Title
Co-speech gestures do not originate from speech production processes: Evidence from the relationship between co-thought and co-speech gestures

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 31(31)

ISSN
1069-7977

Authors
Chu, Mingyuan
Kita, Sotaro

Publication Date
2009

Peer reviewed

Co-speech gestures do not originate from speech production processes: Evidence from the relationship between co-thought and co-speech gestures

Mingyuan Chu (M.Chu.1@bham.ac.uk)
School of Psychology, University of Birmingham,
Edgbaston, Birmingham, B15 2TT, UK.

Sotaro Kita (s.kita@bham.ac.uk)
School of Psychology, University of Birmingham,
Edgbaston, Birmingham, B15 2TT, UK.

Abstract

When we speak, we spontaneously produce gestures (co-speech gestures). Co-speech gestures and speech production are closely interlinked. However, the exact nature of the link is still under debate. To addressed the question that whether co-speech gestures originate from the speech production system or from a system independent of the speech production, the present study examined the relationship between co-speech and co-thought gestures. Co-thought gestures, produced during silent thinking without speaking, presumably originate from a system independent of the speech production processes. We found a positive correlation between the production frequency of co-thought and co-speech gestures, regardless the communicative function that co-speech gestures might serve. Therefore, we suggest that co-speech gestures and co-thought gestures originate from a common system that is independent of the speech production processes.

Keywords: co-thought gestures; co-speech gestures; speech production.

People often spontaneously gesture when they speak (co-speech gestures). There is a consensus in the literature that co-speech gesture and speech production are tightly linked. They are highly coordinated semantically and well synchronized temporally (McNeill, 1992). For example, a speaker may draw circles in the air with an extended index finger and say “rotating” simultaneously when describing a rotational movement. When speech is dysfluent, gesture is interrupted as well. Mayberry and Jaques (2000) showed that co-speech gestures were held motionless during stuttering in speech. Furthermore, speaking and gesturing can influence each other. The way people verbally expressed a motion event had an effect on the way they gesturally depicted it (Kita & Özyürek, 2003) and prohibiting or allowing gesture could alter children’s explanations of Piagetian conservation tasks (Alibali & Kita, under review).

However, there is a lack of consensus regarding the exact nature of the link between co-speech gesture and speech production. One class of theories claims that these two systems are inherently inseparable. According to McNeill (1992), speech and gesture are a single-integrated system and they both arise from a “growth point”, which is the speaker’s minimal idea unit that combines image and word. Meanwhile, some other researchers suggest that co-speech gesture originates from subprocesses of speech production. For example, gesture is generated from the lexical retrieval process (Butterworth & Hadar, 1989; Rauscher, Krauss & Chen, 1996) or the “conceptualizer”, which specifies the pre-linguistic message to be verbalized in the next utterance (de Ruiter, 2000). Generally speaking, this class of theories holds that co-speech gesture production is inseparable from the speech production process.

An alternative view is that co-speech gesture and speech production are two interactive but independent systems (Kita, 2000; Kita & Özyürek, 2003). Kita and Lausberg (2008) found that the linguistically non-dominant hemisphere alone in split-brain patients can generate co-speech gestures based on spatial imagery. This result indicates that co-speech gesture and speech production are dissociable processes. In addition, co-speech gestures can express different information from the concurrent speech (e.g., Perry, Church, & Goldin-Meadow, 1988; Garber & Goldin-Meadow, 2002). The semantic mismatch between co-speech gesture and speech indicates that at least some gestures are produced independently of the speech production process. Kita (2000) further proposed that co-speech gesture is generated from spatio-motoric thinking (or an “Action Generator” in Kita & Özyürek, 2003), which organizes information with action schemas and their modulation according to the environmental information. In other words, co-speech gestures originate from a cognitive system that is independent of the speech production system and responsible for generating body movements in the physical environment.

