children integrate information from cohesive use of space in gesture with information from spoken discourse. The studies of children's gesture comprehension have focused on the processing of a single gesture at a time. For example, it has been shown that adults and children can pick up information conveyed uniquely in gestures (e.g. Broaders & Goldin-Meadow, 2010; Goldin-Meadow & Sandhofer, 1999; Goodrich & Hudson-Kam, 2009; Kelly & Church, 1998; Namy, Cambell, & Tomasello 2004; Tomasello, Striano, & Rochat, 1999), facilitate comprehension of semantically co-expressive words (McNeil, Alibali, & Evans, 2001; Morford & Goldin-Meadow, 1992), and integrate the information with semantically co-expressive words (Cocks, Sautin, Kita, Morgan, & Zlotowitz, 2009; Kelly et al., 2010). Other studies have shown that children start integrating gesture and speech that each contributes unique information to the unified interpretation from 5-year-olds when stimuli are presented by video (Kelly, 2001; Sekine, Sowden, & Kita, under review). In other words, children after 5-year-olds can interpret a combination of iconic gesture and speech that have semantic synergy effect, in which the interpretation of speech-gesture combination goes beyond the sum of speech interpretation and gesture interpretation.

Previous studies focused on how people integrate a single gesture and speech

in comprehension, but few studies have investigated how well children and adults integrate information from spoken discourse and a sequence of gestures, in comprehension. Thus, the current study investigated this question. Given the previous studies on gesture production in discourse (Cassell, 1991; Sekine & Furuyama, 2010) and those on semantic integration of speech and gesture in comprehension (Kelly, 2001; Sekine, Sowden, & Kita, under review), we focused on 5-, 6-, 10-year-olds, and adults. We investigated whether these age groups can integrate spoken discourse and cohesive use of space in gestures. Our stimuli were passages consisting of three sentences with accompanying gestures. Each passage referred to two protagonists. The first two sentences indicated two protagonists by conjoined subject noun phrases, and accompanying gestures consistently assigned one protagonist to the right and the other to the left in gesture space. The third sentence was ambiguous without an overt subject noun phrase (it is grammatical in Japanese to omit a subject noun phrase; Shibatani, 1990) and could refer to one of the protagonist's (or both protagonists') action, but an iconic gesture was produced in either the right or left space and made it clear which protagonist (always one protagonist) performed the action. Participants were asked to indicate which protagonist performed the action referred to in the third sentence in a forced choice task.

1. Method

1.1. Participants

24 five-year-olds (mean age: 5;03, range: 4;10 to 5;09), 24 six-year-olds (mean age: 6;3, range: 5;10 to 6;09), 24 ten-year-olds (mean age: 10;03, range: 9;10 to 10;08), and 24 adults (mean age: 23, range: 18 to 31). All participants were monolingual speakers of Japanese, participated. Each age group has 12 females and 12 males.

1.2. Material

An actor was filmed producing combination of gestures and a short passage. 17 vignettes were made in total (two for practice, 15 for the main experiment). The actor's lower part of the face was covered by a mask to conceal lip movements (see Figure 1). Each vignette consisted of three short sentences and gestures. In the first sentence, the actor introduced two protagonists by two conjoined full nouns or proper names in the subject position and an event involving them. In the second

sentence, she referred to the same two protagonists by two conjoined full nouns or proper names in the subject position again and to an event. In the third sentence, she described a protagonist's movement as a result of the event, but omitted the subject and did not explicitly mention any characters. Thus, participants cannot know which character did the movement if they took only the speech into account.

In the first sentence, gestures were produced to assign the two protagonists to the actor's right and left sides of frontal space with her right and left hand respectively when each protagonist was mentioned, as if she places two entities in the space ((2)-(3) in Figure 1). In the second sentence, two gestures placed the protagonists in the same locations as in the first sentence ((4)-(5) in Figure 1). The two hands were being held in the air at the beginning of the third sentence ((6) in Figure 1). In the third sentence, either her right or left hand iconically depicted one of the protagonists' movement within the right or left space, respectively. The stationary hand was held until the other hand finished the gesture ((7) in Figure 1). In other words, the gesture specified which protagonist did the movement. Finally, both hands retracted to the lap.

In each vignette, we made four versions to counterbalance the location of gestures: the location (left or right) in which a gesture assigned the first protagonist in the first two sentences, and the location (left or right) in which depicting a protagonist's movement appeared in the third sentence. Each video lasted about 20 seconds. An example is in Figure 1.

