Acta Psychologica 179 (2017) 89–95

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

The listener automatically uses spatial story representations from the speaker's cohesive gestures when processing subsequent sentences without gestures

Kazuki Sekine[a,b,*], Sotaro Kita[c]

[a] Radboud University, The Netherlands
[b] Max Planck Institute for Psycholinguistics, The Netherlands
[c] University of Warwick, United Kingdom

ARTICLE INFO

Keywords:
Gesture
Speech
Discourse
Multimodal communication
Speech comprehension
Speech-gesture integration

ABSTRACT

This study examined spatial story representations created by speaker's cohesive gestures. Participants were presented with three-sentence discourse with two protagonists. In the first and second sentences, gestures consistently located the two protagonists in the gesture space: one to the right and the other to the left. The third sentence (without gestures) referred to one of the protagonists, and the participants responded with one of the two keys to indicate the relevant protagonist. The response keys were either spatially congruent or incongruent with the gesturally established locations for the two participants. Though the cohesive gestures did not provide any clue for the correct response, they influenced performance: the reaction time in the congruent condition was faster than that in the incongruent condition. Thus, cohesive gestures automatically establish spatial story representations and the spatial story representations remain activated in a subsequent sentence without any gesture.

People often produce gestures while speaking. Research on such co-speech gestures has revealed that the listener/observer can take up information from the speaker's gestures and use it to comprehend an underlying overall message (Beattie & Shovelton, 1999; Cassell, McNeill, & McCullough, 1999; Kendon, 1994). This study examined whether a listener uses spatial information expressed in gestures even after the gesture has disappeared.

Most of the previous research on gesture comprehension focused on the processing of a single gesture at word or sentence level. For example, some studies shown that adults and children can pick up information conveyed exclusively in gestures (e.g., Broaders & Goldin-Meadow, 2010; Kelly & Church, 1998; Namy, Campbell, & Tomasello, 2004; Morford & Goldin-Meadow, 1992). Other studies reported that adults and children integrate gesture and speech, each of which contributes unique information to the unified interpretation (e.g., adults: Cocks, Sautin, Kita, Morgan, & Zlotowitz, 2009; Kelly, Özyürek, & Maris, 2010; children: Kelly, 2001; Sekine, Sowden, & Kita, 2015). Thus, the findings from these studies suggest that adults and children can pick up information conveyed by gesture and integrate it with information from the concurrent speech. However, these studies have focused on comprehension of speech and a single gesture, and thus integration of speech and gesture at word or sentence level. Comprehension of speech and a sequence of gestures at the discourse level is under-studied.

Studies on gestures in discourse have revealed that during a narrative, an adult speaker builds coherent discourse by using linguistic devices and speech-accompanying gestures (Gullberg, 2006; McNeill, 2005; McNeill & Levy, 1993; Yoshioka, 2005). As McNeill (1992) argued, adult speakers often use gestures to indicate continuity of a topic by repeating the same form or the same location. For example, when a new protagonist is introduced in a story, an adult speaker locates the referent to specific space in front of him or her by a pointing gesture or an iconic gesture (iconic gestures are gestures that depict objects, actions and movements on the basis of similarity). When mentioning the same referent again later, (s)he gesturally indicates the same location (Gullberg, 2006; So, Kita, & Goldin-Meadow, 2009). McNeill and Levy (1993) argued that the assigned spaces for referents were gestured more frequently when characters were re-introduced with explicit referring form such as a noun phrase than when the narrative maintained focus on one character with a less explicit form such as a pronoun. So et al. (2009) found that speakers tended to produce gestures in a particular location in gesture space to identify referents that were also uniquely specified in speech (e.g., two different gender protagonists were referred to by the pronouns “he” and “she”), rather than referents that

* Corresponding author at: Max Planck Institute for Psycholinguistics, Room 308, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands.
E-mail address: kazuki.sekine@mpi.nl (K. Sekine).

http://dx.doi.org/10.1016/j.actpsy.2017.07.009
Received 30 September 2016; Received in revised form 10 July 2017; Accepted 17 July 2017
Available online 24 July 2017
0001-6918/

were ambiguous in speech (e.g., two same gender protagonists were referred to by the same pronoun "he"). These findings suggest that speakers tend to use locations in gesture space to indicate referents that are lexically specified by the concurrent speech.

Gestural reference-tracking is attained when specific gestural behaviors, whose features (e.g., location, handedness, movement, orientation, hand shape) are repeated, are systematically associated with referential expressions in speech (Gullberg, 2006). Such an association establishes explicit, visual co-reference, and thus, enhances the cohesiveness of the discourse (McNeill, 2005). Previous studies have shown that referential space is constructed and used not only by individuals (So et al., 2009) but also by conversation participants (Stec & Huiskes, 2014). Gestures used in establishing locations for referents and tracking the references in the discourse are called cohesive gestures (McNeill, 1992), as they contribute to the discourse cohesion.

There have only been seven studies that investigated how listeners process cohesive gestures. Two studies investigated how spatial information in cohesive gesture influenced subsequent speech and gesture production. First, Cassell et al. (1999) showed that listeners take up information from the cohesive use of space in gesture. Cassell et al. presented a video-recorded narrative to adult participants, who then retold the story to a listener. In the stimulus narrative, a narrator located two protagonists in his frontal space with deictic gestures, and then linguistically referred back to one of the protagonists while pointing to the wrong space (the space for the other referent). When retelling the narrative, participants incorporated information from gesture and speech even when they were incongruent with each other.