In addition to co-speech gestures, people also spontaneously gesture when they solve problems without speaking (co-thought gestures). Schwartz and Black (1996) asked participants to verbally explain their solutions of some simple gear problems. The authors found that many participants produced co-thought gestures before their verbal response. Furthermore, people spontaneously produce co-thought gestures during problem solving even in a task that does not involve any use of language. Chu and Kita (2008) found that people spontaneously produced co-thought gestures in a mental rotation task, in which participants seated alone in a room and only needed to make left or right judgments by pressing the correspondent foot pedal. These co-thought gestures presumably originate from an action generation system that does not involve any speech production process. Therefore, if one assumes that co-speech gestures originate from a part of speech production processes (e.g., Butterworth & Hadar, 1989; de Ruiter, 2000; McNeill, 1992), then co-speech and co-thought gestures must originate from two different processes. Consequently, there should be no systematic relationship between the two behaviours (see Figure 1a). In contrast, if one assumes that co-speech and co-thought gestures both originate from an action generation system (outside of the speech production process), then there may be a systematic relationship between the two behaviours (see Figure 1b).

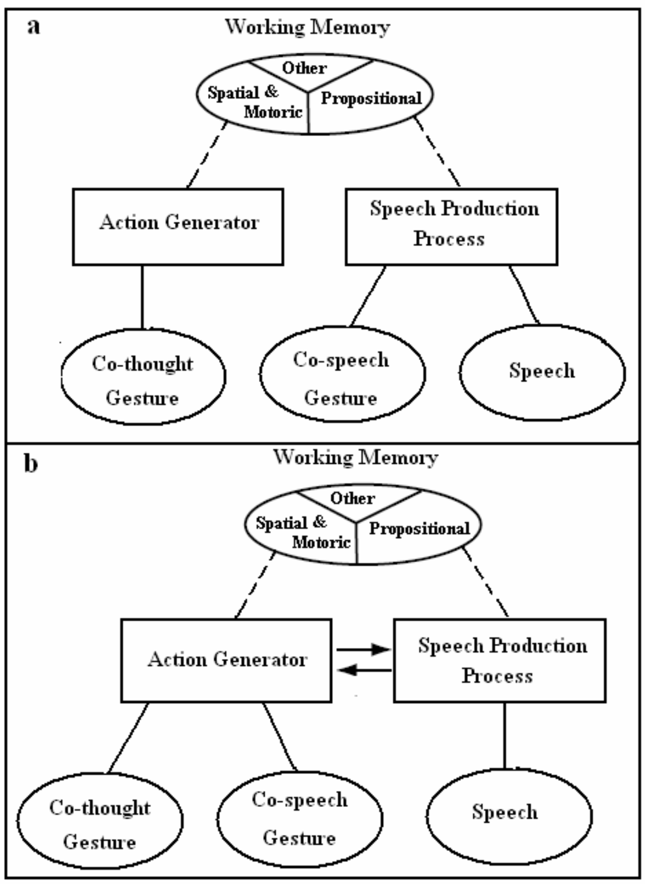

Figure 1. Two possible mechanisms underlying the production of co-thought and co-speech gesture.

Chu and Kita (2008) indeed found a systematic relationship between co-thought and co-speech gestures, suggesting that they may originate from a single system. Over the course of the experiment, the rate of both co-thought and co-speech gestures decreased, and the representational contents of both co-thought and co-speech gestures changed in the same pattern. These parallel findings between co-thought and co-speech gestures suggest that they may originate from the same system, which is independent of the speech production system.

The present study aims to investigate whether co-speech gestures originate from the speech production system or from a system that is independent of the speech production processes. We used a within-participant design to examine the relationship between the rates of co-thought gestures and co-speech gestures. Individuals differ greatly as to how often they produce co-speech gestures (Hostetter & Alibali, 2007) and co-thought gestures (Chu & Kita, 2008). If co-thought and co-speech gestures originate from a common system, we should expect a correlation between the rates of co-thought and co-speech gestures. However, if they originate from different systems, there should not be any systematic relationship.