Note that the third sentence did not have an overt subject. It is grammatical in Japanese to omit arguments of a sentence (Shibatani, 1990). As Japanese does not have subject-verb agreement (e.g., based on number and person), it was not clear from the speech whether protagonist A or protagonist B or both protagonists performed the action. However, the accompanying gesture provided a cue to disambiguate who performed the action. Note also that the use of the full noun phrases in the second sentence was natural in Japanese. This is because Japanese does not use third person pronouns in everyday discourse (third person pronouns are used mainly in translation of European languages) and omitting an overt subject would have made the story too unclear.

After the video stimulus was presented, participants were asked the following forced three-choice question; *correct choice*, *incorrect-protagonist choice*, and *both-protagonists choice*. In case of the example in Figure 1, after participants watched the clip, the experimenter asked the participants "Did the mouse fall down, did the frog fall down, or did both of them fall down?" The iconic gesture with the

third sentence was produced in the space associated with the frog in the first two sentences. Therefore, the choice that the frog fell down was coded as a *correct choice*. The choice that the mouse fell down was coded as an *incorrect-protagonist choice*. The choice that both did it was coded as *both-protagonists choice* (this was also an incorrect choice).

1. (2)**Nezumi**-to (3)**Kaeru**-ga kouen-de asondeimasu
 mouse-and frog-NOM park-DAT play.PROG.Polite
 "(2)**Mouse** and (3)**frog** are playing in park."
2. (4)**Nezumi**-to (5)**Kaeru**-wa (6)buranko-ni norimashita
 Mouse-and frog-TOP swing-DAT get.Polite.PST
 "(4)**Mouse** and (5)**frog** got on swing."
3. Demo, hayasugite (7)**ochichaimashita**
 but too.fast-and fall.down-regrettably.Polite.PST
 "But, because too fast, (7)**fell down**."



Figure 1. Example stimulus. The top panel (speech): Word in boldface indicates a word in which a gesture accompanied, and underlines indicate periods in which gesture(s) was held in the air. The abbreviations in the interlinear gloss are ACC (accusative), DAT (dative), NOM (nominative), PST (past tense), PROG (progressive aspect) and TOP (topic marker). The numbers in parentheses indicate where gestures occurred, and correspond to the numbers in the bottom panel. Note that Japanese does not have articles on nouns or commonly used third person pronouns, and it allows omission of arguments as in the third sentence. **The bottom panel (gesture):** Gestures that accompanied the speech stimulus in the top panel.

1.3. Procedure

Participants were tested individually. Participants were asked to watch a video stimulus embedded in a PowerPoint presentation on a laptop with a 12 inch screen. Before watching each vignette, an experimenter told participants (except for adults participants) what protagonist would appear in the next vignette to make it easier for children to remember the protagonists. After each vignette, participants were asked to pick one of three choices about who did the movement in the third sentence. No matter what choice they picked, the experimenter gave them a positive feedback. Two practice trials were followed by 15 experimental trials. Each participant was presented with one of the eight counterbalancing sets for the experimental trials: four gesture locations (as described in the materials section) x two vignettes orders (forward vs. backward). The experiment lasted about 10 minutes.

2. Results

The pattern of responses did not statistically differ between the eight counterbalancing sets. The data was therefore collapsed across counterbalancing sets.

2.1. Correct choices for each age group

To examine age differences in accuracy, an analysis of variance (ANOVA) was conducted on the mean number of correct choices with the age groups as a between-subject factor (see Figure 2 for the means and SEs). A main effect of age group was found, $F(3, 92) = 13.35, p < .001$. Post hoc comparisons (Bonferroni, $p < .05$) showed that adults chose the correct answer significantly more often than 5- and 6-year-olds did, and that 10-year-olds selected the correct answer significantly more often than 5-year-olds did. This indicates that it is relatively difficult for 5- and 6-year-olds to integrate information from both speech and cohesive gestures.

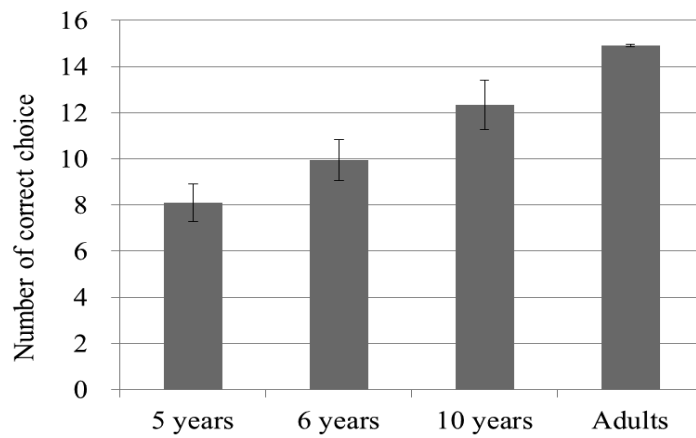


Figure 2. Mean number of trials with a correct choice for each age group (out of 15 trials). The error bars indicate standard errors.