Second, an EEG study by Gunter, Weinbrenner, and Berndt (2012) found that the brain prepares to produce cohesive gestures even when producing the gestures was not required for the task. Participants watched video clips where a narrator tells stories and establishes two locations for two referents (e.g., left side for cats and right side for dogs) with cohesive gestures. Then the participants were asked to respond verbally to a question like "Which animal barks?" (no gesture was produced). Event related potentials (ERPs) revealed that they covertly prepared to produce a cohesive gesture by the left or right hand that was compatible with the location (left or right) where the gestures in the stimulus placed the relevant referent.

Five previous studies investigated how cohesive gestures influence comprehension of discourse, which is the topic of the current study. Three EEG studies (Gunter & Weinbrenner, 2017; Gunter, Weinbrenner, & Holle, 2015; Weinbrenner, 2017) showed that cohesive gestures influenced listener's comprehension of a sentence even when the gesture was not crucial for interpreting the sentence. In Gunter et al.'s (2015) study, participants were presented with video clips of an interview between an interviewer and an interviewee. In each video clip, the interviewee talked about a topic consisting of two opposing referents (e.g., "Donald vs. Mickey") and assigned the two referents to two locations in gesture space with cohesive gestures (e.g., left space for Donald and right space for Mickey). With the target sentence at the end of each topic, the interviewee produced a cohesive gesture that was either congruent (pointing to the left while saying "Donald") or incongruent (pointing right while saying "Donald") to the previously established location with a sentence like "As far as I know, Donald was created later". The target sentence was unambiguous and fully interpretable without the accompanying gesture. Participants were asked to pay attention to the video clips as they were given a memory task about the video content, which was neither about the content of the target sentence nor was it related to gesture. The result showed that the congruency between speech and cohesive gesture influenced ERPs recorded from the participants: N400 and P600 components were larger in the incongruent condition than in the congruent condition. This indicates that it was more difficult to process the interviewee's message when the cohesive gesture was incongruent with the speech, even though participants were not asked to pay attention to gestures while watching the stimulus videos. The same pattern of results was found by other studies (Gunter & Weinbrenner, 2017; Weinbrenner, 2017).

Fourth, Goodrich Smith and Hudson Kam (2012) showed that cohesive gestures can influence the interpretation of otherwise ambiguous sentences. They investigated whether cohesive gestures influenced interpretation of ambiguous pronouns. Participants watched video clips of a narrator telling stories that ended with a sentence with an ambiguous pronoun: e.g., "*Annie and Sarah are having a picnic in the park. They have a lot of food with them. Annie is carrying the picnic basket, and Sarah has a blanket to sit on. She is excited about the cookies*". In the first two sentences, the narrator consistently located the two protagonists to either her right or left side with cohesive gestures. In the last sentence, the narrator's gesture was manipulated: she either produced no gesture, indicated the location of the first-mentioned protagonist or the second-mentioned protagonist. After each clip, the participants were presented with a question (without any gestures) about the referent of the ambiguous pronoun (e.g., "*Who is excited about the cookies?*"). The participants tended to respond with the referent that was consistent with the location indicated by the gesture in the third sentence. This indicates that cohesive gestures influence people's interpretation of the ambiguous pronounce.

Fifth, Sekine and Kita's (2015) study showed that cohesive gestures influence comprehension of discourse in elementary school children. They examined how well Japanese 5-, 6-, 10-year-olds and adults integrated information from spoken discourse and cohesive gestures in comprehension. The participants were presented with three-sentence stories. In the first two sentences, a narrator referred to two protagonists by full nouns and an event involving them (e.g., "Nori-kun and Yuuto-kun are crossing a pedestrian bridge. Nori-kun and Yuuto-kun are ascending stairs." Note that Nori-kun and Yuuto-kun are Japanese boys' names). The narrator produced gestures to consistently locate each of the two protagonists in two distinct locations (e.g., left space for Nori-kun and right space for Yuuto-kun). In the third sentence, she described a protagonist's movement without explicitly mentioning any protagonists, which is grammatically possible in Japanese (e.g., "and suddenly, (one) tumbled down"). In addition, she iconically depicted one of the protagonists' movements within the right or left space. Thus, participants could infer which character did the movement only if they took the gestures into account. Then, they were asked to indicate which protagonist performed the action in the third sentence. The result showed that 6- and 10-year-olds, and adults consistently selected the protagonist consistent with the location indicated by the iconic gesture in the third sentence, but not 5-year-olds, whose choice was at chance.

These five studies on the impact of cohesive gestures on discourse comprehension showed that cohesive gestures influence processing of the concurrent sentence. However, it is not clear if cohesive gestures influence processing of a subsequent sentence *without* any accompanying gestures.

Hudson Kam and Goodrich Smith (2011) showed that spatial information encoded in gestures persists beyond the sentence that the gestures co-occurred with. In their study, participants were presented with video clips where an actor located two protagonists in the left and the right side of the gesture space with cohesive gestures (e.g., Andrea on the right and Bobby on the left, from the participant's perspective). After watching each clip, participants chose one of two pictures that best represented the story they heard. One picture showed one protagonist on the right and the other on the left, and the other picture flipped the left-right positions of the two protagonists. Participants systematically picked the picture with the two protagonists located in the left-right positions compatible with where the gestures located the two protagonists (e.g., Andrea dancing on the right, and Bobby singing on the left, from the participant's perspective). However, in this task, because locations indicated by gestures were the only clue that allowed participants to select the response, the task required participants to pay attention to the gestures and try to remember the locations even after the story. Thus, it is still not clear whether the spatial representation created by gestures is activated *automatically*, that is, in a situation