In the present study, we also manipulated the communicative context of the tasks that were used to elicit co-speech gestures. A robust finding in the literature is that the rate of co-speech gesture varies according to the communicative context. Speakers produce more co-speech gestures in a face-to-face interaction than in conditions in which gestures cannot be seen by the addressee (Alibali, Heath, & Meyer, 2001; Bavelas, et al., 2008; Cohen, 1977). One possible explanation of this rate difference may be that co-speech gestures originate from different mechanisms in different communicative contexts. For example, in a face-to-face interaction, co-speech gestures might purely serve for communicative purpose, whereas in the gesture non-visible condition, co-speech gestures might merely be triggered by the cognitive demands from the speakers themselves. Therefore, the present study obtained the rates of co-speech gestures in situations in which the speaker talks to an addressee face-to-face and to a tape-recorder alone in the room. We correlated the co-thought gesture rates with the co-speech gesture rates in both conditions to see if the correlation was robust across communicative contexts in which co-speech gestures were produced.

Method

Participants

Forty one native English speakers (37 females and 4 males) at the University of Birmingham took part in this experiment. All participants had normal or corrected-to-normal vision. The participants' age ranged from 18 to 28 years (M = 19.05, SD = 1.73).

Tasks

Mental rotation task. Shepard and Metzler (1971) type of three-dimensional objects was used (see Figure 2) to elicit co-thought gestures. The upper left and right objects were mirror images of each other. The lower object was rotated in four angles (60°, 120°, 240° and 300°) around the bisector that went through the object's centre between the horizontal and vertical axis, the horizontal and in-depth axis, and the vertical and in-depth axis. The lower object was rotated from the upper left object in half of the trials and from the upper right object in the other half of the trials.

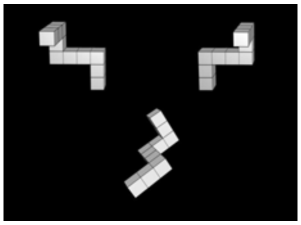

Figure 2. An example of a stimulus (Lower object, 60 degrees on bisector of x-axis and y-axis rotation; Upper left and right objects in the canonical position).

Participants seated alone in a room and their task was to make a judgment on whether the lower object was the same as the upper left or right object by pressing the correspondent left or right foot pedal. They were told that accuracy was the first priority, and it was not important to respond quickly. We de-emphasized quickness of responses so that spontaneous gestures were not suppressed due to the time pressure. Each trial began with a white fixation cross in the center of the screen for 1000ms, followed by the stimulus. When the response was given, the next trial started automatically. No feedback was given concerning the accuracy of the response. During the experiment, participants responded with two foot-pedals silently, leaving their hands free for spontaneous gestures. Their behaviours were video-recorded by a hidden camera (Sony PAL DV).

Speaking tasks. Four speaking tasks were used to elicit co-speech gestures. In the geometric shape motion task, eight movie clips depicting movements of two geometric shapes (circle, triangle or cube) along a horizontal "floor" were presented to the participants on the computer screen. Each video clip was 4 seconds in duration and played once. Then participants were instructed to recount the scene in the movie clip. The other three description tasks were not analyzed in the current study.

Each participant described half of the movie clips of the geometric shape motion task in the face-to-face condition and the other half in the tape recorder condition. In the face-to-face condition, the experimenter faced the participant, and the video camera recording participants' responses (PAL DV camera) was placed next to the experimenter. In the tape recorder condition, participants were left alone in the room and spoke to a tape recorder. Their responses were video-recorded by a hidden camera (Sony PAL DV). There were no practice trials preceding the main trials.

Procedure

Participants were tested individually. After filling out the informed consent form, participants first completed the mental rotation task. Then they were then given half of the speaking tasks either in the face-to-face condition or in the tape recorder condition. Next they were given some personality questionnaires which took about 30 minutes and the results are not going to be reported in this paper. Last, they completed the other half of the speaking tasks in the other condition. The order of the two conditions and the speaking tasks was counterbalanced across participants. At the end of the session, participants were debriefed of the hidden video camera and its purpose, and they were given the opportunity to request erasing the recording. None of the participants reported that they were aware of the hidden camera.