2.2. Error analysis

We analyzed the number of the two types of errors: an *incorrect-protagonist choice* or a *both-protagonists choice*. We calculated the mean number of each type of error for each age group (Table 1).

Table 1. Mean number (*SD*) of two types of error for each age group.

Type of error	5 years	6 years	10 years	Adults
Number of incorrect-protagonist choice	6.0 (3.7)	3.1 (2.4)	0.7 (2.2)	0.04 (0.20)
Number of both-protagonists choice	0.9 (3.2)	2.0 (4.4)	2.0 (5.0)	0.04 (0.20)

The participants rarely selected the *both-protagonists choice*. This is perhaps not surprising as the key gesture in the third sentence was produced by just one hand. In addition, most of adults did not make any errors. To examine age differences, we conducted analyses of variance (ANOVA) on the mean number of each error type with the age groups as a between-subject factor (Table 2). A main effect of age group was found only for the number of *incorrect-protagonist choices*, $F(3, 92) = 29.31$, $p < .001$. Post hoc comparisons (Bonferroni, $p < .05$) showed that the 5-year-olds selected the *incorrect-protagonist choice* significantly more often than

the other three age groups, and that 6-year-olds selected this *incorrect-protagonist choice* more often than the 10-year-olds and adults. This indicates that 5- and 6-year-olds tended to pick the *incorrect-protagonist choice* when they made an error, and suggests that they did not understand which one of the two protagonists was intended.

2.3. Proportion of correct choice between the two one-protagonist choices

As the participants rarely picked the *both-participant choice*, we examined whether the proportion of correct choices was above the chance level (50%) when they picked one of the two one-protagonist choices (*correct choice* vs. *incorrect-protagonist choice*) (Table 2), by excluding trials in which the participants picked the both-protagonist choice. Some participants were excluded from the analysis because they selected both-protagonist choice in all trials. Note that the chance level was 50% regardless of how the first and second sentences set up a linguistic bias for a particular referent being the intended subject in the third sentence. This is because for each passage two pairs of stimuli were created that counterbalanced the right-left locations of the gestures in the first and second sentence and in the third sentence. Thus, in half of the trials, one protagonist was the correct choice, in the other half, the other protagonist was the correct choice. The proportions of trials with the correct choice were significantly higher than the chance level (0.5) for all age groups except for 5-year-olds. This indicated that it was difficult for 5-year-olds to integrate information from spoken discourse and cohesive gestures to pick the correct protagonist.

Table 2. The mean (SD) proportion of correct choices among the two one-protagonist choices.

	Mean	(SD)	T value
5 years (N=23)	.57	(.24)	1.42
6 years (N=23)	.76	(.19)	6.53***
10 years (N=21)	.95	(.16)	13.01***
Adults (N=24)	1.00	(.01)	170.43***

*** $p < .001$ for one sample t-tests against the chance level (.50)

3. Discussion

This study examined how well children and adults integrated information from spoken discourse and cohesive use of space in gesture, in comprehension. There are two main findings. First, we did not find evidence that 5-year-olds could integrate information from spoken discourse and cohesive use of space in gestures, but 6-year-olds could perform above the chance level, although their performance was not good as adults. Previous studies (Kelly, 2001; Sekine, Sowden, & Kita, under review) showed that when participants were shown video recordings of a combination of a single utterance and a gesture, 5-year-olds could select correct choices above the chance level. Thus, the integration of speech and gesture at discourse level develops later than the utterance/single-gesture level integration.

Second, we showed, for the first time, that adults can integrate information from spoken discourse and cohesive gestures in comprehension. Research on gesture has revealed adult speakers produce gestures to create coherent discourse with speech (Gullberg 2006; McNeill 1992, 2005; McNeill & Levy 1993; So, Kita & Goldin-Meadow, 2009; Yoshioka, 2005). However it was previously not clear that the listener could actually integrate cohesive use of space in gestures with spoken discourse to arrive at a unified interpretation. In this regard, the current studies added new empirical finding to the adult literature as well.

Some theorists claimed that information from gesture and speech is always integrated (Kelly et al., 2010) based on data from young adults and suggested (Kelly, Creigh, & Bortolotti, 2010) that such automaticity has foundation in phylogenetically old action interpretation system, namely, mirror neurons, which are hypothesized to be linked to language evolution (Rizzolatti & Arbib, 1998). If this was the case, one would expect that young children can integrate speech and gesture. However, the current results indicate that older children and young adults do so at the discourse level, but not young children. The present study showed that obligatory speech-gesture semantic integration in discourse develops in late childhood. This is more compatible with the studies about production of gestures that structure discourse. Children start producing gestures that co-occur with spoken referential expressions, and locate the referents in abstract space frequently from 10 or 11 years old (Cassell, 1991; Sekine & Furuyama, 2010). Thus, an ability to integrate speech and gesture at discourse level in both production and comprehension may develop in parallel in late childhood.

References

- Bellugi, U., & Klima, E. (1982). From gesture to sign: Deixis in visual-gestural language. In R. J. Jarvella and W. Klein (Eds.), *Speech, place and action: Studies in deixis* (pp.297-313). New York: John Wiley & Sons.
- Broaders, S. C., & Goldin-Meadow, S. (2010). Truth is at hand: How gesture adds information during investigative interviews. *Psychological Science, 21*(5), 623-628.
- Cassell, J. (1991). *The development of time and event in narrative: Evidence from speech and gesture*. Unpublished doctoral dissertation, University of Chicago.
- Cassell, J., McNeill, D., & McCullough, K.E. (1999). Speech-gesture mismatches: Evidence for one underlying representation of linguistic and non-linguistic information. *Pragmatics and Cognition, 7*(1): 1-33.
- Cocks, N., Sautin, L., Kita, S., Morgan, G., & Zlotowitz, S. (2009). Gesture and speech integration: An exploratory study of a man with aphasia. *International Journal of Language and Communication Disorders, 44*, 795-804.
- Goldin-Meadow, S., & Sandhofer, C. M. (1999). Gestures convey substantive information about a child's thoughts to ordinary listeners. *Developmental Science, 2*(1), 67-74.
- Goodrich, W., & Hudson-Kam, C. L. (2009). Co-speech gesture as input in verb learning. *Developmental Science, 12*(1), 81-87.
- Gullberg, M. (2006). Handling discourse: Gestures, reference, tracking, and communication Strategies in early L2. *Language Learning, 56*(1), 155 -196.
- Kelly, S. D. (2001). Broadening the units of analysis in communication: Speech and nonverbal behaviours in pragmatic comprehension. *Journal of Child Language, 28*, 325-349.
- Kelly, S. D., Creigh, P., & Bartolotti, J. (2010). Integrating speech and iconic gestures in a stroop-like task: Evidence for automatic processing. *Journal of Cognitive Neuroscience, 22*(4), 683-694.
- Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science, 21*, 260-267.
- McNeil, N. M., Alibali, M. W., & Evans, J. L. (2001). The role of gesture in children's comprehension of spoken language: Now they need it, now they don't. *Journal of Nonverbal Behavior, 24*(2), 131-150.
- McNeill, D. (1992). *Hand and mind*. Chicago: University of Chicago Press.
- McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.
- McNeill, D., Cassell, J., & McCullough, K. E. (1994). Communicative effects of speech-mismatched gestures. *Language and Social Interaction, 27*, 223-237.
- McNeill, D., & Levy, E. T. (1993). Cohesion and gesture. *Discourse Processes, 16*(4),

363-386.

- Morford, M., & Goldin-Meadow, S. (1992). Comprehension and production of gesture in combination with speech in one-word speakers. *Journal of Child Language*, 19(3), 559-580.
- Namy, L. L., Campbell, A. L., & 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.
- Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, 21(5), 188-194.
- Tomasello, M., Striano, T., & Rochat, P. (1999). Do young children use objects as symbols? *British Journal of Developmental Psychology*, 17, 563-584.
- Sekine K., & Furuyama, N. (2010). Developmental change of discourse cohesion in speech and gestures among Japanese elementary school children. *Rivista di psicolinguistica applicata*, 10(3), 97-116.
- Shibatani, M. (1990). *The languages of Japan*. Cambridge: Cambridge University Press.
- So, W. C., Kita, S., & Goldin-Meadow, S. (1990). Using the hands to identify who does what to whom: Gesture and speech go hand-in-hand. *Cognitive Science*, 33, 115-125.
- Yoshioka, K. (2005). *Linguistic and gestural introduction and tracking of referents in L1 and L2 discourse*. Nijmegen, Radboud University.