Coding

Participants' verbal descriptions were transcribed verbatim, and all hand gestures were identified. Each gesture was categorised into the following two types (developed on the basis of the classification system in McNeill, 1992): (1) *Representational gestures* were the hand movements that (a) represented the perceptual information of the referent entity; (b) represented the movement of the referent entity; (c) pointed at the referent entity. For example, in the mental rotation task, if a participant simulated manipulating the stimulus object with the index finger and the thumb opposed, this would be counted as a representational gesture. In the geometric shape motion task, if a participant drew circles with her right index finger and moved her hand horizontally while saying "the ball rolled towards a square", this hand movement would be counted as a representational gesture. (2) *Non-representational gestures* are those gestures that could not be classified as representational gestures, including (a) emblem gestures that conveyed conventionalized meanings, such as "maybe" (e.g., a flat hand with the palm down, wavering), "you know" (e.g., a flat hand with the palm up, possibly with a shoulder shrug); (b) beat gestures that were small, baton like gestures produced along with the rhythm of speech; (c) unclear gestures that were unable to be interpreted.

In order to establish inter-coder reliability of gesture coding, 15% of all gesture coding was randomly selected, and a second independent coder classified the selected gesture (N = 287). The two coders' decisions matched 96.17% for the gesture type coding (Cohen's k = .79, p < .001).

Results

In the analysis, we focused on representational gestures. We examined the relationship between the rates of co-thought gestures (number of gestures per minute) in the non-communicative mental rotation task and the rates of co-speech gestures (number of gestures per minute) in the geometric shape motion in both face-to-face and tape recorder conditions. In the mental rotation task, the participants produced overall 259 gestures, 92.66% of which (N = 240) were representational gestures. In the geometric shape motion description task, the participants produced overall 721 gestures in the face-to-face condition, in which 99.58% (N = 718) were representational gestures and overall 480 gestures in the tape recorder condition, in which 98.96% (N = 475) were representational gestures. In line with previous research (Cohen, 1977; Bavelas, et al., 2008), the participants produced more representational gestures in the face-to-face condition (*M* = 25.54 gestures per minute, *SD* = 15.89) than in the tape recorder condition (*M* = 12.68 gestures per minute, *SD* = 12.46), t (40) = 6.07, $p < .01$.

Because the distribution of the co-thought gesture rates data was highly skewed, the Spearman *rho* correlation was applied. The scatter plots of the correlation matrix for the gesture rates of the mental rotation task and the geometric shape motion description tasks is presented in Figure 3. The rates of co-thought gestures were significantly positively correlated with the rates of co-speech gestures both in the face-to-face condition, *rho*(41) = .35, $p < .05$ and in the tape recorder condition, *rho*(41) = .46, $p < .01$. Furthermore, the gesture rates in the face-to-face and tape recorder condition were significantly correlated *rho*(41) = .59, $p < .01$.

Since gesture and spatial thinking are closely linked to each other (Kita & Özyürek, 2003; Chu & Kita, 2008; Ehrlich, Levine, & Goldin-Meadow, 2006), it is possible that the correlation between co-thought and co-speech gestures was attributed to individuals' spatial ability, regardless whether co-thought and co-speech gestures originate from the same or different systems. Therefore, we examined the correlation between individuals' performance in the mental rotation task (as an indicator of their spatial ability) and the rates of co-thought and co-speech gestures. To eliminate possible time-accuracy trade-off, we first transformed participants RTs and error rates into Z scores, and then used the sum of their RT Z scores and error rates Z scores to index their mental rotation performance. Therefore, the higher the sum of the Z scores, the worse their performance was. There was no significant correlation between the sum of the Z scores and the rates of co-thought gestures (*rho*(41) = .22, *ns*.), the rates of co-speech gestures in the face-to-face condition (*rho*(41) = -.07, *ns*.), the rates of co-speech gestures in the tape recorder condition (*rho*(41) = .17, *ns*.).

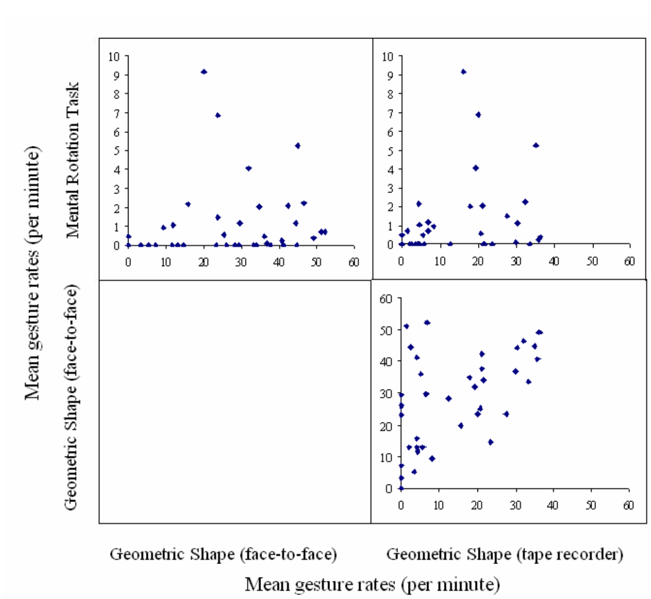

Figure 3. Correlation matrix of co-thought gesture rates (per minute) in the mental rotation task and co-speech gesture rates (per minute) in the geometric shape motion description task

Discussion

The present study investigated whether co-speech gestures originate from the speech production processes or a system that is independent of the speech production processes by examining the relationship between co-speech gestures elicited in a verbal description task and co-thought gestures elicited in a non-communicative mental rotation task. We found a significant correlation between the rates of co-thought gestures and the rates of co-speech gestures. The correlation was robust across different communicative context, and was not attributed to individuals' spatial ability.

As co-thought gestures in the non-communicative mental rotation task presumably originate from a system that does not involve any speech production processes, the positive correlation between the rates of co-thought and co-speech gestures indicates that co-thought and co-speech gestures originate from a common mechanism that is independent of speech production processes. This finding is in line with the idea that co-speech gestures originate from an action generation system that is independent of the speech production processes (Kita, 2000; Kita & Özyürek, 2003). The correlation between co-thought and co-speech gestures is also compatible with the findings from Chu & Kita (2008), in which they found a parallel pattern of results in co-thought and co-speech gestures. These pieces of evidence cannot be explained by the theories which claim that the co-speech gesture production is inherently inseparable from the speech production process. For example, co-speech gestures originate from a "growth point" consisting of a combination of an image representation and a linguistic category (McNeill, 1992) or from one of the stages of the speech production process (Butterworth & Hadar, 1989; de Ruiter 2000).

The results of this study does not exclude the possibility that only a subset of co-speech gestures originates from the same mechanism as co-thought gestures, and the rest of co-speech gestures originate from a part of speech production processes. However, the most parsimonious interpretation is that all (representational) co-speech and co-thought gestures originate from a common mechanism.

In addition, the lack of correlation between participants' performance in the mental rotation task and the rates of co-thought and co-speech gestures may shed some light on the study of individual differences in gesturing. To our knowledge, only one study so far has directly test the relation between spatial skill and gesture production, in which the authors also report a non-significant correlation between individuals' spatial ability and the rates of co-speech representational gestures (Hostetter & Alibali, 2007). Therefore, although gestures and spatial thinking are tightly linked (Kita & Özyürek, 2003; Chu & Kita, 2008; Ehrlich, Levine, & Goldin-Meadow, 2006) and speakers gesture more often when they produce spatial contents than non-spatial contents (Krauss, 1998), the link between individuals' spatial ability and gesture production may not be straightforward, and possibly mediated by other factors such as language abilities (Hostetter & Alibali, 2007), personality, or the social situation.

If the correlation between co-thought and co-speech gesture rates cannot be attributed to variability in spatial abilities, what can the correlation be attributed to? Based on Hostter and Alibali's (2008) theory on gesture production, we suggest that the correlation can be attributed to individual differences in the following two aspects of the action generation system. First, individuals may vary as to how strongly they activate actional representation when they process visuo-spatial information for thinking or speaking. Second, individuals may vary as to the "threshold" for producing overt gestures, that is, the minimum activation level of actional representation that triggers overt gestures.

In sum, the present study provides further understanding of the link between gesture and speech production. We suggest that co-speech gestures originate from a system that is independent of speech production processes. However, it should be noted that the present study only focused on representational gestures, it is possible that only representational co-speech gestures originate from a system that is independent of speech production processes, and other type of gestures such as beat gestures (baton like gestures highlighting some aspects of discourse structure, McNeill, 1992) or emblem gestures (conventional gestures such as a "ok" sign) may instead originate from speech production processes. Of course, further studies need to be done to examine their link with speech production. In addition, to our knowledge, there are very few studies investigating co-thought gestures, which are the spontaneous representational arm and hand movements produced during silent thinking. We suggest that future studies on gestures should not only focus on the speech-accompanying gestures, but also co-thought gestures, which also play important roles in learning and problem solving.

References

Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory & Language, 44,* 169-188.

Alibali, M. W., & Kita, S. (under review). On the role of gesture in thinking and speaking: Prohibiting gesture alters children's problem explanations.

Bavelas, J., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language, 58,* 495-520.

Butterworth, B., & Hadar, U. (1989). Gesture, speech, and computational stages: A reply to McNeill. *Psychological Review, 96*, 168-174.

Chu, M, & Kita, S. (2008). Spontaneous gestures during mental rotation tasks: Insights into the microdevelopment of the motor strategy. *Journal of Experimental Psychology: General, 137,* 706-723.

Cohen, A. A. (1977). The communicative function of hand gestures. *Journal of Communication, 27,* 54-63.

De Ruiter, J. P. A. (2000). The production of gesture and speech. In D. McNeill (Eds.), *Language and gesture* (pp. 284-311). Cambridge: Cambridge University Press.

Ehrlich, S. B., Levine, S., & Goldin-Meadow, S. (2006). The importance of gesture in children's spatial reasoning. *Developmental Psychology, 42,* 1259-1268.

Garber, P., & Goldin-Meadow, S. (2002). Gesture offers insight into problem-solving in adults and children. *Cognitive Science, 26,* 817-831.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin and Review, 15,* 495-514.

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

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Eds.), *Language and Gesture* (pp. 162-185). Cambridge, UK: Cambridge University Press.

Kita, S., & Lausberg, H. (2008). Speech-gesture discoordination in split brain patients' left-hand gestures: Evidence for right-hemispheric generation of co-speech gestures. *Cortex, 44*, 131-139.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language, 48,* 16-32.

Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science, 7,* 54-60.

Mayberry, R. I., & Jaques, J. (2000). Gesture production during stuttered speech: Insights into the nature of gesture-speech integration. In D. McNeill (Eds.), *Language and Gesture* (pp. 199-213). Cambridge: Cambridge University Press.

McNeill, D. (1992). *Hand and Mind.* Chicago: University of Chicago Press.

Perry, M., Church, R. B., & Goldin-Meadow, S. (1988). Transitional knowledge in the acquisition of concepts. *Cognitive Development, 3,* 359-400.

Rauscher, F. B., Krauss, R. M., & Chen, Y. (1996). Gesture, speech and lexical access: The role of lexical movements in speech production. *Psychological Science, 7,* 226-231.

Schwartz, D. L., & Black, J. B. (1996). Shuttling between depictive models and abstract rules: Induction and fallback. *Cognitive Science, 20,* 457-497

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171,* 701-703.